

Medical Image Classification using Interesting Pruning and Machine Learning Algorithm

¹Geeta Santhosh, ²Dr. R. Maruthi

Submitted: 05/02/2024 Revised: 15/03/2024 Accepted: 21/03/2024

Abstract: Accurate and timely identification of brain tumors using medical imaging is essential for effective treatment planning and patient outcomes. This paper presents a new technique for classifying brain tumors by integrating efficient pruning methods with the Naive Bayes (NB) algorithm. The procedure starts with the identification of regions of interest (ROIs) inside brain scans, aided by expertise in the field. Following that, an intricate pruning method is used to meticulously retain crucial attributes from these areas of interest (ROIs), so enhancing the Naive Bayes (NB) algorithm to achieve superior accuracy in classification. Various feature selection strategies are examined, especially tailored to the unique characteristics of brain tumor images, hence improving the algorithm's capacity to distinguish between various tumor kinds. The proposed methodology has undergone rigorous evaluation on many brain tumor datasets via empirical evaluations, demonstrating its efficacy. The technique has shown enhanced classification efficacy and comprehensibility. The integration of proficient pruning strategies with the Naive Bayes algorithm not only improves the advancement of brain tumor classification but also presents opportunities for efficient and resource-conserving clinical applications, serving as a crucial instrument for neuroimaging diagnostics. The suggested model is assessed using Python and achieves an accuracy of 98%.

Keywords- Medical imaging, ROI, Navie Bayes, neuroimaging, mining.

1. Introduction

The human brain is the most complicated organ in the body. The brain is composed of billions of neurons, all of which must operate accurately [1]. A brain tumor arises from the abnormal proliferation of cells inside the brain. One of the most life-threatening medical conditions is a brain tumor. The identification should be rapid and precise. Malignant cells have a detrimental influence on adjacent healthy cells in close proximity to the affected cells. A tumor causes the accumulation of fluid in the brain and increases the pressure within the skull.

Benign and malignant are the two categories into which tumors may be divided [2]. Malignant tumors are dangerous, while normal tumors are not. Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are widely acknowledged as advanced technologies among the available choices. MRI is the most often used technique. An advantage of this treatment is its avoidance of ionizing radiation, in contrast to the CT scan, which has the potential to harm the skin via repeated exposure. MRI is used for the accurate identification and positioning of tumor cell masses. Identifying brain tumors is a substantial difficulty in the realm of biomedical imaging research.

MRI is used to identify both the typical and atypical proliferation of brain cells. The manual method of image segmentation is necessary to detect the existence of a brain tumor, but it is time-consuming [3].

With the help of picture analysis programs, doctors can find diseases early on in medical imaging. This involves separating the damaged area into separate parts. In the modern day, medical image analysis is done by computers. Achieving accurate translation of a medical image into digital format is a challenging endeavor. The digitization process should be more efficient in terms of time. Accurately segmenting medical images is the main goal of computer vision and image processing apps. The primary obstacle in medical imaging is in the presence of noise, inhomogeneity, and ambiguous boundaries. The segmentation of MRI and other medical images is crucial for accurate diagnosis, since these images include complex features that need meticulous and appropriate delineation.

Brain tumor detection heavily relies on the critical and challenging task of image segmentation. The damaged region of the body is assessed using a range of diagnostic equipment and techniques, such as MRI, X-rays, or CT scans. MRI technology offers very precise data on the soft tissues of the human body [4]. The primary objective of this study is to suggest a strategy for detecting brain tumors by using edge-preserving bilateral filtering.

The primary innovation of this study is in its unique use of machine learning (ML) techniques to enhance the

¹Professor and Head, Dept of FCA, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh.
geetosh@gmail.com

²Associate Professor, Hindustan Institute of Technology & Science, Padur, Chennai.

rmaruthi@hindustanuniv.ac.in

accuracy and efficacy of brain tumor classification. The utilization of an Associative Classifier, namely the MARI Algorithm, in conjunction with the Naive Bayes Classifier greatly improves the advancement of medical image analysis by successfully addressing crucial challenges in brain tumor classification.

