Scientia Iranica D (2022) 29(1), 121–134

Sharif University of Technology
Scientia Iranica
Transactions D: Computer Science & Engineering and Electrical Engineering
http://scientiairanica.sharif.edu

Modified tumor diagnosis by classification and use of canonical correlation and support vector machines methods

H. Samadi Ghuoshchi and Y. Pourasad*

Faculty of Electrical Engineering, Urmia University of Technology, Urmia, Iran.

Received 8 July 2020; received in revised form 5 July 2021; accepted 25 October 2021

KEYWORDS
Tumor detection;
Classification;
CCA;
SVM;
Image processing.

Abstract. The main objective of this research is to investigate techniques for classifying tumor grades based on image processing. The algorithms used to classify tumors are introduced, and their performance in the experimental results are evaluated. In the proposed algorithm, first, the scan images of the lung are pre-processed and then, the histogram, texture, and geometric features are extracted. These characteristics are then employed in Support Vector Machines (SVM) and Canonical Correlation Analysis (CCA) classifiers to diagnose tumors and classify benign and malignant types. These integrated approaches to investigation of medical images are vital tools for improving the diagonalization accuracy. In the current research, experimental and simulated medical images are employed. The outcomes of the developed techniques in this research are compared with those found in the literature review to confirm the efficacy and reliability of the proposed approach in diagnosing and classifying tumors. In addition to high accuracy in diagnosis, this method is also a low-cost and low-risk method. Owing to its very high sensitivity, this method has the desired values of two criteria of precision and specificity as well as the small number of features used for classification; therefore, the developed method was proposed as an efficient and appropriate one for tumor classification.

1. Introduction

There are two common ways to diagnose a tumor: diagnosis based on computational and physical characteristics. In order to determine computational features such as standard deviation, color intensity, entropy, etc., applied algorithms are used to detect tumors [1–3]. Computational features can be used to check the presence or absence of heterogeneous tissues in the brain. Medical images mainly include the appearance of the brain, such as cerebrospinal fluid, cranial bone tissue, and gray and white particles, as well as information about tissue contrast, resolution, and location of tissue masses. Diagnosis based on these features has gained significance due to the possibility of tumor detection in its initial stages [4,5]. In [3], identification of the brain tumor area was made feasible by predicting the tumor type and boundary box. Magnetic Resonance Imaging (MRI) brain tumor images are trained faster than the baseline using Region-based Convolutional Neural Networks (R-CNN). The faster R-CNN combines the AlexNet and proposed region network models.

In [6], tissue features and possible neural network were employed to diagnose malignant and normal tumors. This algorithm uses wavelet conversion coefficients in different bands. In [7], four tumors were classified as astrocytoma, meningioma, carcinoma, and

*. *Corresponding author.*
E-mail address: y.pourasad@uut.ac.ir (Y. Pourasad)

doi: 10.24200/sci.2021.56370.4694

sarcoma using matrix-based features of co-occurrence of gray surfaces, network classification Neon, and Levenber-Markard nonlinear optimization algorithm. In [8], neural network classifications of radial base function and Support Vector Machine (SVM) were employed to detect normal and abnormal brain tumors using matrix features of co-occurrence of extracted gray surfaces from MRI images and tomography. The reported accuracy of a database of 123 MRI images, without the application of color channels, is 73% with a sensitivity rate of 77% and a specificity of 68%. In [9], researchers proposed an interpretive approach to converting four-dimensional data so that standard Convolution Neural Networks (CNN) and two-dimensional architectures can be utilized to perform brain tumor segmentation. Findings on the brain tumor segmentation dataset point to a dice score of 7.83% for the total tumor region, 6.73% for the tumor core region, and 69% for the active tumor region. In [10], brain tumor segmentation data were obtained using a simple SegNet network. The average dice accuracy reported on the brain tumor segmentation database was 6.87%. The main problem attributed to the methods proposed so far is their low accuracy rate, and different methods have attempted to increase the accuracy of the classification. This study aims to increase the accuracy of classification as much as possible based on the proposed model using different useful features in this regard.

In recent studies, the diagnosis and classification of brain tumor diseases have been presented, e.g., using image processing techniques for diagnosis of brain tumors [11], brain tumor diagnosis using fragment MRI images and K-mean algorithm [12], use of digital image processing in the diagnosis of brain tumors and image feature calculation techniques [13], detection of brain tumors in proximity to the nearest neighborhood and then, fragmentation of images using the waterfall algorithm [14], diagnosis of the location of brain tumor using the intensity of MRI [15], diagnosis of a brain tumor in medical images using K-mean violet conversion [16], MRI image processing and extraction of the best feature by particle swarm optimization, and genetics algorithm. Moreover, fuzzy clustering is employed to diagnose brain tumors [17]. In [18], the comparison of random forest could be used in addition to SVM for the brain tumor segmentation and detection. One of the benefits of random forest is the capability of handling large data sets with higher dimensionality and accuracy. While it handles classification quite professionally, it performs poorly in resolving regression issues as it does not provide an accurate continuous nature forecast. The regression case did not predict beyond the trained data variety and might overfit noisy data sets.

To improve the diagnosis of brain tumors and ensure quick access to the location of the brain tumor mass in this study, the brain tumor was divided into pieces, and the area in the brain images was calculated using the classification method. The proposed method in this paper offers promising results when compared to the brain tumor segmentation detection system. In this study, the proposed architecture was also used for the gastric cancer dataset, and a better performance was achieved. Based on the findings, tumor identification becomes predictable through Canonical Correlation Analysis (CCA), and the efficacy of this prediction is quite acceptable.

2. Methodology

In the proposed method, two types of Flair and T1W brain MRI images were used. They should be preprocessed to normalize their pixel situation and extract extra data such as noise. After preprocessing, both of these images move on to the fenestration state to extract the features already discussed. Based on the previous post-windowing process, there is a quite important step which is extracting the tissue and statistical properties applied to each Flair and T1W brain MRI image. In the next step, which is the innovation of the proposed method, the extracted features of both images are combined using the CCA algorithm to obtain a number of new features and reduce these new features and accept the classification over them. Based on the obtained data, there are two parts of training and trial in the classifiers step. Trained and real images for this plan are also employed. These images are chosen based on a batch K-cross validation algorithm. Finally, the post-processing and benchmarking steps are performed.

The proposed approach in the present study (Figure 1) adds new features, including two CCA and a batch K-cross validation algorithms, in comparison to the reference [3], to overcome its weakness, i.e., accuracy. The batch K-cross validation algorithm produces K values which are stochastic in nature. The process K is reproduced to give great fidelity to the findings.

The brain tumor classification system presented in this study follows an algorithm described in the

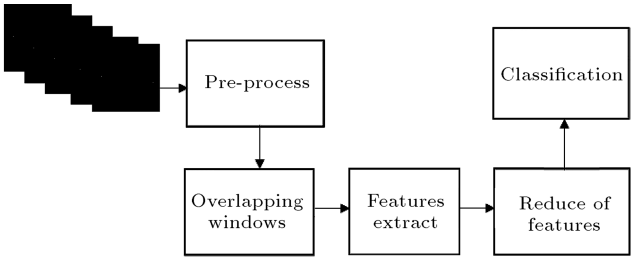


Figure 1. Schematic of image processing.

following. This algorithm includes multiple steps such as preprocessing, windowing, feature extraction, feature size reduction, classifier and post-processing, and performance calculation. Figure 1 shows a general diagram of the tumor classification system.

