Journal of Theoretical and Applied Information Technology
15th July 2025. Vol.103. No.13

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

ENHANCING BRAIN TUMOR CLASSIFICATION THROUGH ADVANCED IMAGE PROCESSING, HYBRID FEATURE EXTRACTION, MRMR FEATURE SELECTION, AND MULTI-CLASS CLASSIFIERS

GOURI SANKAR NAYAK[1], DR. PRADEEP KUMAR MALLICK[1], DR. BHUSHAN MARUTIRAO NANCHE[2], DR. NEELMADHAB PRADHI[3], DR. DILLIP RANJAN NAYAK[4]

[1]School of Computer Engineering, Kalinga Institute of Industrial Technology (KIIIT), Deemed to be University, Bhubaneswar, Odisha, India.

[2] Dept of Information Technology, D Y Patil College of Engineering, Akurdi, Pune, India.

[3] School of Computer Science & Engineering, GIET University, Gunupur, India

[4] Department of CSE,Government College of Engineering,Kalahandi, Bhawanipatna, India.

E-mail: [1]2181047@kiit.ac.in,mr.gouri4u@gmail.com, [1]pradeep.mallickfcs@kiit.ac.in,

2bmnanche@gmail.com, [3]dr.neelamadhab@giet.edu, [4]drnayak@gcekbpatna.ac.in

ABSTRACT

The purpose of this research is to create a holistic methodology for detecting and classifying brain tumors using up to date image processing, feature extraction and machine learning methods. To improve the diagnostic accuracy, this research evaluated the model performance of various classifiers as well as expanding feature representation to select the best available model for Brain tumor classification. Techniques: Such process includes developing techniques of image enhancement using TV-L1 norm and MRI segmentation algorithms etc. Preparation: Using sophisticated MRI segmentation techniques, the areas of tumors were precisely identified and improved image quality by implementing the TV-L1 standard for efficient image augmentation. For robust feature extraction and feature richness, Multi Scale Local Binary Pattern (MSLBP) for texture analysis was integrated with the multi-dimensional feature representation feature through the use of Quaternion Wavelet Transform (QWT). The Histogram of Oriented Gradients (HOG) is leveraged to quickly capture edge and shape data. Selecting Features Effectively: the dimensionality is reduced and hence increased classification effectiveness by selecting and prioritizing the most useful features via the Minimum Redundancy Maximum Relevance (MRMR) technique. Different Classification Techniques: Random Forest has great accuracy and is robust, so a number of classifiers were constructed. When it comes to modeling complex data distributions, Support Vector Machines (SVMs) prove effective. The K-Nearest Neighbors (KNN) is utilized for local patterns because it is user friendly and good. Results: The dataset from China's Nanfang Hospital and General Hospital was used to evaluate the techniques and is available on Kaggle. With low error rate (0.0147), high sensitivity (98.53%), and specificity (99.51%), Random Forest (RF) had the highest accuracy of 98.53%. In the end, Naive Bayes (NB) performed the worst at 96.67% while SVM and KNN produced slightly lower accuracies at 97.22% and 97.50% respectively. More sophisticated processing, feature extraction, and selection procedures used as part of the suite of preprocessing operations contributed significantly to the increased classification accuracy. RF revealed great promise for clinical use in identifying brain tumors, achieving high sensitivity and specificity while decreasing false positives to an accuracy rate of 98.53%. Even though RF consistently outperformed the others like SVM and KNN on all measures, they did very well on their own as well.

Keywords: *Histogram of Oriented Gradients, K-Nearest Neighbors, Multi-Scale Local Binary Pattern, Naïve Bayes Classifier, Quaternion Wavelet Transform, Random Forest Classifier, Support Vector Machine, TV-L1 norm*

1. INTRODUCTION

Early detection and accurate classification of brain tumor is of vital importance for improving patient outcomes and surgery plan. Medical imaging has come a long way but the anatomy of brain tumors in MRIs remain difficult to interpret because they appear so complex and variable. To overcome these challenges, this study develops an advanced methodology using image processing, feature extraction, and machine learning techniques that can improve brain tumor classification accuracy. The research aims to automate this process in order to deliver more dependable and efficient diagnostic tools, helping healthcare professionals establish better educated decisions.

The development of treatment regimens and patient care both depend on the early diagnosis of brain tumors. Radiologists may find it difficult to analyze brain tumor pictures, which might cause delays in treatment planning and even endanger the health of patients. Brain tumors originate from abnormal development of cells, proliferating uncontrollably and potentially damaging brain cells by exerting pressure within the skull. They can arise from brain cells surrounding the membranes (meninges), glands, or nerves. Malignant tumors, particularly Glioblastoma Multiforme (GBM), pose a significant threat, causing approximately fourteen thousand deaths annually with an overall survival rate of only fifteen months despite extensive research efforts. The severity of brain tumors is categorized into different grades [1]:

- Grade 1: Least dangerous tumors associated with prolonged survival, growing slowly and exhibiting a typical appearance under a microscope. Surgical treatment can often be successful for this grade, with examples including Pilocytic astrocytoma, ganglioglioma, and gangliocytoma.
- Grade 2: Tumors that grow slowly but appear anomalous under microscopic examination. Some may spread into neighboring tissues and progress to higher grades.
- Grade 3: Malignant tumors with a tendency to recur as grade 4. While there may not be a significant contrast compared to grade 2 tumors, they are generally more aggressive.
- Grade 4: The most malignant tumors, characterized by rapid growth, abnormal appearance under the microscope, invasion into neighboring tissues, and the formation of new blood vessels. GBM is a notable example.

Problem Statement: Detection and accurate classification of brain tumors early are important factors in planning and treating for successful treatment and improved survival rate of patients. Diagnosing brain tumors from medical images is still a difficult task because tumor characteristics are quite complex and brain structures are variable. In particular, heterogeneous texture and variations in tumor appearance makes the problem difficult, especially when several kinds of tumors are overlapped together in one scan or the appearance of tumors resembles normal brain tissue. Moreover, a limitation of the current diagnostic tools is their dependence on radiologists' subjective visual interpretation which leads to a chance of human error. Another obstacle to overcome in treating such highly heterogeneous cancers is the lack of standardized ways to classify brain tumors using different imaging modalities. In addition, the complexity of MRI images along with the demands for highly specialized expertise to pinpoint accurate diagnosis has proven tough enough for even clinicians in non-specialized medical settings to deliver decisions with the appropriate speed and accuracy. This is an issue that healthcare professionals face major challenge in giving accurate and time effective diagnoses. But these limitations delay or incorrect diagnosis of patients may result in suboptimal treatment decisions for patients that may worsen their prognosis and make their treatment plans ineffective. In addition, tumor classification from MRI scans is slow and difficult to perform reliably, unnecessarily burdening healthcare systems with long wait times before critical treatment can be commenced. This urgency calls for advanced and automated solutions which may improve diagnostic accuracy and assist clinicians to take early informed decisions.

Image processing methods, particularly segmentation, play a crucial role in tumor detection. Segmentation aims to partition an image into homogeneous regions to identify tumor shapes. Brain anatomy is frequently examined using computed tomography (CT) scans or magnetic resonance imaging (MRI), with MRI being safer and more effective because it doesn't involve radiation. However, since tumors consist of various biological tissues, a single type of MRI may not provide complete information. Therefore, integrating different complementary information, such as weighted MRI images (T2, T1, and Proton Density), enhances tumor segmentation accuracy.

Segmentation strategies have been highly effective, especially in detecting and characterizing diseased tissues during the developmental stages [2]. Computer-assisted automated diagnostic approaches have gained popularity as a solution to this issue, especially those that use machine learning techniques to classify brain tumors. Robust machine learning systems improve the accuracy of diagnoses, which helps medical practitioners make decisions.

Rationale of the Study: The main objective of this work is the development of an advanced and accurate solution for detecting and classifying brain tumors using the state of art image processing methods and the machine learning algorithms. The motivation behind this approach is that conventional diagnostic techniques based on obtaining frequently subjective expert opinion is usually time consuming, subjective and prone to human error. Second, the brain tumors are complex with various appearances across different MRI scans and so require sophisticated automated and more accurate diagnosis. State-of-the-art image enhancement techniques, robust feature extraction methods, and powerful machine learning classifiers, are integrated in the proposed methodology to bring about a multi-dimensional solution to the problem of tumor heterogeneity, high dimensionality of MRI data, and inter-subject variability in brain structures. Utilizing Quaternion Wavelet Transform (QWT) and Multi Scale Local Binary Pattern (MSLBP), they guarantee effective capture detail of both texture and structural information at different scales, improving feature richness needed for more reliable classification. Additionally, the Histogram of Oriented Gradients (HOG) contains rich information about both edge and shape, which is necessary for differentiating between multiple tumor types and normal brain. Reason for using MRMR for feature selection is to reduce dimensionality of feature set, hence make learning faster and more efficient without decoding relevancy of the processed features. Using MRMR, this study seeks to select the most informative features retaining as little redundant information as possible in order to ensure the classification problem is computationally feasible and highly accurate. Due to this desire to utilize the strengths of each algorithm, we use multiple classification algorithms such as Random Forest, Support Vector Machines, K Nearest Neighbors and Naïve Bayes. For instance, RF achieves high accuracy for dealing with complex data distributions while SVM and KNN excel in high dimensional spaces, a characteristic of medical image analysis. Because a wide variety of classifiers is used, the system is robust and adaptable to different types of input data and for different tumor classifications. Finally, this paper seeks to offer a better and quicker resolution to the classification of brain tumors that can serve as support to help clinicians make well informed decisions leading to patient outcomes. It is justifiable to choose the proposed approach because it combines the latest computational techniques with domain knowledge, thus a reasonable and feasible solution for real world applications of medical imaging and diagnosis.

The following sections comprise the rest of the paper. Section 2 provides a comprehensive review of relevant literature, exploring previous studies on brain tumor classification methods, feature extraction methods, and classification algorithms. The materials and techniques used are described in Section 3. Section 4 presents the suggested methodology, explaining the various feature extraction techniques as well as the MRMR-based feature selection procedure. The use of several classifiers to classify brain tumors is also covered in this section. Section 5, which also assesses the effectiveness of the proposed method in terms of classification accuracy and other pertinent metrics, presents the findings and comments. The major conclusions, their implications, and suggestions for more research are summarized in Section 6.

2. LITERATURE REVIEW

Since image classification [1] also requires the rigorous feature and classifier selection to get the best results, brain tumor classification is no different. For the purpose of their proposed framework to categorize brain tumors, the authors in [3] considered the effectiveness of three separate feature extraction methods. The density histogram, gradients of locally contrasted magnitudes (GLCM), and the bag of words (BoW) model. For their work, they used classifiers Support Vector Machine (SVMs), K-nearest neighbor (K-NN) and K-means algorithm. When combined with the SVM classifier, the Bow properties collected delivered a remarkable 91.14% accuracy. Their strategy is hindered, however, by the bag of words model's computational complexity when working with large datasets.

