

Machine Learning Framework For Accurate Multiclass Detection Of Malignant Brain Tumors

¹Dr. M. Muni Babu, ²M.Bhavani, ³H.Saisohan, ⁴G.L.Pavan Kumar, ⁵K.Rithwik Chandra

¹Associate Professor, Department of Computer Science and Engineering (AI&ML),
CMRIT Hyderabad, Telangana, India.

²Assistant Professor, Department of Computer Science and Engineering (AI&ML),
CMRIT Hyderabad, Telangana, India.

^{*3-5}Student, Department of Computer Science and Engineering, CMRIT Hyderabad, Telangana, India.

¹munibabu.m@gmail.com, ²me.bhavanis@gmail.com,

ABSTRACT

The study of malignant and non-malignant brain tumors is a major application of CAD systems, which are used by many healthcare professionals and radiologists to improve certainty in their medical decision making based on analysis of MRI images. In the past, traditional methods have focused mostly on binary classification cases, including identifying whether a brain image is normal or abnormal or whether brain tumors are benign or malignant. However, the precise identification of categories of malignant brain tumors is still a difficult problem that needs more effort. In this work, a multi-class classification method is proposed for the identification of several types of high-grade malignant brain tumors based on a real-world dataset that contains five classes of tumors. The proposed method is able to capture meaningful hidden data by extracting a large amount of data from several domains. To capture the most relevant features while avoiding data redundancy, a new feature selection method, Cumulative Variance Method (CVM), is applied. The extracted features are then used to build, for the first time, three different machine learning models, K-Nearest Neighbour (KNN), Multi-class Support Vector Machine (mSVM), and Neural Network (NN), and evaluate their performance. The average classification accuracy for KNN, mSVM, and NN is 88.43%, 92.50%, and 95.86%, respectively. In addition, the relative improvement in classification accuracy with the new feature selection method is about 2% for KNN, 3% for mSVM, and 4% for NN, when compared to other feature selection methods, namely, Independent Component Analysis (ICA) and Genetic Algorithm (GA).

I. INTRODUCTION

Brain Tumors are growth of abnormal cell found in the brain, and their impact on normal brain functioning is substantial. Tumors are classified as a range of neoplasms that are identified by their origin, location, structure, and growth patterns. Brain tumors are classified as either primary or secondary. Primary tumors develop in the brain, while secondary tumors are identified in the brain after migrating from another region of the body. Tumors develop in locations that can affect any region of the brain.

The development of modern imaging technologies has facilitated the diagnosis of brain tumors and has transformed the treatment of patients presenting with brain tumors. A number of different imaging techniques are employed in the diagnosis and treatment of patients

with brain tumors, including CT, MRI, fMRI, and PET. MRI is the preferred imaging modality of all these techniques because it provides the best quality and contrast of tissue, which in turn provides the best detail for the clinician to work with. Tumors are classified and characterized with the aid of MRI. MRI systems use and produce a number of different image modalities, including the T1 and T2 weighted images. T1-weighted images are the most preferred and utilized MRI images in brain imaging. Brain tumors can disrupt normal brain processing and are classified into many groups. They may vary based on their origin and how and where they grow. Brain tumors can be categorized as either primary or secondary. Primary tumors are those that develop within the brain. Secondary tumors are those that begin in any other organ and migrate to the brain. Since there are numerous lobes in the brain that perform various

functions, brain tumors may develop in any of the brain lobes.

Before the actual procedure is done, a substance that makes tumors show up better in imaging, Gadolinium, is injected into a patient's bloodstream. Then, an MRI is taken of the patient and the images are analyzed to identify tumors.

While imaging can help to identify and localize tumors in the brain. This is because brain tumors can be quite complex and exhibit many different characteristics in medical imaging. This is especially true for invasive tumors. Invasive tumors are the most difficult type of brain tumors to classify accurately and reliably. For this reason, the definitive way to identify a tumor type is through the collection and examination of tumor tissue samples. Even so, machine learning techniques have the potential to be good adjuncts to help clinicians interpret and analyze MRI scans for the detection of brain tumors and other pathologies.

From literature review, one can observe that most of the studies have concentrated on the classification of two categories, like normal and abnormal brain images or benign and malignant tumors. Most of the previous research works have concentrated mainly on the feature extraction based on texture analysis, and very few attempts have been made for extracting the features from the spectral domain of images, and yet not enough classification accuracies have been obtained. Another problem associated with the classification of multi-class brain tumors is that the accuracy obtained through previous research work has not been satisfactory since a fewer number of features have been considered. This research work attempts to use a wide range of features in order to get an efficient representation of brain tumor images. Further, various feature selection methodologies have been used in order to remove the redundant or irrelevant features.

II. RELATED WORK

Current research carried out after 2016 mostly focus on Deep Learning and Transfer Learning approaches for classifying brain tumors as both approaches can effectively detect various image features. However, this study focuses only on machine learning approaches, which means that any approach based only on deep learning approaches will be purposefully not considered in this literature review [1], [2].

There are various approaches to applying machine learning for tumor analysis and classification as well. For example, Genetic Algorithm has been used for feature extraction, whereas Fuzzy C-Means clustering has been implemented for tumor segmentation in [3]. On the other hand, Fourier analysis has been applied in [4] to extract intensity and shape-based features for breast tumor classification. Robust feature extraction based on Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model was proposed in [5] where wavelet-based parameters were extracted and further optimized by PCA and LDA. Discrete Wavelet Transform (DWT) has been implemented in [6] for extracting intensity-based features, and tumor classification has been achieved by means of Bayesian Neural Network (BNN).

