



Analyzing Brain Tumor from Structural MR Images Using Kernel Support Vector Machine and Principal Component Analysis

P. Yugander¹ · M. Jagannath² · N. Nathiya³

Received: 12 December 2022 / Accepted: 30 July 2023
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2023

Abstract

Accurate and automated classification of magnetic resonance (MR) brain image is a vital task for medical image interpretation and analysis. This paper proposed a new method for brain tumor detection with the two-dimensional discrete wavelet transform (2D-DWT), kernel support vector machine (KSVM), and principal component analysis (PCA). The abnormalities of the human brain cannot be identified using traditional imaging techniques. Human brain neural architectures can be distinguished and classified using MR imaging techniques. In this paper, structural MR Images are used to classify brain tumors. The proposed algorithm is divided into three stages: pre-processing, classification, and post-processing. In pre-processing stage, 2D-DWT and PCA were employed to obtain the MR image features, and PCA is used to decrease the size of the feature vector. The KSVM is used to classify benign tumors from structural MR images. This employs three types of kernel functions: linear, polynomial, and Gaussian radial basis (GRB) kernels. In the last stage stratified K-fold cross-validation is used to avoid the overfitting problem. The proposed algorithm was performed on 80 structural MR images using three different kernel functions. Among the three kernels, the GRB kernel achieved the highest performance than linear and polynomial kernels in terms of accuracy, sensitivity, and specificity. The GRB kernel achieved an accuracy of 98.75%, sensitivity of 98%, and specificity of 99%.

Keywords Brain tumor · MR image · Kernel support vector machine · Principal component analysis

P. Yugander and N. Nathiya have contributed equally to this work.

✉ M. Jagannath
jagan.faiath@gmail.com

P. Yugander
yug460@gmail.com

N. Nathiya
nadhyan@gmail.com

¹ Department of Electronics and Communication Engineering, Kakatiya Institute of Technology and Science, Opposite Yerrakattu Hillock, Warangal, Telangana 506015, India

² School of Electronics Engineering, Vellore Institute of Technology Chennai, Vandalur to Kelambakkam Road, Chennai, Tamil Nadu 600 127, India

³ Division of Mathematics, School of Advanced Sciences, Vellore Institute of Technology Chennai, Vandalur to Kelambakkam Road, Chennai, Tamil Nadu 600 127, India

Introduction

The brain tumor is an unusual growth of cells in the brain. It is broadly classified into benign and malignant type. The benign tumors are non-cancerous tumors and malignant are cancerous tumors [1]. Basically, these tumors are analyzed according to the grading system. The grade 1 and 2 comes under benign tumors while grade 3 and 4 comes under malignant tumors. The malignant tumors are fast growing and affect the brain. These tumors spread into the different parts of the brain and spinal cord [2]. Identification of tumor position and type in the early stage is very important in medical field. A typical brain image with benign tumor is shown in Fig. 1.

Magnetic resonance (MR) imaging provides high quality brain images. MR image includes structural MRI and functional MRI. The sMRI provides anatomical structure of the brain and fMRI provide metabolic function of the brain [3, 4]. In this paper, sMRI is used to identify the brain tumors. Brain tumor detection approach includes three stages: preprocessing, classification and post processing.

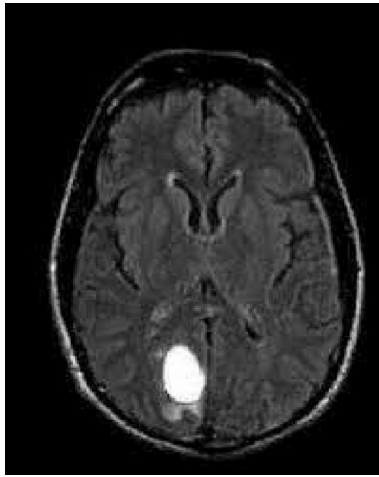


Fig. 1 Typical MR brain image with benign tumor

2D-DWT and PCA is used in preprocessing stage. Wavelet transform plays a vital role in MR image feature extraction. MR images can be analyzed at different levels of resolution using wavelet transform [5]. The features of the MR images can be extracted using 2D-DWT. But, this technique requires large storage [6]. The principal component analysis is used to eliminate the above limitation. PCA efficiently reduces the dimensions of the feature vector and provides most important features to classifier [7].