First, MRI images taken from the patient's brain require preprocessing of images in multiple classification algorithms. This stage (preprocessing) incorporates several parts such as noise cancellation, normalization of the brightness of brain MRI images, smoothing of the size of brain MRI images, and removal of extra information lying in the background of brain MRI images.

The next step in this classification system is brainstorming the MRI images. This step consists of two parts: overlapping windows and non-overlapping windows.

In the third step, which is one of the main steps of this system, according to the windows formed in the previous step, the features given as input to the classifier are extracted. Of note, a feature vector is provided for each window.

At the fourth stage, the dimensions of the feature are reduced due to a large number of extracted features. In the next and last step, the available attribute vectors are trained as input to the classifier. Each of these feature vectors is labeled either healthy or tumorous. To determine whether it is healthy or tumorous at the window testing phase, if a window is in the healthy category, the center pixel of that window is repeated as a tumor. Upon completing these steps, post-processing is employed to reduce the error of this method and measure its efficiency.

One of the characteristics of first-time statistics is image histogram. In other words, the information related to the first-order statistics is obtained through image histogram. In addition, it provides a summary of the statistical data of an image [19]. Among the characteristics of the first-order statistics, energy, entropy, mean, elongation, asymmetry, and mean contrast (variance) can be mentioned.

1. Energy:

$$E = \sum_{i=0}^{G-1} [p(i)]^2. \tag{1}$$

2. Entropy: The degree of uncertainty in light intensity values.

$$H = -\sum_{i=0}^{G-1} p(i) \log_2 [p(i)]. \tag{2}$$

3. Average: The average brightness of the whole image

$$\mu = \sum_{i=0}^{G-1} ip(i). \tag{3}$$

4. Elongation: How flat the histogram is

$$\mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (i-\mu)^4 p(i) - 3. \tag{4}$$

5. Asymmetry: The standard deviation of the histogram around the mean:

$$\mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i-\mu)^3 p(i). \tag{5}$$

6. Variance: Changes in light intensity around the mean.

$$\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 p(i). \tag{6}$$

In the above relations, (G) is the number of gray surfaces of the image, and $P\ (i)$ the probable density of light intensity levels, expressed in the following:

$$p(i) = \frac{h(i)}{N}, \tag{7}$$

where $h(i)$ is the number of all pixels whose brightness value is I, and N represents the total pixel numbers.

2.1. Characteristics of second-order statistics

Gray-Level Run Length Matrix (GLRLM) is a second-order statistic and spatial method. This matrix is formed in such a way that the primary texture, called the gray surface row length, is considered, and a large number of pixels in a row have the same gray surface. These rows have a row length and direction for a gray value [20]. To calculate this matrix, first, the number of the rows of a gray area with row length should be determined; for example, Figure 2 shows a 4-4 image with four gray areas between the numbers 0-3. In order to create the GLRLM matrix for this image, we must first consider four different directions: 0, 45, 90, and 135 degrees are the directions that should be used. According to Figure 3, calculation of this matrix is clear. For example, in Figure 3(a), we first look at the gray surface row; the first row takes the number 'zero', thus reaching the length of the row. Based

0	1	2	3
0	2	3	3
2	1	1	1
3	0	3	0

Figure 2. Hypothetical image matrix.

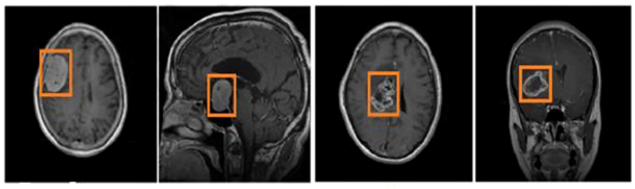
Row length/ gray surface	1	2	3	4
0	4	0	0	0
1	4	0	0	0
2	0	0	1	0
3	3	1	0	0
Row length/ gray surface	1	2	3	4
0	4	0	0	0
1	4	0	0	0
2	3	0	0	0
3	5	0	0	0

(a)

Row length/ gray surface	1	2	3	4
0	4	0	0	0
1	1	0	1	0
2	3	0	0	0
3	3	1	0	0
Row length/ gray surface	1	2	3	4
0	2	1	0	0
1	4	0	0	0
2	3	0	0	0
3	3	1	0	0

(b)

Figure 3. Gray surface length matrix.



(a) (b) (c) (d)

Figure 4. Brain tumors of weighing T1: (a) The area in the rectangle includes the tumor, (b) the meningioma is nearby the skull, (c) glioma contains edema and necrosis, and (d) glioma is surrounded by inflammation.

on the image matrix, the zero pixels in the 0-degree direction are repeated four times individually. In other directions, no duplication of zero pixels is observed in the 0-degree direction. Therefore, in rows 2, 3, and 4, the gray area becomes zero. This trend continues for other angles.

2.1.1. Gray Surface Length Matrix

In Figure 4, (a) is for 0 degrees, (b) for 45 degrees, (c) for 90 degrees, and (d) for 135 degrees. The GLRLM matrix has the following properties:

1. Short length rows where according to the following equation, the length value of each row is divided by the power of two squares of each row:

$$RF_1(R(\theta)) = \frac{1}{T_p}\sum_{i=0}^{G-1}\sum_{l=1}^{N_R}\frac{r'(i,l\,|\theta)}{l^2}. \tag{8}$$

2. Long length rows where according to the following equation, the length of each row is multiplied by the power of two squares of each row:

$$RF_2(R(\theta)) = \frac{1}{T_p}\sum_{i=0}^{G-1}\sum_{l=1}^{N_R} j^2\, r'(i,l\,|\theta). \tag{9}$$

3. Gray area distribution: where according to the following equation, first, the lengths of rows are calculated for each gray area. These values are then added, and they must be normalized to calculate the gray area distribution:

$$RF_3(R(\theta)) = \frac{1}{T_p}\sum_{i=0}^{G-1}[\sum_{l=1}^{N_R} r'(i,l\,|\theta)]^2. \tag{10}$$

4. The row length distribution is calculated using the following equation for the image:

$$RF_4(R(\theta)) = \frac{1}{T_p}\sum_{i=0}^{N_R}[\sum_{l=1}^{G-1} r'(i,l\,|\theta)]^2. \tag{11}$$

Row percentage according to the following equation (the ratio of the number of all rows to the number of possible rows) is obtained.

$$T_p = \sum_{i=0}^{G-1}\sum_{l=1}^{N_R} r'(i,l\,|\theta), \tag{12}$$

where G is the number of gray surfaces; P the total number of pixels; N_R the number of longest rows (number of row lengths); T_P the total number of rows in images. T_P is also used to normalize images.