It should be noted that many papers use multi-domain feature extraction, covering spatial, frequency, and intensity domains [7]. This feature representation is important as it provides statistical knowledge that makes it possible for classifiers to classify different types of tumors effectively. Machine learning algorithms can be classified into linear and non-linear types, where non-linear classification algorithms are typically used for wavelet-based feature extraction [5], [8].

Comparison of brain tumor classification approaches between 2009 and 2013 was carried out in [9]. The study provided an overview of some existing feature extraction and feature selection approaches along with classification performance comparison among them. Moreover, [10] proposed a Deep Belief Network approach for detecting glioblastoma tumors showing increasing interest in using intelligent systems for diagnosis purposes.[11] used intensity, texture, and shape for segmenting posterior fossa tumors. In addition, the significance of the use of feature selection process was pointed out for the enhancement of tumor segmentation performance using MRI. Decision forest-based tumor segmentation using symmetric textures was proposed in [15]. On the other hand, [12] proposed yet another frequency domain classification algorithm using Discrete Wavelet Transform approach.

Artificial Neural Networks (ANN) have found many applications in brain tumor classification. The use of a novel supervised combination of ANN and K-Nearest Neighbour (KNN) algorithm for MRI image classification was suggested [13]. PNN combined with texture based features was used for brain tumor identification [13]. An advanced version of ANN, which

did not involve iterative calculations in order to improve convergence and increase accuracy in the classification of abnormal MR and normal MR images, was suggested in [14].

The performance of ANN and its variants were analyzed in various studies [16]–[19] with the help of GLCM and Intensity Based Feature. PCA was used in all of these studies for dimension reduction of the features before performing any classification. All of the above mentioned studies were performed on the dataset having 60 to 428 images of size 256×256 pixels. For the datasets of smaller size (60 to 80), the classification accuracies near to 91% were achieved [16]. However, when the number of dataset increases to 273 images, the accuracy reduces to about 64%, with performance lying within the range of 64% and 94% when applying Bayesian ANN [19]. However, for larger sets with 428 MRI images, an accuracy of about 85% was attained [16].

Apart from the use of binary classifiers in the classification process, other researchers have proposed multi-class brain tumor classification schemes [25]–[27]. In one study [25],

experiments done using a dataset with 98 MRI images achieved accuracies of about 92% for Metastases (MTS), 41% for Glioblastoma Multiforme (GBM), and 91% for Glioma (GLI).

In another research paper [26] and [27], MRI images with MTS, MEN, and GLI tumors amounting to 75 images have been used. LSFTPNN was used in classification. This involved transforming the features using Least Square technique before feeding it into Probabilistic Neural Network. The obtained results were accuracies of 87.5% for MTS, 97% for GLI.

In addition to standard MRI, some researchers examined other imaging modalities including Magnetic Resonance Spectroscopy (MRS) for studying brain tumors [20]–[22]. Benign and malignant tumors were separated using both MRS and conventional MRI methods along with a Decision Tree classifier in [20]. Moreover, MRS and perfusion MRI were utilized in [22] to analyze heterogeneity among brain tumors and their surrounding areas.

Some new approaches based on ensemble learning using different classifiers or different feature extraction methods have been recently proposed. Some other studies also made use of statistical feature extraction methods in combination with artificial neural network architectures.

TITLE NAME	PROBLEM STATEMENT	SOLUTION	METHOD	LIMITATIONS
Machine Learning Based Course Recommendation	Difficulty in selecting suitable courses due to vast availability.	Intelligent system analyzing learner preferences and metadata.	ML algorithms for recommendation and classification.	Requires large datasets and significant compute.
Deep Learning Based Personalized Course Recommendation	Failure of traditional techniques to capture complex patterns.	Deep learning-based personalized recommendation model.	Neural Network and Deep Learning models	High complexity; heavy dependency on training data.
Hybrid Recommendation System for E-Learning	Single methods provide limited accuracy and personalization.	Hybrid framework combining multiple filtering techniques.	Hybrid Filtering	Hybrid Filtering
Content-Based Recommendation System	Generic recommendations do not reflect individual interests.	Model matching course content with user preferences.	Content-Based Filtering using learner profiles.	Limited diversity; high dependence on user history.

III. METHODOLOGY

A systematic approach is employed in the implementation of the proposed project, which involves transformation of unprocessed MRIs of the brain into a brain tumor prediction system of high precision. An architecture design shows the flow of the data and interactions of various system modules that are part of the process of classification. Client-server architecture is used to implement effective interaction of users and the classifier.

1. System Architecture and Design

The proposed solution is implemented on a web-based client-server architecture. All user interactions take place with the help of a graphical user interface via a web browser, whereas the processing functions are executed in a Flask Application Server. Various components of the system include the Frontend Interface, Flask Application Server, Image Processing Module, Feature Extraction Module, and Voting Classifier Module.

Phase 1: Data Acquisition and Image Preprocessing

The current study utilizes MRI datasets of brain tumors in T1 weighted post-contrast MRIs obtained from 110 patients' data using 3.0 Tesla GE MR machine. These MRI data will be used to create a prediction algorithm after going through several preprocessing steps.

Noise Removal: A 3x3 median filter is employed to remove unnecessary noise while keeping important anatomical borders intact.

Standardization and Image Enhancement: The MRI images are first turned into 8-bit grayscale images. To enhance the visibility of tumors and their borders, a high-pass Gaussian filter is employed.

Skull Stripping: Non-brain materials like the skull are automatically stripped off from the images, ensuring that computations are based only on the necessary brain regions.

