Received: August 14, 2019. Revised: November 20, 2019. 40

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Brain Diagnoses Detection Using Whale Optimization Algorithm Based on Ensemble Learning Classifier

Amal Fouad[1]* **Hossam M. Moftah[1]** **Hesham A. Hefny[2]**

[1]*Faculty of Computers and Information, Beni Suef University, Egypt*
[2]*Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt*
* Corresponding author's Email: amalfouad@fcis.bsu.edu.eg

Abstract: Brain cancer importance emanates from the importance of the brain as an organ and its functions. It has a great effect on the whole human body. Identification brain cancer according to its type, it refers to a multiclass classification problem in the machine learning world. In the real-world, object detection and classification face numerous challenges. The object has a large variation in appearances. In this research, a Haar Discrete Wavelet transforms hybrid with the Histogram of Oriented Gradients (HDWT-HOG) features descriptors are proposed by the local gradients in MR image as shape information. The whale optimization algorithm (WOA) plays a great role to reduce the numbers of HOG and Harr features from 38,640 to 120 features only which are less than .01% from all features. This reduction doesn't affect the system performance but it saves time in the classification phase. The test image is matched with its learned class by performing a Bagging ensemble learning classifier. Bagging achieves 96.4% in average accuracy but when Boosting is used, it achieves 95.8%.

Keywords: Features extraction, Features reduction, Whale optimization algorithm, Haar wavelet transforms, Histogram of oriented gradients, Bagging ensemble learning, Boosting ensemble learning.

1. Introduction

Cancer is a serious worldwide community health problem. According to the statistics of the world health organization for 2018, 18.1 million new cases are affected by cancer and 9.6 million people meet death [1]. The annual affected rate and death rate by cancer are rising. Cancer early detection is the best solution to improve the reliability of cancer diagnoses.

Brain tumor interrupts the brain functions by the uncontrolled division of cells. MR is a dominant and flexible modality to investigate the pathological conditions of the human brain and other body parts [2]. The MR scan is a more common practice and comfortable scan than other scans for diagnosis to identify the brain abnormalities. It is just a magnetic field and radio waves so it does not affect the human body. MR produces a high resolution of texture image with anatomical details which is used to distinguish between the diseased and healthy anatomy [3-5].

MR images have three types of visualization: T1, T2, and PD which differ in the contrast of the body tissues and three orientations, namely, corona, sagittal and axial. The sagittal orientation is taken from ear to ear where the coronal orientation is taken from the nose tip to the back of the head and axial orientation is taken from neck to head.

The task of tumor detection from MRIs are the most challenging and upcoming field, Computer-Aided Diagnose (CAD) system is an automatic tool, it can help and support Medical experts and Radiologists to decide the correct therapy at the right stage for tumor-infected. This research will try to develop a computational model to be a step towards automated cancer diagnoses.

On our proposed system is constructed on three phases. Firstly, we extract texture features based on a Haar Discrete Wavelet transforms hybrid with the Histogram of Oriented Gradients, this phase followed by the whale optimization algorithm as an optimizer algorithm to select the most efficient features, and finally, ensemble learning is used for the classification process. The great advantage of our proposed approach is that

the detection of the brain tumor type occurs with high accuracy.

This research study is organized as follows: Section 2 shows the literature survey and related works, Section 3 displays the methods and materials with the stages used in the proposed technique, Section 4 illustrates the results and the performance analysis, and finally Section 5 contains the conclusion of our work.

2. Literature review

The objective of those studies is the multiclass classification problem of brain tumors [2, 3, 4]. The features are extracted by discrete wavelet transformation (DWT) using Haar wavelet. In the next stage, PCA used to reduce the execution time of the classification process. In the last stage, The Support Vector Machine (SVM) for multiclass data is employed.

Fuzzy logic is used for brain image fusion and GLCM features are extracted from the fused brain images [4]. The adaptive neuro-fuzzy classifier helps to identify the tumor whether it is benign or malignant tumor.

In paper [6], it employs three features extraction techniques namely, Gray-Level Co-Occurrence Matrix (GLCM), Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG). The obtained features vectors of each technique are passed through a K-Nearest Neighbor (K-NN) classifier. It based on 226 images as its dataset.

In [7], it compares three classification models which are conducted using features extracted using local binary patterns, the histogram of gradients, and a pre-trained deep network. Three common image classification methods, including support vector machines, decision trees, and artificial neural networks are used to classify features vectors obtained by different features extractors. 90.52% is the highest accuracy is achieved by SVM based on local binary patterns.

In [8], a system for melanoma skin cancer is developed mainly by application of a SVM (Support Vector Machine) model on an optimized set of HOG (Histogram of Oriented Gradient) based descriptors of skin lesions. It is achieving a high-level accuracy of 97.32%.

Lung diseases such as Emphysema, Pneumonia, and Bronchitis are classified by using the support vector machine (SVM). The classification task carried out with classifier using Histogram of Oriented Gradient (HOG) –global descriptors and Local Binary Pattern (LBP) – local descriptors. The prediction performance of the lung disease achieved 98% predictive accuracy [9].

Oral images cancer detected fed to the KNN classifier after many steps [10]. Firstly, the preprocessed images with the median filtering technique are segmented by watershed. There are three techniques applied for features extraction such as GFE, HOG and LBP. The results proved that the LBP feature extraction technique achieved high accuracy which is 97 %, but HOG is 95% and GFE is 93%.

In that paper [11], the Multiclass SVM (Support Vector Machine) Classifier achieves effective results in MRI images with accuracy 94%. The results come from the segmented image to obtain the region of interest by Region growing method. Texture features (GLCM) are extracted from the segmented region. Almost 22 features are extracted in the region of a segmented area.