2.1.2. Linear Binary Pattern (LBP) features

The LBP performs windowing throughout the image that can sweep the entire image. It then compares the center pixel value with the surrounding pixel and assigns binary numbers (0 and 1) to the surrounding pixels. In the next step, the LBP multiplies the numbers zero and one by the powers of two clockwise or counterclockwise, and finally adds them up. Finally, the resulting histogram represents the target tissue. For the pixel neighborhood, R and P are used, where P is the number of sample points on a circle with radius

R. The LBP pixel values YC and XC is calculated using the following equations:

$$LBP_{P,R} = \sum_{P=0}^{P-1} S(g_P - g_c)\, 2^P, \tag{13}$$

$$S(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{14}$$

2.2. Features of Oriented Gradients (OG)

The operation of this actuator is such that once the gradient intensities are disrupted, it is able to reveal the appearance and shape of the tumor or the desired objects. This operator, one of the most important operators in object detection, is calculated on a space known as a cell and has a uniform network. Local contrast normalization is employed to increase the accuracy of this operator. In general, a summary of feature extraction from brain MRI images is as: Extracted and statistical features (image histogram features, features of second-order statistics, features of LBP, features of OG).

2.3. System classification

In the tumor classification system, the classification stage is one of the most important stages. This stage of the system is followed by the stage of reducing the number of extracted features. In this step, attributes are given to the classifier input. The classifier converts the input data, which is quantitative, into qualitative output data. This output is a real value or a vector.

2.3.1. *Classification of brain tumors for MR images using accurate transfer learning and quantification*

Accurate classification of MRI of brain tumors plays an essential role in clinical analysis and decision-making for therapy. The critical issue in analyzing MRI is the semantic gap in the low-level visual data recorded using the MRI machine and the high-level data recorded by human evaluator. Traditional machine learning systems for classification focus only on high- and low-level characteristics. They employ some manual elements to overcome this gap and require proper characteristic classification and extraction techniques.

Recent advances in deep learning have shown tremendous progress, and deep CNNs have emerged successful in task classification. Deep learning has become influential in providing features that can fully display data and embed the characteristics analysis in learning, but it generally requires data set [11]. For most medical learning scenarios, this training dataset is small. Therefore, in a small dataset, application of deep learning and training to CNN is challenging from the beginning. In addition, the hardware and computational time constraints are the other factors that justified usage of machine learning over deep learning.

As shown in Figure 4(a) and (b), meningioma is generally squares of gray matter and cerebrospinal fluids. The pituitary tumor was placed nearby the sphenoid sinus, internal carotid arteries, and optic cysts. Glioma appears to be in an unequal pattern, as depicted in Figure 4(c) and (d). Therefore, the most crucial data and diagnostic characteristics of brain tumors were associated with the area of the tumor in the MRI image along with its border, tissue, intensity, and pattern [5].

SVM represents one of the supervised learning methods used for classification and regression. This technique is relatively novel and enjoys enhanced proper and acceptable efficacy compared to older methods for classification, including perceptron neural networks. The basis of the SVM classifier is the linear classification of data, and in linear segmentation of data, the line with the highest confidence margin is chosen. Quadratic Programming (QP) methods were employed to solve the optimal line equation for data. They are also known methods for solving constrained problems.

The simplest interpretation of the operation of the backup vector machine function is that in the backup vector machine algorithm, the distance between the nearest data sample and separator line (boundary between categories) is calculated (these data samples may belong to any category). SVM uses a technique called kernel trick to convert the available data and then, the optimal boundary between possible outputs (see Figure 5) is determined based on that conversion. In simple terms, it performs very complex conversions and shows how to separate the available data based on the tags or outputs defined. One of the methods that is currently widely used for the classification problem is the SVM method. Perhaps, the current popularity of the SVM method can be comparable to that of neural networks over the past decade mainly because the former enjoys the ability to solve different problems, while techniques like the decision tree cannot be easily utilized in different types of applications.

CCA algorithm was developed in [4] as a statistical method dealing with the interaction between two

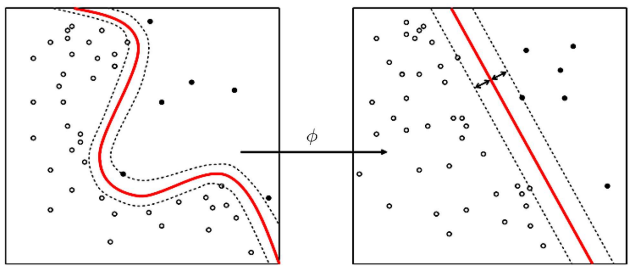


Figure 5. Data scatter diagram in support vector.

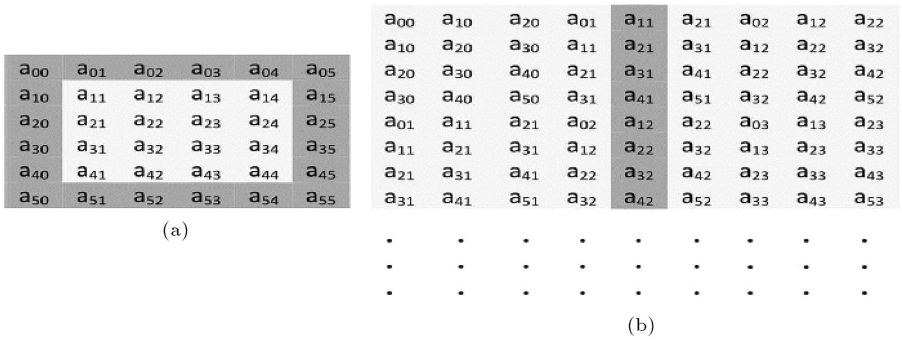


Figure 6. Formation of a mean skull-k patch: (a) Original image (light gray) with zero plug (dark gray) and (b) patch matrix.

random vectors used in multivariate statistical analysis. This algorithm measures the linear relationship between two multidimensional random variables. For each of those multidimensional variables, CCA finds a base vector that has optimal conditions in terms of cross-correlation and, at the same time, calculates the cross-correlation of two multidimensional variables. In other words, this algorithm seeks to find two base vectors in which the correlation matrix between the variables is diagonal whose diameter has maximum values.