The authors of [4] presented a novel brain tumor categorization model that uses statistical data together with neural network categorization of MRI images. They successfully differentiated among three different brain cancer types (pituitary, glioma, and meningioma) employing a backpropagation neural network classifier which formed a combination of 2D DWT and 2D Gabor filter techniques with an impressive characterization accuracy of 95.66%. The however, one flaw they might have is they need to tweak extensive parameters such as the architecture of the neural network, which is a problem in terms of complexity of the model and processing requirements. Just like above, the authors of the work presented in [5] proposed a brain tumor classification hybrid feature extraction technique based on a regularized extreme learning machine (RELM). They achieved 94.233% success rate when they used L2 normalization and combined both PCA with GIST identification to use with RELM classifier. However, the reliance of their strategy on the RELM algorithm's sensitivity to hyperparameter selection could be limiting, requiring careful hyperparameter tuning to guarantee good performance on a mix of datasets. The authors studied in [6] categorized T1-weighted MR images that gave different sorts of brain tumor such as astrocytoma (AS), glioblastoma multiform (GBM), childhood tumor-medulloblastoma (MED), meningioma (MEN), secondary tumor-metastatic (MET) and normal regions (NR). Their strategy extracted features of texture and density, which were then selected by feature selection using PCA, with an astounding 91% success rate using ensemble learning. Their approach, however, is not without its disadvantages, including the possible uninterpretability of the resulting features, which may hurt the resulting biological importance of the selected traits. Before categorizing brain tumors, the authors of [7] utilized kernel based SVM technique for segmentation. They were able to classify with a success rate of 91.4% by obtaining first and second order statistical characteristics for regions segmented using the SVM technique. A potential drawback of their method may be its susceptibility to tuning of the SVM algorithm's parameters, requiring either domain expertise or extensive testing to achieve good performance over a spectrum of different datasets. As in the study referenced in [8], background reduction, feature extraction, and classification with a multi-layer perceptual neural network are used in the authors' multi-phase brain tumor classification. A sophisticated whale optimization technique was used to provide an optimal feature selection etc. Their dependence on heuristic optimization approach, e.g. whale optimization algorithm, depends on parameter change and cannot guarantee the convergence on global optimum is one possible drawback to this approach. Finally, kernel extreme learning machine (KELM) was used to classify convolutional neural network (CNN) features extracted from different brain MR images. Authors of [9] used the KELM CNN structure to get a 93.68% success rate. The reason their approach may be flawed is because they would need a lot of annotated data to properly train the CNN model; they will have issues with datasets having fewer labeled instances. The study in [10] tackled preprocessing, tumor classification using extreme learning machine local receptive field (ELM LRF), and image processing to extract tumor area for tumor segmentation using morphological operators. While their technique is effective, their reliance on manually generated features and stages of preprocessing may be disadvantageous since it could potentially sacrifice the ability to scale and apply it to other datasets. The study in [11] provides a semiautomatic mutli-step classification. The tumor area recognition process was first performed by a content based active contour technique so that the ROI can be manually defined by the radiologist. Then 71 texture and intensity characteristics are extracted from the segmented ROI, the best of which are chosen using a Genetic Algorithm (GA). Finally, the comparison of selected characteristics were classified using SVM and ANN classifiers. To evaluate this method, two datasets with 428 and 260 MR image datasets, each for six tumor categories were tested. A potential disadvantage of their method is that they need help with user participation for ROI selection. Because of this, categorizing the dataset may prove to be more subjective, and could potentially be more unpredictable between different radiologists or datasets. In this case, the discussed studies show that the feature extraction methods and the classifier chosen play a significant role in the classification of brain tumours, based on the stated advancements in the research taken place in the literature. Of special interest in this work, several of the feature extraction methods are specifically designed for use in the classification of common types of brain tumors, including gliomas, meningiomas, or pituitary tumors, in particular – Quaternion Wavelet Transform (QWT) [12] combined with Multi-Scale Local Binary Patterns (MSLBP) [13], Speeded-Up Robust Features (SURF) [14], and HOG features [15]. Machine learning methods including Random Forest Classifier, SVM, KNN, and Naïve Bayes are used by the study to assess feature matrices generated from using QWT-MSLBP, HOG, and SURF techniques. Different feature extraction techniques should be able to extract the texture and spatial difference among the different kinds of brain tumors by concentrating on the peculiar features of these images. Each of these various approaches might show discriminative aspects overlooked by traditional approaches. In addition to the MRMR (Minimum Redundancy Maximum Relevance) based feature selection, the study used a number of feature extraction methods to ascertain from the extracted sets which characteristics have the most information [16]. In order to achieve appropriate classification of brain tumors, MRMR yields a brief, highly discriminative feature set by measuring relevance to the target variable and minimizing feature redundancy. The strategy offers a mechanical method to boost the performance of the subsequent categorical schemes. The study also examines a range of powerful classification algorithms intended to be able to work with the selected feature sets. The Random Forest Classifier is one that uses an ensemble learning technology that creates a potent decision model that can handle sophisticated interaction among features. The SVM method is also used, which is known for highly effective performance in high-dimensional feature fields and for the ability to build non-linear decision boundaries. Along with the KNN algorithm presented, the paper also includes a straightforward but effective instance based learning algorithm that groups the samples based on their proximity to neighboring samples. Finally, given its effectiveness and scalability to manage large feature set, Naïve Bayes classifier founded on the ideas of Bayesian probability is used. By combining a diverse ensemble of classification algorithms with the MRMR based feature selection, this research work develops a comprehensive framework for accurate and effective brain tumor classification. This method ultimately uses the unique benefits of each algorithm and reduces the inherent deficits of any single methodology to arrive at a diagnosis and treatment to patients with brain tumors.

3. MATERIALS AND METHODS

3.1 Quaternion Wavelet Transform (QWT)

A mathematical tool for processing signals and images is called QWT. It is a kind of wavelet transform that works with signals that are represented in the complex number system's extension, the quaternion number system. QWT works with quaternion signals, which may represent signals with many orientation components, as opposed to real or complex signals, which are the domain of standard wavelet transformations. For processing signals and pictures with directional information, such texture or shape, QWT is hence especially well-suited. QWT may be applied to image processing applications as texture analysis, feature extraction, denoising, and image compression.

The primary benefit of QWT above conventional wavelet transforms is its ability to record an image's orientation and magnitude information, enabling a more thorough depiction of the picture. The QWT is superior to the DWT because it examines 2D signals in more detail. Like the Fourier transform, its coefficients are expressed in terms of amplitude and phase, and unlike the DWT, it maintains translational invariance. The quaternion adds three imaginary portions (i, j, and k) to complex numbers. It may be expressed in polar notation as $\mathrm{q} = |\mathrm{q}|\mathrm{e}^{\mathrm{i}\varphi}\mathrm{e}^{\mathrm{j}\theta}\mathrm{e}^{\mathrm{k}\psi}$ and in Cartesian form, which is similar to complex exponential notation, as $\mathrm{q} = \mathrm{a} + \mathrm{bi} + \mathrm{cj} + \mathrm{dk}$. It is differentiated by an argument with three angles as well as a modulus. The phase of 2D signals may be determined using the quaternionic argument.

The Local Phase of the Signal: Bülow [17] provided an example of how to use the Hilbert transform (HT) and the argument of complex numbers to recover the local phase of a one-dimensional signal. Thus, the quaternionic phase associated with a two-dimensional function is obtained, which is characterized by partial Hilbert transforms $(\mathcal{H}_1, \mathcal{H}_2)$ and total Hilbert transforms $(\mathcal{H}_\mathrm{T})$, which together comprise the analytical quaternionic signal [17]:

$$f_A(x,y) = f(x,y) + i\mathcal{H}_1 f(x,y) + j\mathcal{H}_2 f(x,y) + k\mathcal{H}_T f(x,y) \tag{1}$$

This signal's quaternionic argument reflects the quaternionic phase of f, defining its local structures at each point in time.

Wavelet Integration: Phase is incorporated into the wavelet decomposition process in the QWT context. This entails the generation of quaternionic coefficients using an analytical quaternionic mother wavelet, which use phase information to precisely describe the encoded structures. The phase information further refines the image's features, which are already powerfully described by the breakdown into sub-bands. The one-dimensional analytical signal's "local amplitude" (module) and "local phase" (argument) will remain unchanged. The presence of a component in each frequency sub-band at each spatial position is measured by the amplitude of a QWT coefficient $|q|$, which is independent of picture translation. In the meanwhile, the three angles (φ, θ, and ψ) that define the phase give a thorough explanation of the structure of these elements. In the conversation that follows, this point will be clarified further.

Establishment: According to the operational principle, two-dimensional Hilbert Transforms (HTs) mirror one-dimensional HTs along the image's rows and columns if the principal wavelet is divisible, as in this case, $\psi(x,y) = \psi_h(x)\psi_h(y)$. Therefore, by looking at the Hilbert pairs $\left(\psi_h, \psi_g = \mathcal{H}\psi_h\right)$ (for wavelets) and $\left(\phi_h, \phi_g = \mathcal{H}\phi_h\right)$ (for scaling functions), the analytical two-dimensional wavelet may be represented by independent products.

$$\begin{aligned}
\psi^D &= \psi_h(x)\psi_h(y) + i\psi_g(x)\psi_h(y) \\
&\quad + j\psi_h(x)\psi_g(y) + k\psi_g(x)\psi_g(y) \\
\psi^V &= \phi_h(x)\psi_h(y) + i\phi_g(x)\psi_h(y) \\
&\quad + j\phi_h(x)\psi_g(y) + k\phi_g(x)\psi_g(y) \\
\psi^H &= \psi_h(x)\phi_h(y) + i\psi_g(x)\phi_h(y) \\
&\quad + j\psi_h(x)\phi_g(y) + k\psi_g(x)\phi_g(y) \\
\phi &= \phi_h(x)\phi_h(y) + i\phi_g(x)\phi_h(y) + j\phi_h(x)\phi_g(y) \\
&\quad + k\phi_g(x)\phi_g(y)
\end{aligned} \tag{2}$$

This suggests that such a decomposition produces a non-isotropic wavelet and is significantly dependent on the image's orientation in the frame (x, y) (rotation variance). By using distinct filter banks and introducing the concept of phase, the QWT overcomes this limitation.

3.2 Multi-Scale Local Binary Pattern (MSLBP)

As a development of the Local Binary Pattern (LBP) method, MSLBP provides a multi-scale depiction of the subtleties of texture in brain MRI data. By using varying neighborhood sizes around the center pixel, MSLBP calculates texture properties over many scales, acting as a multi-scale extension of the traditional LBP process. A thorough multi-scale texture representation of the brain MRI picture is subsequently created by combining these texture elements from each scale. By employing a variety of scales, MSLBP effectively captures texture information at various granularities, increasing its usefulness for brain tumor classification tasks. MSLBP has proven to be effective in a variety of image processing applications, including character identification, face recognition, and texture analysis. Its ability to offer a thorough and reliable depiction of textural differences within brain MRI images emphasizes its significance in brain tumor categorization. Researchers have investigated the use of several LBP extractions inside an image in order to overcome the drawbacks of standard LBP technique, which confines the operator's calculations to a narrow area (usually a 3×3 zone). Changes in the number of samples (P) or the size of the regions (R) are possible with this modification. The necessity for more robustness in managing local texture variations brought on by things like rotation or illumination changes is what drives this kind of approach. In order to increase its applicability and flexibility to the many picture properties seen in brain tumor classification tasks, the classic LBP approach has been expanded into a multiscalar variant that differs in the radius size (R) in each extraction.

$$LBP_{PR} = \sum_{n=0}^{P-1} s(g_n - g_c) \times 2^n \tag{3}$$

Where P is the number of samples (pixels $g_0, g_1, \dots, g_{P-1}$) that are evenly distributed inside the neighborhood of the center pixel g_c, contained in its neighborhood. Figure 1 shows some regions across this data variation.