Two recent studies using the convolution neural network (CNN) technique, the first CNN architecture is consisting of 2 layers of convolution, it uses (ReLu) for activation function, and maxpool, followed by one hidden layer contains 64 neurons [12]. It achieves 84.19% for validation accuracy. The second study [13], it applies a pre-trained CNN model with a block-wise fine-tuning strategy, that paper achieves an average accuracy of 94.82%.

3. The proposed methodology

Our proposed methodology contains processing of MRIs of brain tumors to distinguish between three types of human brain tumors which are meningioma, glioma or pituitary. The main process is features extraction which is used to extract MRI features. Features are extracted by Harr- discrete wavelet transform and histogram of oriented gradients. Then, The Whale optimization algorithm is used in features reduction. After features reduction, ensemble machine learning is applied to the classification process.

3.1 Feature extraction

Feature extraction is the process of extracting qualitative information from an image such as texture features, color, contrast, and shape. Here, we have used The Histogram of Oriented Gradients (HOG) to extract the oriented gradient features and discrete wavelet transform (DWT) for extracting wavelet coefficients.

3.1.1. Discrete wavelet transform

Discrete wavelet transform (DWT) as expression is described term by term, The transform of a signal as a term is just another form of representing the signal, It does not change the information contents in the signal. The second term is Discrete, which is sampled signal at discrete intervals. Finally, the wavelets term comes; it was developed to overcome the shortcoming of the Short-Time Fourier Transform (STFT) in 1989 [14]. Wavelet Transform is not like (STFT), it depends on the multi-resolution technique which different frequencies are analyzed with different resolutions. It is chosen for analysis to extract the most discriminative multi-scale properties. This process is occurred by captures both location and frequency information. It captures not only a notion of the frequency content of the image, by examining it at temporal content, but also different scales. The DWTs transform an image to a set of sub-images. These sub-images describe better behavior and reveal more information about the process than the original image. DWT as a mathematical research study presents a multi-scale image processing in tree decomposition [14], Fig. 1 clarifies the tree decomposition which is called the Mallat –tree algorithm [15, 16].

DWT achieves multi-level decomposition of the pre-processed image. These features coefficients are effective features which are insensitive to arbitrary environmental variations. It is designated by using two functions; the scaling function ϕ (t) and mother wavelet (t). DWT analyzes the signal by passing it through filters with different cut-off frequencies at different scales.

$$C_{a,b}=\int_{-\infty}^{\infty} a\, x(t)\psi_{a,b}(t)\, dt \quad (1)$$

$$\psi_{a,b}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \text{ with } a\neq 0 \quad (2)$$

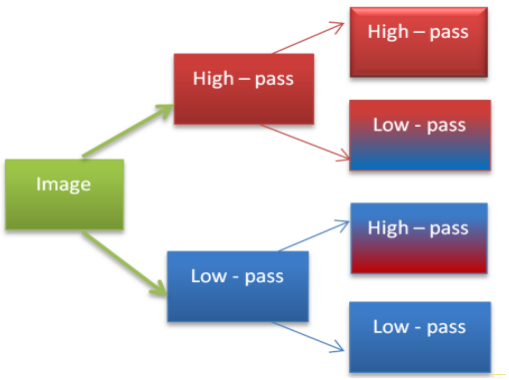


Figure. 1 Discrete wavelet transform tree decomposition

Where

$C_{a,b}$ of a signal x(t) is the wavelet coefficient, a is the scale or frequency parameter and b is the shift or time parameter.

Image is decomposed into four different frequency bands namely HH (high–high), HL (high–low), LH (low–high), LL (low–low). HH, HL, LH, LL which contain diagonal contents, vertical contents, horizontal contents, and approximate contents respectively. The approximation component (LL) is used for decomposing the image in the next level [17].

Fig.4 represents the schema of DWT decomposition that is done recursively for 3 levels of decomposition. The rows of the array are processed first with only one level of decomposition. This essentially divides the array into two vertical halves, with the first half storing the average coefficients, while the second vertical half stores the detail coefficients. This process is repeated with the columns, resulting in four sub-bands (see Fig. 1) within the array defined by filter output. Fig. 2 shows a three-level 2D DWT decomposition of the image.

The Haar-wavelet was introduced by Haar in 1910[18]. It is one of the simplest filter function. Basically, it is one period of a square wave. Because of its simplicity, it is often the wavelet to be chosen. The Haar transform decomposes a discrete signal into two sub-signals of half its length as shown in Fig. 3. One sub-signal is a running average or trend; the other sub-signal is a running difference or fluctuation.

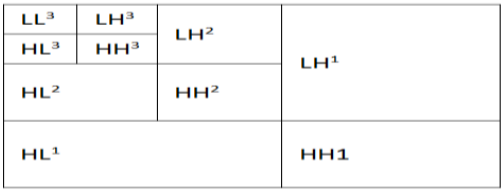


Figure. 2 DWT image decomposition scheme

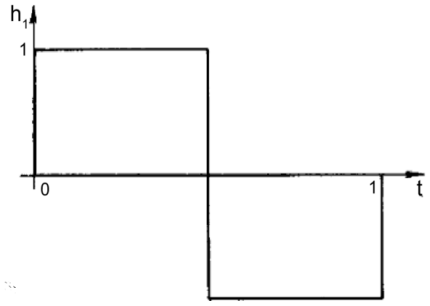


Figure. 3 Haar wavelet functions [16]

3.1.2. The histogram of oriented gradients (HOG)

Dalal and Triggs first introduced the Histogram of Oriented Gradients to recognize a person in an image in 2005 [19]. HOG are features descriptors used in computer vision and image processing for object detection purpose [20, 21]. HOG features extract the oriented gradient and it has the availability to catch the difficult edge or shape conditions. A feature descriptor intends to generalize the object in such a way that the same object (in this case a tumor brain) produces as close as possible to the same feature descriptor when viewed under different conditions. This makes the classification task easier.