Suppose that $X = (x_1, x_2, ..., x_n) \in \mathbb{R}^{p\times n}$ and $Y = (y_1, y_2, ..., y_n) \in \mathbb{R}^{q\times n}$. There are two sets of feature vectors with assumed dimensions that are extracted from n experimental samples. CCA attempts to find two base vectors w_x and w_y for the two variables X and Y such that the correlation between the reflections of two vectors on those canonical bases ($x_i^* = w_x^T(x_i - \bar{x})$ and $y_i^* = w_y^T(y_i - \bar{y})$ as $i = 0, 1, ..., n$ $\bar{x}$, where $\bar{x}$ is the mean of X and $\bar{y}$ mean of Y), is mutually maximized. In other words, Eq. (15) should be maximized as follows:

$$\rho = \frac{E[xy]}{(E[x^2]E[y^2])^{1/2}} = \frac{E[w_x^T XY^T w_y]}{(E[w_x^T XX^T w_x]E[w_y^T YY^T w_y])^{1/2}}. \quad (15)$$

Through simplifying Eq. (15), Eq. (16) is obtained as:

$$\rho = \frac{w_x^T C_{xy} w_y}{(w_x^T C_{xx} w_x w_y^T C_{yy} w_y)^{1/2}}, \quad (16)$$

where C_{xx} and C_{yy} are the covariance matrices of x and y, respectively, and the C_{xy} is the mutual covariance matrix of x and y. After optimizing Eq. (16), Eqs. (17) and (18) are obtained as:

$$Z_1 = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} w_x & 0 \\ 0 & w_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (17)$$

$$Z_2 = X^* + Y^* = \begin{pmatrix} w_x \\ w_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (18)$$

where Z_1 and Z_2 are called distinctive Canonical Correlation Discriminant Features (CCDFs), which are actually a combination of features extracted by both aggregation and addition methods. The method used in this study enjoys new characteristics, compared to that in [3], which are the application of two CCA algorithm as well as the K-category cross. The batch K-cross validation algorithm produces stochastic K values which are stochastic. As shown in Figure 6, the process K is reproduced to give great fidelity to the findings. The patch-based K-mean method is used to empty the skull (brain tissue extraction) as a preprocessing step, which has the added property of classifying the element depending on its neighborhood patterns. Following the removal of the skull, a classifier is applied to the brain tissue and finally, brain tumors are removed using shape-based topological properties.

In the windowing stage, the obtained properties contained 315 features: 32 related to the directional histogram, 20 to the row matrix of gray surface, 256 to the linear binary features, and the remaining 7 to the first-order ones. Therefore, it can be concluded that each window provides a property vector with 315×1 dimensions. In addition, as mentioned earlier, each brain MRI image has 4,096 windows with a size of 10×10. As a result, for each brain MRI image, we have a characteristic matrix with dimensions of 4096×315.

3. Results

As mentioned earlier, the available images contain simulated and real MRI ones. It should be noted that the application of the pre-processing step is distinct for all the simulated and real images. Simulated images were first required to convert the format of the images taken from the database to the standard one such as the bmp. Next, the pixel location of the MRI images was normalized. For this purpose, the size of the image should be changed into 200×200 pixels to remove the extra pixels. In order to pre-process real images, in addition to the steps done for the simulated images, two more actions are required. Normalizing the light

intensity and implementing the histogram matching are the same steps mentioned.

3.1. Normalization of the light intensity

Normalization of the light intensity is an important step in preprocessing which is based on the standard deviation of MRI. The basis of this method is to calculate the standard deviation of the MRI image columns and then, to divide the pixels of each column using the corresponding standard deviation.

3.2. Application of histogram algorithm

Histogram algorithm begins functioning by considering an image from a group of real images as the reference ones and calculating and matching them. The reference images are selected based on researchers' personal experience as well as measuring relevant criteria. Figure 7 depicts the actual MRI image and calculated histogram as graphs. Figure 8 illustrates the actual MRI image with its histogram with/without employing the algorithm.

3.2.1. Results of SVM classification

The backup vector machine classifier has four main kernels: linear kernel, radial base function kernel, Gaussian kernel, and polynomial kernel. Each of these kernels is taken into account so that the SVM classifier can be applied to the entire category of textural and geometric features and to each category of features. In addition, the best performance of each recording mode as well as the best kernel, in which the largest number of criteria is obtained, will be selected as the winning category. Figure 9 shows the criteria values for the backup vector machine classifier with all features. In all of the diagrams in this chapter, the values of the three criteria of accuracy, sensitivity, and specificity are presented in three colors of blue, orange, and yellow, respectively.

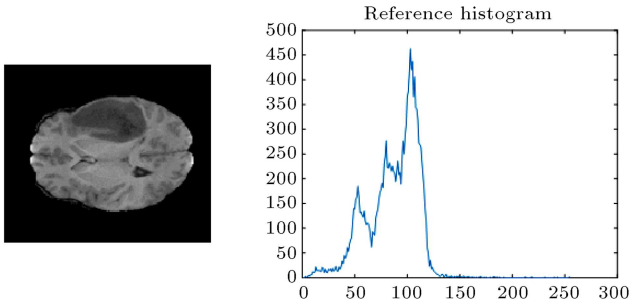


Figure 7. Images and its references histogram.

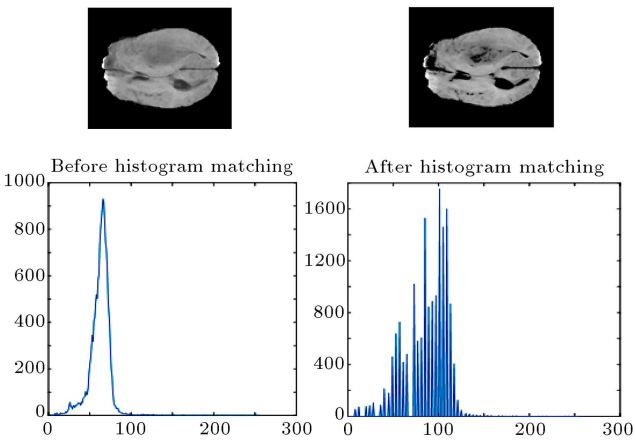


Figure 8. Images and the use of algorithms.

Table 1. Results of the criteria of the top two categories of backup vector machines with all features.