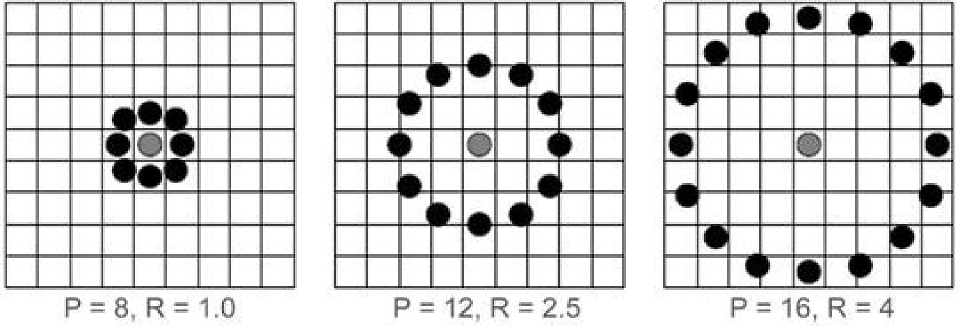


Figure 1: Circular region for three diverse values of P and R

The authors of [22] presented two variations of the Multiscale Local Binary Patterns (MSLBP) method. The first case, called MSLBP$_1$, uses LBP by extracting 8 samples from different regions of the image, after Gaussian filtering, defined by a radius R_n, where n is the respective scale and is denoted by LBP_{8,R_n} . The second type, called MSLBP2, is an alternative way to the first in which a set of average filters is applied in the region before the extraction of the $LBP_{8,1}$ pattern. The method used here is entirely based on MSLBP, in which a Gaussian filter of size G_n is applied to each scale, with the evolution of the scales given by changing the values of the radii. This change occurs respecting the optimal distribution in relation to the application of Gaussian filters. The objective of using low-pass Gaussian filters is so that in each sample in the neighborhood a larger amount of information can be collected than the one with a single pixel. The circles represent the area for extraction of the LBP standard around each sample after filtering. The outer radius of the "useful area", r_n, is used to calculate the radius R_n for extracting the LBP_{P_n,R_n} and is specified by:

$$r_n = r_{n-1} \cdot \left(\frac{2}{1 - sin\left(\frac{\pi}{P_n}\right)} - 1 \right), \quad n \in \{2, \dots, N\} \tag{4}$$

Where P_n, which in the instance of the MSLBP was thought to be $P_n = 8 \; \forall n \in N$, is the sampling of each scale and N is the number of scales. The radius r_n establishes the distance between the pixel and the edge of a neighborhood, so r_1 is defined as 1.5, which is the smallest distance between the pixel and the edge of a 3×3 region. Consequently, the radius of operator LBP_{8,R_n} is defined on scale $n(n \geq 2)$, as the midpoint between r_n and r_{n-1}, like this:

$$R_n = \frac{r_n + r_{n-1}}{2}, \quad n \in \{2, \dots, N\} \tag{5}$$

R_1 is specified as 1 since it is the lowest distance between the central pixel and the pixels in its neighborhood (3×3). This is similar to the situation of r_n, where the radii R_n of the operators indicate the distance between the central pixel and the pixels in its neighborhood. The r_n are used to calculate the window size of Gaussian filters, G_n, on the n scale as shown in Equation (6):

$$G_n = 2.round\left(\frac{r_n - r_{n-1}}{2}\right) + 1 \tag{6}$$

In which the function round: $\mathbb{R} \rightarrow \mathbb{N}^+$ assigns the smallest non-negative integer greater than it to a value. At each level, the δ_n value needed to apply the Gaussian filter is given by:

$$\delta_n = \frac{G_n}{\sqrt{-2\, ln(1-p)}} \tag{7}$$

Where p was set to 0.95.

3.3 SURF

The Speeded-Up Robust Features (SURF) algorithm is designed for efficient detection of distinctive points within images, focusing on optimizing both computational speed and robustness against transformations like scaling, rotation, and varying lighting conditions. SURF achieves these improvements by using approximations for operations that would otherwise be computationally heavy.

To begin, SURF constructs a scale space using an approach where Gaussian filters are approximated with box filters. This simplification allows convolutions to be performed rapidly using integral images, where each pixel at coordinates (x, y) represents the cumulative sum of all pixel values up to that point:

$$I_\Sigma(x, y) = \sum_{i \leq x} \sum_{i \leq y} I(i, j) \tag{8}$$

Where I_Σ is the integral image of the original image I. The advantage of integral images is that they allow the algorithm to compute sum values over any rectangular region in the image using only a few operations, regardless of the region's size, enabling efficient handling of large images.

To locate potential feature points, SURF relies on the Hessian matrix, which provides an indication of intensity variations around a point. For a point (x, y) at scale s, the Hessian matrix $H(x, y, s)$ is defined by:

$$H(x, y, s) = \begin{bmatrix} L_{xx}(x, y, s) & L_{xy}(x, y, s) \\ L_{xy}(x, y, s) & L_{yy}(x, y, s) \end{bmatrix} \tag{9}$$

Where L_{xx}, L_{yy}, and L_{xy} denote the second-order partial derivatives of the image's intensity at that point, calculated with approximated Gaussian derivatives (using box filters). The determinant of this Hessian matrix approximates the likelihood that a point is a feature, and for computational efficiency, SURF defines it as:

$$det(H_{aprox}) = D_{xx} D_{yy} - (w \cdot D_{xy})^2 \tag{10}$$

Where D_{xx}, D_{yy}, and D_{xy} are responses from the box filters approximating the Gaussian second derivatives, and w is a constant weighting factor, often set around 0.9. Filter sizes are incrementally increased in "octaves" to detect features across multiple scales, where each octave expands the filter size to capture larger structures within the image.

Following feature point detection, SURF applies non-maximum suppression within a neighborhood to refine the selection to the strongest candidates. Once these key points are established, SURF calculates Haar-wavelet responses in the x and y directions within a circular area centered at each point. These Haar responses are sampled within a radius proportional to the scale, and each is weighted by a Gaussian function centered on the feature point, enhancing orientation robustness.

To assign an orientation, SURF computes the summed responses in sliding angular sections around the feature point. The section with the largest sum determines the primary orientation. With this orientation established, the algorithm aligns a square window around the feature point, which is subdivided into smaller regions (e.g., 4×4 cells). Within each cell, SURF calculates the sum and absolute sum of Haar-wavelet responses along x and y directions:

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|\right) \tag{11}$$

Where d_x and d_y represent the directional wavelet responses within each cell. Concatenating the vectors from all cells yields a descriptor of 64 dimensions, which is robust to variations in geometry and illumination due to the weighted responses and consistent orientation alignment.

Feature matching between images is then performed using a nearest-neighbor approach, where descriptors with the smallest Euclidean distances are paired. A threshold, typically set around 0.7, screens out matches that may not be reliable, thereby prioritizing the closest, most accurate feature correspondences.

4. PROPOSED METHODOLOGY

Figure 2 shows the block diagram for the proposed approach and its detailed description is described in the following headings.

4.1 Preprocessing

4.1.1 Image Enhancement with TV-L1 Regularization

For improved visualization and analysis in complex imaging data, TV-L1 regularization serves as an effective technique for image enhancement. By combining Total Variation (TV) with L1 norm constraints, this method refines image quality, reducing noise while retaining essential edges and features, which is essential for accurate analysis and classification.

Total Variation measures the intensity variation across an image, with the TV-L1 model applying this concept to control noise levels without blurring critical details. Constraining the L1 norm of the image gradient encourages a balance, where regions with little variation appear smooth while preserving edge definition, allowing for the extraction of meaningful features even from noisy data [20].

Suppose an image I of dimensions $\mathrm{M} \times \mathrm{N}$ (height M and width N). TV-L1 regularization enhances this image by minimizing the following function [20]:

$$\mathrm{E}(\mathrm{I}') = \|\mathrm{I}' - \mathrm{I}\|^2 + \lambda \|\nabla \mathrm{I}'\|_1 \tag{12}$$

Where:

- $\|\mathrm{I}' - \mathrm{I}\|^2$ enforces similarity between the enhanced image I' and the original I,
- $\|\nabla \mathrm{I}'\|_1$ denotes the L1 norm of the gradient, preserving sharp transitions,
- and λ controls the trade-off between noise reduction and fidelity to the original.

Adjusting λ allows the technique to balance detail retention and noise suppression, making TV-L1 regularization ideal for producing clearer, analysis-ready images.

The pre-processing stages of the brain tumor detection methodology proposed are depicted in Figure 3.