HOG divides the image into small connected regions called cells. It uses a sliding detection cell or window which is moved around the image. At each position of the detector cell, a HOG descriptor is computed for the detection cell [19]. The descriptor is the concatenation of these histograms. After that, the local histograms can be contrast normalized by computing the intensity through a larger region of the image, called a block. These normalization outcomes are in higher invariance to variations in illumination and shadowing. The steps involved in the HOG features extracting process is shown in Fig. 4.

The extraction of a HOG feature vector of an image is done according to the following four steps:

Step1: Gradient computation

It uses the gradient filter [1; 0; 1] to compute the horizontal fx(x, y) and vertical fy(x, y) gradient of an image.

$$fx(x, y) = I(x + 1, y) - I(x - 1, y) \quad (3)$$

$$fy(x, y) = I(x, y + 1) - I(x, y - 1) \quad (4)$$

Where I(x, y) is the intensity value for image coordinates x and y.

The magnitude M(x, y) and angle θ(x, y) of the gradient are computed as follow

$$M(x, y) = (fx(x, y)2 + fy(x, y)2)1/2 \quad (5)$$

$$p(x, y) = \tan -1 (fy(x, y) / fx(x, y) \quad (6)$$

Step2: The orientation bins

This step divides the image into cells of q × q pixels. Then, the histogram with nine orientation bins in 0o −180o will be computed. Magnitude |M(x, y)| whose angle θ(x, y) belongs to the same bin will be added up as the value of this bin.

Step3: Descriptor block

Four connected cells combined into a block, each block contains 2×2 cells (i.e., (2q) × (2q) pixels) for each HOG feature. The main advantage of calculating the histogram on the blocks of an image is that it makes the image more robust to local variations in illumination.

Step4: Block overlapping

The HOG detector depends on the window overlap principle by default, the HOG blocks typically overlap such that each cell contributes more than once to the final feature descriptor. Adjacent neighboring blocks are overlapping by eight pixels both horizontally and vertically [9, 10, 19].

3.2 Features reduction and optimization

When features optimization techniques are used to remove any irrelevant or redundant information from machine learning data, they increase the performance of diagnosis and prognosis. In these techniques, the relevant subset and uncorrelated attributes are selected. In bioinformatics and statistical data, there are many algorithms used for selecting relevant features. The Whale optimization algorithm is selected to optimize our features in our proposed system

3.2.1. The whale optimization algorithm (WOA)

The optimization algorithm (WOA) is a nature-inspired meta-heuristic optimization algorithm to solve optimization problems by imitating the biological behavior of humpback whales. It was proposed by Mirjalili and Lewis in 2016 [22]. Besides, the use of WOA was described in different areas, such as electrical and electronics engineering, automatic control system, civil engineering, fuel energy, and medical engineering [23, 24].

It mimics the hunting mechanism of Humpback whales, the hunting is done by two main attacking mechanisms; first, by chasing the prey with random or the best search agent, second by simulating the bubble net hunting strategy So, whales swim around the target inside and alongside a thin circle to make a winding shaped way. This winding shaped way is

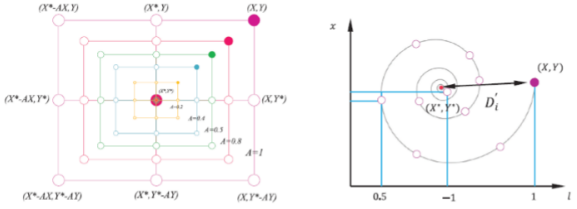


Figure. 5 Mathematical models for prey encircling and bubble-net hunting [24]

creating distinct blebs along a circle or '9' shaped ways altogether as shown in Fig. 5.

This algorithm consists of two main phases; in the first phase, encircling prey and spiral updating position are implemented (exploitation phase). However, searching for prey is done randomly in the second phase (exploration phase). The mathematical model of each phase is illustrated in the following subsections.

3.2.1.1. Encircling prey

Humpback whales can recognize the location of prey and encircle them by bubble-net feeding technique [20]. Since the start location of the optimal value in the search space is unknown yet, the WOA supposes that the best target prey or closing to the optimum current candidate solution. When the best search agent is detected, the other search agents will modify their positions in the direction of the best search agent [23, 24]. This manner is symbolized by the following Eqs. (7) and (8) [8].

$$\vec{D} = \left|\vec{C} \,.\, \vec{X}^{*}(t) - \vec{X}\,(t)\right| \tag{7}$$

$$\vec{X}\,(t+1) = \vec{X}^{*}(t) - \vec{A}.\,\vec{D} \tag{8}$$

Where the meanings of t, X, A, | |, C, X* and . are displayed in Table 1. D is the distance that calculated in (7).

The vectors A and C are computed by the following Eqs. (9) and (10) respectively [24].

$$A = 2ar - a \tag{9}$$

$$C = 2r \tag{10}$$

Where a is linearly decreased from 2 to 0 over the range of iterations (in both exploration and exploitation phases) and r is a random vector in [0, 1] according to [22, 23].

Table 1. The meanings of *t, X, A,* | |, *C, X**, and .

Symbol	Meaning
T	the present iteration
X	the position vector
A	coefficient vectors
| |	the absolute value
C	coefficient vectors
.	an element-by-element multiplication
X *	the position vector of the best solution achieved so far

Table 2. The meanings of b, l, . and $\vec{D}$

Symbol	Meaning
B	a constant for defining the shape of the logarithmic spiral
L	a random number in [−1,1]
$\vec{D}$	the distance of the ith whale to the prey (best solution obtained so far
.	an element-by-element multiplication.

3.2.1.2. Bubble-net attacking method (Exploitation phase)

Two methodologies are prepared for the Bubble-net behavior in mathematically way of the Humpback whales:

Shrinking encircling mechanism: This method is achieved by reducing the value of $\vec{a}$ in Eq. (9). In this approach, the new position of a search agent can be defined by the historical position and it achieves the best one.

