Original Article

Analysis of Hybrid Feature Optimization Techniques Based on the Classification Accuracy of Brain Tumor Regions Using Machine Learning and Further Evaluation Based on the Institute Test Data

Soniya Pal[1,2], Raj Pal Singh[1], Anuj Kumar[3]

[1]Department of Physics, GLA University, Mathura, [3]Department of Radiotherapy, S. N. Medical College, Agra, Uttar Pradesh, [2]Batra Hospital and Medical Research Center, New Delhi, India

Abstract

Aim: The goal of this study was to get optimal brain tumor features from magnetic resonance imaging (MRI) images and classify them based on the three groups of the tumor region: Peritumoral edema, enhancing-core, and necrotic tumor core, using machine learning classification models. **Materials and Methods:** This study's dataset was obtained from the multimodal brain tumor segmentation challenge. A total of 599 brain MRI studies were employed, all in neuroimaging informatics technology initiative format. The dataset was divided into training, validation, and testing subsets online test dataset (OTD). The dataset includes four types of MRI series, which were combined together and processed for intensity normalization using contrast limited adaptive histogram equalization methodology. To extract radiomics features, a python-based library called pyRadiomics was employed. Particle-swarm optimization (PSO) with varying inertia weights was used for feature optimization. Inertia weight with a linearly decreasing strategy (W1), inertia weight with a nonlinear coefficient decreasing strategy (W2), and inertia weight with a logarithmic strategy (W3) were different strategies used to vary the inertia weight for feature optimization in PSO. These selected features were further optimized using the principal component analysis (PCA) method to further reducing the dimensionality and removing the noise and improve the performance and efficiency of subsequent algorithms. Support vector machine (SVM), light gradient boosting (LGB), and extreme gradient boosting (XGB) machine learning classification algorithms were utilized for the classification of images into different tumor regions using optimized features. The proposed method was also tested on institute test data (ITD) for a total of 30 patient images. **Results:** For OTD test dataset, the classification accuracy of SVM was 0.989, for the LGB model (LGBM) was 0.992, and for the XGB model (XGBM) was 0.994, using the varying inertia weight-PSO optimization method and the classification accuracy of SVM was 0.996 for the LGBM was 0.998, and for the XGBM was 0.994, using PSO and PCA-a hybrid optimization technique. For ITD test dataset, the classification accuracy of SVM was 0.994 for the LGBM was 0.993, and for the XGBM was 0.997, using the hybrid optimization technique. **Conclusion:** The results suggest that the proposed method can be used to classify a brain tumor as used in this study to classify the tumor region into three groups: Peritumoral edema, enhancing-core, and necrotic tumor core. This was done by extracting the different features of the tumor, such as its shape, grey level, gray-level co-occurrence matrix, etc., and then choosing the best features using hybrid optimal feature selection techniques. This was done without much human expertise and in much less time than it would take a person.

Keywords: Hybrid optimal feature selection method, machine learning, support vector machine, tumor region classification, extreme gradient boosting model

Received on: 15-06-2023 **Review completed on: 23-02-2024** **Accepted on: 23-02-2024** **Published on: 30-03-2024**

Introduction

The brain is a complex organ made up of billions of nerve cells that control the entire nervous system of the body, and any problems with it can deteriorate a person's health as a whole.[1,2] Brain tumors are very common, and both young and old people

Access this article online

Quick Response Code:

Website: www.jmp.org.in

DOI: 10.4103/jmp.jmp_77_23

Address for correspondence: Dr. Anuj Kumar, Department of Radiotherapy, S. N. Medical College, Agra - 282 002, Uttar Pradesh, India.
E-mail: toaktyagi@gmail.com

How to cite this article: Pal S, Singh RP, Kumar A. Analysis of hybrid feature optimization techniques based on the classification accuracy of brain tumor regions using machine learning and further evaluation based on the institute test data. J Med Phys 2024;49:22-32.

can have them.[3] The first step in treating this kind of disease is to investigate the type and extent of the tumor. There are two types of brain tumors: Cancerous and noncancerous. Cancerous brain tumors cannot be cured, but noncancerous brain tumors almost never cause problems that are life-threatening if they are found early. The enhancing tumor core (ETC), the peritumoral edema (ED), and the necrotic/non-ETC (NCR/NETC) are some of the tumor subregions. Gliomas and meningiomas are two of the most dangerous types of primary tumors,[4] because they are life-threatening, if found too late. Images from a brain magnetic resonance imaging (MRI) are mostly used to find brain tumors. Medical image processing algorithms, deep learning algorithms, and machine learning algorithms can all be used to make accurate disease classification and prediction.[5,6]

Intensity normalization[7] plays a pivotal role in the domain of medical image analysis, serving as a fundamental preprocessing step that addresses inherent variations in image intensities. Medical images acquired from diverse sources and modalities often exhibit substantial inconsistencies, resulting from dissimilar imaging protocols, equipment variability, and patient-specific factors.[8]

Feature extraction constitutes a pivotal stage in medical image analysis, encompassing a sophisticated process that delineates informative patterns and discriminative characteristics from complex medical images.[9] With the advent of advanced imaging modalities and the proliferation of high-dimensional image data, extracting robust and representative features is imperative for effective analysis, diagnosis, and prognosis in the field of medical image analysis.[10] Feature extraction involves the transformation of raw pixel intensities into a compact and expressive representation, capturing salient visual cues and intrinsic structural attributes. By discerning discriminative patterns, textures, shapes, and spatial relationships, this procedure enables the extraction of essential image descriptors that encapsulate relevant anatomical and pathological information.[11-13] The process of feature extraction is characterized by a multitude of advanced methodologies, including but not limited to statistical descriptors, morphological operations, spatial and frequency domain transformations, local binary patterns, texture analysis, wavelet analysis, and deep learning-based approaches.[14]

Feature optimization is also a critical aspect of medical image analysis, playing a pivotal role in refining and enhancing the discriminative power of extracted features.[15,16] This process involves a systematic exploration and fine-tuning of feature representations to maximize their utility, relevance, and discriminatory capacity. The overarching objective is to identify a subset of features that are most pertinent to the underlying medical problem, while minimizing redundancy, noise, and irrelevant information. Wrapper, filter, and embedded methods enable efficient exploration of the feature space, facilitating the identification of discriminative features that significantly contribute to accurate diagnosis and prognosis. To address the curse of dimensionality inherent in medical image data, dimensionality reduction techniques are employed. Principal component analysis (PCA),[17,18] linear discriminant analysis,[19,20] and manifold learning methods transform the original high-dimensional feature space into a lower-dimensional subspace while preserving the discriminative information. Feature ranking algorithms assign scores or ranks to each feature based on their relevance and discriminative power. Statistical measures, information-theoretic approaches, and machine learning-based techniques are utilized to estimate feature importance.

Therefore, we put forward this approach to tumor region classification using machine learning methods. This paper aims to develop a mechanized strategy for facilitating the localization of distinct tumor regions. To achieve this goal, features were extracted from MRI images after applying intensity normalization. To get the most accurate results, we tweaked the extracted features using hybrid optimal feature selection techniques, the particle swarm optimization (PSO) combined with PCA feature optimization technique. We also used a couple of different classification algorithms, as support vector machine (SVM),[21,22] light gradient boosting model (LGBM),[23] and extreme gradient boosting model (XGBM),[23] for the classification task.