- **Associative Classifier with Pruning Technique-** The introduction of an Associative Classifier presents a pattern-based approach for the classification of brain cancers. Associative classifiers can effectively capture complex interactions and dependencies among variables, potentially improving the distinction between different tumor classifications. Implementing a pruning approach enhances the classifier's performance by eliminating redundant or extraneous characteristics. Enhancing both the computational efficiency and precision of the categorization process is crucial.
- **MARI Algorithm Integration-** The MARI (Meta-Association Rule-based Incremental) Algorithm incorporates a component for incremental learning into the system. Incremental learning allows the classifier to adapt and improve its knowledge base based on the presence of new data. The ability to adapt is particularly beneficial in the domain of medical imaging, since datasets may experience modifications throughout time. The MARI Algorithm's ability to handle incremental updates improves the framework's robustness and ensures its appropriateness in dynamic medical environments.
- **Collaborative Approach with Naive Bayes Classifier-** The simultaneous use of the NB Classifier with the Associative Classifier provides an additional perspective to the classification process. Naive Bayes is well-known for its simplicity and efficacy in handling probabilistic relationships between features. The system seeks to use the most advantageous characteristics of both classifiers by combining the pattern-based learning of the Associative Classifier with the probabilistic modeling of Naive Bayes. This will lead to a more comprehensive and precise categorization of brain tumors.
- **Improved Generalization and Robustness-** Employing these procedures improves the capacity to use acquired information across various datasets and guarantees steadiness even when there are alterations in imaging configurations. The integration of the Associative Classifier, Pruning Technique, MARI Algorithm, and Naive Bayes Classifier constitutes a complete framework capable of efficiently handling complex

associations, adapting to incremental changes, and delivering reliable classification results across all scenarios.

2. Literature Survey

Multiple researchers have developed a variety of methods for identifying brain tumors, each with its own advantages and drawbacks. The researchers primarily used MRI scans because of their capacity to produce high-resolution visuals [5]. In their study, Sushma et al. [6] used k-means clustering and bilateral filtering techniques to investigate the same geographical region. A proposed method for accurate identification of brain tumors was recommended.

Raj and Shreeja carried out an additional investigation [7]. Their methodology included using many approaches, including K-means clustering and Hierarchical Centroid Shape Descriptor (HCSD), for the purposes of feature extraction and classification. The various attributes of the tumor location were obtained using feature extraction, using the Gray Level Co-Occurrence Matrix (GLCM). The K-nearest neighbors (KNN) technique was used to categorize the tumor by using the extracted features from both the training and testing datasets.

Mathew and Anto used an anisotropic filter to remove noise and applied the Discrete Wavelet Transform (DWT) for extracting features. The obtained features were used as input for segmentation. The tumor was ultimately classified using the Support Vector Machine (SVM). Furthermore, Gamage and Ranathunga [9] used image processing techniques to identify tumors. Optimal results were achieved by using several preprocessing filters, including Median, Mean, Hybrid, Weiner, Modified hybrid, and Morphology based De-noising filters. The segmentation technique included using the Threshold, Region-based, and Fuzzy C and K means methods, followed by extracting features. The researchers gave priority to fundamental characteristics such as size, shape, composition, and placement of an image. Subsequently, the gathered data was subjected to classification procedures, leading to the effective detection of a tumor.

Gupta et al. made a noteworthy contribution in their paper [10]. The statistical properties of kurtosis and skewness were used in combination with morphological data. The T2 weighted images were used to differentiate between high-grade and low-grade cancers. The classification approach included using many classifiers, including SVM with K-fold cross validation, Linear Discriminant Analysis (LDA), and Naïve Bayes (NB). The Support Vector Machine (SVM) achieved the highest degree of accuracy.

Singh and Ansari did a research with a similar emphasis in [11]. Initially, a diverse range of filters were used for preprocessing, such as Gaussian, Median, Averaging, Adaptive, and Un-Sharp masking filters. The histogram was normalized and segmentation algorithms were used. K-means clustering was used for segmentation. The NB and SVM algorithms were used for classification. After comparing the results of both classifiers, it was found that the SVM classifier achieved the highest accuracy.