Spiral updating position: the Humpback whales plunge about 12 meters then initiate bubbles in a spiral shape around the prey, then hunt them [24]. The spiral equation calculates the distance between the Humpback whale and the location prey. The following equation explains it.

$$X(t+1) = D'.ebl.cos(2\pi) + X^{*}(t) \tag{11}$$

Where $\vec{D} = |(X^{*})\ \vec{t} - \vec{X}\,(t)|$, the meanings of b, l, $\vec{D}$ and . are displayed in Table 2. D is the distance that calculated in Eq. (7).

The position of whales will be updating during the optimization process, the shrinking encircling mechanism or the spiral model will be chosen depending on the probability 50% as the algorithm assumption. The mathematical model is in Eq (12):

$$\vec{X}(t+1) = \begin{cases} \vec{X} * (t) - \vec{A}.\vec{D} & if\ p < 0.5 \\ D'.e^{bl}.Cos(2\pi l) + \vec{X} * (t) & if\ p \geq 0.5 \end{cases} \tag{12}$$

Where p is a random number in the range of [0, 1]. In addition to the bubble-net technique, the humpback whales search for prey randomly.

3.2.1.3. Search for prey (Exploration phase)

Moreover to the bubble-net method, the Humpback whales search randomly for the prey in the iteration. The mathematical model declares in the following Eqs. (13) and (14) [24].

$$D = |C.X_{rand} - X| \tag{13}$$

$$X\,(t+1)=X_{rand}-AD \quad (14)$$

Where (X) is a random position vector (a random whale), it is chosen from the current population. ($Xrand^{\rightarrow}$) is the best search agent.

3.3 Classification phase based on ensemble learning

In recent years, machine learning techniques have been widely employed in the medical domain to support decision-making [4-6, 8-11]. Moreover, medical decision support systems are in high demand to automatically detect these unexpected changes correctly and classify the brain MRI as normal or any class of disease [2-6]. CAD system requires specific conditions of the brain MRI and develops the diagnostic abilities of the medical personnel. The radiologists can use these automated systems as a tool for diagnosis, pre-surgical and post-surgical procedures.

Image classification is supervised learning which refers to the labeling of images into one of a set of predefined instances before applying a learning method. Classifier can be applied later in determining the right category of new instances. Image classification is a vital and challenging task in different application fields.

In the statistics and machine learning world, Ensemble learning was originally proposed in supervised learning for classification tasks in1965 [27]. The ensemble learning has a basic principle, the multiple base learners as ensemble members are trained altogether and combine their predictions into a single output. It improves the robustness over a single model [28, 29, 30]. Better performance is achieved by voting in classification cases or average in regression case. When component classifiers are built with uncorrelated errors, it leads to high predictive accuracy. Ensemble learning has essential benefits, which can be used when data is large volume or too little. There are two main methods applied on a sub-classifiers parallel ensemble like Bagging or sequential ensemble like Boosting. Ensemble learning's concept and its branches are illustrated in the next sub-sections.

3.3.1. Bagging based ensemble learning

It is one of the machine learning ensemble meta-algorithms construction technique which is also known as Bootstrap AGGregatING [31]. Bootstrap establishes the foundation technique to improve the stability and accuracy of machine learning algorithms used in statistical classification and

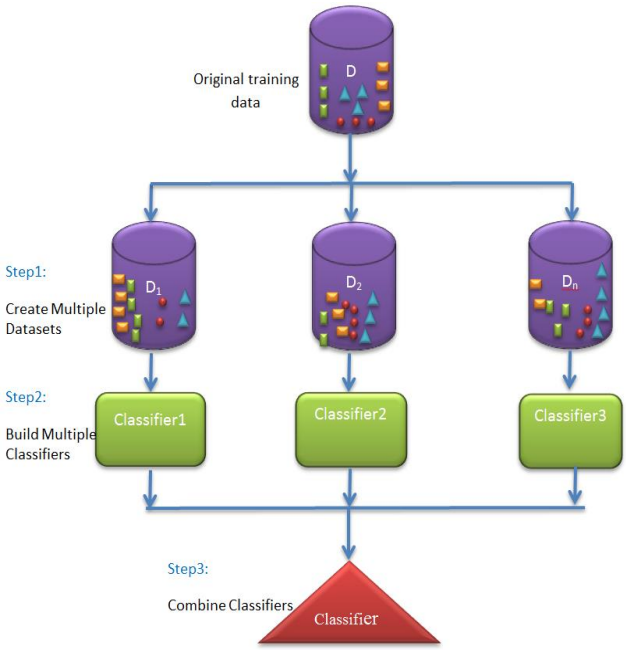


Figure. 6 Bagging ensemble learning model

regression. Fig. 6 can leap out the Bootstrap principal. Bagging model can decrease errors by reducing the variance term in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original dataset [31]. The individual predictions from each classifier are aggregated (either by voting or by averaging) to form a final single output prediction and helps to avoid overfitting. Although it is usually applied to decision tree methods, it can be used with any type of method.

Bagging Steps:

- When there are N observations and M features in the training data set. A sample from the training data set is taken.
- N observation is splitting into (N1, N2, N3, …, Nn) randomly with replacement.
- Each subset is fed to the classification sub-model in the training phase.
- In the test phase, the tested case with M features is given for each sub-model in the machine learning ensemble model then, the final output prediction is given based on the aggregation of sub-models predictions.