Materials and Methods

The survival rate of brain tumors varies depending on the type and extent of the tumor. As a result, categorizing these regions is critical. Machine learning has enormous potential for this task and can efficiently classify different tumor regions with high accuracy in a short amount of time.[24,25] However, training a model can take a long time and a lot of resources. Therefore, optimal feature selection[26] is a great technique for reducing the time required to train the model with limited resources. In this paper, we used hybrid optimization methods for the classification of brain tumor regions. For this, we used two different datasets, online test dataset (OTD) and (ITD) institute test data, which were intensity normalized using contrast limited adaptive histogram equalization (CLAHE) to enhance the visual features of the image. The features were extracted using the pyRadiomics library and optimized using the hybrid optimization technique using varying inertia-weight PSO and PCA. Finally, three different classification models, SVM, LGBM, and XGBM, were used to classify the three tumor regions to evaluate the effect of optimization techniques.

SVM[27] is very less prone to over fitting, especially effective in high-dimensional spaces. It has versatile kernal functions such as linear, polynomial, and radial basis function, enabling flexibility in modeling complex relationships. At the same time, it is very sensitive to noisy data, and outliers can have a significant impact on the model and is computationally expensive, especially with large datasets. On the other hand,

LGBM is designed to be efficient and can handle large datasets. It uses a histogram-based approach for splitting nodes, making it faster than other algorithms. LGBM[28] supports parallel and distributed training, making it suitable for scalable and efficient training on large datasets. However, with this efficiency, it has limited interpretability and requires tuning parameters for optimal performance. Extreme gradient boosting (XGB)[29] has a built-in feature importance score, aiding in understanding the contribution of each feature to the model. It can handle both regression and classification problems and can be used with custom loss functions. As like SVM, it is computationally expensive, especially with a large number of trees and depth. It has multiple parameters that require tuning and finding the optimal combination can be time-consuming.

Dataset

In this study, a set of three-dimensional (3D) multi-modal images from the brain tumor segmentation challenge was used. This dataset was made available by the University of Pennsylvania's Center for Biomedical Image Computing and Analytics (CBICAs) and downloaded from the CBICA image processing portal.[30-32] It had a total of 599 scans in the neuroimaging informatics technology initiative format. Each scan includes a fluid-attenuated inversion recovery (FLAIR), T1, T2-weighted, and T1-contrast enhanced (T1-CE) series with its own mask. All the imaging datasets have been segmented manually, by four different people following the same annotation protocol, and their annotations were approved by experienced neuroradiologists. Each annotation comprises three different regions of the tumor: The ETC, the peritumoral edema (ED), and the necrotic/non-ETC (NCR/NETC). The data were divided into: Training, validation, and testing. Out of 599 images, 569 were used for training and validation in the ratios of 70% and 30%, respectively. A subset of the remaining 30 scans out of 599 total scans was used for testing as OTD. To further evaluate the model, in addition to the OTD, another set of 30 patients was taken from our own institute as institutional test data (ITD). These ITD scans were taken before surgery or any kind of radiation or chemotherapy. This dataset contains the MRI images of the FLAIR, T1, T1-CE, and T2 types. All these images were segmented for different distinct tumor regions, which were, peritumoral edema, enhancing-core, and necrotic tumor core, by two different persons, and their annotations were approved by an expert neuroradiologist.

Intensity normalization

In the context of scientific literature, intensity normalization acts as a vital preprocessing step that mitigates the confounding effects of inherent intensity variations.[33] The purpose of intensity normalization is to rectify disparities found in medical images by ensuring consistent intensity scales across different medical images. This approach improves the comparability and reliability of image characteristics by harmonizing the intensity distribution, allowing for more robust and accurate image analysis. This technique equips researchers and clinicians with a standardized framework for extracting meaningful information from medical images, enabling objective and reproducible analysis. Leveraging advanced normalization algorithms and statistical models, researchers can achieve optimal data normalization, thereby improving the accuracy and consistency of subsequent analytical methodologies. The different intensity normalization methods include minimum–maximum normalization[34] which linearly maps the intensity range of an image to a predefined range, typically (0,1) or, (−1,1) *Z*-score normalization[35] which transforms the intensity values of an image to have zero mean and unit standard deviation and quantile normalization[36] that aims to align the intensity distributions of different images by matching their quantiles.

In our study, we utilized an alternative method for intensity normalization known as CLAHE.[37] This technique is an advanced approach that combines the advantages of adaptive histogram equalization with contrast limitation. The primary goal of CLAHE is to enhance image contrast while mitigating the risk of amplifying noise and artifacts that can occur with traditional histogram equalization methods. CLAHE operates by dividing the image into smaller overlapping sections called tiles. For each tile, a histogram equalization process is performed to improve the local contrast within that specific region. However, to avoid excessive amplification of noise, a contrast limitation mechanism is applied. The CLAHE intensity normalization method offers several benefits over other methods. First, as it operates on smaller regions, therefore allowing for localized contrast enhancement. This aspect is particularly advantageous when dealing with medical images that exhibit variations in contrast and intensity across different regions. The statistical properties of the image, such as intensity distribution, provide valuable information for the diagnosis or analysis. Second, CLAHE incorporates a contrast limitation mechanism to prevent excessive amplification of noise and artifacts. This ensures that the enhanced image maintains the local statistical characteristics of the original image to some extent, preserving important information. It also offers control over various parameters, such as the size of the tiles, histogram bins, and the contrast limit which helps in fine tuning the process of intensity normalization. CLAHE proves to be a valuable technique for intensity normalization, providing localized contrast enhancement while effectively managing noise and maintaining important statistical properties of the image.

Feature extraction

To transform unstructured data into a form amenable to further analysis is known as feature extraction.[38] For this study, statistical features, shape features, gray-level co-occurrence matrix (GLCM),[39] and gray-level run length matrix (GLRLM)[40] were extracted using the pyRadiomics library.[14] It include first order statistics, 2D shape-based, 3D shape-based, GLCM, GLRLM, (NGTDM) Neighbouring Gray Tone Difference Matrix, etc.

Statistics-based features

We had first-order statistics features in this, which describe the distribution of voxel intensities within the image region defined by the mask using commonly used metrics. Total energy is the value of the energy feature multiplied by the voxel volume in cubic mm. The entropy of an image is a measure of the uncertain or random of its values. It is a standard for gauging how much data is typically needed to encode the image values. Skewness is defined as the degree of deviation from the normal distribution in a set of data and can be either positive, negative, zero, or undefined. Below equation represents the skewness mathematically, where X is the number of voxels included in the region of interest (ROI) and X_m is the average gray level intensity within the ROI.

$$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(X(i) - Xm)^3}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(i) - Xm)^3}}$$

Shape-based features

Every tumor has different features based on the shape of tumor or ROI. The extracted shape-based features were flatness, sphericity, elongation, surface area, and voxel volume. The shape-based features are independent from the gray level intensity distribution in the ROI and are therefore only calculated on the nonderived image and mask.