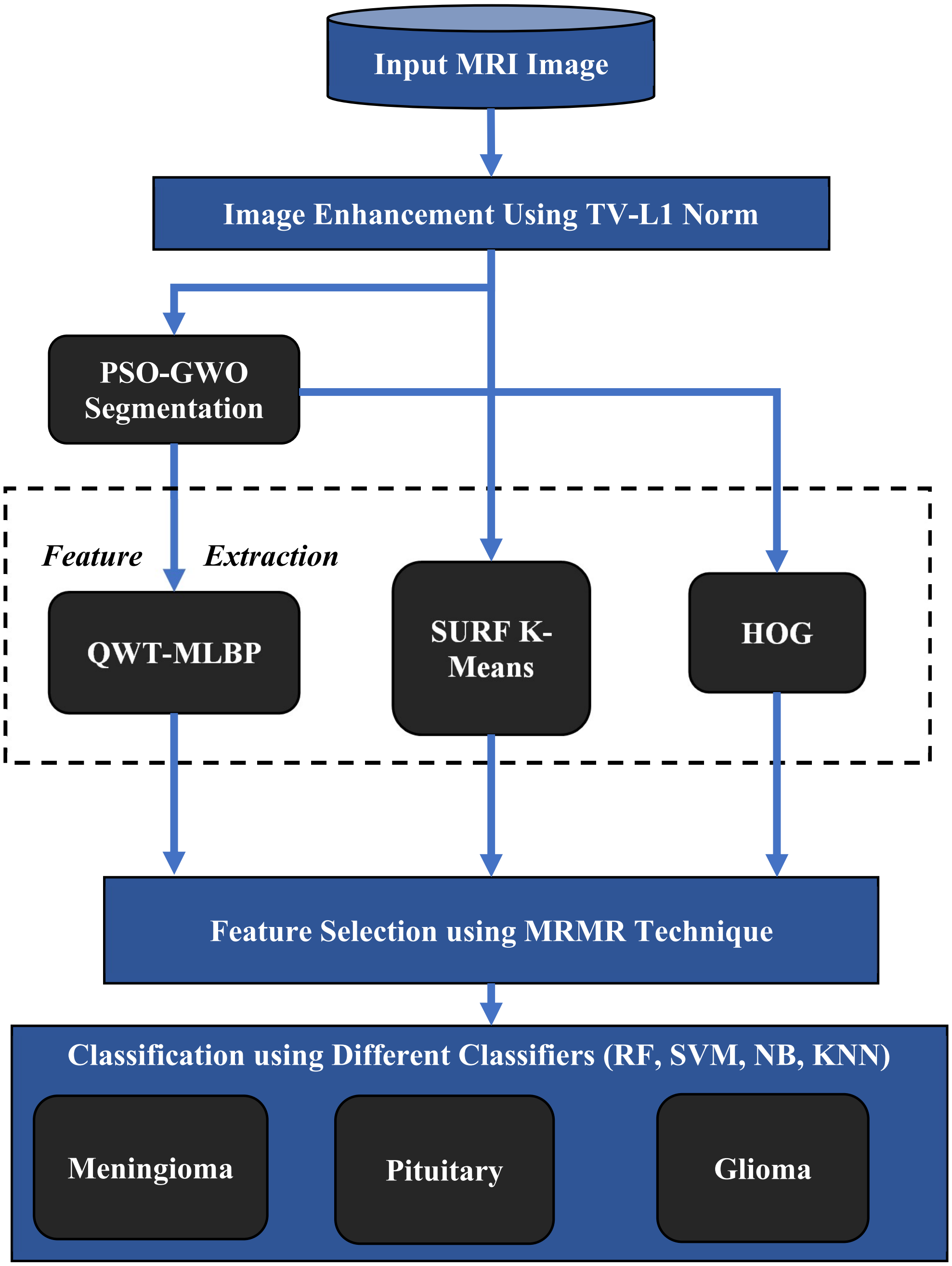


Figure 2: Block diagram for proposed brain tumor detection and classification

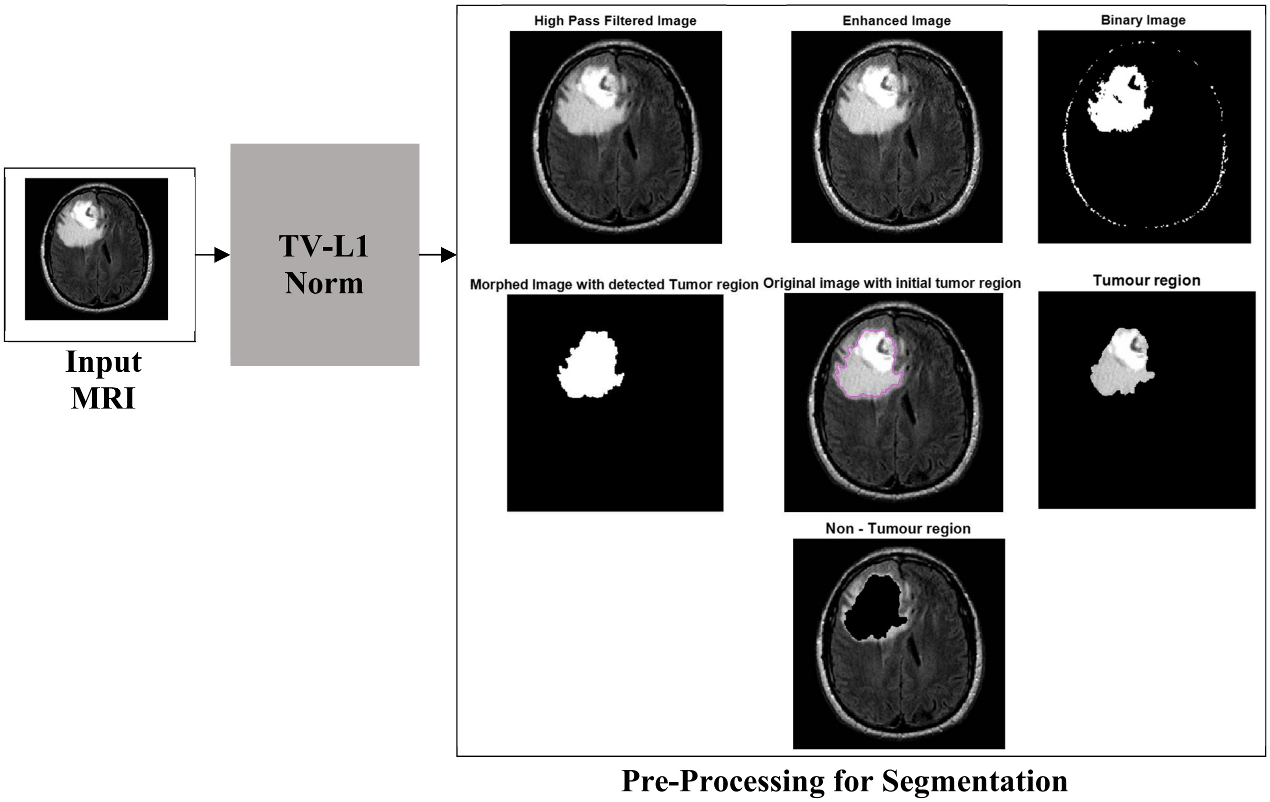


Figure 3: Pre-processing stage for proposed brain tumor detection

4.2 MRI Segmentation Algorithm Using Hybrid PSO-GWO

A crucial stage in medical image processing is MRI segmentation, which aims to precisely locate and isolate areas of interest in brain images. Using a hybrid method that combines Particle Swarm Optimization (PSO) and Gray Wolf Optimization (GWO), this work improves segmentation quality and convergence speed by utilizing the advantages of both optimization approaches. By iteratively optimizing parameters, the hybrid PSO-GWO technique improves clustering-based segmentation and makes it possible to handle the intricate intensity distributions of MRI.

Particle Swarm Optimization (PSO): The social behavior of fish schools and avian flocks served as inspiration for this evolutionary algorithm. Particles—a "swarm" of potential solutions—explore the solution space in PSO. Converging toward an ideal solution, each particle modifies its location in response to its own experience as well as the experience of nearby particles. The PSO algorithm's location and velocity updates for each particle i are explained mathematically as follows:

Velocity Update:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 \big(p_i - x_i(t)\big) + c_2 r_2 \big(g - x_i(t)\big) \tag{13}$$

Where:

- $v_i(t)$ is the velocity of particle i at iteration t,
- ω is the inertia weight, controlling exploration,
- c_1 and c_2 are cognitive and social coefficients, respectively, that determine the influence of individual and global best positions,
- r_1 and r_2 are random values in $[0,1]$,
- p_i is the particle's personal best position, and
- g is the global best position in the swarm.

Position Update:

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{14}$$

Where $x_i(t)$ symbolizes the existing location of particle i at iteration t.

PSO is an effective optimizer because it can balance exploration and exploitation through velocity and position updates. However, it occasionally experiences premature convergence, especially in complicated problem environments like MRI segmentation.

Gray Wolf Optimization (GWO): It mimics gray wolves' social structure and hunting tactics, especially their method. Alpha (α), beta (β), delta (δ), and omega (ω) are the four groups into which GWO divides the population; each group has a unique function in directing the optimization. The ω wolves trail after the α, β, and δ wolves, which stand for the finest answers and guide the investigation. In GWO, the main mathematical expressions are:

- Position Update: Each wolf adjusts its position toward the leading wolves α, β, and δ by calculating the distances as follows:

$$D_\alpha = |C_1 \cdot \alpha - X|, D_\beta = |C_2 \cdot \beta - X|, D_\delta = |C_3 \cdot \delta - X| \quad (15)$$

  Where C_1, C_2, and C_3 are coefficient vectors to control exploration and exploitation.

- New Position Calculation:

$$X(t+1) = \frac{X_\alpha + X_\beta + X_\delta}{3} \quad (16)$$

  Where X_α, X_β, and X_δ are positions updated based on the respective distances from the α, β, and δ wolves.

GWO is useful for fine-tuning solutions because of its architecture, which enables it to dynamically balance exploration and exploitation. However, in the early phases, GWO could show slower convergence than PSO.

PSO-GWO hybrid algorithm: This technique improves segmentation accuracy by combining the robust local search capabilities of GWO with the quick convergence of PSO in the early stages. Particles and wolves that update locations using both PSO and GWO techniques are used to initiate the algorithm. The following is a summary of the procedure:

1. Initialization: Randomly initialize positions and velocities for each particle (solution) and assign initial ranks for GWO components.
2. Position and Velocity Update:
   a. Perform PSO updates to rapidly explore the solution space.
   b. Use GWO principles to refine solutions, focusing on the leading particles (analogous to alpha, beta, delta wolves).
3. Fitness Evaluation: Calculate a fitness function based on segmentation accuracy metrics, such as intensity homogeneity within segmented regions.
4. Convergence Check: Iterate until convergence criteria (such as a predefined maximum number of iterations or minimal change in fitness value) are met.

Mathematical Formulation of Fitness Function: For MRI segmentation, the objective function typically minimizes inter-region similarity while maximizing intra-region similarity, defined as:

$$\text{Fitness} = \sum_{i=1}^{N} \sum_{j \in \Omega_i} |I(j) - \mu_i| + \lambda \sum_{k=1}^{N-1} \sum_{l=k+1}^{N} |\mu_k - \mu_l| \quad (17)$$

Where,

- $I(j)$ is the intensity of pixel j,
- μ_i is the mean intensity of segment i,
- N is the total number of segments,
- Ω_i is the set of pixels in segment i,
- and λ is a balancing parameter between intra-region and inter-region similarity.

This fitness function encourages segmentation outcomes that are both compact (homogeneous within segments) and distinct from other segments.

Figure 4 illustrates a flow diagram outlining the proposed segmentation method. It begins with the Input RGB Image, which undergoes Gray Conversion to simplify subsequent processing. The image then enters the FCM/PSO-GWO stage, representing the application of fuzzy c-means clustering or PSO combined with GWO for segmentation. The resulting segmentation is converted to binary format for clarity in distinguishing object boundaries. Morphological operations are applied to refine the segmentation by adjusting object shapes and sizes. Small objects are removed using the BWAREOPEN operation to eliminate noise or irrelevant features. The Masked XOR original operation is performed to compare the segmented result with the original image, highlighting areas of interest. Finally, the output of the segmentation process is achieved, providing a clear delineation of objects within the MRI image.

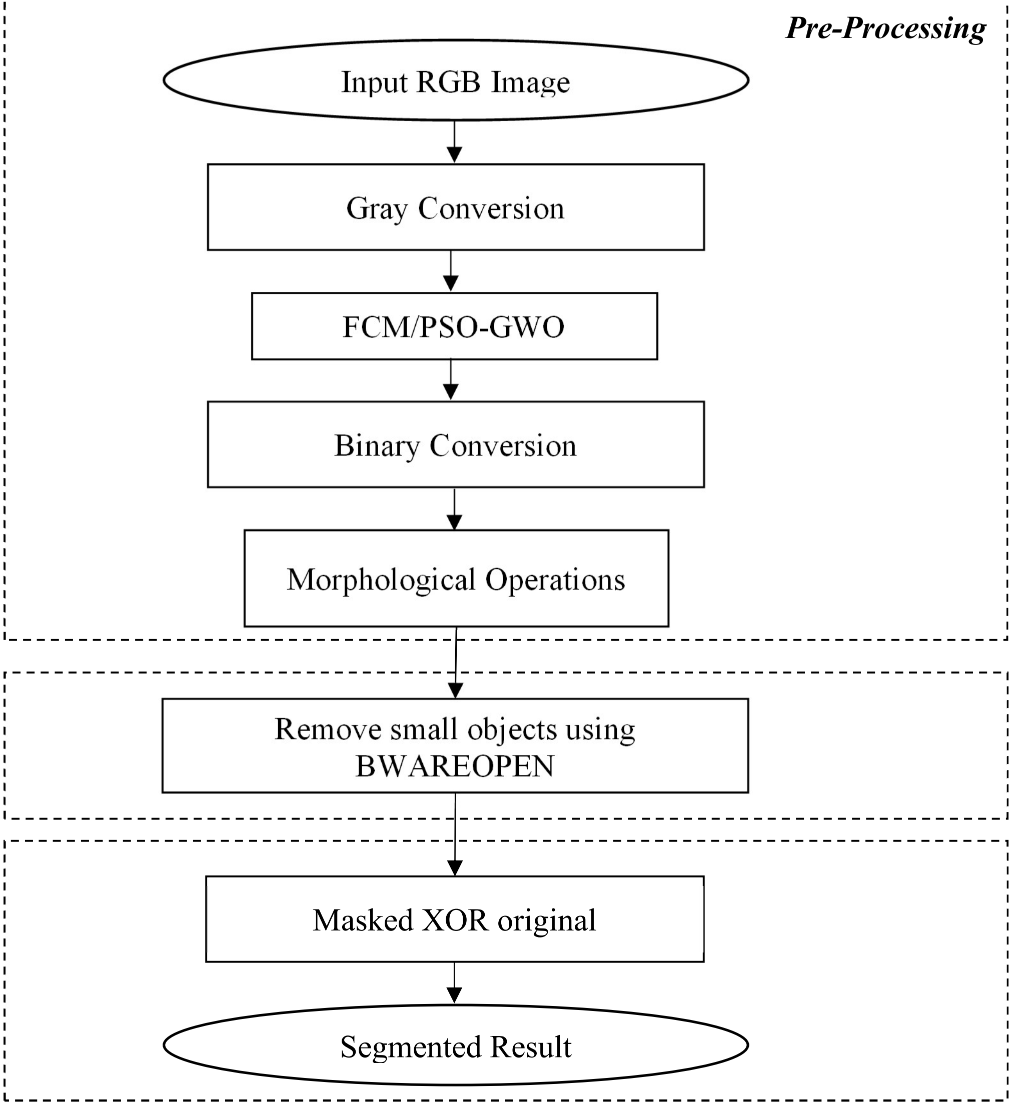


Figure 4: Flow diagram for proposed segmentation method

4.3 Post Processing

Post-processing procedures are essential for improving and fine-tuning the segmented MRI data after hybrid PSO-GWO segmentation, guaranteeing precision in locating regions of interest. By thresholding intensity levels, the segmentation mask created during the post-processing step aids in separating the foreground (area of interest) from the background.