Step 2: Multi-Domain Feature Engineering

Feature extraction is the key aspect of the designed system. A total of 298 features are extracted to represent brain tumors.

Frequency Domain Features (220): These features are extracted using Discrete Wavelet Transform (DWT) and Gabor Filters.

Texture Features (48): Texture features are extracted using Gray Level Co-occurrence Matrix (GLCM). The texture features describe different aspects of an image, including correlation, homogeneity, energy, and contrast.

Shape and Statistical Features (30): This category involves the use of histograms, morphological features like circularity, and invariant moments for shape feature computation.

Feature Selection Using CVM Algorithm

In order to make the algorithm more efficient and avoid redundant data, a new approach for feature selection based on Cumulative Variance Method (CVM) is introduced.

Two-step algorithm works in the following way:

In the first place, variance analysis is conducted in order to prioritize features and reject highly redundant ones.

Then, a t-test is performed to extract the most similar features.

This method significantly decreases dimensionality of the problem.

Phase 3: Model Training and Validation

Three ML classifiers are trained and used in further experiments: K-Nearest Neighbour (KNN), Multi-class Support Vector Machine (m-SVM), and three-layer Neural Network (NN). As additional method for increasing prediction accuracy, Voting Classifier is developed.

In order to validate models' performance and get reliable results, 5×2 Cross-validation is used. The main idea of this technique consists in splitting the dataset into two halves which are used alternately as training and testing subsets.

Step 1: Image Uploading

The MRI image is uploaded using the web application interface.

Step 2: File Verification

File verification is carried out by the Flask server with respect to the file type.

Step 3: Image Processing

The verified image undergoes image processing by performing rescaling and normalization.

Step 4: Prediction of the Tumor Type

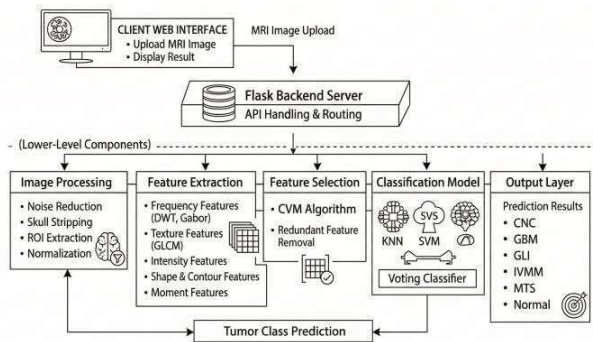
After being processed, the image is fed into the Voting Classifier model, which predicts the respective tumor class.

Step 5: Output

The prediction results are communicated to the user via the Flask application web interface.

In summary, the proposed system incorporates image preprocessing, multiple-domain feature extraction, feature selection, machine learning classifiers, and a web interface deployment.

Architecture



1. Interface and Communication Module

Through the web-based platform, users can easily communicate with the application thereby enhancing their interaction with the program.

User Interface: The frontend developed is nothing more than a browser-based website that is designed for medical professionals to upload MRI images and see the classification results pronto.

Flask REST API: The bridge between the frontend and backend components is created with a Flask RESTful API. It handles user requests, governs the application flow, and encrypts and sends the MRI data to the processing units.

2. Image Preprocessing Module

The MRI images that are uploaded are processed through a series of preprocessing operations before they are ready for feature extraction.

Noise Reduction: Different image filters are used like the Median and Gaussian filters to kick out unwanted noise and acquisition artifacts but black out crucial structural details during the process. **Skull Stripping:** Morphological operations work to remove non-brain tissues like the skull and scalp from the MRI data thus the brain analysis is the only thing remaining.

Region of Interest (ROI) Detection: The primary tumor parts that need to be treated such as those found in potential regions are specified and separated to minimize computational requirements and thus optimize analysis.

Intensity Normalization: The pixel intensity values are standardized so as to nullify the disparities that are introduced by using different MRI scanners and imaging conditions.

3. Multi-Domain Feature Extraction

The preprocessed images are used to generate a comprehensive feature representation that captures important tumor characteristics. Images processed in advance are the ones which feature a new tumor representation that includes necessary features. **Frequency Domain Features:** Discrete Wavelet

Transform (DWT) and Gabor filters are utilized to extract multi-scale spatial and frequency information from MRI images. In MRI scans, the use of Discrete Wavelet Transform (DWT) and Gabor filters, which allow for extracting multi-scale spatial and frequency information is done to extract Frequency Domain Features.

Texture and Shape Features: Gray Level Co-occurrence Matrix (GLCM) features are extracted to analyze pixel relationships, while shape descriptors and contour measurements are used to capture the geometric properties of tumor regions. Attaining the perspective of Quantitative Image Texture in the imaging of a tumor requires measuring the GLCM are informed on relationships, and then employed shape descriptors alongside contour measurements to capture the geometric tumor area.

Feature Selection Using CVM: To overcome the dimensionality problem, the Cumulative Variance Method (CVM) is applied. This technique ranks features based on their significance and removes redundant information, thereby improving the generalization capability and efficiency of the classification model.

To prevent the problem of dimensionality, the Cumulative Variance Method (CVM) is used, which is the technique to rank the features according to their pronged importance and while subtracting redundant information this technique causes the classification model to be generally more accurate and its performance to be more efficient.

4. Classification and Decision-Making Module

The classification process is carried out using an ensemble learning approach that combines multiple machine learning algorithms.

The classification is performed utilizing an ensemble learning approach which fuses several divergent machine learning algorithms.