Gray-level co-occurrence matrix

The GLCM, also called the gray-level spatial dependence matrix, is a statistical way to look at texture that takes into account how pixels are placed in space. The GLCM functions describe an image's texture by figuring out how often pairs of pixels with certain values and in a certain location appear in an image and then extract statistical measures from this matrix. Gray-level size zone matrix quantifies gray level zones in an image. A gray level zone is defined as the number of connected voxels that share the same gray level intensity.

Gray level run length matrix

Features are a class of features based on creating a histogram of co-occurring pixel intensities at a given run length and orientation. Specifically, run length refers to the consecutive pixels with the same gray level value in a particular direction. GLRLM features have been found to be effective in capturing texture information in images and have been used in a variety of applications, including medical imaging analysis.

Optimal feature selection

As the size and variety of datasets grow, it is critical to reduce their size so that it can fit into system memory for further analysis. Feature selection techniques are used to reduce the number of input variables by removing redundant or irrelevant features and narrowing the set of features down to those which are most relevant to the machine learning model.[41,42] Irrelevant, redundant, and noisy features can clog an algorithm, lowering learning performance, accuracy, and computational cost.

Particle-swarm optimization method with varying inertia weight

The original PSO algorithm, also known as the "bird swarm algorithm," was developed by Kenny and Eberhart[43] back in 1995.[44] Mathematically speaking, PSO is an approach to optimization problems. According to PSO, if a bird is flying around aimlessly searching for food, the other birds in the flock can benefit from the bird's discovery by hearing about it and then going on to find even more. Mathematical calculations for the particle's speed and location send "solutions" (particles) whizzing over each issue. Particles' velocities are controlled by their fitness values, which are evaluated using the fitness function that needs optimizing. It is a random optimization algorithm and used in computational techniques for feature selection and classification.

The mathematical representation for updating velocity of particles in PSO is shown in the following equation

$$F_i^{t+1}(j+1) = \omega.F_i^t(j) + n_1 .r_1^t\left(A_{bi}^t(j) - X_i^t(j)\right)$$
$$+n_2 .r_2^t (B_{bi}^t(j) - X_i^t(j))$$

Where ω is the inertia weight, a constant that is always positive. This parameter is important for balancing the global search (exploration, when higher values are set) and the local search (exploitation, when lower values are set). In each iteration, two "best" values are added to each feature. F is the initial speed, which is between I_{min} and I_{max}, and the solution is X. t is the size of the search space, j is the number of iterations, n_1, n_2 are acceleration factors, and r_1, r_2 are two random numbers between 0 and 1. A_b is the personal best solution (the best solution that has been found so far), and B_b is the global best solution that the particle swarm optimizer keeps track of.

However, the above formula is likely to cause the local optimization with premature convergence phenomenon, and to accomplish the desired objective, the value of w should be gradually decreased as the optimization process progresses, aligning with the advancement of iterations. This gradual reduction of w over time proves more advantageous than employing a fixed value. We used three different strategies in this study to gradually decrease the w over time as follow.

Inertia weight with linearly decreasing strategy

Shi and Eberhart[45] proposed this strategy based on the idea that exploration is preferred at the beginning of the optimization process and exploitation is preferred at the conclusion. W is defined at each iteration as follows:

$$W_1(t) = w_{max} - (w_{max} - w_{min}) \times (t/T)$$

In this strategy, w is decreased linearly from w_{max} at the early iterations to w_{min} at the later. In the early stages, a higher inertia weight encourages exploration, allowing particles

to cover a broader search space. As optimization advances, a linear reduction in inertia drives exploitation, with an emphasis on local optima to improve. This technique seeks to maintain a balance between global exploration and local exploitation throughout the optimization process. Figure 1a shows the behavior of this strategy for out ITD dataset over 100 iterations.

Inertia weight with nonlinear coefficient decreasing strategy

Yang *et al*.[46] presented the nonlinear coefficient reducing technique to increase the performance of PSO that employs linear updating strategy, W at each iteration is defined as;

$$W_2(t) = w_{max} - (w_{max} - w_{min}) \times (t/T)^{\alpha}$$

Where α is suggested by the authors to be $\alpha = 1/\pi^2$. This method enables for a more adaptable change of the inertia weight, potentially responding more dynamically to the optimization environment. Nonlinear techniques may be adjusted to individual optimization circumstances, giving them the flexibility to address a wide range of problem characteristics. It can be tailored to the task at hand, potentially leading to improved performance in various instances. Figure 1b shows the behavior of this strategy.

Inertia weight with logarithmic strategy

Gao *et al*.[47] proposed a logarithmic decreasing technique to update the w value across repetitions using equation;

$$W_3(t) = w_{max} + (w_{max} - w_{min}) \times \log_{10}[a + 10(t/T)]$$

Where a is a constant that set to 1 by the authors. Logarithmic strategies involve using logarithmic functions to decrease the inertia weight, which provides a slow decrease at the beginning, promoting exploration in the early stages when the algorithm is searching broadly. As the search advances, the inertia weight drops more quickly, promoting exploitation, and convergence on favorable locations. This method can help to balance the exploration-exploitation trade-off by allowing the algorithm to first explore the solution space completely before focusing on refining the results. Figure 1c shows the behavior of this strategy.

These different strategies were used to select the optimal number of features using PSO. In this study, for all of these strategies, we had taken w_{min} as 0.1 and w_{max} as 0.9, n_1 as 0.8 and n_2 as 0.9.

Fitness function

In machine learning models, the fitness function, also known as the objective function, quantifies how well the model performs its task.[48,49] The fitness function varies depending on the specific machine learning task, such as classification, regression, or clustering, and the desired outcome. The common fitness functions for classification include cross-entropy loss, hinge loss, and log loss.

For this study, we had log loss as the fitness function. In the context of machine learning, log loss (also known as cross-entropy loss)[50] is a widely used fitness function for multi-class classification problems. It measures the performance of a model by evaluating the difference between the predicted class probabilities and the true class labels. The goal is to minimize this difference, indicating a better fit between the predicted and actual outcomes.[51] To understand log loss, let's consider a multi-class classification problem with N classes. Each data point in the training set has a set of features and belongs to one of the N classes. The model assigns probabilities to each class for a given input, and the sum of these probabilities is equal to 1. Log loss is defined as the negative logarithm of the predicted probability assigned to the correct class. The formula for log loss is as follows:

$$\text{log loss} = -(1/N) \times \Sigma [\Sigma (y_{ij} * \log (p_{ij}))]$$

where: N is the number of data points in the training set. y_{ij} is an indicator function that equals 1 if data point i belongs to class j, and 0 otherwise. p_{ij} is the predicted probability of data point i belonging to class j.