Most of the researchers used machine learning algorithms for detecting brain tumors from MR images. Machine learning algorithms like support vector machine (SVM), decision trees (DT), random forest (RF), Naive Bayes (NB), and hidden Markov models (HMM) play major roles in biomedical image processing. SVM is a popular linear classifier in machine learning. It provides higher accuracy with less data [8]. If the dataset contains less data, it classifies the images efficiently [9]. But it fails to classify images, if the dataset is too large. Moreover, the SVM classifier performs well for linearly separable data. Most of the tumor features like intensity, shape, and texture are nonlinear. In this paper, the proposed kernel SVM (KSVM) classifies the nonlinear MR image features by mapping extracted features into a multi-dimensional space. This mapping feature allows KSVM to obtain a hyperplane that extracts tumor features. Also, it integrates multiple extracted features into a single classification framework. It uses various kernels like linear, polynomial, and GBR kernels. These kernels are used to obtain the nonlinear relationship between the MR image features. Unlike SVM, the proposed KSVM is robust to noise. In order to overcome overfitting, stratified K-fold cross-validation is used in KSVM [10].

This paper is organized in 7 sections. The two-dimensional discrete wavelet transform is presented in Sect. "Discrete Wavelet Transform". The dimensionality reduction using

PCA is introduced in Sect. "Principal Component Analysis". The kernel support vector machine classification is interpreted in Sect. "Kernel Support Vector Machine". Section "K-Fold Cross-validation" presents Stratified K-fold cross validation. The results are discussed in Sect. "Results and Discussion". Section "Conclusion" concludes the study with quantitative analysis.

Discrete Wavelet Transform

Feature extraction is procedure of extracting statistical data from MR image like edges, contrast, region of interests, color features, shape, ridges, contrast, texture, etc. [11]. In this paper, DWT is used to obtain the qualitative features like gray-level co-occurrence matrix and wavelet coefficients. In general, it is used to convert spatial domain image into frequency domain. It provides Frequency information along with time information. MR image wavelet coefficients can be extracted using DWT. The primary advantage of DWT is that it gives high temporal resolution for high frequency components and good frequency resolutions for low frequency components. DWT extracts the MR image features using analysis filter banks and decomposition operation. It contains a high pass and low pass filters for every decomposition level. The high pass and low pass filter give the detail and approximation coefficients of a MR image respectively. The 2D-DWT is derived from two separate 1D DWT [12]. In multi-resolution analysis, 2D-DWT decomposition MR image decomposed into two levels. In every decomposition level DWT obtains four sub bands Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH) respectively. The LL subband contains approximation coefficients, where LH, HL, and HH sub-bands contains detailed coefficients. The accurate approximate coefficients can be obtained by extending the number of decomposition levels.

$$\xi(k, l) = \xi(k)\xi(l) \quad (1)$$

$$\psi^H(k, l) = \psi(k)\xi(l) \quad (2)$$

$$\psi^V(k, l) = \xi(k)\psi(l) \quad (3)$$

$$\psi^D(p, q) = \psi(p)\psi(q) \quad (4)$$

$\xi(\cdot)$ and $\psi(\cdot)$ are 1-D scaling function and wavelet function, respectively. ψ^H obtains variation along the columns, ψ^V obtains variation along the rows, and ψ^D obtains variation along the diagonals (refer Eqs. (1–4)) [13]. Equations (5, 6) show the 2D-DWT scaled and translating basis functions.