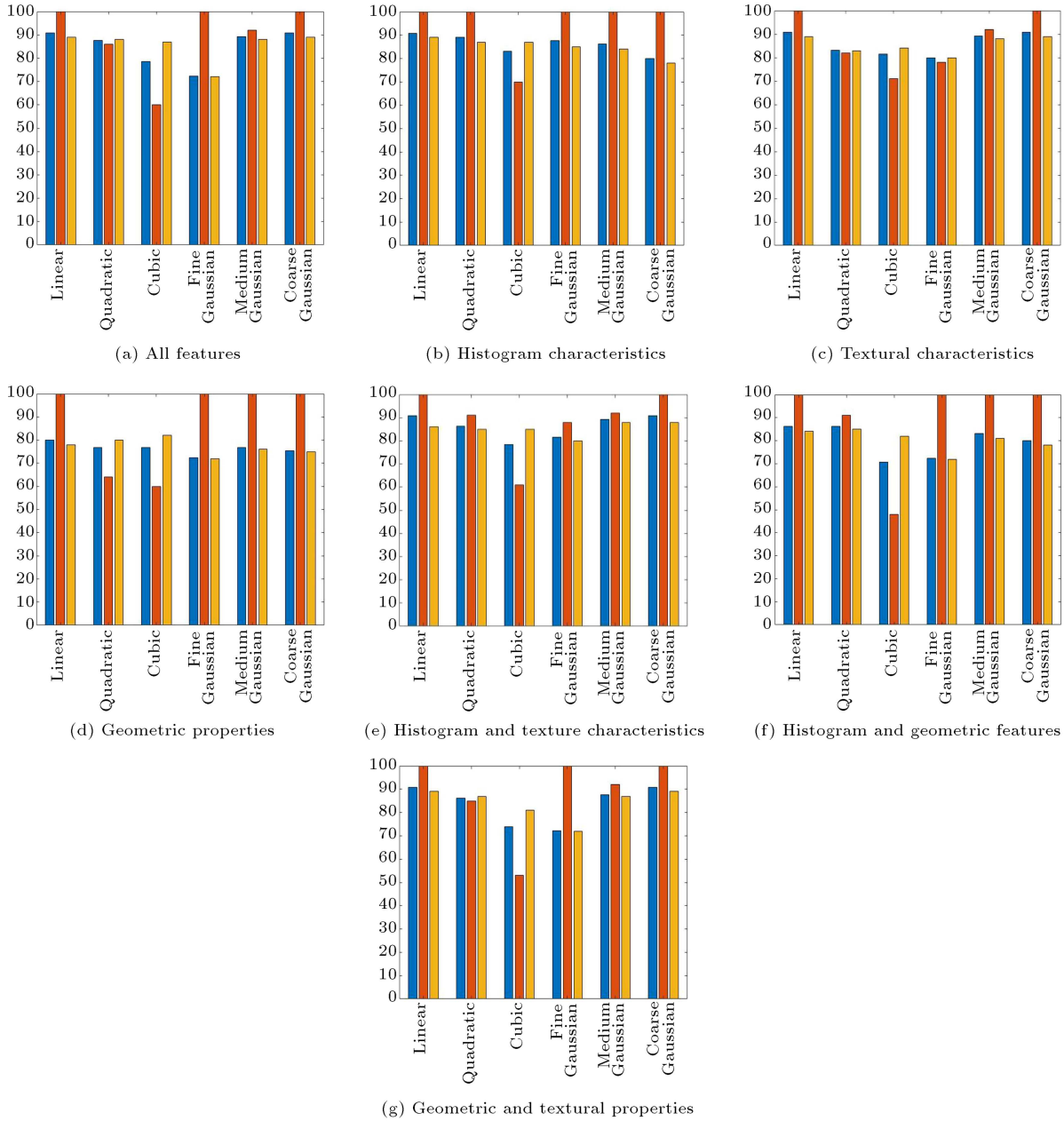
Criteria	Coarse Gaussian	Linear
Accuracy	90.8	90.8
Sensitivity	100	100
Specificity	89	89

Figure 9 shows six SVM kernels. First, each feature set and a dual combination of feature sets are also considered. The criteria for superior classification in each of these cases are given in Tables 1–3. According to the diagrams, the two linear and Gaussian coarse classifications yielded the best results. Therefore, in the case of selecting all the features, these two classifiers are selected as the top SVM classifiers. In the case of selecting each feature category, the histogram feature category with linear kernels and texture feature category with linear kernels and Gaussian coarse exhibited the best performance. Finally, in order to select a dual combination of features for these two modes, i.e., histogram texture and texture-geometric patterns with linear and coarse Gaussian kernels, the highest accuracy, sensitivity, and specificity criteria should be obtained.

3.2.2. Results of K-Nearest Neighbor (KNN) classification

In this section, properties are classified using the KNN classifier, and the criteria of accuracy, sensitivity, and degree of specificity are calculated. In the KNN classifier, the four values of 3, 5, 7, and 9 are selected for the K parameter, and the distance criterion is of Euclidean type. This section, like the SVM classifier, lists the following features selected for classification in order: all the features, each feature, and a dual combination of features, and the top classifier (K-top) is selected.

The classification results are shown in Figure 10. These graphs were investigated according to which, in the case of selecting all the features, selection of 9 nearest neighbors, i.e., $K = 9$, yields the best criteria. In addition, in the case of selecting each feature category alone, texture properties yield the best results by selecting $K = 9$. Finally, in the selection of two feature categories, the combination of texture and geometric feature categories for $K = 7.9$ results in the highest accuracy, sensitivity, and specificity, which



(a) All features (b) Histogram characteristics (c) Textural characteristics

(d) Geometric properties (e) Histogram and texture characteristics (f) Histogram and geometric features

(g) Geometric and textural properties

Figure 9. Comparison of the results of six support vector classifiers.

Table 2. Results of the criteria of the top two categories of backup vector machines with individual characteristics.

Criteria	Histogram	Tissue	Geometry
Top classifier	Linear	Linear, coarse Gaussian	Linear
Accuracy	90.8	90.8	80
Sensitivity	100	100	100
Specificity	89	89	78

Table 3. Results of the criteria of the top two classifications of the support vector with a dual combination of features.

Criteria	Histogram + tissue	Histogram + geometry	Geometry + tissue
Top Classifier	Linear, Coarse Gaussian	Linear	Linear, Coarse Gaussian
Accuracy	90.8	86.2	90.8
Sensitivity	100	100	100
Specificity	89	84	89