The log loss penalizes the model for both confidently incorrect predictions and uncertain predictions. A confident incorrect prediction, where the model assigns a high probability to the wrong class, will result in a high log loss value. Similarly, an uncertain prediction, where the model assigns low probabilities to all classes, will also contribute to a high log loss value. By minimizing the log loss, the model aims to improve its predictions and optimize the class probabilities for each input. This process involves adjusting the model's internal parameters, such as weights and biases, through techniques

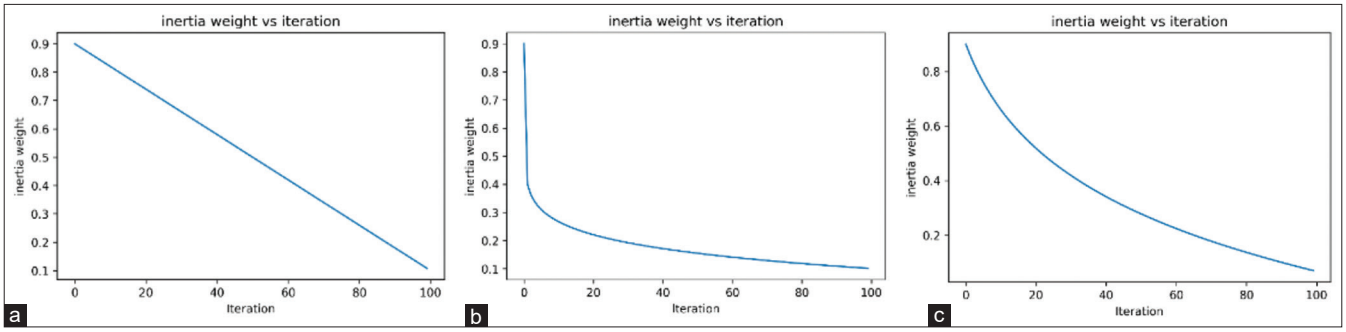


Figure 1: Represents the different weight varying strategies with iteration over time for hundred iterations. (a) Progression of Inertia weight with linearly decreasing strategy (b) Progression of Inertia weight with nonlinear coefficient decreasing strategy (c) Progression of Inertia weight with logarithmic strategy

such as gradient descent or other optimization algorithms. Log loss has become a popular choice for multi-class classification tasks because it provides a continuous, differentiable measure of the model's performance.[52]

Principal component analysis

PCA is an advanced statistical technique that plays a pivotal role in exploratory data analysis and dimensionality reduction.[15] It serves as a powerful tool to extract the most salient features from high-dimensional datasets, revealing the fundamental patterns and relationships inherent in the data.[53] PCA achieves this by transforming a set of correlated variables into a new set of uncorrelated variables known as principal components. These components, organized in the descending order of importance, are the linear combinations of the original variables. The primary objective of PCA is to capture the maximum variance in the data through the first principal component, with subsequent components capturing decreasing amounts of variance. By representing the data in terms of these principal components, it becomes possible to effectively reduce the dimensionality of the dataset while retaining the most significant information. This reduction is valuable in uncovering hidden structures and underlying dimensions within complex datasets, providing a fresh perspective for visualization and comprehension.[54]

Hybrid optimization technique

In this study, we also explored the combination of PCA and PSO. The integration of PCA and PSO leverages the strengths of both methods to tackle complex data analysis and optimization problems. PSO's ability to search and explore the solution space efficiently makes it well-suited for a wide range of optimization tasks. By iteratively updating a population of particles based on their individual and collective experiences, PSO aims to find the optimal solution in a timely manner. For classification task, we had used SVM, LGBM, and XGBM. The complete workflow of the paper is shown in Figure 2.

Algorithm: The code used to select the optimal features using hybrid optimization technique

Input: Set of MRI images and mask (training dataset)

Output: Subset of optimal features

1. Four types of MRI images stacked together, with corresponding mask (I, m)
2. Intensity normalization: CLAHE methods for intensity normalization of MRI images (I_{Norm})
3. Input - I_{Norm}, m to the pyRadiomics Library to extract Radiomics features (A_{Set})
4. Initialize PSO: Set w_{min}=0.1, w_{max}=0.9, n_population=4000, n_1=0.8 and n_2=0.9
5. Fitness function: log_loss = (− (1/N) × Σ [Σ (yij×log (pij))])
6. PSO=Particle Swarm Optimization (Fitness function, n_iteration=100, n_population)
7. fit (model, X_train, y_train, X_valid, y_valid)

ω = W1(iteration)/W2(iteration)//W3(iteration)

update position and velocity

return best_feature_list

8. B_{subset}^{A}=best_feature_list
9. PCA.fit (B_{subset}^{A})
10. Return subset of optimal features

CLAHE: Contrast limited adaptive histogram equalisation, PSO: Particle-swarm optimization, MRI: Magnetic resonance imaging, PCA: Principal component analysis

Results

The model's performance was assessed using the various classification metrics, including accuracy, classification error, precision, F1 score, recall, sensitivity, false negative rate, false positive rate (FPR), and Matthew's correlation coefficient (MCC). False positives (FP) occur when an

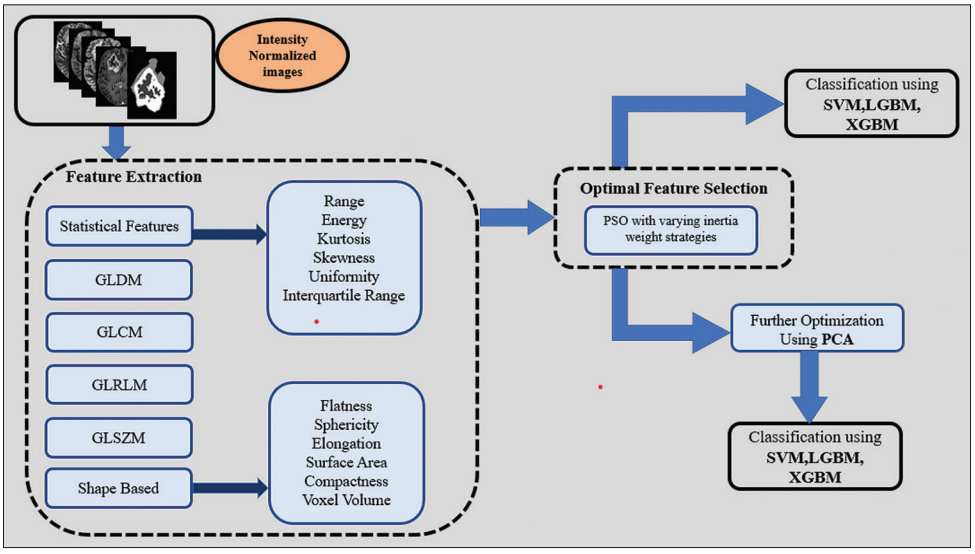


Figure 2: Represents the complete workflow for this study. PSO: Particle-swarm optimisation, PCA: Principal component analysis, GLCM: Gray-level co-occurrence matrix, GLRLM: Gray-level run length matrix, SVM: Support vector machine, LGBM: Light gradient boosting model, XGBM: Extreme gradient boosting model, GLSZM: Gray level size zone matrix, GLDM: Gray level dependence matrix

observation is incorrectly predicted as belonging to a certain class (Type I error), while false negatives (FN) occur when an observation is incorrectly predicted as not belonging to a certain class (Type II error). The MCC[52] is a reliable statistical measure that considers all the four categories of the confusion matrix (true positives, true negative, false negatives, and false positives) in proportion to the positive and negative elements in the dataset. It produces a high score only if the prediction performs well in all categories. The area under the receiver operating characteristic curve (AUC-ROC)[53] quantifies the overall performance of the model by measuring the area beneath the ROC curve, which represents the trade-off between true positive rate and FPR across different classification thresholds. The images before intensity normalization and after intensity normalization with CLAHE method are shown in Figure 3. Table 1 shows the different types of features extracted and corresponding selected features by using PSO and hybrid optimization method.