- Initial Mask Creation: An initial binary mask is created, setting pixels to 1 for the targeted regions and 0 for the background, in order to isolate the regions of interest. Based on intensity and clustering results, this mask enables effective region extraction.
- Morphological Operations: The segmented areas are refined using morphological techniques like "area opening" and "closing," especially to get rid of tiny noisy patches that might not make a significant contribution. Only regions that are larger than a predetermined threshold (such as 400 pixels or more) are kept.

- Feature Calculation: To validate and improve the segmentation, variables such as area and mean intensity are used to compute area and pixel intensity statistics for each detected region.
- Logical Operations: When applied to binary and grayscale pictures, logical XOR operations help to clearly emphasize the segmented region of interest and guarantee a clear distinction from the backdrop.

$$J_{seg}(U, V) = \sum_{k=1}^{K} \sum_{i=1}^{N} \mu_{ik}^{m} \cdot d^2(x_i, v_k) \tag{18}$$

Where,

- For fuzzy clustering, U stands for membership degrees, V for cluster centroids, and m for the fuzzifier parameter governing cluster softness.
- $d(x_i, v_k)$ is the distance between data point x_i and cluster center v_k,
- K and N are the total number of clusters and data points, respectively.

By integrating morphological and logical operations, combined with PSO-GWO's parameter-tuned segmentation, the proposed process reliably isolates and refines brain regions, ensuring robust segmentation even in the presence of noise or variable MRI attributes.

4.4 Feature Extraction

The technique of collecting important and pertinent information from digital images is known as feature extraction, and it is an essential step in image editing. Its primary objective is to convert unprocessed image data into a collection of attributes that may be applied to classification or additional research. The SURF, DCT, and MSLBP methodologies are combined in this study's hybrid methodology.

4.4.1 SURF with K-means

The following are the descriptions of the proposed SURF algorithm (Figure 5):

A well-received clustering approach known as the k-means algorithm proves effective in grouping SURF features, obtained from images, into a predetermined quantity of representative clusters. The mathematical expression outlining the k-means algorithm is articulated as follows:

1. Initialization: From the SURF feature space, select K cluster centers at random.
2. Assignment: Assign each SURF feature to the cluster center that is closest to it.
3. Update: Determine the mean of all SURF characteristics assigned to each distinct cluster in order to reevaluate the cluster centers.
4. Continue repeatedly through stages 2 and 3 until convergence is reached, which can be found by hitting the maximum iteration threshold or by stationary cluster centers.

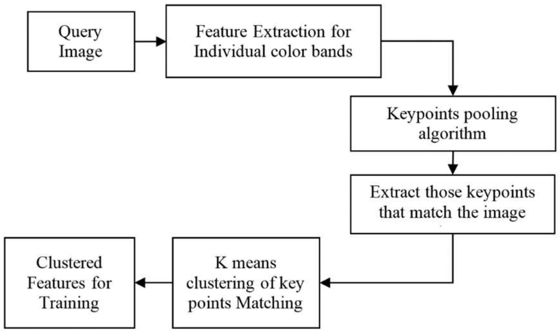


Figure 5: Block diagram for proposed SURF-K-Means approach

Following the clustering of SURF features into K representative clusters, the histogram depicting cluster assignments for each image serves as a feature vector for both training and testing the KNN classifier. Specifically, for each image, the histogram is constructed by tabulating the occurrences of SURF features assigned to individual clusters. The resulting histogram may be normalized to produce a feature vector and guarantee independence from the number of SURF features that were taken from the picture. Once the feature vectors for each picture in the training and testing datasets have been calculated, they are used as inputs for the KNN classifier, which is trained using standard machine learning methods. The k-means clustering method stands as a prevalent approach for grouping features in computer vision applications, including the clustering of SURF or SURF features. Below presents the mathematical formulation for k-means clustering:

In clustering problems, a set of data points is separated into several groups, each of which possesses similar characteristics. K-means clustering divides a set of N data points, $\{x_1, x_2, \ldots, x_N\}$, each represented by a d-dimensional feature vector, into K clusters with the goal of minimizing the dispersion of data points inside each cluster. The method does this by reducing the variance between the data points and the center of their assigned clusters.

Throughout the optimization process, K centroids, $\{c_1, c_2, \ldots, c_K\}$—which represent the centers of each cluster—are selected. These centroids are regularly adjusted to lower the total squared divergence of each data point from its assigned centroid. The following is one way to express the optimization goal:

$$\text{Minimize:} \sum_{i=1}^{N} \left\| x_i - c_{j(i)} \right\|^2 \tag{19}$$

Where $\left\| x_i - c_{j(i)} \right\|$ denotes the Euclidean distance between the data point and its corresponding centroid, and $c_{j(i)}$ is the centroid given to the data point x_i.

The method takes two main stages at each iteration:

- Step of Assignment: Using the Euclidean distance, each data point is paired with the closest centroid. Each data point is assigned to the cluster whose centroid is closest to it after the data is divided into K clusters in this stage.
- Update Step: After the assignment is finished, the mean of the data points in each cluster is used to recalculate the centroids. For every cluster C_j, the updated centroid is provided by:

$$c_j = \frac{1}{\left|C_j\right|} \sum_{x_i \in C_j} x_i \tag{20}$$

When the total is calculated across all of the data points in cluster C_j, and $\left|C_j\right|$ indicates the number of data points in that cluster.

Until convergence—that is, when there are no appreciable changes in the assignment of data points or when a predetermined stopping condition is satisfied—this method repeatedly repeats the assignment and update phases. There are several variations of the conventional k-means algorithm that are intended to handle particular data properties and increase efficiency. While hierarchical k-means constructs clusters in a hierarchical form, which might be helpful in some applications, mini-batch k-means minimizes computing time by using tiny random samples of the data for centroid updates.

Once the clustering process is finished, the centroids that are produced may be utilized to create a representation of a histogram, with each cluster's centroid corresponding to a histogram dimension. The histogram is filled in using the frequency of data points corresponding to each centroid. In image processing, this histogram is sometimes referred to as a "bag-of-visual-words." It functions as a reliable feature vector for machine learning classifiers like k-Nearest Neighbors (KNN). Since the centroids' initialization has a significant impact on k-means' performance, other initialization strategies, like k-means++, can be used to increase the likelihood of discovering a global minimum. Because the number of clusters, or K, affects how well k-means works, experimenting with different hyperparameters is crucial to tailoring the algorithm to a certain job.

4.4.2 MSLBP Features Extraction of QWT Coefficients

When local binary pattern (LBP) and quaternion wavelet transform are used together, both techniques are used to extract features from pictures. This combination is expressed mathematically as decomposing the picture into frequency subbands using the quaternion wavelet transform, then computing LBP features on each subband. This is a high-level summary of the procedure:

- Quaternion Wavelet Transform: The QWT extends the traditional wavelet transform to handle quaternion-valued signals or images. It decomposes the image into different frequency subbands in multiple dimensions (e.g., spatial and scale) using quaternion wavelet filters.
- Local Binary Pattern (LBP): By comparing pixel intensities with those of nearby pixels, this method acts as a texture descriptor, making it easier to extract image information. By thresholding the intensity levels of nearby pixels in relation to the center pixel, it encodes local texture patterns in images.
- Combination: After obtaining the quaternion wavelet coefficients, LBP features are computed independently on each subband. These LBP features capture texture information at different frequency scales and orientations.

Here's a more detailed mathematical formulation:

Let I be the input image.

1. ***Quaternion Wavelet Transform***:
    - Perform quaternion wavelet decomposition on the input image I to obtain quaternion wavelet coefficients at different scales and orientations.
    - Denote the resulting coefficients as $\{W_{j,k}^{d}(I)\}_{d=1}^{4}$, where d represents the quaternion dimension (real, i, j, k), j denotes the scale, and k represents the orientation.
2. ***Local Binary Pattern (LBP):***
    - For each quaternion wavelet coefficient $W_{j,k}^{d}(I)$, compute LBP features.
    - Define a local neighborhood $N_r(p)$ around each pixel p within the coefficient image, where r denotes the radius of the neighborhood.
    - Compute the LBP value for each pixel p within the neighborhood $N_r(p)$ based on intensity comparisons.
    - Concatenate the LBP histograms computed for all pixels within the coefficient image to form the LBP feature vector for that coefficient.
3. ***Combine Features:***
    - Concatenate the LBP feature vectors obtained from all quaternion wavelet coefficients $\{W_{j,k}^{d}(I)\}_{d=1}^{4}$ to construct the ultimate feature vector depicting the input image.

The specific implementation details, such as the choice of wavelet filters, LBP parameters (radius, neighborhood size, etc.), and feature vector concatenation method, may vary depending on the application and desired performance.

4.4.3 HOG Features

HOG features are a powerful method for identifying subtle local gradients in MRI images, which is useful for brain tumor detection. Because these characteristics capture unique patterns and configurations linked to tumor zones, they are crucial for precisely detecting and categorizing brain malignancies. To analyze MRI pictures, the feature extraction process's steps are meticulously planned:

- ***Gradient Calculation:*** To capture variations in intensity and edge orientations suggestive of tumor areas, compute the size and direction of gradients at each pixel across the MRI picture. Calculate the gradient magnitude M and direction θ at each pixel in an MRI picture I using gradient operators designed specifically for MRI images:

$$M(i,j) = \sqrt{\left(G_x(i,j)\right)^2 + \left(G_y(i,j)\right)^2} \qquad (21)$$

$$\theta(i,j) = \arctan\left(\frac{G_y(i,j)}{G_x(i,j)}\right) \qquad (22)$$

Where G_x and G_y stand for the gradients in the x and y directions, respectively.

- ***Orientation Binning:*** Highlight characteristics pertinent to tumor borders and structures by dividing the MRI image into cells and accumulating gradient orientations into histograms inside each cell:

$$H_{cell}(k) = \sum_{\text{pixels in cell}} M.\delta(\theta - \theta_k) \qquad (23)$$

Where θ_k is the center orientation of the k^{th} bin, δ is the Dirac delta function, and $H_{cell}(k)$ is the k^{th} bin of the cell's histogram.

- ***Block Normalization:*** To improve resilience against noise and fluctuations in MRI intensity, normalize the histograms inside each block. This ensures that features are represented consistently across various MRI scans.

$$v = \frac{v}{\sqrt{\|v\|_2^2 + \epsilon}} \qquad (24)$$

Where ϵ is a minute constant that prevents division by zero, and v is the histogram vector inside a block.