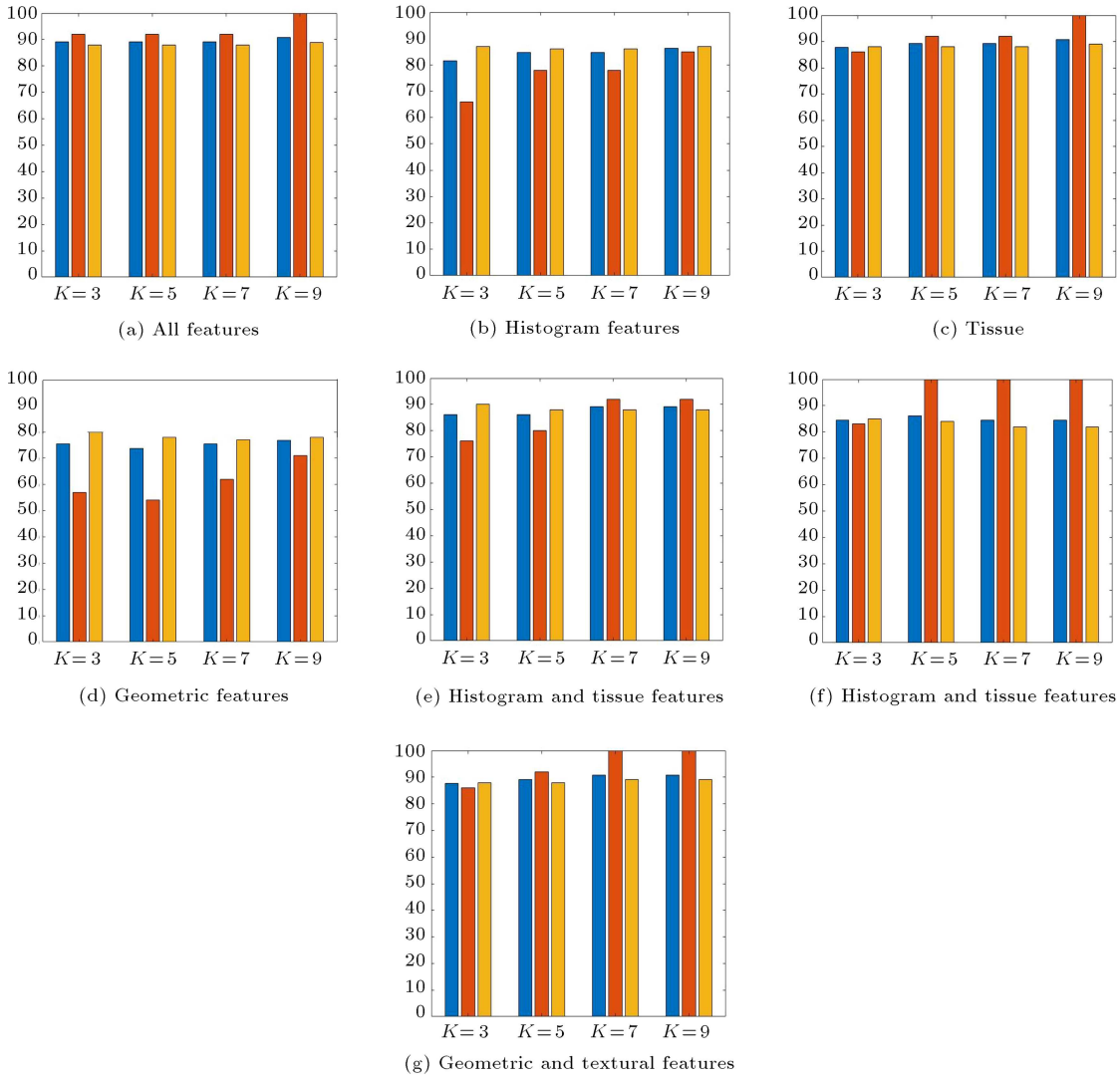


Figure 10. Comparison of system performance in K classifier of the nearest neighbor and weighted at distances $k = 3, 5, 7, 9, 11$.

Table 4. Results of KNN top classification criteria with all features.

Criteria	$K = 9$
Accuracy	90.8
Sensitivity	100
Specificity	89

Table 5. Results of KNN top classification criteria with individual characteristics.

Criteria	**Histogram**	**Tissue**	**Geometry**
K-number	$K = 9$	$K = 9$	$K = 9$
Accuracy	86.2	90.8	76.9
Sensitivity	85	100	71
Specificity	87	89	78

are equal to the selection of all feature properties for $K = 9$. Therefore, in the KNN classifier, a combination of two categories of texture and geometric properties with $K = 7, 9$ is selected as the superior classifier. The values of the criteria for each different feature category and K are given as percentages in Tables 4–6.

3.3. Database and simulated parameters

The objective of this study is to provide a method for classifying real and simulated brain MRI images in order to improve the three criteria for evaluation. Brain MRI images share some similarities in the brightness of their pixels, thus making it difficult to distinguish healthy tissue from tumor tissue and performing a functional classification of the tumor. According to the routine of the classification system, the statistical features of these images are obtained by applying the aforementioned pre-processing and windowing. The dynamic range of the obtained values includes different conditions due to a variety of the derived properties,

Table 6. Results of KNN superior classification criteria with a dual combination of features.

Criteria	Histogram + tissue	Histogram + geometry	Geometry + tissue
K-number	$K = 7.9$	$K = 5$	$K = 7.9$
Accuracy	89.2	86.2	90.8
Sensitivity	92	100	100
Specificity	88	84	89

causing an imbalance in the property matrix. Based on the aforementioned intent, the dynamic range is normalized to eliminate or minimize these differences. First, the method functions in a way that in each property of the matrix column, the mean value of the column is subtracted from the related column and then, it is divided by the column standard deviation.

3.4. Simulation results

In the initial step, the real brain MRI images are evaluated. Then, to make a comparison between the classifier performances, the findings of the simulated MRI images are reported, as shown in the next step. These results are obtained from three classifiers: K-means, SVM, and CCA. The focal correlation is employed to enhance the representation of the proposed approach. Table 7 presents the outcomes of the K-means, SVM, and CCA algorithms.

Given that $K = 2$, the procedure is duplicated, and the rates attributed to each model in the determination of the offered approach are twice the mean of the process. The first group comprises the trained brain MRI image, and the other the trial group. Since a multiple of 72 ($K = 2, 3, 4, 6, 9$) is used for the selection, K is assigned to Number 3 in the order where it exists. In this case, 24 images should be assigned to the training group, and 12 to the experimental group. The procedure is repeated three times, and the trained and real groups are transferred once the procedure is completed. Table 8 presents the outcomes for $K = 3$, which are presented in Table 8.

The findings for the simulated brain MRI images are obtained and investigated for different Ks based on the aforementioned process. These outcomes are presented in Tables 9 and 10 for the simulated images.

3.5. Integration of extracted features to increase accuracy

In studies on the classification and diagnosis of diseases with medical image processing, the results are usually

Table 7. Various algorithms performance for MRI in the case of $K = 2$.

	Classification method			
Criterion	**Ref**	**CCA[a]**	**SVM[b]**	**K-means**
Percentage of sensitivity	—	92.5	91.5	91.5
Percentage of specificity	—	95.1	94.2	95.1
Percentage of accuracy	92.1	95.0	93.7	953

[a]CCA: Canonical Correlation Analysis; [b]SVM: Support Vector Machine.

Table 8. The performance of various algorithms for MRI images in the case of $K = 3$.

	Classification method			
Criterion	**Ref**	**CCA[a]**	**SVM[b]**	**K-means**
Percentage of sensitivity	—	94.1	92.1	92.0
Percentage of specificity	—	94.3	94.2	93.5
Percentage of accuracy	93.0	94.2	94.1	93.4

[a]CCA: Canonical Correlation Analysis; [b]SVM: Support Vector Machine.

Table 9. The performance of various algorithms for MRI images in the case of $K = 2$.

	Classification method			
Criterion	**Ref**	**CCA[a]**	**SVM[b]**	**K-means**
Percentage of sensitivity	96.2	97.2	97.8	97.1
Percentage of specificity	95.7	97.6	97.5	96.3
Percentage of accuracy	96.7	97.5	97.6	96.4

[a]CCA: Canonical Correlation Analysis; [b]SVM: Support Vector Machine.

Table 10. The performance of various algorithms for MRI images in the case of $K = 3$.

Criterion	Classification method			
	Ref	CCA[a]	SVM[b]	K-means
Percentage of sensitivity	96.2	96.9	97.8	98.1
Percentage of specificity	95.7	97.7	97.4	96.1
Percentage of accuracy	96.7	97.5	97.5	96.4

[a]CCA: Canonical Correlation Analysis; [b]SVM: Support Vector Machine.

expressed by three criteria: accuracy, sensitivity, and degree of specificity:

- TP: The number of malignant glands classified as malignant glands.
- FP: The number of benign glands classified as malignant.
- FN: The number of malignant glands classified as benign glands.
- TN: The number of benign glands classified as benign glands.

The above relationships can be easily extracted from the confusion matrix which is a matrix used for elaborating the performance of the classification model and displaying the distribution of the categories in terms of correctness or incorrectness. Figure 6 shows the binary deformation matrix (two classes).