Features selected using different inertia weight varying methods

We have trained three different models with three different inertia weight varying strategies. Table 2 shows the results of accuracy for W1, W2, and W3 inertia weight varying strategies and number of features selected with each strategy. As we can see from the table, inertia weight with linearly decreasing strategy gives maximum accuracy of 0.995 for classification when used with XGB model. Furthermore, if we combine the PCA and PSO (inertia weight with linearly decreasing strategy) to further optimize the feature, the accuracy was around 0.999. The different classification metrics for the ITD dataset are shown in Figure 4. As we can see from the graphs, the maximum classification accuracy achieved was 0.998, with an F1 score of 0.998, MCC of 0.997, specificity of 0.997, sensitivity of 0.995, and ROC-AUC of 0.993 using the XGB model with PCA combined with PSO (inertia weight with linearly decreasing strategy) feature optimization methods.

We had also compared the proposed methods performance using the OTD dataset and the ITD dataset. Figure 5 shows the comparison of the accuracies of ITD and OTD datasets. It shows the accuracy of the proposed method on the unseen dataset was comparable to the trained dataset, which shows the high efficacy of the proposed technique.

Table 1: The number of different types of features extracted and number of chosen features using particle-swarm optimization and hybrid optimization method

Type of feature	Extracted	PSO			PSO + PCA (hybrid optimization)		
		W1	W2	W3	W1 + PCA	W2 + PCA	W3 + PCA
First order statistics	19	8	10	11	6	9	10
Shape based 2D	16	8	7	8	7	7	7
Shape based 3D	10	4	5	4	4	5	4
GLCM	8	6	5	2	6	5	2
GLSZM	18	9	9	8	7	8	8
GLRLM	10	5	4	6	5	4	6
NGTDM	5	3	2	3	3	2	3
Morphological features	58	8	16	15	5	10	10
Total	144	51	58	57	43	50	50

PSO: Particle-swarm optimization, PCA: Principal component analysis, GLCM: Gray-level co-occurrence matrix, GLRLM: Gray-level run length matrix, NGTDM: neighbouring gray tone difference matrix, GLSZM: Gray level size zone matrix

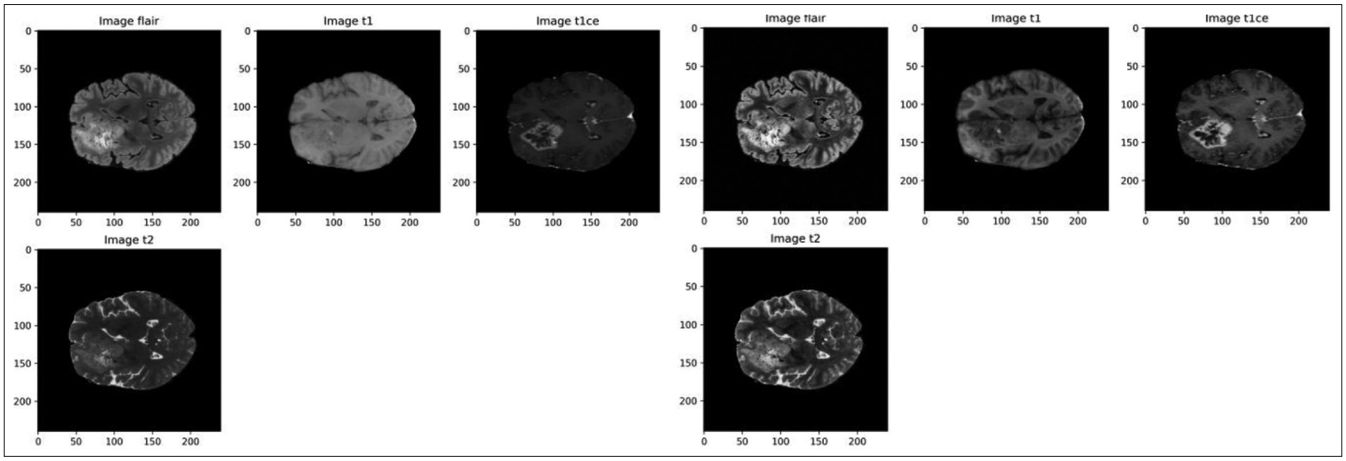


Figure 3: Images before applying any intensity normalisation (left), images after applying contrast limited adaptive histogram equalization intensity normalization methods (right)

Table 2: The accuracy and number of features selected using different optimization methods

Dataset	Model	PSO with different strategies (number of selected features)			PSO+PCA (number of selected features)		
		W1 (51)	W2 (58)	W3 (57)	W1 + PCA (43)	W2 + PCA (50)	W3 + PCA (50)
OTD	SVM	0.990	0.991	0.987	0.996	0.997	0.991
	LGB	0.992	0.993	0.991	0.996	0.997	0.994
	XGB	0.994	0.995	0.992	0.998	0.999	0.997

PCA: Principal component analysis, PSO: Particle-swarm optimization, OTD: Online test dataset, SVM: Support vector machine, LGB: Light gradient boosting, XGB: Extreme gradient boosting