- ***Descriptor Formation:*** The normalized histograms from each cell in each block are concatenated to generate the final feature vector, which captures the distinct gradient patterns associated with the sites of brain tumors.

4.4.4 Feature Set

The feature vectors derived from SURF using the K-means, MSLBP, and HOG approaches are denoted by the letters F_{SURF}, F_{MSLBP}, and F_{HOG}, respectively. Next, the following formula is used to create the combined feature set $F_{combined} = [F_{SURF}, F_{MSLBP}, F_{HOG}]$:

$$F_{combined} = [F_{SURF}, F_{MSLBP}, F_{HOG}] \qquad (25)$$

A thorough representation of MRI images tailored for accurate brain tumor diagnosis and classification is provided by the combined feature set $F_{combined}$, which combines the unique information obtained by SURF, MSLBP, and HOG approaches.

4.5 MRMR Based Feature Selection

In tasks like classification and clustering, where dimensionality can greatly affect the system's accuracy and efficiency, feature selection is essential to enhancing the performance of machine learning models. A popular feature selection technique in this regard is Minimum Redundancy Maximum Relevance (MRMR), which seeks to choose a subset of characteristics that are minimum redundant and highly informative. Finding a feature subset that minimizes duplication between chosen features while preserving the most pertinent information about the target variable is the aim of MRMR.

The MRMR criteria seeks to concurrently optimize two goals:

- Maximum Relevance: Maximizing the relevance between each feature and the class label.
- Minimum Redundancy: Minimizing the redundancy between the features selected.

The MRMR feature selection approach can be formulated mathematically as follows:

1. Maximum Relevance

Mutual information is commonly used to assess a feature's relevance to the target variable, in this instance the class label. The degree to which the feature and the target variable share information is measured by mutual information. The mutual information between a feature f_i and the target variable Y is provided by:

$$I(f_i, Y) = H(Y) - H(Y|f_i) \tag{26}$$

Where:

- $I(f_i, Y)$ is the mutual information between feature f_i and the target variable Y,
- $H(Y)$ is the entropy of the target variable,
- $H(Y|f_i)$ is the conditional entropy of Y given f_i.

Maximizing the importance of each chosen feature is the aim. Therefore, the following is a definition of the MRMR criteria for maximum relevance for a feature set S:

$$R(S) = \sum_{f_i \in S} I(f_i, Y) \tag{27}$$

This guarantees that as much information as feasible about the class label Y is contained in the characteristics in the collection S.

2. Minimum Redundancy

MRMR seeks to reduce the mutual information between any two features in the subset in order to prevent repetition between the features that were chosen. The following provides the redundancy between two features, f_i and f_j:

$$I(f_i, f_j) = H(f_i) - H(f_i|f_j) \tag{28}$$

Where:

- $I(f_i, f_j)$ represents the mutual information between features f_i and f_j,
- $H(f_i)$ is the entropy of feature f_i,
- $H(f_i|f_j)$ is the conditional entropy of feature f_i given feature f_j.

The redundancy of a feature set S can be defined as:

A feature set S's redundancy may be described as follows:

$$\text{Redundancy}(S) = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i, f_j) \tag{29}$$

Where the summation covers every pair of features in the set, and $|S|$ is the number of features in the set S.

3. MRMR Criterion

Finding a feature subset that optimizes relevance and reduces duplication is the goal of the MRMR objective function. Thus, the MRMR criteria may be expressed as follows:

$$\text{MRMR}(S) = R(S) - \lambda \cdot \text{Redundancy}(S) \tag{30}$$

Where the trade-off between relevance and redundancy is controlled by the regularization parameter λ. One can balance the significance of redundancy and relevance to meet the particular requirements of the work by varying λ.

Iteratively choosing characteristics that optimize the MRMR criteria is how the MRMR algorithm works. Usually, the algorithm does the following actions:

1. Start with feature set S empty.
2. Choose the feature f_i that, when added to the current feature set, maximizes the MRMR criteria in each iteration.
3. Until a certain number of features are chosen or convergence is achieved, update the feature set and repeat the procedure.

Feature Selection for $F_{combined}$

The MRMR approach would assess each feature's relevance to the target variable from these three sets (SURF, MSLBP, and HOG) in the case of the combined feature set $F_{combined}$. Additionally, it would assess the redundancy of characteristics both inside and between sets. In order to minimize repetition across the three feature sets and increase the total relevance, the MRMR algorithm chooses a subset of features.

Improved classification performance results from MRMR's selection of the most pertinent and non-redundant features from these several sources, which guarantees that the final feature set F_{final} is compact and informative. The classification algorithm then uses the chosen characteristics as input, increasing efficiency and accuracy.

5. Final Feature Set

A subset of the initial combined feature set $F_{combined}$ that has been optimized using MRMR is the final set of features chosen, represented by the symbol F_{final}. The most pertinent and unique characteristics from this set can be applied to further machine learning tasks, such classifier training.

An efficient method of reducing dimensionality while guaranteeing that the features chosen preserve the most information and show the least amount of redundancy is MRMR-based feature selection. A feature set is produced that improves the brain tumor classification model's performance by using this method on the combined feature set derived from SURF, MSLBP, and HOG.

4.6 Classification Algorithms

Multiple classification algorithms are used to precisely assign tumor areas based on the characteristics chosen from the extraction and selection processes in order to detect brain tumors in MRI images. These algorithms, which are applied to the features derived from techniques like SURF, MSLBP, and HOG, provide unique ways for classification jobs, each with unique benefits. The main classifiers utilized in this study are described below:

4.6.1 Random Forest Classifier

During the training phase, the Random Forest Classifier, an ensemble learning approach, generates a large number of decision trees. Random feature subsets are used to construct each tree, and the sum of all the trees' forecasts yields the final prediction. Voting is usually used for this aggregation in classification tasks, with the predicted class being the one that obtains the most votes throughout the trees.

Algorithm Overview:

- A random subset of features is chosen from the entire feature set.
- Decision trees are constructed by recursively splitting the data based on feature thresholds.
- Repeat steps 1-2 to create multiple decision trees (forest).
- When making predictions, each tree within the forest offers its classification for the input data, and the ultimate prediction is ascertained through either a collective voting process for classification tasks or an averaging technique for regression tasks across all trees.

Mathematical Formulation: Let X be the feature vector input, Y be the assigned class label, and T be the number of decision trees in the forest. The following formula can be used to calculate the expected class, represented by $\hat{Y}$, for a given input X:

$$\hat{Y} = \mathrm{mode}\left[\left\{f_i(X)_{i=1}^{T}\right\}\right] \quad (31)$$

Here, the i^{th} decision tree's prediction is denoted by $f_i(X)$.

4.6.2 SVM

The Support Vector Machine (SVM) is a popular supervised learning method for categorization. SVM looks for the feature space hyperplane that maximizes the margin between different classes. When the data cannot be separated linearly, SVM employs a kernel approach that converts the data into a higher-dimensional space where a linear separation is possible.

Method:

1. The procedure maximizes the margin between the classes to determine which hyperplane best separates them.

2. The kernel function converts the data into a higher-dimensional space so that a separating hyperplane can be found if linear separation is not possible.
3. SVM uses the location of a new data point in relation to the hyperplane to classify it for prediction.

Given an input vector X, a weight vector w, and a bias term b, the SVM classification decision function is written as follows:

$$\hat{Y} = \mathrm{sign}(w \cdot X + b) \tag{32}$$

Where X is the feature vector to be classified.

4.6.3 KNN

A new data point is classified using the K-Nearest Neighbors (KNN) algorithm, which is an easy-to-understand classification technique that uses the majority class of its closest neighbors in the feature space. A user-defined parameter is the value of k, which is the number of neighbors to take into account.

Procedure:

1. All training data is stored by KNN together with the class labels that correspond to it.
2. Using a selected distance metric (such as the Euclidean distance), it predicts a new point in the feature space by identifying its k-nearest neighbors.
3. The predicted class is the most prevalent class among the k-nearest neighbors.

The predicted class $\hat{Y}$ may be calculated mathematically given an input feature vector X and a training set $\{X_i, Y_i\}$, where X_i are the feature vectors and Y_i are the corresponding class labels.

$$\hat{Y} = \mathrm{mode}[\{Y_i | X_i \in \text{nearest neighbors of } X\}] \tag{33}$$

4.6.4 Naïve Bayes

Based on the provided input characteristics, the Naïve Bayes classifier calculates the posterior probability of each class using Bayes' Theorem. In order to simplify the computation of probabilities, Naïve Bayes relies on the fundamental premise that the characteristics are conditionally independent given the class.

Method:

1. Based on Bayes' Theorem, Naïve Bayes determines the posterior probability for each class given the training data, which consists of feature vectors and their matching class labels.
2. The result for the input data is then projected to be the class with the highest posterior probability.

Assuming a class label Y and an input feature vector $X = \{x_1, x_2, \ldots, x_n\}$, the predicted class $\hat{Y}$ is the one that maximizes the posterior probability, according to mathematics:

$$\hat{Y} = \arg\max_{y \in (1,2,\ldots C)} \{P(Y = y|X)\} \tag{34}$$

Where $P(Y = y|X)$ is the posterior probability of class y given the input characteristics X, and C is the number of classes.

Based on the retrieved feature sets, each of these classification methods has special advantages for locating and categorizing tumor areas in MRI images. The kind of data, the available computing power, and the required level of accuracy for the task all influence the classifier selection.

5. RESULTS AND ANALYSIS

5.1 Dataset Description

This dataset, which includes a cohort of 233 individuals with brain tumors [21], was thoroughly compiled from medical imaging data from Nanfang Hospital and General Hospital in China [26]. The dataset highlights three different tumor classifications among these patients: meningioma, pituitary, and glioma tumors, each of which has a distinct clinical appearance. The photos in this dataset are all of the same size, measuring 512×512 pixels. The dataset includes a comprehensive collection of 3064 picture slices, each of which provides a window into the complex terrain of brain disease. With 708 cases of meningioma, 1426 cases of gliomas, and 930 cases of pituitary tumors, the tumor distribution among these slices is varied, providing a multidimensional view of the range of brain tumor pathologies [21].

5.2 Results and Analysis

The results of the developed approach for detecting and categorizing brain tumors are displayed in Figure 6.