Figure 11 presents a correlation diagram of sensitivity patterns upon the application of a focal correspondence interpretation to simulating brain MRI images based on different K values in the K-fold cross-validation procedure, using the K-means classifier. Based on the number of simulated brain MRI images, if $K = 2$, 20 images are chosen as trained images and the remaining as trial ones, and the number of iterations is 2. Since $K = 3$, 26 images are picked as trained images and the others as trial ones, and the number of iterations is 3. If $K = 6$, 30 images are chosen as trained images and the others as trial ones; the number of iterations is 6 and so on for the $K = 9$.

As presented in Figure 11, the K-means algorithm is used to show that application of CCA to the simulated brain MRI improves the sensitivity criteria from 0.6% to 1.56%, hence the apprehension ratio of tumor pixels rises. Figure 10 shows the comparative diagrams of sensitivity using the focal correlation procedure for classifying brain tumors in real images with different K values in batch cross-validation utilizing K-means. Under such condition, the number of actual brain MRI images handled is 40. In this experiment, it is assumed that $K = 2$, $K = 3$, $K = 6$, and $K = 9$. Tumor classification outcomes reveal that the sensitivity pattern increases depending on CCA employed in the characteristic extraction step, especially for real brain ones.

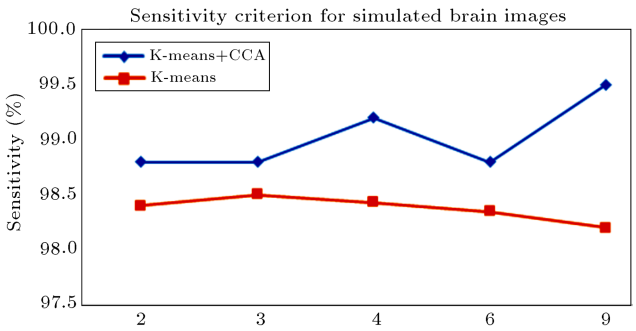


Figure 11. Sensitivity for K-means classifier using canonical correlation.

Application of the CCA algorithm developed the sensitivity pattern up to 22%. The three parameters of specificity, sensitivity, and accuracy criteria were evaluated for the K-means classifier in the stated and reference approaches [20]. The total number of images applied in [21] was 26, while 70 images were used in this study. Therefore, the database employed here is more extensive than the reference; therefore, it can be concluded that the obtained results are reliable [21].

In [21], only the accuracy factor was taken into consideration, and two other factors were not listed. In the present research, only the most beneficial outcomes from Reference [21] to the SVM classifier were discussed, and the consequences characterized by lower accuracy than the standard accuracy were withdrawn.

In [21], to enhance safety, the K-batch cross-validation procedure was only used for $K = 10$, i.e., most of the data were employed as trained images, and only about 15% were employed as experimental ones. However, this study selected $K = 2$, $K = 3$, $K = 6$, and $K = 9$, which utilized 50%, 34%, 25%, 15%, and 10% of the data as experimental ones, respectively, thus making the intended approach in the current study more authentic.

Figure 12 presents a sensitivity comparison of the classifiers of simulated brain MRI images in the current developed method with those of [21]. Since the sensitivity and specificity factors were not presented in [21] for classifying real brain MRI, just the presented approaches charts in the current research utilize K-means, CCA, and SVM. The real brain MRI for the precision factor is presented in Figure 10.

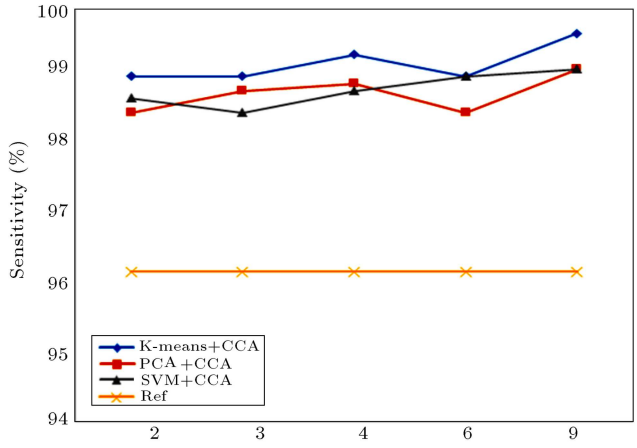


Figure 12. Accuracy factor for three classifiers (simulated).

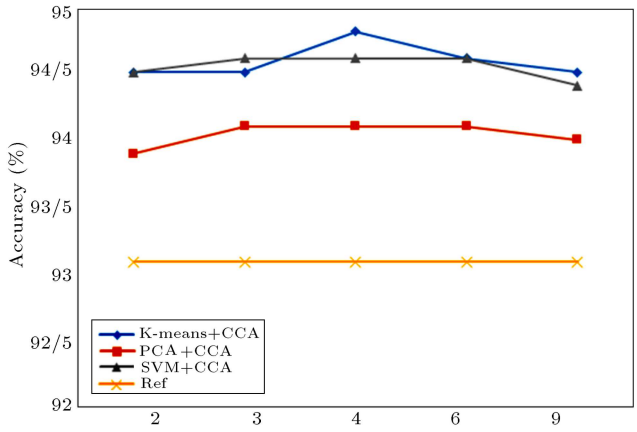


Figure 13. Accuracy factor for three classifiers (experiment).

As presented in Figure 12, the sensitivity factor in simulated brain MRI images procedure increased by up to 3% using the CCA, compared to that in [20] with regard to the number of various batches (various K) in the K batch cross-validation algorithm. According to Figure 12, the K-means outperforms its counterparts. This research concentrates more on the effectiveness of classifier in the intended CCA formation.

As depicted in Figure 13, the accuracy factor in real brain MRI images using the CCA procedure improved up to 2%, compared to [21]. Furthermore, by estimating the precision of real images and linking the three intended purposes for classification via [21], it can be concluded that applying the CCA algorithm enhances the system performance. Additionally, the analysis of the three applied approaches to classification in this study revealed that K-means could outperform SVM and CCA and upgrade system efficiency.

4. Conclusion

It is essential that a disease be diagnosed by a physician through scanning MRI images and determining the exact location of tumors to improve the treatment process. Computerized techniques provide more accurate and comprehensive results. In this regard, this study managed to provide image processing to aid physicians in making decisions about this issue and achieved general results obtained from preprocessing. Finally, the desired features were extracted, and the tumor area was obtained. In this study, a new and efficient method for classifying tumors from images was proposed. Then, the mentioned characteristics were taken into consideration to classify tumors into two categories: benign and malignant. The proposed classification algorithm could classify tumors with accuracy of 90%, sensitivity of 100%, and specificity of 89%. Application of this method in medical centers helps doctors and radiologists increase the accuracy of diagnosis and save time.