3.3.2. Boosting based ensemble learning

Boosting is a form of sequential ensemble learning technique. That technique works by adjusting the weight of an observation based on the last classification training [32, 33, 34, 35]. The subsequent models are constructed by fitting the

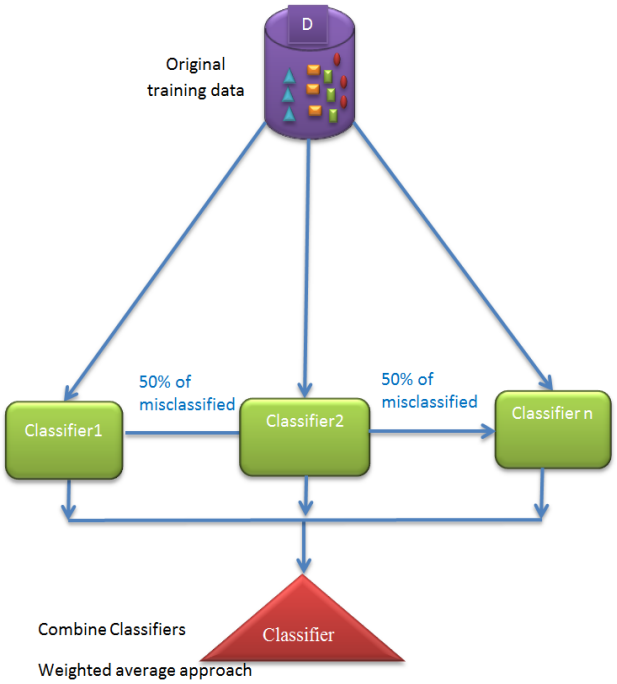


Figure. 7 Boosting ensemble learning model

residual error values of the initial model. If an observation was classified incorrectly, it tries to increase the weight of this observation. The final output is calculated by several classifier models and then the averages result is taken by a weighted average approach. Fig. 7 represents the sequential method which is tracked by the boosting model upon training data.

Boosting steps

Step1: The base learner takes all the distributions and assigns equal weight or attention to each observation.

Step 2: If there is any prediction error caused by first base learning algorithm, then we pay higher attention to observations having prediction error. Then, we apply the next base learning algorithm.

Step 3: Iterate Step 2 until the limit of the base learning algorithm is reached or higher accuracy is achieved.

Step 4: Combine all the weak learners via the majority weighted average.

4. Experimental results and discussion

The main objective of this study is to improve the accuracy of a multi-class brain MRI classifier, which is adept to detect the tumor class in brain MRIs. The proposed of multi-class classifier is to classify three different brain diseases in high detection. These tumor brain diseases include meningioma, glioma, and pituitary tumors. The proposed approach of this research technique is

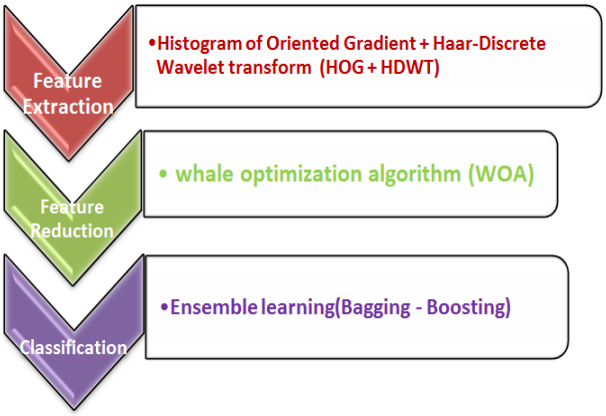


Figure. 8 the overall architecture of the proposed approach phases

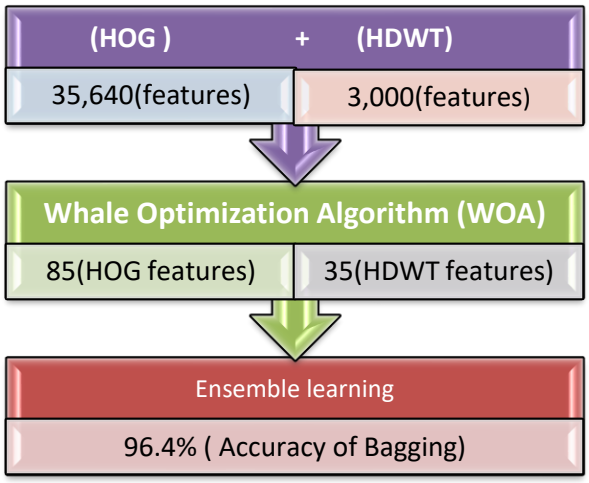


Figure .9 The architecture of the proposed approach phases in details

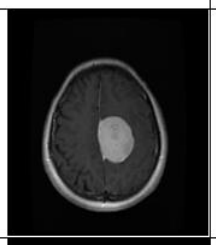
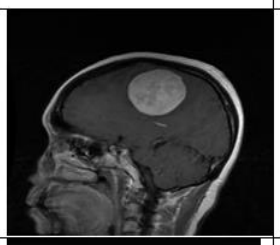
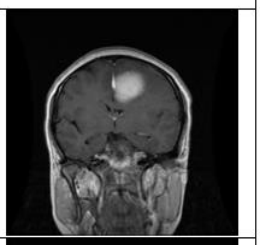
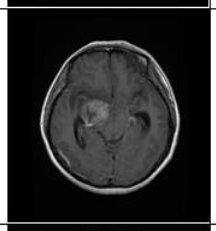
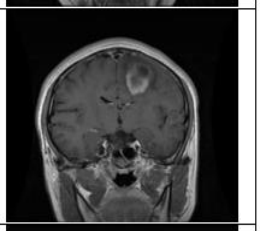
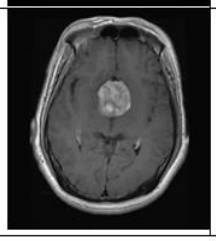
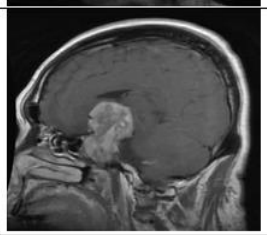
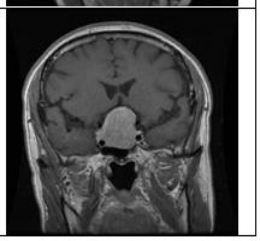
composed of three main Phases which are: (a) Features extraction phases (b) optimization Phases and (c) decision model Phases. Image feature extraction Phases uses two techniques: the first one is the Haar-Discrete Wavelet Transform (HDWT) and the second one is the Histogram of Oriented Gradient (HOG) followed by the second stage which is whale optimization algorithm (WOA).