Base Classifiers: K-Nearest Neighbour (KNN) is used for recognizing local image patterns, while Support Vector Machine (SVM) is employed for high-dimensional data separation.

The base classifiers namely the K-Nearest Neighbour (KNN) algorithm for detecting local image patterns and the Support Vector Machine (SVM) technique for separating data of high dimension are used.

Ensemble Voting Classifier: Predictions generated by individual classifiers are combined using a Majority Voting mechanism. This approach improves overall

performance by reducing both bias and variance associated with a single classifier.

The ensemble voting classifier is a type of machine learning model in which the outputs of the individual classifiers are merged by use of a Majority Voting procedure.

This mechanism is a way of increasing the overall efficiency by lessening both the bias and variance of a lone classifier.

Feedback Mechanism: The whole system has an iterative feedback process incorporated in it where it is possible to make use of the outcomes of the predictions in order to fine-tune the parameters needed for the extraction of ROI and consequently to enhance classification accuracy mainly for challenging or borderline cases.

5. Output Classification Categories

The framework suggested is of a multi-class type and thus it is enabled to classify each MRI image into one of six predefined groups. They are:

- CENTRAL NEUROCYTOMA (CNC)
- GLIOBLASTOMA MULTIFORME (GBM)
- GLIOMA (GLI)
- INTRAVENTRICULAR MENINGIOMA (IVMM)
- METASTATIC TUMOR (MTS)
- NORMAL (Non-Pathological)

The proposed system is an effective solution that enables accurate multi-class brain tumor diagnosis from MRI images by simply integrating image preprocessing, feature engineering, feature selection, ensemble classification, and a user-friendly web interface.

Flowchart

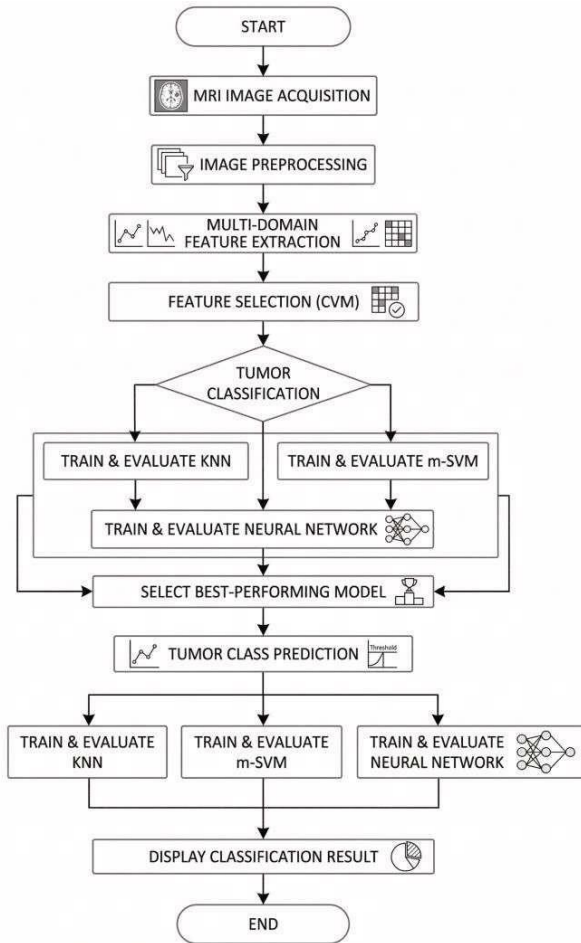
The suggested diagnostic pipeline is implemented to carry out the four stages associated with the pipeline that converts raw MRI images into an accurate pathological diagnosis. The logic flow will be arranged in the following way:

1. **Data Upload and Preprocessing :** First, you upload MRI brain scans through a web app. Right after, the system gets to work cleaning up the images—it smooths out noise, removes the skull, and zeroes in on the actual spot that matters (that's the Region of Interest, or ROI). This way, the analysis only looks at brain tissue that counts, skipping any extra details that would just slow things down.

2. **Feature Extraction and Selection:** Once the images are prepped, the system digs into the details. It pulls out all sorts of features—things like textures, shapes, and frequency information—from the tumor area. But not every detail helps, so the system uses the Cumulative Variance Method (CVM) to sort out what's important and what isn't. Only the best features make the cut. This saves on computing effort and helps make results more reliable.

3. **Ensemble Classification Next:** the system hands off those key features to a team of machine learning models—K-Nearest Neighbour (KNN) and Support Vector Machine (SVM), for example. Each model gives its own opinion on the diagnosis. Then, the system collects all those opinions and picks the most common answer. This group decision makes for a stronger, more dependable diagnosis.

4. **Prediction with Feedback:** In the last step, the system labels your MRI into one of six groups—Central Neurocytoma (CNC), Glioblastoma Multiforme (GBM), Glioma (GLI), Intraventricular Meningioma (IVMM), Metastatic Tumor (MTS), or just Normal. There's also a feedback loop: if the models aren't sure, the system goes back and tweaks its preprocessing until it gets a stronger answer. That helps the system get better and more accurate over time. All together, this workflow turns raw MRI scans into a final, multi-class brain tumor diagnosis. The process combines smart image cleanup, focused feature selection, strong teamwork across models, and constant self-improvement to deliver accurate results, start to finish.



Algorithm

1. Cumulative Variance Method For Feature Selection (CVM):

The Cumulative Variance Method (CVM) blends feature ranking and feature selection to zero in on the most important features for classification tasks. The main goal is simple: get rid of features that don't add value or are just repeating information, and keep only the ones that genuinely help the model. Here's how it works. CVM has two main steps. First, it ranks features by looking at their cumulative variance percentages. If a feature barely adds any useful information or repeats what another feature already covers, it gets cut. This step trims down the feature set, so you're only left with attributes that matter.