A comparison of the results from the proposed method with those of previous methods revealed that the proposed method obtained the highest sensitivity and, as in many studies, the highest classification accuracy. Given that the criterion of susceptibility implies the ability of the classifier to correctly diagnose the disease [22,23], and that the criterion of the degree of specificity means the ability of the classifier to diagnose the absence of the disease correctly, it can be concluded that the criterion of susceptibility in the study of endocrine diagnosis medical images is of great importance. This is because correct diagnosis of the existence of a disease or cancer is more vital than diagnosing its absence in a subject. Therefore, the proposed method can be regarded as an efficient and appropriate method for tumor classification due to its very high sensitivity, the desired values of two criteria of accuracy and specificity, and small number of features used for classification. The following recommendations are provided for future studies:

1. To increase accuracy, other classifiers such as artificial neural network or KNN can be used.
2. To increase the criterion of sensitivity in this study, a combination of features extracted from two types of brain MRI images T1-w and T2-flair based on CCA algorithm was used. To increase the accuracy, all four types of images can be used along with their integrated features.
3. One of the features of brain MRI images is the wavelet transform feature. In addition to extracting the statistical features used in this study, accuracy can be augmented through wavelet transform.
4. There are different methods for reducing the dimensions of the feature matrix, which can be used by the independent component analysis.
5. One of the methods for changing the three variables

and evaluation criteria is to change the size of overlapping windows; they can change from 10×10 to 20×20, which reduces the computational complexity by as much as possible.

References

1. Ahmed, S. and Sara A. "Tumor volume fuzzification for intelligent cancer staging", *Applied Soft Computing,* **35**, pp. 227–236 (2015).
2. Hashemi, M.M., Nikfarjam, A., and Raji, H. "Novel fabrication of extremely high aspect ratio and straight nanogap and array nanogap electrodes", *Microsystem Technologies*, **25**(2), pp. 541–549 (2019).
3. Saraswathi, D., Sharmila, G., and Srinivasan, E. "An automated diagnosis system using wavelet based SFTA texture features", *IEEE, International Conference on Information Communication and Embedded Systems (ICICES)*, pp. 1–5, Chennai-India (2014).
4. Zabihi, O., Khodabandeh, A., and Mostafavi, S. M. "Preparation, optimization and thermal characterization of a novel conductive thermoset nanocomposite containing polythiophene nanoparticles using dynamic thermal analysis", *Polymer Degradation and Stability*, **97**(1), pp. 3–13 (2012).
5. Chen W., Smith R., Nabizadeh N.K., Ward, C., et al. "Texture analysis of brain CT scans for ICP prediction", *SPRINGER, Image and Signal Processing*, **6134**, pp. 568–575 (2010).
6. Pauline, J. "Brain tumor classification using wavelet and texture based neural network", *International Journal of Scientific & Engineering Research,* **3**(10), pp. 34–43 (2012).
7. Zulpe, N. and Pawar, V. "GLCM textural features for brain tumor classification", *IJCSI International Journal of Computer Science*, **9**, pp. 354–359 (2012).
8. Jain, S. "Brain cancer classification using GLCM based feature extraction in artificial neural network", *International Journal of Computer Science & Engineering Technology*, **4**(7), pp. 966–970 (2013).
9. Thamaraichelvi, B. and Yamuna, G. "Gray level co-occurrence matrix features based classification of tumor in medical images", *ARPN Journal of Engineering and Applied Sciences*, **11**(19), pp. 11403–11414 (2016).
10. Zikic, D., Ioannou, Y., Brown, M., et al. "Segmentation of brain tumor tissues with convolutional neural networks", *MICCAI Workshop on Multimodal Brain Tumor Segmentation Challenge (BRATS)*, Boston, MA, USA (2014). https://www.researchgate.net/publication/303703706
11. Li, B., Chui C., Chang S., et al. "Integrating special fuzzy clustering with level set methods for automated medical image segmentation", *Computers in Biology and Medicine*, **41**(1), pp. 1-10 (2014). DOI: 10.1016/j.compbiomed.2010.10.007
12. Korfiatis, P., Kline, T., and Erickson, B. "Automated segmentation of hyperintense regions in FLAIR MRI using deep learning", *Tomography*, **2**(4), pp. 334–40 (2016). DOI: 10.18383/j.tom.2016.00166
13. Iqbal S., Ghani M., Saba T., et al. "Brain tumor segmentation in multi-spectral MRI using Convolutional Neural Networks (CNN)", *Microscopy Research and Technique*, **81**(4), pp. 419–27 (2018). [DOI:10.1002/jemt.22994]
14. Hu, A. and Razmjooy, N. "Brain tumor diagnosis based on metaheuristics and deep learning", *International Journal of Imaging Systems and Technology*, **31**(2), pp. 657–669 (2021).
15. Hamid, M.A. and Khan, N.A. "Investigation and classification of MRI brain tumors using feature extraction technique", *Journal of Medical and Biological Engineering*, **40**(2), pp. 307–317 (2020).
16. Eshghi, S., Rajabi, H., Darvizeh, A., et al. "A simple method for geometric modelling of biological structures using image processing technique", *Scientia Iranica*, **23**(5), pp. 2194–2202 (2016). DOI: 10.24200/sci.2016.3948
17. Lidschreiber, K., Jung, L.A., von der Emde, H., et al. "Transcriptionally active enhancers in human cancer cells", *Molecular Systems Biology*, **17**(1), 9873 (2020).
18. Subhranil, K., Anup, K., Pabitra, M., et al. "Delineation and diagnosis of brain tumors from post contrast T1-weighted MR images using rough granular computing and random forest", *Applied Soft Computing*, **41**, pp. 453–465 (2016).
19. Kanniappan, S., Samiayya, D., Vincent, P.M.D.R., et al. "An efficient hybrid fuzzy-clustering driven 3D-modeling of magnetic resonance miagery for enhanced brain tumor diagnosis", *Electronics*, **9**(3), p. 475 (2020).
20. Aravindhan, S., Younus, L.A., Hadi Lafta, et al. "P53 long noncoding RNA regulatory network in cancer development", *Cell Biology International*, **45**(8), pp. 1583–1598 (2021).
21. Gupta, K.K., Dhanda, N., and Kumar, U. "A novel hybrid method for segmentation and analysis of brain MRI for tumor diagnosis", *Advances in Science, Technology and Engineering Systems Journal*, **5**(3), pp. 16–27 (2020).
22. Comlekciler, I., Gunes, S., and Irgin, C. "Three-dimensional repositioning of jaw in the orthognathic surgery using the binocular stereo vision", *Scientia Iranica* (In press). DOI: 10.24200/sci.2017.4351
23. Nikfarjam, A., Raji, H., and Hashemi, M.M. "Label-free impedance-based detection of encapsulated single cells in droplets in low cost and transparent microfluidic chip", *Journal of Bioengineering Research*, **1**(4), pp. 29-37 (2019).

Biographies

Hamed Samadi Ghoushchi is a Researcher at Urmia University of Technology. He received an MSc in

Electrical Engineering in 2018. His current research interests include image processing, deep learning, and bioelectrical engineering.

Yaghoub Pourasad is an Assistant Professor at Urmia University of Technology. He received his BSc and MSc degrees in Electrical Engineering from the K. N. Toosi University of Technology in 2002 and 2004, respectively, and his PhD degree in Electrical Engineering from the Amirkabir University in 2011. He is the author of more than 10 journal papers and five book chapters. His current research interests include image processing, signal and systems, neural networks, and Internet of things.