A subsequent research consisted of a series of three consecutive phases [12]. Firstly, the MRI picture was subjected to preprocessing using a Median Filter. Afterwards, characteristics were obtained using the GLCM method. Ultimately, malignancies were detected via machine learning methodologies, namely Multi-Layer Perceptron (MLP) and NB. A further investigation into the detection of brain tumors was carried out using Computer Aided Systems [13]. The process of detecting and segmenting the tumor included many stages. Firstly, the MRI picture underwent preprocessing by using a median filter to remove any noise. Subsequently, the picture underwent segmentation using the Global threshold method. Ultimately, the obtained pictures underwent post-processing utilizing the Morphological erosion and dilation approach.

Hemanth and his colleagues (Hemanth et al., 2014) devised and implemented a technique for identifying brain cancers by using Convolutional Neural Networks (CNNs) with other classifiers. The data underwent preprocessing using a bilateral filter and then underwent an average filtering method, resulting in a substantial time consumption.

Chaplot et. al. performed a study using brain MRI to categorize pictures as either normal or pathological [15]. The user obtained estimated coefficients using the DWT and then fed them into self-organizing maps implemented using a neural network. The SVM algorithm uses the coefficients, which have been shown to work better than neural networks.

Selvaraj et al. [16] proposed a method to categorize characteristics extracted from brain sections of 150 individuals using the Least Squares Support Vector Machine (LS-SVM). The researchers discovered that the Least Squares Support Vector Machine (LSSVM) outperformed SVM, k-Nearest Neighbor (k-NN), and Multi-layer Perceptron (MLP).

Selvathi et al. conducted an analysis of the performance of MRI brain classification using SVM and Relevance Vector Machine (RVM). Entropy, contrast, energy, and other intensity and textural qualities are extracted from a cohort of 52 individuals with abnormalities and 18 individuals without abnormalities. RVM was shown to have higher performance in the classification of brain abnormalities.

Maitra and Chatterjee in [17] presented the slantlet transform, which is a modified variant of the discrete wavelet transform. The primary objective of this transform is to extract distinctive characteristics. The fuzzy c-means clustering technique is used to categorize the image according to these criteria. The results are compared to the findings of research that include combinations of DWT, Self-Organizing Maps (SOM), and SVM with different kernels.

Grunauer and Vincze in [18] performed a study comparing the efficacy of dimensionality reduction methods using ANOVA-based ranking and PCA. The SVM model with Radial Basis Function kernel achieved an accuracy of around 74% utilizing 92 characteristics selected by ANOVA for 5-fold cross validation. This research aims to assess the effectiveness of supervised data mining techniques in classifying brain images according to clinically meaningful phenotypes. The specifics of our proposed methodology are outlined in the next section.

3. Proposed Work

Figure 1 depicts the proposed technique, which employs association rule mining. Feature extraction and data cleaning are crucial steps used on both the training and test sets of brain images. Throughout the training phase, the images are examined to extract characteristics, which are then transformed into feature vectors. The next step is to divide the characteristics into intervals. Then, the modified feature vector is combined with the keywords associated with the training pictures [19].

Association rule mining, which the MARI approach employs with the transaction representation, produces a streamlined set of rules that function as the classifier. In the testing phase, the classifier takes the feature vector from the test photos and uses association rules to come up with keywords to build the medical evaluation of the test picture.