Afterward, the reduced features are fed to the ensemble learning model at decision model Phases. The three Phases in the proposed system methodology are illustrated in Fig. 8 overall but Fig. 9 in details.

4.1 Brain image dataset

The image dataset used in this work is collected from [36]. The total number of images is 3064 with representations from three categories (meningioma, glioma and pituitary). These images are T1-weighted contrast-enhanced images from 233

Table 3. MRIs-T1 brain tumor types and orientations

T1-Images orientations / Tumor types	Axial	Coronal	Sagittal
Meningioma			
Glioma			
Pituitary			

patients with three kinds of brain tumor: meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices). The images are gray level images with intensity value ranges from (0 to 255) in Dicom form, its size 256 *256 by type int16. The Dicom form can't be used directly in Matlab but convert it to jpg form then used in this system. The brain images dataset is divided into two sets. Training dataset and testing dataset. Some samples of the MRI dataset have been displayed in Table 3.

4.2 Feature extraction phase dimensionalities

We have used HDWT and HOG features for features extraction process on the three types of tumors brain images. The first one is haar-wavelets which provide localized frequency information. This property is the main advantage of classification. In HOG features extraction as a second method is used in this study, the image is divided into cells, where each cell gathers a local 1-D histogram of edge orientations over the pixels called gradient directions and the whole object is characterized by combining these cells histograms. The local energy (histogram) is calculated over a block which is a bit greater spatial regions.

4.2.1. HDWT features

Three-level DWT is applied to the MR brain

Figure. 10 Tree complex wavelet transform on MR brain image

Images, using the Haar wavelet family which describe in Fig. 10. The transformation creates one approximate and three detailed (horizontal, vertical, and diagonal). The features are computed from the approximation and the detail sub-images at each decomposition. In This study, 3,000 features are selected from the middle to avoid the border features which are obtained from three decomposition levels.

4.2.2. HOG features

The image size is 256 x 256 pixels. The detection window will be divided into 45 blocks vertically and 22 blocks across, for a total of 990 blocks. Each block contains 4 cells with a 9-bin

histogram for each cell, for a total of 36 values per block. The final vector size is 45 blocks vertically 22 * blocks across 4* cells per block 9 *bins per histogram = 35,640 values for each image [8, 9 ,17]. The feature values can be used directly for training the classifiers, but 35,640 values for one image are huge data for manipulating or feeding to the classifier. Therefore the phase of feature optimization presents to select the efficiency features.

4.3 Features reduction and optimization phase with WOA

Features optimization is a powerful factor in enhancing the classifier capacity in the classification problem [24]. The WOA algorithm has the ability to search and select the most informative features in the features space for the classification task. WOA is considered as one of the most current optimization algorithms [25]. in the Matlab environment, The results were computed fifteen times. Overall, WOA outperformed features selection, which approved the ability finding for optimal features subset in the search space, which lead to saving time and hardware resources. the following diagram (Fig. 11) describes the steps which are followed to produce the selected features on both HDWT and HOG features.

The input data is a matrix (M, N) where M is the number of observation (training or testing) data and N is the number of features. After WOA is applied to this matrix to optimize these features, a loop is created over the entire matrix. Each column is rescaled to be in range between (0 -1), then the average is calculated when the column (feature) is increased a specific threshold, that column is chosen to be in the selected columns (features) bins.

WOA successfully decreases the feature vector which has a great effect overall the performance of the system and reduces the overall time in the classification process. When these steps are applied on HDWT, 35 features are obtained from 3000 features. 35 features represent .0117% of Haar features. This percentage is decreased to =.0025% in HOG features which means 85 features from 35,640 Hog features.

4.4 Classification phase

We applied our technique on the same dataset which was used in previous publications [12, 13].

The dataset is separated randomly into three parts. The classification training process is done by the training part, and the validation part is used to asses the classification ability. Finally, the test part was necessary for evaluating the selected features.

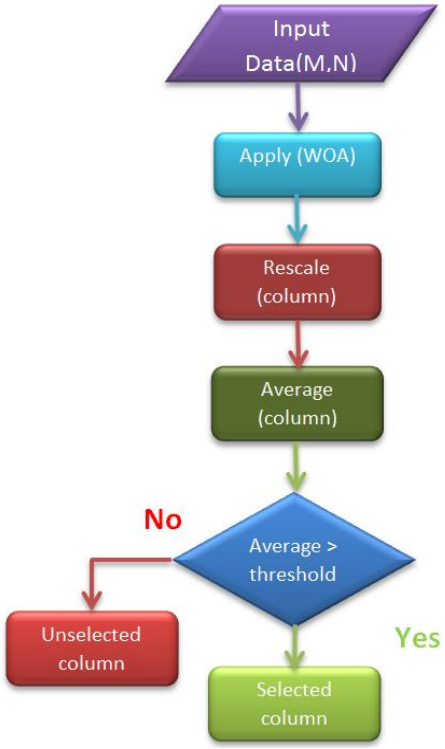


Figure. 11 Selected features steps diagram

The decision tree as a traditionally classification model is applied to our data. It didn't achieve the desired success. In Table 4, the higher accuracy achieved by the decision tree is lower than the minimum of Boosting or Bagging. Tables 4 and 7 show these results.