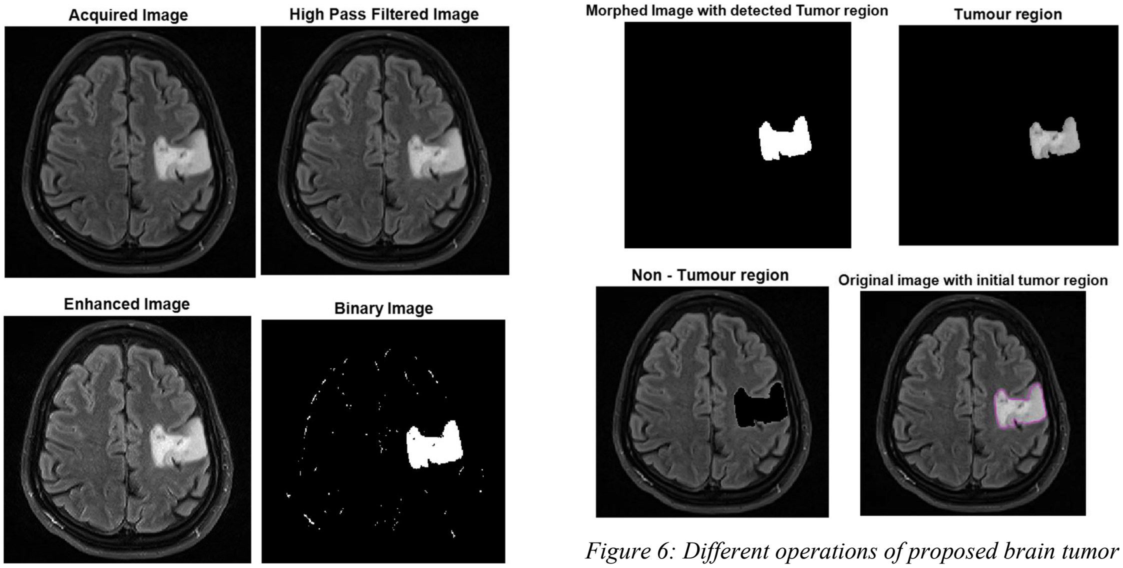


Figure 6: Different operations of proposed brain tumor detection

Table 1: Performance Evaluation for Different Feature Extraction Methods

Parameters	Surf-K-Means	QWT+MSLBP	HOG	Hybrid Features
Accuracy	94.44%	96.43%	94.64%	98.53%
Error Rate	5.56%	3.57%	5.36%	1.47%
Sensitivity	94.44%	96.43%	94.64%	98.53%
Specificity	98.15%	98.81%	98.21%	99.51%
Precision	95.45%	96.43%	95.21%	98.61%
False Positive Rate	1.85%	1.19%	1.79%	0.49%
F-Score	94.56%	96.43%	94.47%	98.53%
MCC	93.05%	95.24%	93.09%	98.08%
Kappa Statistics	85.19%	90.48%	85.71%	96.08%

The results of the investigation of several feature extraction techniques are shown in Table 1, which shows notable variations in performance metrics. Among the methods evaluated, Hybrid Features demonstrate the highest accuracy at 98.53%, closely followed by QWT+MSLBP at 96.43%. While Surf-Kmeans and HOG techniques exhibit respectable accuracies at 94.44% and 94.64% respectively, the Hybrid Features approach outperforms them by a notable margin. In terms of error rates, Hybrid Features again display the lowest value at 1.47%, indicating its robustness in classification tasks. The specificity scores further emphasize the dominance of Hybrid Features, attaining the peak score of 99.51%, showcasing its proficiency in precisely recognizing true negatives. Precision metrics across all methodologies are notably elevated, with Hybrid Features marginally surpassing the rest. Furthermore, diverse matrices bolster the exceptional efficacy of the Hybrid Features methodology, consistently exhibiting superior metrics in contrast to the alternative techniques assessed.

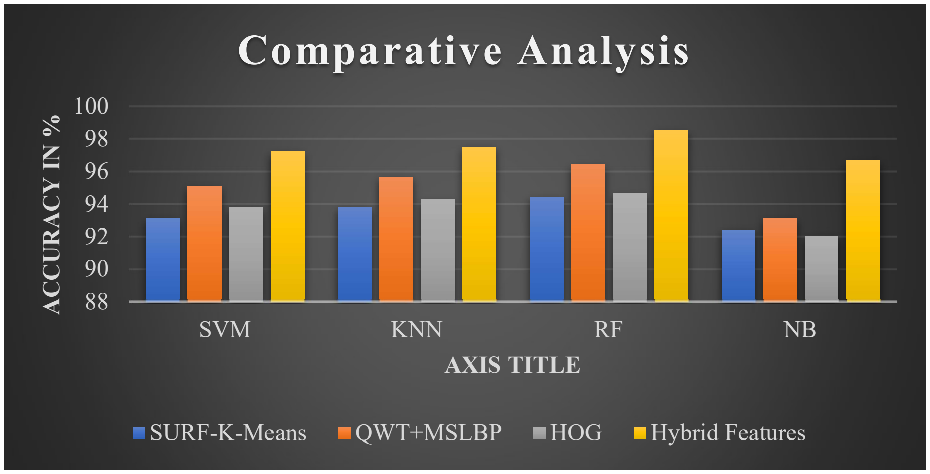


Figure 7: Comparative Assessment of Accuracy on Dataset Trained and Tested with Different Classifiers Using Different Features

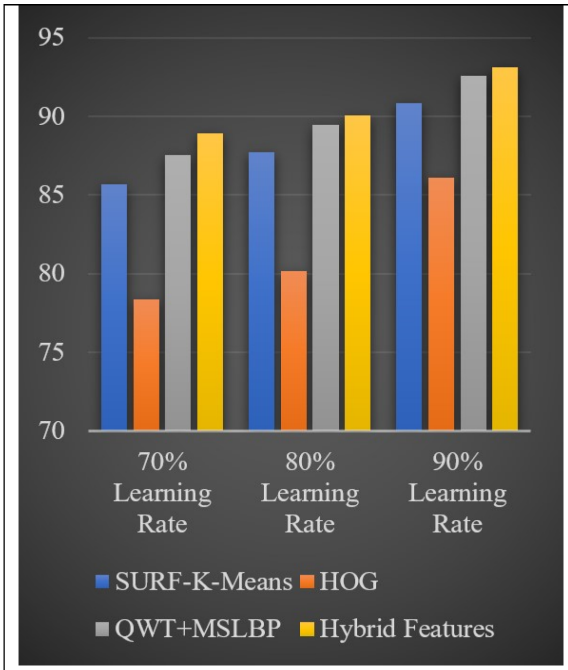
Fig. 7 displays a range of classification algorithms utilized in the categorization of brain tumors using various feature extraction techniques, displays the accuracy performance metrics. According to the visualization, the Naïve Bayes (NB) classifier with HOG features gets the lowest accuracy level of 92.02%, while the hybrid feature set combined with the Random Forest (RF) classifier obtains the greatest accuracy rate of 98.53%.

Table 2: Hybrid Feature Extraction Performance Comparison Using Various Classifiers

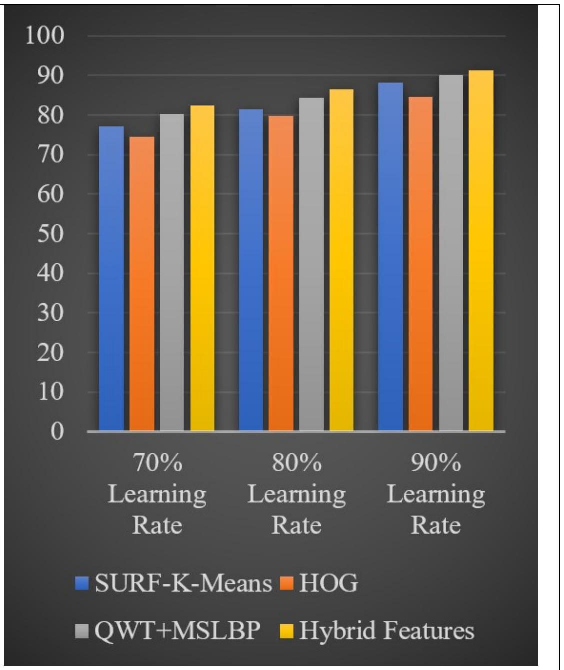
Parameters	SVM	KNN	RF	NB
Accuracy	97.22%	97.50%	98.53%	96.67%
Error Rate	2.78%	2.50%	1.47%	3.33%
Sensitivity	97.22%	97.50%	98.53%	96.67%
Specificity	99.07%	99.17%	99.51%	98.89%
Precision	97.75%	97.53%	98.61%	96.82%
False Positive Rate	0.93%	0.83%	0.49%	1.11%
F-Score	97.21%	97.49%	98.53%	96.69%
MCC	96.42%	96.68%	98.08%	95.62%
Kappa Statistics	92.59%	93.33%	96.08%	91.11%

Table 2 provides hybrid feature extraction performance comparison for different classifiers. A variety of metrics, including SVM, KNN, RF, and NB, are assessed for each classifier in the table. With the lowest error rate of 0.0147 and the highest accuracy rate of 98.53%, RF stands out as the most effective method for correctly classifying brain tumor data. Additionally, RF performs better on a variety of criteria, which confirms its efficacy in classification approaches.

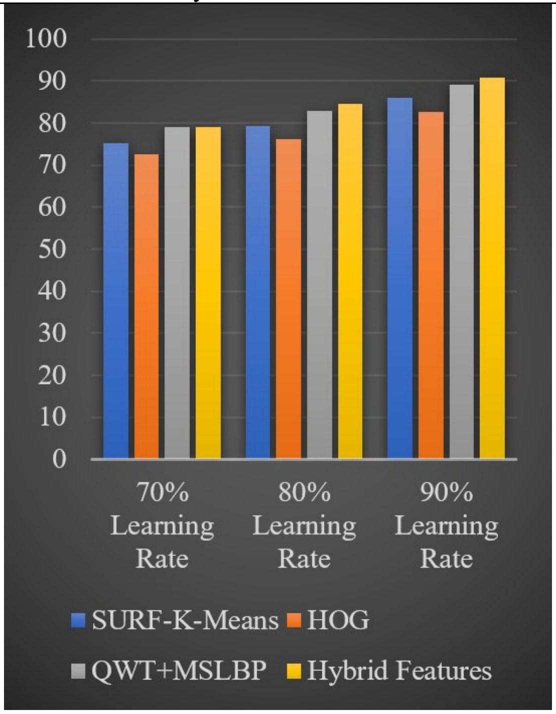
Table 3: Comparative Visualization of Accuracy across Different Classifiers



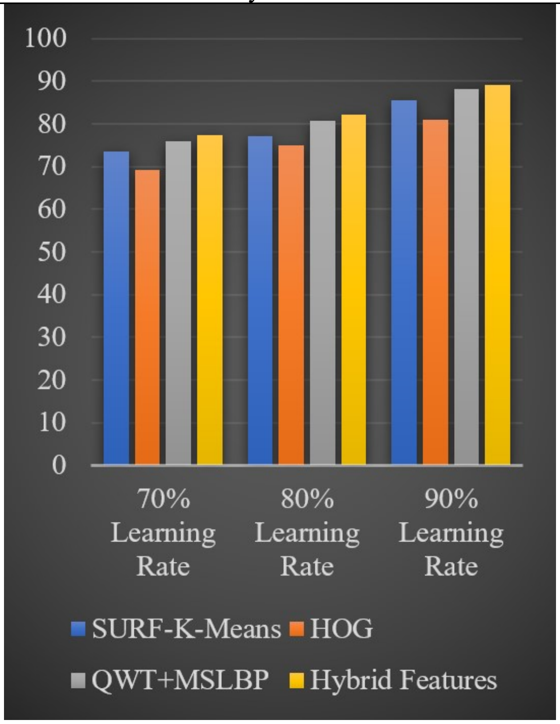
Comparative Analysis for Various Features Classified by Random Forest Classifier



Comparative Analysis for Various Features Classified by SVM Classifier



Comparative Analysis for Various Features Classified by KNN Classifier



Comparative Analysis for Various Features Classified by Naive Bayes Classifier

Table 4: Comparison of the Proposed Method's Performance with Previous Methods

Method	Dataset Source	Technique Used	Accuracy
Singh et al. (2023) [3]	MICCAI Brain Tumor Segmentation (BRATS) Challenge 2020	Convolutional neural network	92.50%
Kaplan et al. (2020) [22]	Kaggle Dataset	nLBP + Knn	95.56%
		LBP + Knn	93.28%
		αLBP + Knn	90.57%
Rinesh et al. (2022) [23]	Open-access brain tumor dataset from Kaggle	Multilayer feedforward neural network	96.47%
Ayadi et al. (2022) [24]	Figshare dataset	Hybrid features and SVM with LIN kernel	90.27%
Şahin et al. (2024) [25]	Brain Tumor MRI Dataset from Kaggle	Bayesian Multi-Objective (BMO) optimization	96.61%
Proposed Methodology	**Kaggle Dataset**	**Naïve Bayes with Hybrid features**	**96.67%**
		SVM with Hybrid features	**97.22%**
		KNN with Hybrid Features	**97.5%**
		Random Forest with Hybrid features	**98.53%**

A thorough comparison of the suggested methodology with earlier methods for classifying brain tumors is shown in Table 4. Singh et al. (2023) [3] achieved a 92.50% accuracy rate using a convolution neural network (CNN) on the MICCAI Brain Tumor Segmentation (BRATS) Challenge 2020 dataset. Using a Kaggle dataset, Kaplan et al. (2020) [22] tested several Local Binary Pattern (LBP) variations in conjunction with K-nearest neighbor (KNN) classifiers, achieving accuracy rates ranging from 90.57% to 95.56%. Using a multilayer feed forward neural network on an open-access brain tumor dataset from Kaggle, Rinesh et al. (2022) [23] achieved an accuracy of 96.47%. Using hybrid features and SVM with a linear kernel, Ayadi et al. (2022) [24] achieved an accuracy of 90.27% on a Figshare dataset. With a 96.61% accuracy rate, Şahin et al. (2024) [25] used Bayesian Multi-Objective (BMO) optimization on a Brain Tumor MRI Dataset from Kaggle. Comparatively, the suggested approach used Naïve Bayes, SVM, KNN, and Random Forest classifiers using Hybrid features on the Kaggle dataset, producing accuracy rates of 96.67%, 97.22%, 97.5%, and 98.53%, respectively.