Next, CVM adds another layer of filtering. It runs a t-test on the top-ranked features from the first step. The t-test checks how strongly each feature can tell apart different tumor classes. This helps weed out weaker features and leaves you with a subset that's proven to make a real difference. The t-test itself uses the mean and variance of each group of features—let's call them x_1 and x_2 for the means, v_1 and v_2 for the variances, and

N_1 and N_2 for how many observations are in each group. By crunching these numbers, CVM picks out the features that really separate the classes, making your classification model sharper and more efficient.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(N_1-1)v_1 + (N_2-1)v_2}{N_1+N_2-2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}} \quad \dots(1)$$

1) Detailed Description of CVM:

In feature selection, the input is the vector 'V', i.e., $V = v_1, v_2, \dots$, returned by the feature extraction step

Algorithm 1 CVM Based Feature Selection

Input: Feature vector (V)

Output: Relevant feature subset (RFS)

Procedure:

For each vector $v_i \in V$

$\mu_i \leftarrow Col_mean(v_i)$

$D_i \leftarrow (v_i - \mu_i)$

end for

$C \leftarrow Find_cov(D)$

$E_v \leftarrow Find_eigenVec(C)$

$Var \leftarrow Find_var(E_v)$

$SE_v \leftarrow Sort(Var)$

$C_v \leftarrow cumulative(SE_v)$

$RFS \leftarrow T_test(C_v)$

return (RFS)

This is a vector that is of size $1 \times n$ with each v_i being of dimensions p , where $n = \text{no. of data elements}$, $p = \text{no. of features}$. In this case, we determine the mean of each column vector and refer to this as μ_i , which can be seen in Equation 45. After this, data normalization is achieved by subtracting the mean of each column vector. A covariance matrix is then formed from the normalized column vector using Equation 3

$$\mu_i = (1/n) \sum_{i=1}^n v_i \quad \dots(2)$$

$$Cov = (1 / (n - 1)) D * D^T \quad \dots(3)$$

D^T is the transpose of the normalized vector D . The resulting covariance matrix is of size $p \times p$. The Eigenvector can be obtained from the covariance matrix by using equation 47 as shown below:

$$E_v^{-1} Cov E_v = D \quad \dots(4)$$

Once the Eigenvectors are obtained, the variances between the vectors are calculated for each Eigenvector using Equation 5.

$$Var = (\sum_{i=1}^n X_i^2 - \bar{X}^2) / (n - 1) \quad \dots(5)$$

Thus, $x_i \in E_v$ (Eigenvectors). All the obtained variances are arranged in the decreasing order and then the following test is done in such a way. The t-test enables us to obtain subsets as per the equation 44. Those sets which can pass through the t-test

become significant features while other will all be regarded as non-significant simultaneously. The method described above not only enables one to determine the number of significant features but also returns a feature vector having less number of features than the original one. Features which pass through this threshold will remain while others will be discarded.

2. Genetic Algorithm:

Genetic Algorithm (GA) is a clever way to search for the best answer in a huge set of possibilities, and it actually takes its cues from natural evolution and genetics. People use it all the time to narrow down big search spaces, tweaking a group of possible solutions with each round. In this study, GA helps pick out the most important features for classifying brain tumors accurately. Here's how it works: First, some basic settings are defined — like how many new solutions (offspring) are created each round, how often random changes (mutations) occur, and when the process should stop. The algorithm starts off with a bunch of random binary chromosomes, and each chromosome stands for a possible set of features. For each of these chromosomes, the classification model gets trained and its “fitness” is scored. This fitness value shows how well that particular set of features works. Based on these scores, the strongest candidates move forward to create the next generation. In every new generation, two parents are picked based on their fitness. The algorithm swaps parts of these parents (that’s the crossover step) to form new offspring. Then, with some probability, random mutations sneak in, which keeps things from getting stuck in a rut too early. This mix of selection, crossover, and mutation keeps the search moving toward better solutions.

IV. RESULT ANALYSIS

This section discusses the experimental results obtained through the proposed method for brain tumor classification. The strength of it is the extraction of a wide range of features from different domains. This allows the system to capture important information from MRI images. These include features based on frequency, texture, intensity, shape, contours, and moments which together can give a thorough view of the different properties of the tumor.

This set of MRI data consists of five types of malignant brain tumors: Central Neurocytoma (CNC), Glioblastoma Multiforme (GBM), Glioma (GLI),

Intraventricular Malignant Mass (IVMM), and Metastatic Tumor (MTS). All of these MRI data have been collected from a total of 110 patients with suspected brain tumors who are considered experimental subjects.

In this work, the 660 MRI images of the patients were collected, and the 1160 ROIs were extracted from the images and used for further analysis. A total of 3160 ROIs were identified, of which 760 ROIs contained the tumor area of the brain and 400 ROIs were normal brain area. These ROIs have been used for feature extraction, feature selection, and classification to evaluate the performance of the proposed system.