Table 4. Decision tree accuracy

Accuracy / Trails	HOG	HDWT	HOG+HDWT
1	92.363	65.459	92.871
2	90.332	65.459	92.363
3	90.332	65.459	91.855
4	89.825	65.459	91.855
5	89.317	64.952	91.855
6	89.317	63.937	91.855
7	89.317	62.921	90.840
8	89.317	62.921	90.332
9	88.810	62.921	89.825
10	88.302	62.921	89.825
11	88.302	62.414	89.825
12	88.302	61.906	89.825
13	88.302	61.398	89.317
14	87.794	61.398	88.810
15	87.287	61.398	88.810
Average	89.148	63.395	90.671

Table 5. Bagging accuracy

Accuracy / Trail	HOG	HDWT	HOG+ HDWT
1	95.409	96.424	96.931
2	95.409	95.916	96.931
3	95.409	95.409	96.424
4	94.901	94.901	96.424
5	94.901	94.393	96.424
6	94.393	94.393	96.424
7	94.393	93.886	95.916
8	94.393	93.886	95.409
9	94.393	93.886	95.409
10	93.886	93.886	95.409
11	93.886	93.378	95.409
12	93.886	93.378	95.409
13	93.886	93.378	95.409
14	93.886	93.378	95.409
15	93.378	93.378	95.409
Average	94.427	94.258	96.424

Table 6. Boosting accuracy

Accuracy / Trail	HOG	HDWT	HOG+HDWT
1	95.409	95.916	98.454
2	95.409	93.886	96.931
3	94.901	93.886	96.424
4	93.886	93.886	96.424
5	93.886	93.886	96.424
6	93.378	93.886	95.916
7	93.378	93.378	95.916
8	93.378	92.871	95.916
9	93.378	92.363	95.916
10	93.378	92.363	95.409
11	93.378	92.363	94.901
12	92.871	91.855	94.901
13	92.871	91.855	94.901
14	92.871	91.855	94.901
15	92.871	91.348	94.393
Average	93.683	93.040	95.849

Bagging and Boosting are ensemble learning models. They based on decision tree algorithm but in different architecture. Tables 5 and 6 observe fifteen trails upon our data.

Fig. 12 shows the accuracy of different classifiers based on different features descriptors.

HOG and HDWT are tested separately and hybrid. The hybrid features have high accuracy across all classifiers.

Table 7 shows the results of the proposed approach which is based on the bagging ensemble classifier. It shows that the performance of the proposed approach achieves better performance compared with the Decision tree and Boosting. Bagging as a voting classifier had a good efficiency than boosting which in the same classifier family. In all cases the high accuracy is achieved by Bagging except in the case of maximum, Boosting provides 98.5% in one trail over fifteen trails. According to these results, we can't conceder boosting better than Bagging.

Our hybrid proposed system based on bagging ensemble learning achieves maximum accuracy of 96.9% and an average accuracy of 96.4%. These results are comparable to the performance of deep learning [12, 13], which accuracies reached 84.39 and 94.68%, respectively on identical dataset [36].

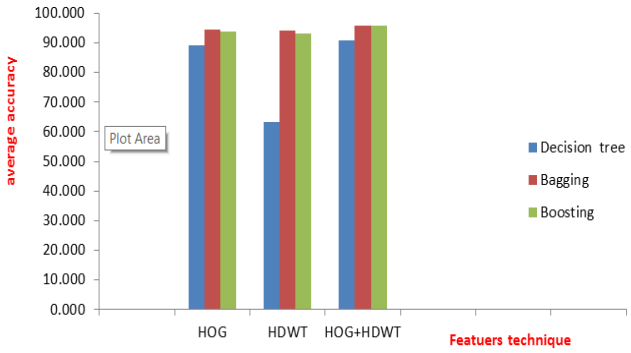


Figure. 12 Average accuracy chart for our proposed system

Table 7. Results of the proposed approach based on accuracy

Accuracy / Model	Average			Maximum			Minimum		
	HOG	HDWT	HOG+ HDWT	HOG	HDWT	HOG+ HDWT	HOG	HDWT	HOG+ HDWT
Decision tree	89.148	63.395	90.671	92.363	65.459	92.871	89.148	63.395	90.671
Bagging	94.427	94.258	96.424	95.409	96.424	96.931	93.683	93.378	95.409
Boosting	93.683	93.04	95.849	95.409	95.916	98.454	92.871	91.348	94.393

5. Conclusion

This paper focused on the detection of human brain cancer types. It distinguishes between three types which are meningioma, glioma, and pituitary. Tumor shape information based on MRI is taken from HDWT and HOG as feature extraction techniques. WOA plays a great role when it reduces the features to .01% from all features. Bagging ensemble learning can decide the brain tumor types with 96.4% average accuracy.

Our future work direction will be intended to further expand the approach to combine computer vision and deep learning to achieve more reliable results.

References

[1] World Health Organization, Web: data available at https://www.who.int/cancers/en/ accessed May 2019.

[2] E. Mohamed, M. Haggag, and A. Seddik, "Using Data Mining Techniques for Children Brain Tumors Classification based on Magnetic Resonance Imaging", *International Journal of Computer Applications*, Vol. 975, pp.8887-8893,2015.

[3] R. Sentamilselvan and M. Manikandan, "Discrete Wavelet Transform Based Brain Tumor Detection using Haar Algorithm", *International Journal of Data Mining Techniques and Applications*, Vol. 5, No. 1, pp.83-86, 2016.

[4] P. Bhosale, P. Lalge, A. Dhandekar, P. Gaykar, and P. Pate, "Brain Tumor Automated Detection and Segmentation", *International Journal of Innovative Research in Science*, Engineering and Technology, Vol. 6, No. 3, 2017.

[5] A. Pathak and R. Sunkaria, "Multiclass Brain Tumor Classification using SVM", *International Journal of Computer Applications*, Vol.97, No.23, 2014.

[6] M. Khalila, H. Ayada, and A. Adiba, "Performance evaluation of feature extraction techniques in MR-Brain image classification system", *Procedia Computer Science*, Vol. 127, pp. 218-225, 2018.

[7] T. Alhindi, S. Kalra, K. Ng, A. Afrin, and H. Tizhoosh, "Comparing LBP, HOG and Deep Features for Classification of Histopathology Images", In: *Proc. of International Conf. on Neural Networks (IJCNN)*, pp. 1-7, 2018.