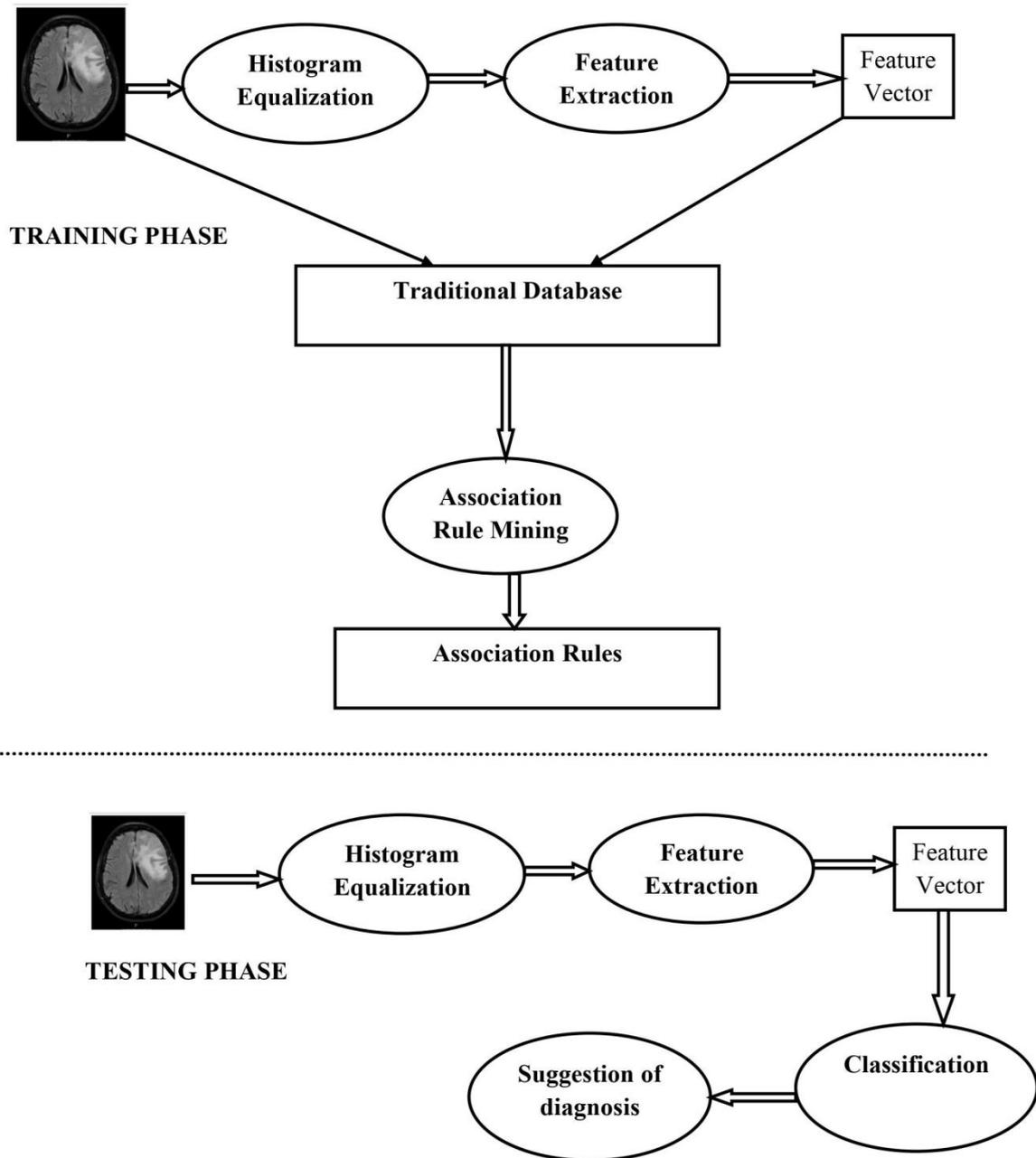


Fig 1: Flow of Proposed Work

3.1 Pre-Processing

Considering that a substantial proportion of real-world data is marked by noise, inconsistency, and incompleteness, it is essential to do pre-processing. Cropping may be used to remove the backdrop, while image enrichment can be utilized to widen the dynamic range of certain characteristics in order to boost detectability.

3.2 Feature Extraction

Texture feature extraction is essential for precise categorization of brain tissues, since the obstacles in identifying them solely based on form or intensity

level are significant. Numerous methods exist for characterizing and analyzing medical picture textures, such as run length encoding, fractal dimension, discrete wavelet transform, and co-occurrence matrices [20].

Spatial Gray Level Dependent features (SGLDF) may be used to determine the distance between samples, which improves the accuracy of diagnosis, even if medical image classification has made use of several texture characteristics. The use of texture information is employed in the construction of association rule mining algorithms with the aim of detecting abnormalities in medical pictures. The data

may be categorized into categories according on the spatial arrangement of pixel levels [21]. In order to examine the distribution of gray levels in a specific region, we may use 2D co-occurrence matrices to get both the overall and individual properties of each pixel.