$$\xi_{j,m,n}(k, l) = (2)^{\frac{b}{2}}(2^b k - p, 2^b l - q) \tag{5}$$

$$\psi_{i,m,n}^j(k, l) = (2)^{\frac{b}{2}}(2^b k - p, 2^b l - q) \tag{6}$$

Where k, l are the translation quantites, p, q are the spatial quantities, and j is the scale. The transformed image of $f(k, l)$ of size $P \times Q$ is represented in Eqs. (7, 8)

$$Z_{\phi}(j_0, p, q) = \frac{1}{\sqrt{(PQ)}} \sum_{i=0}^{a-1} \sum_{j=0}^{b-1} f(k, l) \xi_{j_0,m,n}(k, l) \tag{7}$$

$$Z_{\psi}(j, p, q) = \frac{1}{\sqrt{(PQ)}} \sum_{x=0}^{p-1} \sum_{y=0}^{q-1} f(k, l) \psi_{j_0,m,n}^i(k, l) \tag{8}$$

Where $Z_{\phi}(j_0, p, q)$ are the approximate coefficients, $Z_{\psi}(j, p, q)$ are the detailed coefficients, and j_0 is an arbitrary scale.

Principal Component Analysis

Excessive features increase the computation complexity. Also, excessive features lead to curse of dimensionality. Number of features must be reduced to avoid the above stated problems [14]. The dimensionality reduction can be done using principal component analysis (PCA), linear discriminate analysis (LDA), and multifractal detrended fluctuation analysis (MFDFA). Among the above algorithms PCA is an efficient and powerful tool. It is a popular non-parametric method for dimension reduction. PCA is an unsupervised learning algorithm and it does not require any knowledge about the data. But, LDA requires labelled data for each data point. LDA can be applied to multivariate normal distribution data only. Also, it assumes that the covariance matrices of all the labels are equal. MFDFA reduces the data by calculating the fitting power-law function and fluctuation function of the input data. These calculations required a lot of time compared to PCA. Moreover, MFDFA is noise sensitive. Due to these limitations, LDA and MFDFA are not used in data reduction. Dimensionality reduction is achieved by transforming the extracted features into a new reduced dataset according to the variance of the input data [15]. Because of using variance parameter, it extracts the most of the relevant information from confusing data sets. The input data must be normalized with unity variance and zero mean before performing dimensionality reduction. After performing PCA the below mentioned features were formed [16].

Contrast (C): The contrast measurement of MR image given by Eq. (9).

$$C = \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} (k - l)^2 f(k, l) \tag{9}$$

Energy (E): The energy of the image is given in Eq. (10).

$$E = \sqrt{\sum_{i=0}^{p-1} \sum_{j=0}^{q-1} f^2(k, l)} \tag{10}$$

Correlation (COR): The correlation finds the spatial features dependencies between the neighbor pixels. It is given in Eq. (11).

$$\gamma_{fg} = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (f_{min} - \bar{f})(g_{min} - \bar{g})}{\sqrt{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (f_{min} - \bar{f})^2 \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (g_{min} - \bar{g})^2}} \tag{11}$$

Where f_{min} is the minimum pixel in the image $f(x, y)$, g_{min} is the minimum pixel in the image $g(x, y)$, \bar{f} and \bar{g} are the mean of the input and output images respectively.

Homogeneity (HOM): The homogeneity is a measurement of local uniformity in MR image and shown in Eq. (12).

$$H = \sum_{k=0}^{P-1} \sum_{l=0}^{Q-1} \frac{1}{1 + (k - l)^2} f(k, l) \tag{12}$$

Entropy (E): The entropy is used to calculate the designated interference of the MR image and as shown in Eq. (13).

$$ENT = \sum_{i=0}^{P-1} \sum_{j=0}^{Q-1} f(k, l) \log_2 f(k, l) \tag{13}$$

These extracted qualitative features were given to the Kernel SVM for classification of brain tumor images.

Kernel Support Vector Machine

Brain tumor classification using machine learning algorithms such as support vector machine (SVM), decision trees (DT), random forest (RF), Naive Bayes (NB), and hidden Markov models (HMM) plays a major role in biomedical image processing. In general, MR images have a greater number of image features or voxels, resulting in high-dimensional data. In general, the RF algorithm fails to classify the high-dimensional data. HMMs are mostly used in the analysis of temporal or sequential data such as time-series or brain activity patterns. It is more suitable for functional MR images rather than structural MR images. In this paper, structural MR images were used for identifying brain tumors. The tumor’s presence in the brain is decided by the imaging features like shape, intensity, and texture. These image features are complex and nonlinear.