[8] S. Bakheet, "An SVM Framework for Malignant Melanoma Detection Based on Optimized HOG Features", *Computation*, Vol.5, No.1, pp. 4-17, 2017.

[9] R. Pradeebha and S. Karpagavalli, "Classification of Lung Disease using Local and Global Descriptors", *International Journal of Computer Applications*, Vol. 135, No. 6, pp. 19-23, 2016.

[10] X. Stella, T. Sivagangai and N.Sujatha, "Performance Analysis of GFE, HOG and LBP Feature Extraction Techniques using KNN Classifier for Oral Cancer Detection", *Journal of Network Communications and Emerging Technologies (JNCET)*, Vol. 6, No. 7, 2016.

[11] J. Bethanney, G. Umashankar, D. Sindu, M. Shelcy, and S. Nancy, "Classification of Cervical Cancer from MRI Images using Multiclass SVM Classifier", *International Journal of Engineering & Technology*, Vol. 7, No. 2, pp. 1-5, 2018.

[12] N. Abiwinanda, M. Hanif, S. Hesaputra, A. Handayani, and T. Mengko, "Brain Tumor Classification Using Convolutional Neural Network", In: *Proc. of International Conf*. Medical Physics and Biomedical Engineering, Springer Nature Singapore Pte Ltd, Vol. 68, No. 1, 2019.

[13] Z. Nawab, K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Brain tumor classification for MR images using transfer learning and fine-tuning", *Computerized Medical Imaging and Graphics*, Vol. 75, pp. 34-46, 2019.

[14] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Transactions on Pattern Recognition and Machine Intelligence*, Vol. 11, No. 7, pp. 674-693,1989.

[15] K. Usman and K. Rajpoot, "Brain tumor classification from multi-modality MRI using wavelets and machine learning", *Pattern Analysis and Applications*, Vol. 20, No. 3, pp. 871-881,2017.

[16] A. Primer, C. Burrus, and A. Gopinath, "Introduction to wavelets and wavelet transforms", In: *Proc. of International Conf.* Prentice-Hall Upper Saddle River, 1998.

[17] K. Singh, V. Shukla, S. Tiwari, and R. Biradar, "Wavelet Based Histogram of Oriented Gradients Feature Descriptors for Classification of Partially Occluded Objects", *International Journal of Intelligent Systems and Applications*, Vol. 7, No.4, pp. 54-62, 2015.

[18] A. Haar, "Zur Theorie der orthogonalen Funktionensysteme", *Mathematische Annalen*, Springer, Vol. 69, No. 3, pp. 331-371, 1910.

[19] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", In: *Proc. of International Conf. on Computer Vision and Pattern Recognition*, 2005.

[20] T. Kobayashi, "BoF meets HOG: Feature Extraction based on Histograms of Oriented p.d.f Gradients for Image Classification", In: *Proc. of International Conf. on Computer Vision and Pattern Recognition*, pp. 747-754, 2013.

[21] Y. Iwahoria, A. Hattoria, Y. Adachi, M. Bhuyanb, R. Woodhamc, and K. Kasugaid, "Automatic Detection of Polyp Using Hessian Filter and HOG Features", In: *Proc. of International Conf. 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems - KES*, pp. 730 – 739, 2015.

[22] S. Mirjalili and A. Lewis, "The whale optimization algorithm", *Advances in Engineering Software*, Elsevier, Vol. 95, pp. 51-67, 2016.

[23] H. Hu, Y. Bai, and T. Xu, "Improved whale optimization algorithms based on inertia weights and theirs applications", *International Journal of Circuits, Systems, and Signal Processing*, Vol. 11, pp. 12-26, 2017.

[24] M. Mafarja and S. Mirjalili, "Hybrid Whale Optimization Algorithm with simulated annealing for feature selection", *Neurocomputing*, Vol. 260, pp. 302-312, 2017.

[25] R. Masadeh, A. Alzaqebah, and A. Sharieh, "Whale Optimization Algorithm for Solving the Maximum Flow Problem", *Journal of Theoretical and Applied Information Technology*, Vol. 96, 2018.

[26] E. Chong and S. Zak, "An Introduction to Optimization", 4th Edition, Wiley India, Vol. 76, 2017.

[27] N. Nilsson, "*Learning machines: Foundations of trainable pattern-classifying systems*", 1st Edition edition, McGraw-Hill, 1965.

[28] I. Barandiaran, "The random subspace method for constructing decision forests", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 8, pp. 832–844, 1998.

[29] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting", *Annals of Statistics*, Vol. 28, No. 2, pp. 337–407, 2000.

[30] T. Hastie, R. Tibshirani, and J. Friedman, "*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*", 2nd Edition, Springer series in statistics, Vol. 1, No. 10, 2001.

[31] L. Breiman, "Bagging Predictors", *Machine Learning*, Vol. 24, pp. 123–140, 1996.

[32] R. Schapire, Y. Freund, P. Bartlett, and W. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods", *Annals of Statistics*, Vol. 26, No. 5, pp. 1651–1686, 1998.

[33] J. Friedman, "Greedy function approximation: A gradient boosting machine", *Annals of Statistics*, Vol. 29, No. 5, pp. 1189–1232, 2001.

[34] M. Warmuth, J. Liao, and G. Ratsch, "Totally corrective boosting algorithms that maximize the margin", In: *Proc. of International Conf. on Machine Learning*, New York, pp. 1001–1008, 2006.

[35] C. Seiffert, T. Khoshgoftaar, J Hulse, and A. Napolitano, "RUSBoost: Improving classification performance when training data is skewed", In: *Proc. of International Conf. on Pattern Recognition*, pp. 1-4, 2008.

[36] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang, Y. Zhou, R. Yang, J. Zhao, Y. Feng, Q. Feng, and W. Chen, "Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector", *Plos One*, Vol. 11, No. 6, 2016.