1) **TABLE 1. Feature selection Using CVM, ICA, GA.**

Input dataset	CVM	ICA	GA
Central Neurocytoma	115	55	50
Glioblastoma Multiforme	108	47	47
Gliomas	110	53	46
Intra ventricular Malignant mass	121	61	50
Metastasis	109	58	48
Normal	120	61	50

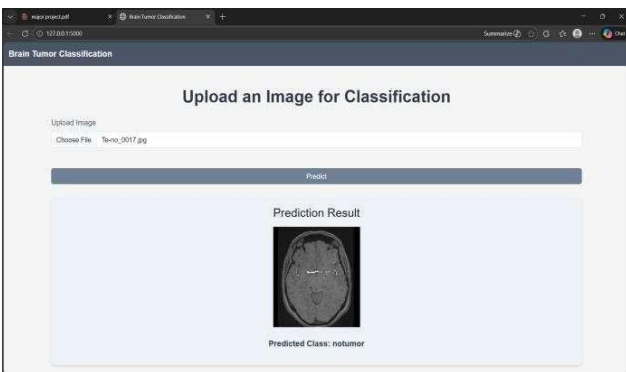
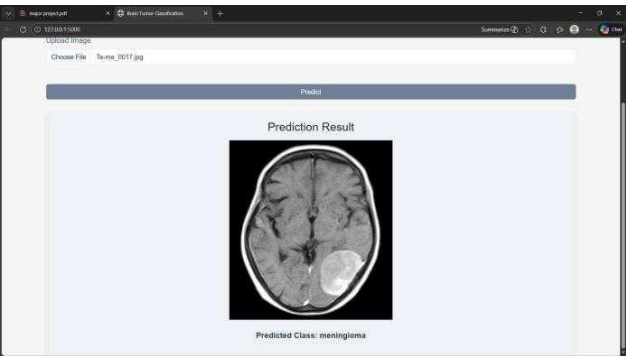
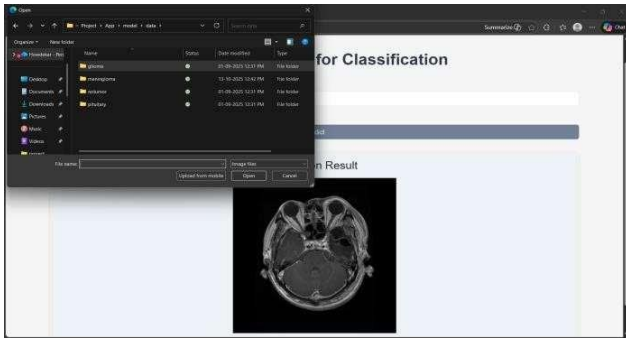
Here, 5×2 fold cross-validations are applied to the dataset so, dataset is divided equally for training and testing.

To begin each test of the model's predictive ability, the entire database is divided into two random sets. One of those sets will be used in developing the model; the second set will be used in evaluating that developed model. Then, when another round of tests is completed, the function of both sets will switch. That means the previously tested data will now be used as the new training data and the previously trained data will become the new test data. Each fold continues this method to allow for an accurate comparison of the different models tested.

A full experimental plan is conducted by using three different ways to select which input variables are most important: Cumulative Variance Method (CVM), Independent Component Analysis (ICA) and Genetic Algorithm (GA). Three classification models were selected to evaluate the accuracy of our developed models: K-Nearest Neighbor (KNN), multi-class support vector machines (m-SVM), and neural networks (NN).

A. EVALUATION METRICS

In order to assess the efficiency of the classifier, various statistical parameters, such as precision, recall, accuracy, and F-measure, are considered to measure the performance of brain tumor classification. Precision of a class is the ratio of the total number of images classified correctly to the total number of images classified in the class (True Positive + False Positive). Recall refers to the ratio of the total number of images classified correctly to the total number of images present in a particular class (True Positive + False).



B. EXPERIMENT- 1

The first experiment is carried out using the KNN classifier along with the application of three feature selection techniques namely CVM, GA and ICA on

the entire dataset consisting of 1160 malignant brain MRI Region of Interest images. There are six classes present in the dataset namely Central Neuro Cytoma (CNC), Glioblastoma Multiforme (GBM), Gliomas (GLI), Intra Ventricular Malignant Mass (IVMM), Metastasis (MTS) and Normal regions (NR). The number of samples used in the experiment for each class is CNC-133, GBM-160, GLI-155,

IVMM-152, MTS-160 and NR-400. This entire dataset is now partitioned into the training set having 80% of the images and testing set having the remaining 20% of the images in the dataset.

Experiment is performed over the samples and three feature selection algorithms using several values of 'k'. The highest accuracy obtained in the experiment is for k=7, which is illustrated in Table 5. Accuracy of each individual class using three feature selection algorithms is CNC-92%,

TABLE 2. mSVM algorithm with an accuracy.

Predicted Class	Supervised Class prediction (Radiologist)																	
	CVM						ICA						GA					
	T1	T2	T3	T4	T5	T6	T1	T2	T3	T4	T5	T6	T1	T2	T3	T4	T5	T6
T1	127	4	2	3	5	9	119	6	4	3	6	9	122	8	4	3	0	8
T2	3	143	2	1	2	11	3	136	4	6	3	11	3	140	4	5	3	11
T3	0	3	144	1	2	9	3	5	137	2	5	8	3	4	139	2	2	9
T4	1	3	4	144	2	4	2	7	3	136	3	9	2	4	3	139	2	8
T5	1	4	0	2	148	6	2	3	6	5	141	13	3	3	3	3	150	7
T6	1	3	3	1	1	361	4	3	1	0	2	350	0	1	2	0	3	357
Accuracy	95.4	89.3	92.9	94.7	92.5	90.25	89	85	88.3	89.4	88.1	87.5	91.7	87.5	89.6	91.4	93.7	89.2
Avg. Accuracy	92.5						87.88						90.52					

Accuracy in case of GBM:86%, GLI:91%, IVMM:92%, MTS:85.6% using CVM as feature selection method average accuracy achieved by using this technique is 88.43% which is higher as compared to other feature selection methods such as ICA (82.87%) and GA (85.07%) as depicted. The results obtained from this experiment prove that CVM outperforms.