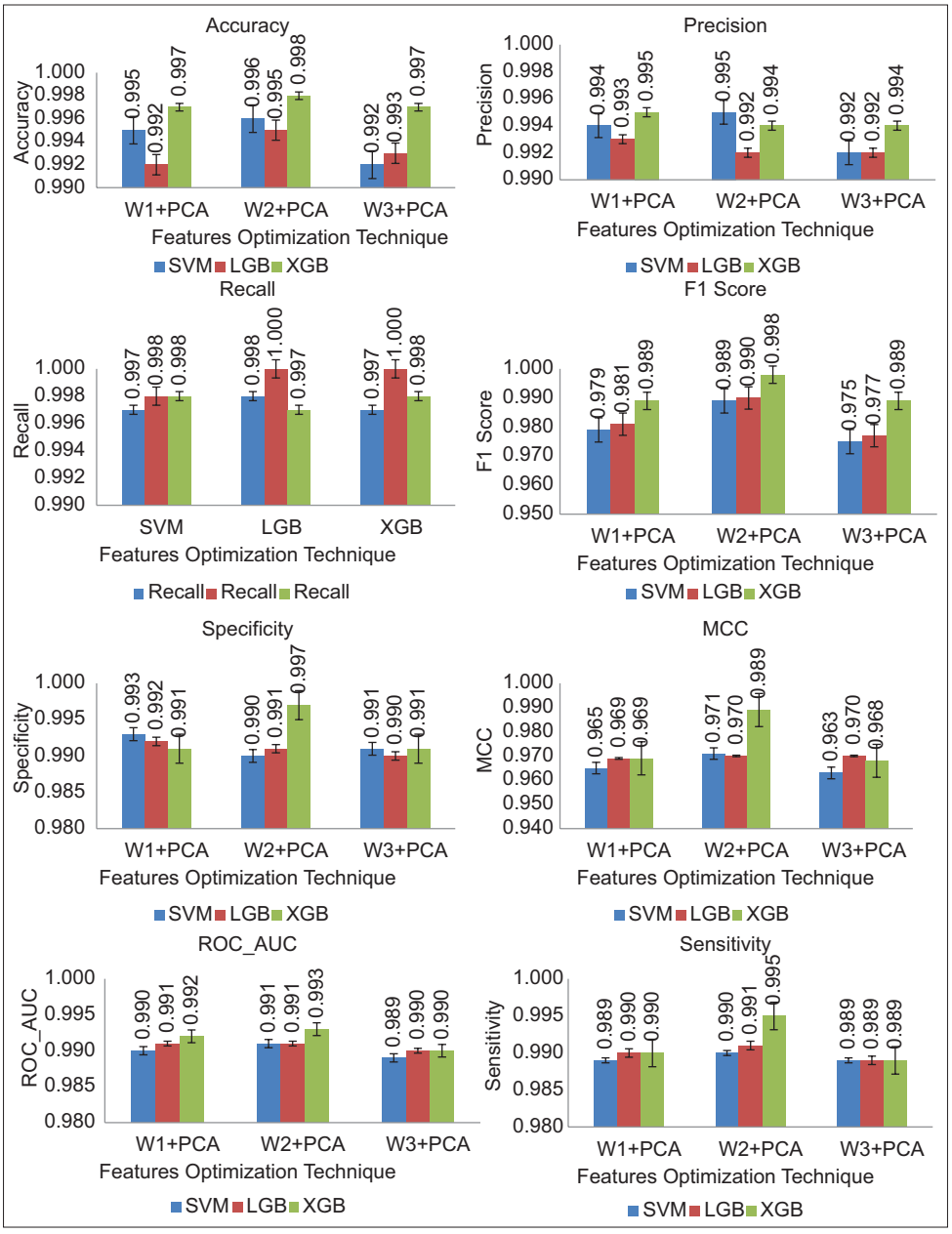


Figure 4: Represents the accuray, precision, recall, F1 score, specificity, area under the receiver operating characteristic curve, and sensitivity for the ITD dataset. PCA: Principal component analysis, SVM: Support vector machine, LGB: Light gradient boosting, XGB: Extreme gradient boosting, MCC: Matthew's correlation coefficient, ROC-AUC: Area under the receiver operating characteristic curve

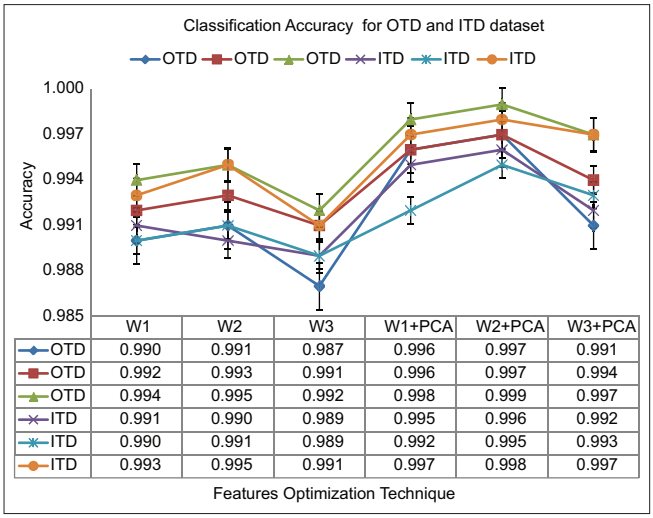
Comparison with existing work

We also compared the outcomes of our methods to those of other methods. Table 3 shows that the proposed method outperforms the other on every classification metric.

Table 3: Comparison of the proposed method with existing methods

Work done by	Accuracy	Sensitivity	Specificity	ROC_AUC	F1 score
Acquitter *et al*.[55]	0.750	0.750	0.760	NA	NA
Gao *et al*.[56]	0.903	0.947	0.817	0.958	-
Bacchi *et al*.[57]	0.823	NA	NA	0.800	0.860
Ramtekkar *et al*.[58]	0.989	NA	NA	NA	0.990
Vijithananda *et al*.[59]	0.845	NA	NA	NA	0.890
Noreen *et al*.[60]	0.947	NA	0.947	NA	0.940
Kumar *et al*.[61]	0.959	NA	0.957	NA	NA
Zahid *et al*.[62]	0.959	NA	0.959	NA	0.959
Hossain *et al*.[23]	0.970	NA	0.980	NA	0.968
Proposed	0.998	0.995	0.997	0.997	0.998

NA: Not reported, AUC_ROC: Area under the receiver operating characteristic curve



	W1	W2	W3	W1+PCA	W2+PCA	W3+PCA
OTD	0.990	0.991	0.987	0.996	0.997	0.991
OTD	0.992	0.993	0.991	0.996	0.997	0.994
OTD	0.994	0.995	0.992	0.998	0.999	0.997
ITD	0.991	0.990	0.989	0.995	0.996	0.992
ITD	0.990	0.991	0.989	0.992	0.995	0.993
ITD	0.993	0.995	0.991	0.997	0.998	0.997

Figure 5: Shows the comparison of accuracies of the ITD and online test dataset datasets for different optimization techniques. ITD: institute test data

Discussion

This study explored the robust classification model using different combinations of optimization strategies. Due to the strong impact of image normalization on the performance of a machine learning classifier, special attention should be provided in the image preprocessing step before typical radiomic analysis are performed. To remove this impact due to normalization, we had used the CLAHE intensity normalization method.

For every population-based metaheuristic algorithm, the optimization process comprises two fundamental stages: Exploration and exploitation. Ideally, an algorithm should initially prioritize the utilization of the exploration operator over the exploitation operator. A specific parameter or set of parameters within the algorithm is assigned this pivotal role. In the case of PSO, inertia weight plays the role of that parameter. Therefore, the proper variation of inertia weight with each iteration is crucial to achieving better classification accuracy. In our study, the different weight-varying strategies produced a very good classification result.

To further select the component with the highest variance from the optimized parameters, the combination of PCA and PSO offers several advantages. First, it enables more effective optimization by operating on a reduced feature space obtained through PCA, which eliminates redundant and irrelevant features. This improves the efficiency and effectiveness of the optimization process. Second, the optimal features provided by PSO allow PCA to work with a more manageable and interpretable representation of the data, leading to better understanding and insights into the underlying problem. Dimensionality reduction mitigates computational complexity, enhances visualization, and improves classification and clustering performance.

In our study, the hybrid methodology adopted by leveraging PSO for optimization and PCA for further optimization which improved efficiency, interpretability, and effectiveness in solving this problem of classification. Furthermore, when we compared our proposed method with the other available methods, it showed better results on every metric of classification tasks, which shows the novelty of the proposed approach.