The nonlinear relationship between input features and class labels can be effectively handled by SVM. Among all the improved SVM algorithms, kernel SVM (KSVM) is most effective and popular algorithm. KSVM is mostly used in biomedical image processing, bioinformatics and natural language categorization. It provides global and unique solution for convex quadratic optimization using tunable parameters [17]. The flowchart of the proposed algorithm is shown in Fig. 2.

In general, SVM uses hyperplane to classify input data. Traditional SVM algorithm fails to classify the data, if the data is located in different locations of the hypersurface. In the above case kernel strategy is used with SVM to classify the data. KSVM is obtained by replacing dot product with a nonlinear kernel function [18]. The kernel function is given in Eq. (14).

$$k(p_i, p_j) = \varepsilon(p_i)\varepsilon(p_j) \quad (14)$$

Where p_i is an n-dimensional vector, $\varepsilon(p_i)$ and $varepsilon(p_j)$ are the mapping function and penalty function, respectively.

Transformed space w is given in Eq. (15).

$$w = \sum_i \eta_i \lambda_i \xi(x_i) \quad (15)$$

Where η_i is a regularization parameter, λ_i is a step length, and $\xi(x_i)$ is a penalty coefficient.

Dot product is represented in Eq. (16).

$$\varepsilon(x) = \eta_i \lambda_i k(p_i, p) \quad (16)$$

In a transformed space KSVM provides maximum-margin hyperplane compared to SVM. The KSVM transformation is nonlinear and input data is transformed to higher dimensional space [19].

KSVM uses:

Linear kernel: It is represented in Eq. (17).

$$k(p_i, p_j) = p_i^T p_j \quad (17)$$

Polynomial kernel: It is represented in Eq. (18).

$$k(p_i, p_j) = (\alpha p_i^T p_j + a)^b \quad (18)$$

Where a, b are constants and α is an error term associated with kernel function.

Gaussian radial basis kernel: It is used when there is no prior knowledge about the input data. It is represented in Eq. (19).

$$k(p_i, p_j) = e^{-\left(\frac{\|p_i - p_j\|^2}{2\sigma^2}\right)} \quad (19)$$

Where $\|p_i - p_j\|$ is the Euclidean distance between p_i and p_j , and σ^2 is the variance. The kernel parameters must be adjusted before the data training. The training data base has abnormal and normal images.

K-Fold Cross-validation

The common problem in ML classifier algorithm is overfitting. In order to eliminate the above problem cross validation is introduced. Cross-validation methods are broadly classified into three types: leave-one-out validation, k -fold cross validation (KFCV) and Random sub-sampling. Second method is widely used because of its simplicity. In KFCV method data is dividing into k subsets. Out of k subsets $k - 1$ folds used for train the input data and one-fold is used for test the input data [20]. Also, calculates the average of errors for k folds.

In k -fold cross validation method all folds cannot have equal samples. The above limitation can be eliminated by using stratified k -fold cross validation (SKFCV). Selection of parameter k plays crucial role in SKFCV. If k is large, the error estimator variance is high and the bias is low. If k is small, the error estimator variance is low and the bias is high. Also, k -value is proportional to the computational time. In these paper k -values randomly selected as five. After number of iterations, optimum k -value can be chosen according to highest classification accuracy.

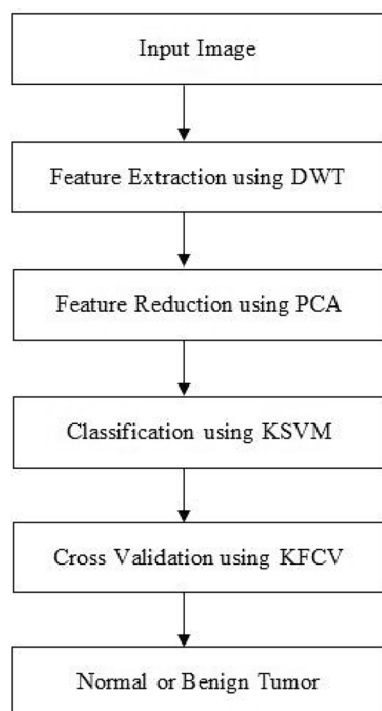


Fig. 2 Workflow of the proposed algorithm

Table 1 Details of training and validation dataset

Total input image	Training (n=64)		Validation (n=16)	
	Normal	Benign Tumor	Normal	Benign Tumor
80	8	56	2	14