Originality and Innovation of the Study: The technique presented here is a new approach of modeling brain tumor classes via a combination of sophisticated image processing and multiple machine learning classifiers in a hybrid scheme. Unique methods form a unique combination of the QWT and MSLBP for feature extraction and MRMR for feature selection, greatly improving classification accuracy. Not a widely explored use of these combined techniques within the context of brain tumor classification, this methodology is highly innovative. Furthermore, the study is distinguished for its use of multi-class classification, which implements ensemble of multiple classifiers to provide robust and reliable tumor classification for different MRI datasets. This method explicitly addresses the complicated nature of MRI scans, which often consist of multi type tumors and overlapping features. The research helps extend the literature on improving automated diagnostic tools in medical imaging by addressing both the challenge of high dimensional data and the problem of feature richness.

6. CONCLUSION

In order to improve the accuracy of diagnosis and treatment planning, this research addresses the significant issue of detecting and classifying brain tumors using medical imaging. By applying state-of-the-art techniques in image processing and machine learning, this study provides a comprehensive methodology aimed at increasing the accuracy of brain tumor identification. A thorough examination of the corpus of recent literature has been done in order to identify the most relevant algorithms and techniques in this area. Important elements in the new process involve segmenting MRI scans using optimal algorithms and refining images using sophisticated regularization techniques. Quaternion

Wavelet Transform (QWT), Multi-Scale Local Binary Pattern (MSLBP), and Histogram of Oriented Gradients (HOG) are among the resilient techniques used in feature extraction, and they all provide important information for efficient tumor identification. Additionally, the most discriminative features are used for future classification tasks when features are selected using Minimum Redundancy Maximum Relevance (MRMR). A dataset from respectable Chinese hospitals was used to evaluate the suggested approach, and it performed well on a number of assessment measures. With an accuracy of 98.53%, the hybrid feature set in conjunction with Random Forest classification significantly outperformed the findings of earlier research. These findings imply that the suggested framework may improve brain tumor categorization efficiency and diagnostic accuracy. In order to further improve feature extraction and classification, future studies can investigate the integration of increasingly complex image processing methods, such as the application of deep learning models like Convolutional Neural Networks (CNNs). Better clinical decision-making in neuro-oncology may also result from the combination of multi-modal imaging data and methods like transfer learning, which may present new chances to enhance model generalization and interpretability in general.

Gaps and Future Directions: Although promising, the results of the proposed approach reveal some gaps for further investigation.

- *Dataset Generalization:* The use of a specific dataset used in the study causes the model to not be that generalizable. In future research, it would be exciting to test the model's performance on datasets coming from other hospitals or regions so to see whether this model is applicable on a larger scale.
- *Multi-Modal Imaging:* In the current study, a single modal MRI is considered. Additional tumor information and higher classification accuracy can be achieved by incorporating multi modal imaging data, e.g., CT or PET scans.
- *Deep Learning Models:* While traditional machine learning algorithms are effective, deeper learning techniques including Convolutional Neural Networks (CNNs) may be able to demonstrate improved results through automatic extraction of higher level features from data.
- *Real-Time Processing:* The methodology runs in a batch mode. Finally, future studies should investigate such classification in real time in order to help clinicians make immediate decisions.
- *Explainability:* The model's decisions have to be more transparent. Exploring explainable AI might allow researchers to both make the model more interpretable and supported by healthcare professionals.
- *Handling Imbalanced Data:* It may be the case that many brain tumor datasets are imbalanced. Expanding on existing work, I investigate alternative methods such as data augmentation or specialized loss functions to improve performance on rare tumor types.

In addition, these areas of future research could help to better develop the proposed approach and make it more applicable and effective in clinical settings.

REFERENCES

[1] Amin, J., Sharif, M., Haldorai, A., Yasmin, M., and Nayak, R.S., "Brain tumor detection and classification using machine learning: a comprehensive survey," Complex & Intelligent Systems, Vol. 8, No. 4, 2022, pp. 3161-3183.

[2] Sravanthi, N.S.R.D.N., Swetha, N., Devi, P.R., Rachana, S., Gothane, S., and Sateesh, N., "Brain tumor detection using image processing," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol, Vol. 7, No. 3, 2021, pp. 348-352.

[3] Singh, R., and Agarwal, B.B., "An automated brain tumor classification in MR images using an enhanced convolutional neural network," International Journal of Information Technology, Vol. 15, No. 2, 2023, pp. 665-674.

[4] Mudda, M., Manjunath, R., and Krishnamurthy, N., "Brain tumor classification using enhanced statistical texture features," IETE Journal of Research, Vol. 68, No. 5, 2022, pp. 3695-3706.

[5] Gumaei, A., Hassan, M.M., Hassan, M.R., Alelaiwi, A., and Fortino, G., "A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification," IEEE Access, Vol. 7, 2022, pp. 36266-36273.

[6] Bhatele, K.R., and Bhadauria, S.S., "Multiclass classification of central nervous system brain tumor types based on proposed hybrid texture feature extraction methods and ensemble learning," Multimedia Tools and Applications, Vol. 82, No. 3, 2023, pp. 3831-3858.

[7] Rao, C.S., and Karunakara, K., "Efficient detection and classification of brain tumor using kernel based SVM for MRI," Multimedia Tools and Applications, Vol. 81, No. 5, 2022, pp. 7393-7417.

[8] Yin, B., Wang, C., and Abza, F., "New brain tumor classification method based on an improved version of whale optimization algorithm," Biomedical Signal Processing and Control, Vol. 56, 2020, p. 101728.

[9] Pashaei, A., Sajedi, H., and Jazayeri, N., "Brain tumor classification via convolutional neural network and extreme learning machines," In 2022 8th International Conference on Computer and Knowledge Engineering (ICCKE), IEEE, October 2022, pp. 314-319.

[10] Dipu, N.M., Shohan, S.A., and Salam, K.M.A., "Deep learning based brain tumor detection and classification," In 2023 International Conference on Intelligent Technologies (CONIT), IEEE, June 2023, pp. 1-6.

[11] Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., and Ahuja, C.K., "Multiclass brain tumor classification using GA-SVM," In 2022 Developments in E-systems Engineering, IEEE, December 2022, pp. 182-187.

[12] Babu, B.S., and Venkatanarayana, M., "AQWT: adaptive quaternion wavelet transform and hybrid heuristic improvement for multi-modal medical image fusion model," Signal, Image and Video Processing, 2023, pp. 1-11.

[13] Vuong, T.T., Song, B., Kim, K., Cho, Y.M., and Kwak, J.T., "Multi-scale binary pattern encoding network for cancer classification in pathology images," IEEE Journal of Biomedical and Health Informatics, Vol. 26, No. 3, 2021, pp. 1152-1163.

[14] Almalki, Y.E., Ali, M.U., Ahmed, W., Kallu, K.D., Zafar, A., Alduraibi, S.K., Irfan, M., Basha, M.A.A., Alshamrani, H.A., and Alduraibi, A.K., "Robust Gaussian and nonlinear hybrid invariant clustered features aided approach for speeded brain tumor diagnosis," Life, Vol. 12, No. 7, 2022, p. 1084.

[15] Sharma, A.K., Nandal, A., Dhaka, A., Polat, K., Alwadie, R., Alenezi, F., and Alhudhaif, A., "HOG transformation based feature extraction framework in modified Resnet50 model for brain tumor detection," Biomedical Signal Processing and Control, Vol. 84, 2023, p. 104737.

[16] Masud Rana, M., and Ahmed, K., "Feature selection and biomedical signal classification using minimum redundancy maximum relevance and artificial neural network," In Proceedings of International Joint Conference on Computational Intelligence: IJCCI 2018, Springer Singapore, 2020, pp. 207-214.

[17] Bulow, T., and Sommer, G., "Hypercomplex signals-a novel extension of the analytic signal to the multidimensional case," IEEE Transactions on Signal Processing, Vol. 49, No. 11, 2001, pp. 2844-2852.

[18] HS, S.K., and Karibasappa, K., "An approach for brain tumour detection based on dual-tree complex Gabor wavelet transform and neural network using Hadoop big data analysis," Multimedia Tools and Applications, Vol. 81, No. 27, 2022, pp. 39251-39274.

[19] Selesnick, I.W., Baraniuk, R.G., and Kingsbury, N.C., "The dual-tree complex wavelet transform," IEEE Signal Processing Magazine, Vol. 22, No. 6, 2005, pp. 123-151.

[20] Priyanka, M., Anusha, T., Srilakshmi, K., Chandana, R., and Srikanth, M.V., "Optimum Weighted Medical Image Fusion Using Non Sub-Sampled Contourlet Transform (NSCT) For Brain Tumour Detection," Journal of Engineering Sciences, Vol. 14, No. 04, 2023.

[21] Huang, M., Yang, W., Wu, Y., Jiang, J., Gao, Y., Chen, Y., Feng, Q., Chen, W., and Lu, Z., "Content-based image retrieval using spatial layout information in brain tumor T1-weighted contrast-enhanced MR images," PLoS One, Vol. 9, No. 7, 2014, p. e102754.

[22] Kaplan, K., Kaya, Y., Kuncan, M., and Ertunç, H.M., "Brain tumor classification using modified local binary patterns (LBP) feature extraction methods," Medical Hypotheses, Vol. 139, 2020, p. 109696.

[23] Rinesh, S., Maheswari, K., Arthi, B., Sherubha, P., Vijay, A., Sridhar, S., Rajendran, T., and Waji, Y.A., "Investigations on brain tumor classification using hybrid machine learning algorithms," Journal of Healthcare Engineering, 2022.

[24] Ayadi, W., Charfi, I., Elhamzi, W., and Atri, M., “Brain tumor classification based on hybrid approach,” The Visual Computer, Vol. 38, No. 1, 2022, pp. 107-117.

[25] Şahin, E., Özdemir, D., and Temurtaş, H., “Multi-objective optimization of ViT architecture for efficient brain tumor classification,” Biomedical Signal Processing and Control, Vol. 91, 2024, p. 105938.

[26] Brain T1-weighted CE-MR dataset, https://doi.org/10.1371/journal.pone.0102754.t001, accessed on 04/01/2025.