On the other hand, this study has several limitations also. First, the small number of patients included in the experiments may be deemed insufficient for effective model validation. Our intention is to address this limitation by pursuing and validating the analysis with the patients' data from different institutions from all over the Indian sub-continent or different continents, which is still in the progress stage. In this study, we utilized two types of datasets, which were relatively small in size, preventing us from conducting a stratified analysis considering each center individually. Imbalanced datasets pose another significant challenge, potentially biasing models toward the majority class and compromising performance on minority classes. The quality of the data and preprocessing steps also influence model performance, as noisy or incomplete data can lead to suboptimal outcomes. Ensuring robustness through rigorous preprocessing, handling missing values, outliers, and addressing mislabeled instances are crucial for reliable results. Moreover, the interpretability of complex models becomes a concern, as high accuracy may come at the cost of understanding

the model's decision-making process. Achieving a balance between accuracy and interpretability is essential, particularly in applications where transparency is crucial. Transferability to other domains adds another layer of complexity, necessitating assessments of a model's adaptability and robustness across diverse datasets and real-world scenarios. Data leakage, experimental setup, algorithm selection bias, and ethical considerations further contribute to the nuanced landscape of challenges. Unintentional data leakage, for instance, can lead to overly optimistic performance estimates, emphasizing the importance of meticulous experimental design.

Future work should prioritize conducting model validation on an external dataset to enhance the study's statistical power. Furthermore, expanding the dataset size would allow for more robust validation of a radiomic signature for discriminating different tumor regions.

Conclusion

This study proposed an intelligent technique for distinguishing between different brain tumor regions. The free python-based library extracted the features, and the optimal feature selection was accomplished using various methods. Our experimental results demonstrate that, although the PSO technique with varying inertia weight strategies selects a robust subset of features, but these features can be further reduced by the use of PCA to achieve promising classification accuracy and other classification metrics for different tumor region classifications. This study also outperforms the other available methods in the literature. This research can be further expanded to other classification tasks also like, differentiation of lung nodules based on their size and other factors, and for classification of tumor for any site and for any type of tumor, etc., In conclusion, we can say that deep learning has shown potential in medical image analysis, but it also has limitations in terms of time and resource requirements and data privacy concerns. Alternative methods, such as feature engineering-based machine learning algorithms, may provide a more effective and efficient solution for real-time medical image analysis with high accuracy and the proposed method does this job very efficiently.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Giovagnoli AR, Tamburini M, Boiardi A. Quality of life in brain tumor patients. J Neurooncol 1996;30:71-80.
2. Kvale EA, Murthy R, Taylor R, Lee JY, Nabors LB. Distress and quality of life in primary high-grade brain tumor patients. Support Care Cancer 2009;17:793-9.
3. Davis FG, Freels S, Grutsch J, Barlas S, Brem S. Survival rates in patients with primary malignant brain tumors stratified by patient age and tumor histological type: An analysis based on Surveillance, Epidemiology, and End Results (SEER) data, 1973-1991. J Neurosurg 1998;88:1-10.
4. McFaline-Figueroa JR, Lee EQ. Brain Tumors. Am J Med 2018;131:874-82.
5. Praveen GB, Agrawal A. Hybrid Approach for Brain Tumor Detection and Classification in Magnetic Resonance Images. 2015 Communication, Control and Intelligent Systems (CCIS); 2016. p. 162-6.
6. Singh S, Singh BK, Kumar A. Magnetic resonance imaging image-based segmentation of brain tumor using the modified transfer learning method. J Med Phys 2022;47:315-21.
7. Jacobsen N, Deistung A, Timmann D, Goericke SL, Reichenbach JR, Güllmar D. Analysis of intensity normalization for optimal segmentation performance of a fully convolutional neural network. Z Med Phys 2019;29:128-38.
8. Zhao B. Understanding sources of variation to improve the reproducibility of radiomics. Front Oncol 2021;11:633176.
9. Chowdhary CL, Acharjya DP. Segmentation and feature extraction in medical imaging: A systematic review. Procedia Comput Sci 2020;167:26-36.
10. Dara S, Tumma P, Rao Eluri N, Rao Kancharla G. Feature extraction in medical images by using deep learning approach. Indian J Pure Appl Math 2018;120:305-312. Available from: https://www.acadpubl.eu/hub/. [Last accessed on 2023 May 31].
11. Kranthi Kumar K, Chaduvula K, Rao Markapudi B. A detailed survey on feature extraction techniques in image processing for medical image analysis. Eur J Mol Clin Med 2022;10:2020.
12. Jasti VD, Zamani AS, Arumugam K, Naved M, Pallathadka H, Sammy F, *et al*. Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis. Secur Commun Networks 2022;2022: Volume 2022, Article ID 1918379, 7 pages.
13. Narayan V, Mall PK, Awasthi S, Srivastava S, Gupta A. FuzzyNet: Medical Image Classification based on GLCM Texture Feature. 2015 Communication, Control and Intelligent Systems (CCIS); 2023. p. 769-73.
14. van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, *et al.* Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104-7.
15. Mohan G, Subashini MM. MRI based medical image analysis: Survey on brain tumor grade classification. Biomed Signal Process Control 2018;39:139-61.
16. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, *et al*. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
17. Zitko V. Principal component analysis in the evaluation of environmental data. Mar Pollut Bull 1994;28:718-22.
18. Raychaudhuri S, Stuart JM, Altman RB. Principal components analysis to summarize microarray experiments: application to sporulation time series. Pac Symp Biocomput. 2000:455-66. doi: 10.1142/9789814447331_0043. PMID: 10902193; PMCID: PMC2669932. [Last accessed on 2023 May 31].
19. Tharwat A, Gaber T, Ibrahim A, Hassanien AE. Linear discriminant analysis: A detailed tutorial. AI Commun 2017;30:169-90.
20. Song F, Mei D, Li H. Feature Selection Based on Linear Discriminant Analysis. Vol. 1. Intelligent System Design and Engineering Application (ISDEA), 2010 International Conference; 2010. p. 746-9.
21. Thai LH, Hai TS, Thuy NT. Image classification using support vector machine and artificial neural network. Int J Inf Technol Comput Sci 2012;4:32-8. Available from: https://www.mecs-press.org/ijitcs/ijitcsv4-n5/v4n5-5.html [Last accessed on 2023 May 31].
22. Chandra MA, Bedi SS. Survey on SVM and their application in image classification. Int J Inf Technol 2018;13:1-11. Available from: https://link.springer.com/article/100.1007/s41870-017-0080-1. [Last accessed on 2023 May 30].
23. Hossain A, Islam MT, Abdul Rahim SK, Rahman MA, Rahman T, Arshad H, *et al*. A Lightweight deep learning based microwave brain image network model for brain tumor classification using reconstructed microwave brain (RMB) images. Biosensors (Basel) 2023;13:238.
24. Teshnehlab M, Nasir K, Siar M. Brain Tumor Detection Using Deep Neural Network and Machine Learning Algorithm. Available from: https://www.researchgate.net/publication/338797226. [Last accessed on 2023 May 30].