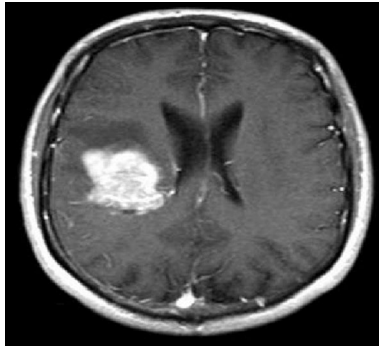


Fig. 3 Input MR image with tumor

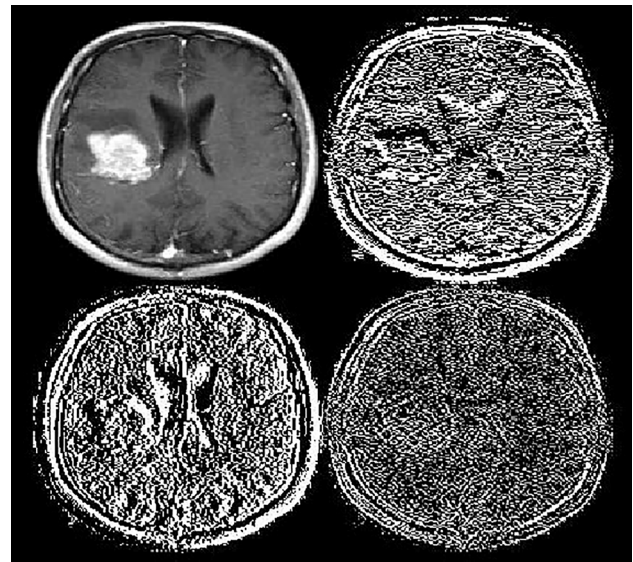


Fig. 4 First level DWT decomposition from MR image with tumor

Results and Discussion

The proposed algorithm tested using intel i7 processor with 8GHz processor, 2GB RAM and Windows 10 operating system. The programs can be run on MATLAB 2022a.

Dataset: The dataset used is downloaded from Harvard Medical School, US. The evaluation dataset contains 80 MRIs, including 70 abnormal and 10 normal images. Dataset details are shown in Table 1.

In the pre-processing, the DWT is applied to extract the features like correlation, homogeneity, entropy, and skewness etc. These features were extracted by performing single level DWT decomposition. Two-dimensional DWT decomposes the MR image into four individual sub bands: Approximate, horizontal, vertical and diagonal coefficients. The input MR image with tumor and all the four coefficients of benign tumor MR image are shown in Figs. 3 and 4, respectively. The input MR image without tumor and the first level decomposed components of No tumor image are shown in Figs. 5 and 6, respectively.

The features extracted using 2D-DWT are reduced using PCA. The 2D-DWT extracted 65536 features from input MR images. By using PCA these 65536 features were reduced to 1024. The confusion matrix for different kernels is shown in Table 2 and Dimensionality reduction details are depicted in Table 3.

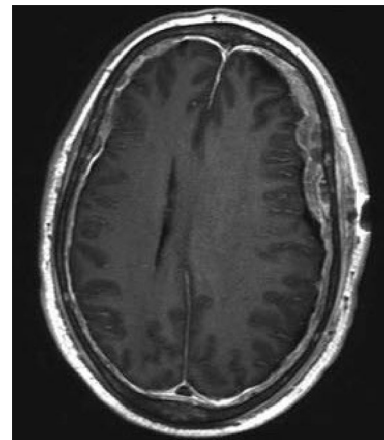


Fig. 5 Input MR image without tumor

The following parameters are used as indices to check the performance and validation of the proposed algorithm and are depicted in Eqs. (20–22).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{20}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{21}$$