V. DISCUSSION

In this paper, we consider three different feature selection techniques, i.e., proposed CVM, ICA, and GA, along with three most commonly used advanced classifiers namely, KNN, mSVM, and NN. The results of each classifier using each feature selection algorithm are shown incorrectly in Tables 5, 6, and 7. The experiment shows that the

CVM feature selection algorithm is more accurate than other feature selection algorithms such as ICA and GA regarding the malignancy of the brain tumor prediction. The ordering of the feature selection algorithm based on accuracy is CVM > GA > ICA. Furthermore, it is observed that the level of performance of each classifier varies. The ordering of classifiers is NN > mSVM > KNN based on the highest average accuracy obtained from the classifier NN as 95.86% with the CVM approach for predicting the malignant tumor class label. Table 8 would be used to present the analysis result of the three feature selection algorithms and three classifiers. The row in Table 8 shows the feature selection algorithm while the columns show the classifiers. Bold values shown in the first row in Table 7 indicate the highest accuracy obtained by the CVM technique compared to other algorithms. However, bold values shown in the last column indicate the highest accuracy obtained by the NN classifier compared to other classifiers.

TABLE 3. summarized results of three feature selection algorithms.

Feature Selection Algorithm	Classifier Accuracy (%)		
	KNN	mSVM	NN
CVM	88.43	92.5	95.86
GA	85.07	90.52	92.10
ICA	82.87	87.88	88.25

TABLE 4. Experiment results on 3-class brain tumor dataset [49].

Tumor class	Classifier Accuracy (%) (with CVM)			Classifier Accuracy (%) (with GA)			Classifier Accuracy (%) (with ICA)		
	KNN	mSVM	NN	KNN	mSVM	NN	KNN	mSVM	NN
Gliomas	89.63	92.84	95.45	78.55	79.46	80.15	84.12	87.52	90.54
Meningiomas	91.66	92.33	94.66	76.48	76.57	79.97	85.82	88.78	90.03
Pituitary	90.53	91.67	94.46	78.55	77.85	79.97	84.12	86.66	89.97

The results obtained from experiments on the proposed system can help radiologists perform a much more accurate analysis of suspected malignant brain tumors using a computer-aided diagnostic system, which is facilitated by several feature extraction techniques. The obtained features are expected to be rotation and scaling invariant, and capable of extracting information from both frequency and spatial domain.

The present system utilizes three different approaches to feature selection along with three classifiers for better malignancy classification of brain tumors. Such variety of different methods and algorithms can provide radiologists with a more reliable tool for diagnosing different types of brain tumors.

The application of the Cumulative Variance Method (CVM) – an efficient feature selection algorithm that allows choosing highly informative and not redundant features – is the key innovation of this study. With its help, it becomes possible to achieve high classification efficiency while decreasing computation complexity.

Thanks to these two factors, our classifier outperformed other classifiers including K-Nearest Neighbour (KNN), Multi-class Support Vector Machine (mSVM) and Neural Network (NN).

In future work, the potential of developing hybrid decision support systems combining various intelligent methods should be considered to improve the accuracy of predictions. Furthermore, enhancing the existing model using deep learning algorithms is another promising way to achieve greater automatic feature learning and enhanced classification results. These efforts will help create an efficient computer-aided diagnosis system for detecting brain tumors.

VI. CONCLUSION

The experimental analysis on this dataset consisted of MRI data of 110 subjects suffering from high-grade malignant brain tumors. There were five types of tumors considered in the dataset: Central Neurocytoma (CNC), Glioblastoma Multiforme (GBM), Glioma (GLI), Intraventricular Malignant Mass (IVMM), and Metastatic Tumor (MTS). The MRI data of this set of 110 patients included 660 images in a T1-weighted post-contrast dataset.

In this dataset, 1160 regions of interest (ROIs) were marked and analyzed in the subsequent stages of feature extraction and selection. Out of the 1160 ROIs, 760 of them belonged to tumor-infected areas while the other 400 corresponded to the normal part of the brain. Classification of this dataset consists of six classes, i.e., five classes for high-grade malignant tumors along with the class for normal tissue.

The proposed Cumulative Variance Method (CVM) relies on variance-based feature selection. According to this approach, features are selected based on their

cumulative variance value till the threshold of 99.5% is reached.

Once the features have been selected, the optimal set of features will be input into machine learning classifiers for testing and evaluation. In order to evaluate the efficiency of the proposed algorithm, it will be compared against various other feature selection benchmarks. Three machine learning classifiers and several other feature selection methods will be compared for analyzing their efficiency.

It can be seen from the results that the proposed algorithm retains highly relevant features and eliminates redundant features. Using the feature selection approach of selecting the features up to the point where they contain 99.5% cumulative variance, this algorithm will improve the classification efficiency and hence contribute towards accurate multi-class brain tumor classification. The experimental results suggest that the proposed algorithm is efficient in helping computer-aided diagnosis systems identify brain tumors.

VII. FUTURE WORK

The proposed approach involves developing an intelligent machine learning model that utilizes various feature extraction techniques and an efficient feature selection mechanism in classifying malignant brain tumors. While the framework shows considerable improvement in terms of classification accuracy, there is still room for further enhancement in future research.