25. Al-Ayyoub M, Husari G, Darwish O, Alabed A, Alabed-Alaziz A. Machine learning approach for brain tumor detection Comparative genomics View project Text analysis of cyber threat intelligence of unstructured text reports View project Machine Learning Approach for Brain Tumor Detection; 2012. Available from: https://www.researchgate.net/publication/262163807. [Last accessed on 2023 May 30].
26. Raj RJS, Shobana SJ, Pustokhina IV, Pustokhin DA, Gupta D, Shankar K. Optimal feature selection-based medical image classification using deep learning model in internet of medical things. IEEE Access 2020;8:58006-17.
27. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273-97. Available from: https://link.springer.com/article/10.1007/BF00994018. [Last accessed on 2024 Feb 03].
28. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, *et al*. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Available from: https://github.com/Microsoft/LightGBM. [Last accessed on 2024 Feb 03].
29. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Available from: https://dl.acm.org/doi/10.1145/2939672.2939785. [Last accessed on 2024 Feb 04].
30. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, *et al.* The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging 2015;34:1993-2024.
31. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, *et al.* Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data 2017;4:170117.
32. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. Sandra Gonzlez Vill 2018;124: 152. Available from: https://arxiv.org/abs/1811.02629v3. [Last accessed on 2023 May 31].
33. Kociołek M, Strzelecki M, Obuchowicz R. Does image normalization and intensity resolution impact texture classification? Comput Med Imaging Graph 2020;81:101716.
34. Haga A, Takahashi W, Aoki S, Nawa K, Yamashita H, Abe O, *et al.* Standardization of imaging features for radiomics analysis. J Med Invest 2019;66:35-7.
35. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the impact of intensity normalization on MR image synthesis. Proc SPIE Int Soc Opt Eng 2019;10949:109493H.
36. Chen CH, Chang CK, Tu CY, Liao WC, Wu BR, Chou KT, *et al.* Radiomic features analysis in computed tomography images of lung nodule classification. PLoS One 2018;13:e0192002.
37. Narayanan A, Rajasekaran MP, Zhang Y, Govindaraj V, Thiyagarajan A. Multi-channeled MR brain image segmentation: A novel double optimization approach combined with clustering technique for tumor identification and tissue segmentation. Biocybern Biomed Eng 2019;39:350-81.
38. Fayaz M, Qureshi MS, Kussainova K, Burkanova B, Aljarbouh A, Qureshi MB. An improved brain MRI classification methodology based on statistical features and machine learning algorithms. Comput Math Methods Med 2021;2021:p 1-15.
39. Haralick RM, Dinstein I, Shanmugam K. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610-21.
40. Galloway MM. Texture analysis using gray level Run lengths. Comput Graph Image Process 1975;4:172-9.
41. Soltanian-Zadeh H, Windham JP, Peck DJ. Optimal linear transformation for MRI feature extraction. IEEE Trans Med Imaging 1996;15:749-67.
42. Saeys Y, Abeel T, Van De Peer Y. Robust feature selection using ensemble feature selection techniques. Lect Notes Comput Sci 2008;5212:313-25. Available from: https://link.springer.com/chapter/10.1007/978-3-540-87481-2_21. [Last accessed on 2023 May 31].
43. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks [Internet]. 1995. p. 1942–8 vol.4. Available from: https://ieeexplore.ieee.org/document/488968. [Last accessed on 2024 Mar 9].
44. Okwu MO, Tartibu LK. Particle swarm optimisation. Stud Comput Intell 2021;927:5-13.
45. Shi Y, Eberhart RC. Empirical Study of Particle Swarm Optimization. Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on Volume: 3; 1999. p. 1945-50.
46. Yang C, Gao W, Liu N, Song C. Low-discrepancy sequence initialized particle swarm optimization algorithm with high-order nonlinear time-varying inertia weight. Appl Soft Comput 2015;29:386-94.
47. Gao YL, An XH, Liu JM. A Particle Swarm Optimization Algorithm with Logarithm Decreasing Inertia Weight and Chaos Mutation 2008 International Conference on Computational Intelligence and Security, CIS 2008 Volume 1; 2008. p. 61-5.
48. Kaushik D, Singh U, Singhal P, Singh V. Medical image segmentation using genetic algorithm. Int J Comput Appl 2013;81:975-8887.
49. Da Silva SF, Ribeiro MX, Batista Neto JD, Traina C Jr., Traina AJ. Improving the ranking quality of medical image retrieval using a genetic feature selection method. Decis Support Syst 2011;51:810-20.
50. Sundgaard JV, Harte J, Bray P, Laugesen S, Kamide Y, Tanaka C, *et al.* Deep metric learning for otitis media classification. Med Image Anal 2021;71:102034.
51. Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, *et al.* Optimization for medical image segmentation: theory and practice when evaluating with dice score or Jaccard index. IEEE Trans Med Imaging 2020;39:3679-90.
52. Ghosh P, Azam S, Quadir R, Karim A, Shamrat FM, Bhowmik SK, *et al.* SkinNet-16: A deep learning approach to identify benign and malignant skin lesions. Front Oncol 2022;12:931141.
53. Buciu I, Gacsadi A. Gabor Wavelet Based Features for Medical Image Analysis and Classification. Applied Sciences in Biomedical and Communication Technologies, 2009. ISABEL 2009. 2nd International Symposium; 2009.
54. Hussain Z, Gimenez F, Yi D, Rubin D. Differential data augmentation techniques for medical imaging classification tasks. AMIA Annu Symp Proc 2017;2017:979-84.
55. Acquitter C, Piram L, Sabatini U, Gilhodes J, Moyal Cohen-Jonathan E, Ken S, *et al.* Radiomics-based detection of radionecrosis using harmonized multiparametric MRI. Cancers (Basel) 2022;14:286.
56. Gao Y, Xiao X, Han B, Li G, Ning X, Wang D, *et al.* Deep learning methodology for differentiating glioma recurrence from radiation necrosis using multimodal magnetic resonance imaging: Algorithm development and validation. JMIR Med Inform 2020;8:e19805.
57. Bacchi S, Zerner T, Dongas J, Asahina AT, Abou-Hamden A, Otto S, *et al.* Deep learning in the detection of high-grade glioma recurrence using multiple MRI sequences: A pilot study. J Clin Neurosci 2019;70:11-3.
58. Ramtekkar PK, Pandey A, Pawar MK. Accurate detection of brain tumor using optimized feature selection based on deep learning techniques. Multimed Tools Appl 2023; vol 28: p-44623-44653.
59. Vijithananda SM, Jayatilake ML, Hewavithana B, Gonçalves T, Rato LM, Weerakoon BS, *et al*. Feature extraction from MRI ADC images for brain tumor classification using machine learning techniques. Biomed Eng Online 2022;21:52.
60. Noreen N, Palaniappan S, Qayyum A, Ahmad I, Imran M, Shoaib M. A deep learning model based on concatenation approach for the diagnosis of brain tumor. IEEE Access 2020;8:55135-44.
61. Kumar RL, Kakarla J, Isunuri BV, Singh M. Multi-class brain tumor classification using residual network and global average pooling. Multimed Tools Appl 2021;80:13429-38. Available from: https://link.springer.com/article/10.1007/s11042-020-10335-4. [Last accessed on 2023 May 31].
62. Zahid U, Ashraf I, Khan MA, Alhaisoni M, Yahya KM, Hussein HS, *et al.* BrainNet: Optimal Deep Learning Feature Fusion for Brain Tumor Classification. Comput Intell Neurosci 2022;2022:1465173.