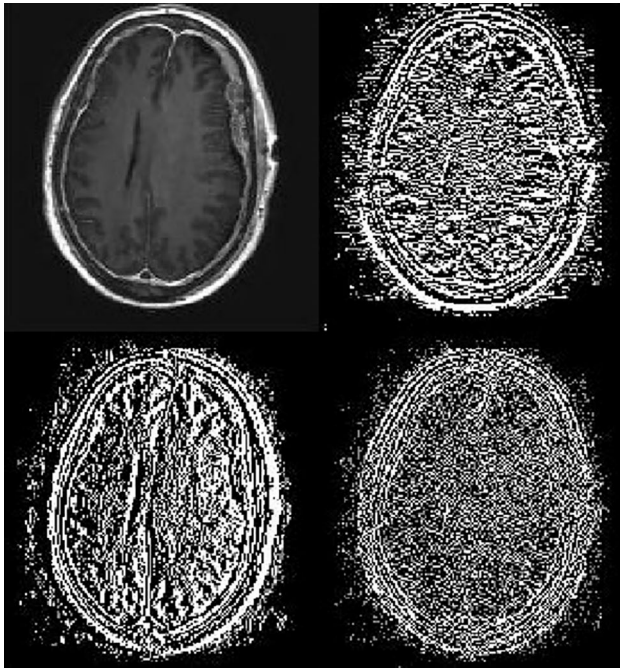


Fig. 6 First level DWT decomposition from MR image without tumor

Table 2 Confusion matrix for different kernels

Linear	Normal (O)	Benign tumor (O)
Normal (T)	8	2
Benign tumor (T)	3	67
Polynomial	Normal (O)	Benign tumor (O)
Normal (T)	9	1
Benign tumor (T)	2	68
Gaussian Radial Basis	Normal (O)	Benign tumor (O)
Normal (T)	10	0
Benign tumor (T)	1	69

$$\text{Specificity} = \frac{TN}{TP + TN} \quad (22)$$

Where TP represents True positive (TP), TN represents True negative, FP represents False-positive and FN represents False-negative. The performance analysis of the chosen algorithm with different kernels is shown in Table 4.

The proposed algorithm obtained good results for training and validation images. Among the three algorithms,

Table 3 Details of dimensionality reduction

Principal component	1	2	3	4	5	6	7
Variance (%)	36.5	43.2	51.8	53.7	60.7	67.6	70.1
Principal component	8	9	10	11	12	13	14
Variance (%)	76.9	79.8	85.5	90.1	92.3	95.4	97.2

Table 4 Performance analysis of the kernel support vector machine with different kernels

Kernel type	Sensitivity (%)	Specificity (%)	Accuracy (%)
Linear	87	63	93.75
Polynomial	96	87	96.25
Gaussian radial basis	99	98	98.75

Table 5 Performance analysis of the proposed algorithm with various classification algorithms

Classification Algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
RF	86.1	83.1	89.6
HMM	67.7	70.2	74.3
SVM	81.6	86.5	93.1
K SVM (GRB kernel)	99	98	98.75

GRB kernel achieved highest sensitivity and specificity. For linear kernel, the classification accuracy was 93.75%; for polynomial kernel, accuracy was 96.25%; and for the GRB kernel, accuracy was 98.75%. The GRB kernel performed well compared to other two kernels.

The proposed algorithm is compared with RF, HMM, and SVM. It obtained good results for training and validation images [21]. Among the four algorithms, K SVM with GRB kernel outperformed with sensitivity of 99%, specificity of 98%, and accuracy of 98.75%. The comparison of various algorithms with the proposed algorithm is shown in Table 5.

Conclusion

Classification of a benign tumor from sMRI using computer aided diagnosis system remains a challenge. The proposed algorithm successfully classified the tumor images from the dataset. The developed algorithm employs DWT, PCA and K SVM. DWT is used to extracted the statistical features of MR images and PCA is reduced the dimensions of the extracted features. K SVM is used to classify the abnormal MR images. Dot product in SVM is replaced by kernels in K SVM. Three kernels were used to identify the highest classification accuracy. The performance analysis showed

that among the three algorithms, GRB kernel achieved 98% accuracy. The limitation of the proposed algorithm is processing time. KSVM processing time is high compare to SVM. Feature work will focus on reducing processing time.