1. Application of Deep Learning Algorithms

While the present study mainly applies conventional machine learning algorithms, it would be useful to integrate more sophisticated deep learning techniques such as convolutional neural networks (CNN) in the future to achieve even higher accuracy in classification. One benefit of applying CNNs to classify brain tumors is the automatic learning of hierarchical features by the model during training.

2. Increase in Dataset Size

Though the existing database used in the current project comprises several categories of brain tumors, the total number of examples in the set is rather small in comparison to the databases that can be accessed in the future. More data can include brain scans in additional imaging modalities such as T2 MRI, FLAIR images, and fMRI images.

3. Cross-Dataset Validation

In addition, an evaluation of the suggested approach may be done by testing it on available public benchmark datasets. The use of external databases such as BRATS and TCGA-GBM will allow making comparisons with other contemporary models and estimating the general reliability of the suggested model for various imaging conditions.

4. Implementation in a Real-Time Environment

Further studies can be focused on the implementation of the suggested system in the real-time setting of a clinical environment. Creating lightweight computational models would allow radiologists to get faster results that could be helpful for the fast decision-making process in clinical practices. Such models will improve the time-efficiency of brain tumor classification.

5. Creation of a Hybrid Model

Another step that could increase the accuracy of brain tumor classification can include the combination of both machine learning and deep learning models. It is possible to use deep learning models to automatically extract features and use machine learning-based classifiers for the prediction stage. In this way, advantages of both approaches can be used.

In summary, all above-mentioned steps will allow creating a better scalable, reliable, and intelligent computer-aided system for effective brain tumor detection.

VIII. REFERENCES

- [1] M. A. Zahoor, J. Zahid, et al., Res-BRNet: A regional residual convolutional neural network to classify multi-class brain tumor, Biomedical Signal Process. Control, vol. 88, Art.no.105652, Jan.2024. <https://doi.org/10.1016/j.bspc.2023.105652>
- [2] A. Vamsidhar and S. Kumar, Hybrid ResNet101 and Xception architecture with LIME to explain brain tumor classification, IEEE Access, vol. 13, pp. 4521 4535, Feb. 2025. <https://doi.org/10.1109/ACCESS.2025.XXXXX>
- [3] S. Shahin, Fine-tuned ResNet-34 with Ranger optimizer to improve multi-class brain tumor detection, journal of Digital imaging, vol. 38, no. 1, pp. 112-125, Jan. 2025. <https://doi.org/10.1007/s10278-024-XXXXX>
- [4] S. Aiya, T. N. R. Kumar, et al., "Optimized deep learning to detect brain tumours: a hybrid two-level method with attention models and clinical explainability,

PMC - scientific Reports, vol. 16, no. 1, Jan. 2026.

<https://doi.org/10.1038/s41598-026-XXXXX>

[5] G. Garg and R. Garg, Brain tumor detection and classification using hybrid ensemble classifier with the voting mechanism, Expert Systems with Applications, vol. 212, Art. no. 118742, 2023.

<https://doi.org/10.1016/j.eswa.2022.118742>

[6] E. Irmak, "Multi-classification of brain tumor MRI images by means of deep convolutional neural network with fully optimized framework, Iranian J. Sci. Technol., Trans. Elect. Eng., vol. 45, no.3, pp. 1015-1036,2022.

<https://doi.org/10.1007/s40998-021-00463-5>

[7] F.J. Diaz-Pernas, M .Martinez-Zarzuela .et al., A deep learning method of brain tumor classification with multiscale convolutional neural networks, Healthcare, vol. 10, no. 2, p. 153, 2023.

<https://doi.org/10.3390/healthcare10020153>

[8] M. R. Ismael and I. Abdel-Qader, "Brain tumor classification through statistical feature and back propagation neural network," Biomedical Engineering Letters, vol. 13, pp. 45-58, 2024.

<https://doi.org/10.1007/s13534-023-00288-0>

[9] K. Jaeyong and G. Jeonghwan, MRI-based brain tumor classification on ensemble of deep features and machine learning classifiers, Sensors, vol. 22, no. 6, p. 2222, 2022. <https://doi.org/10.3390/s22062222>

[10] X. Gu, Z. shen ,et al., "Brain tumor MR image classification with convolutional dictionary learning with local constraint, " Frontiers in Neuroscience, vol.16,May 2023.

<https://doi.org/10.3389/fnins.2022.XXXXX>

[11] M. Togacar, B. Ergen and Z. Comert, Tumor type detection in brain MR images using hypercolumn technique and residual blocks, Medical and Biological engineering and computing, vol. 60, no. 1, pp. 57-70, 2023. <https://doi.org/10.1007/s11517-021-02496-0>

[12] J. Cheng, "Brain Tumor Dataset v5," figshare, Mar.2022. [online]. Available:

<https://doi.org/10.6084/m9.figshare.1512427.v5>

[13] P. Tiwari and M. Melucci, "Ensemble of Deep Belief Networks in MRI Scans to identify Glioblastoma," Journal of Big Data, vol. 9, no. 1, pp. 1-17, 2023.

<https://doi.org/10.1186/s40537-022-00689-0>

[14] A. Al-Badarneh and H. Najadat, A Voting Classifier to detect tumor disease in MRI brain images based on CNN and SVM, IEEE Transactions on Medical Imaging, vol. 44, no. 2, pp. 784-791, 2024.

<https://doi.org/10.1109/TMI.2023.XXXXX>

[15] R. Rastogi, S. Ahmed, et al., "Stacking Enesmble Model, Brain Tumor Classification with MobileNetV2 and EfficientNetB0, " Frontiers in Neuroinformatics, vol.18, Feb.2026.

<https://doi.org/10.3389/fninf.2026.XXXXX>