Funding This study was not funded by any Government or private organization.

Data availability The data that support the findings of this study are downloaded from Harvard Medical School, US.

Declaration

Conflict of interest All authors declare that there is no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Tsui PH, Liao YY, Chang CC, Kuo WH, Chang KJ, Yeh CK. Classification of benign and malignant breast tumors by 2-d analysis based on contour description and scatterer characterization. *IEEE Trans Med Imaging*. 2010;29(2):513–22.
2. Lemay A, Gros C, Zhuo Z, Zhang J, Duan Y, Cohen-Adad J, Liu Y. Automatic multiclass intramedullary spinal cord tumor segmentation on MRI with deep learning. *NeuroImage: Clin*. 2021;31:1–9.
3. Peng B, Wu L, Chen Y. Volumetric changes in amygdala and entorhinal cortex and their relation to memory impairment in patients with medical temporal lobe epilepsy with visually normal MR imaging findings. *Epilepsy Res*. 2015;114(2015):66–72.
4. Matthews PM, Clare S, Adcock J. Functional magnetic resonance imaging: clinical applications and potential. *J Inher Metab Dis*. 1999;22(1999):337–52.
5. Amin J, Sharif M, Gul N, Yasmin M, Ali-Shad S. Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network. *Pattern Recognit Lett*. 2020;129:115–22.
6. Nayak DR, Ratnakar D, Banshidhar M. Brain MR image classification using two-dimensional discrete wavelet transform and Ada-Boost with random forests. *Neurocomputing*. 2016;177:188–97.
7. Uddin MP, Mamun MA, Hossain MA. PCA-based feature reduction for hyperspectral remote sensing image classification. *IETE Tech Rev*. 2021;38(4):377–96.
8. Salvatore C, Battista P, Castiglioni I. Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines. *Curr Alzheimer Res*. 2016;13(5):509–33.
9. Naghash-Almasi O, Rouhani M. A geometric-based data reduction approach for large low dimensional datasets: delaunay triangulation in SVM algorithms. *Mach Learn Appl*. 2021;4:509–33.
10. Amari S, Wu S. Improving support vector machine classifiers by modifying kernel functions. *Neural Netw*. 1999;12(6):783–9.
11. Rinky BP, Payal M, Manikantan K, Ramachandran S. DWT based feature extraction using edge tracked scale normalization for enhanced face recognition. *Procedia Technol*. 2012;6:344–53.
12. Zang Y, Wang S, Huo Y, Wu L, Liu A. Feature extraction of brain MRI by stationary wavelet transform and its applications. *J Biol Syst*. 2010;18(1):115–32.
13. Zheng-Hai H, Li J, Shang J, Wang J, Zhang T. Non-uniform patch based face recognition via 2D-DWT. *Image Vis Comput*. 2015;37:12–9.
14. Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. *Neuroinformatics*. 2014;12(2):229–44.
15. Xie H, Li J, Zhang Q, Wang Y. Comparison among dimensionality reduction techniques based on Random Projection for cancer classification. *Comput Biol Chem*. 2016;65:165–72.
16. Shree NV, Kumar TNR. Identification and classification of brain tumor MR images with feature extraction using DWT and probabilistic neural networks. *Neuroinformatics*. 2014;12(2):229–44.
17. Zang Y, Wu L. An MR brain image classification via principal component analysis and kernel support vector machine. *Prog Electromagn Res*. 2012;130:369–88.
18. Tharwat A. Parameter investigation of support vector machine classifier with kernel functions, Knowledge and information. *System*. 2019;61:1269–302.
19. Wang W, Xu Z, Lu W, Zhang X. Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing*. 2003;55(3):643–63.
20. Wong T, Yeh P. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans Knowl Data Eng*. 2019;32(8):1586–94.
21. Mukhopadhyay S, Das NK, Kurmi I, Pradhan A, Ghosh N, Panigrahi PK. Tissue multifractality and hidden Markov model based integrated framework for optimum precancer detection. *J Biomed Optics*. 2017;22(10):1–8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.