The ten descriptions below were used to pull out texture features:

$$\sum_i^a \sum_j^b x[i, j] \log x[i, j] \quad \text{Entropy} = \quad (1)$$

$$\sum_i^a \sum_j^b x^2[i, j] \quad \text{Energy} = \quad (2)$$

$$\text{Contrast} = \sum_i^a \sum_j^b (i - j)^2 x[i, j] \quad (3)$$

$$\text{Homogeneity} = \sum_i^a \sum_j^b \frac{x[i, j]}{1 + |i - j|} \quad (4)$$

$$\text{SumMean} = \frac{1}{2} \sum_i^a \sum_j^b (i * x[i, j] + j * x[i, j]) \quad (5)$$

$$\text{Variance} = \frac{1}{2} \sum_i^a \sum_j^b ((i - \mu)^2 x[i, j] + (j - \mu)^2 x[i, j]) \quad (6)$$

$$\text{Maximum Probability} = \text{Max } x[i, j] \quad (7)$$

$$\text{Inverse Difference Moment} = \sum_i^a \sum_j^b \frac{p[i, j]}{|i - j|^k} \quad (8)$$

$$\text{Cluster Tendency} = \sum_i^a \sum_j^b (i + j - 2\mu)^k x[i, j] \quad (9)$$

$$\text{Correlation} = \sum_i^a \sum_j^b \frac{(i - \mu)(j - \mu)x[i, j]}{\sigma^2} \quad (10)$$

Consider x to be the adjusted co-occurrence matrix. (i, j) are two pairs of gray level values, and a by b is the co-occurrence matrix's size. The inter-sample distance is computed by estimating the second-order joint conditional probability density function, $x[i, j|d, \theta]$, for angles of 0° , 45° , 90° , and 135° . The function $x[i, j | d, \theta]$ denotes the probability that two pixels are located at a distance of d between samples and in a direction of θ . The calculated joint conditional probability density functions are exactly defined as

$$x[i, j|d, 0^\circ] = \# \left\{ \begin{array}{l} ((p, q), (a, b)) \in [P_x X P_y] X [P_x X P_y]: \\ |p - a| = |q - b| = d, R(p, q) = i, R(a, b) = j \end{array} \right\} / T(d, 0^\circ) \quad (11)$$

$$x[i, j|d, 45^\circ] = \# \left\{ \begin{array}{l} ((p, q), (a, b)) \in [P_x X P_y] X [P_x X P_y]: \\ (p - a = d, q - b = d) \text{ or } \\ (p - a = -d, q - b = d) \\ R(p, q) = i, R(a, b) = j \end{array} \right\} / T(d, 45^\circ) \quad (12)$$

$$x[i, j|d, 90^\circ] = \# \left\{ \begin{array}{l} ((p, q), (a, b)) \in [P_x X P_y] X [P_x X P_y]: \\ (p - a = d, q = b, R(p, q) = i), \\ (p - a = -d, q - b = d) \\ R(a, b) = j \end{array} \right\} / T(d, 90^\circ) \quad (13)$$

$$x[i, j|d, 135^\circ] = \# \left\{ \begin{array}{l} ((p, q), (a, b)) \in [P_x X P_y] X [P_x X P_y]: \\ (|p - a| = d, q - b = -d) \\ R(p, q) = i, R(a, b) = j \end{array} \right\} / T(d, 135^\circ) \quad (14)$$

The sign # denotes the cardinality of a set. The function $R(p, q)$ denotes the brightness of the image at the specific coordinates (p, q) , whereas $T(d, \theta)$ indicates the overall count of pixel pairs inside the image that have a distance of d between them and an orientation of θ .

The co-occurrence matrices for the orientations 0° , 45° , 90° , and 135° have corresponding pixel values of 1, 2, 3, and 4, respectively. By calculating the co-occurrence matrix for every pixel, we may get several matrix properties from the intersample distance, including energy, entropy, homogeneity, variance, and inverse variance. From the co-occurrence matrices, one may construct the feature vectors, which are then stored in the transaction database. The next step is to discretize the continuous valued features; each interval then represents a step in the association rule mining procedure.

3.3 Association Rule Mining

The goal of association rule mining is to find all the connections in a database of transactions. A statement in the form of $X \rightarrow Y$ is an association rule. In this rule, X is called the body or antecedent, while Y is called the head or consequent. In this context, we have a set of transactions, $T_r = \{T_1, T_2, T_3 \dots T_n\}$, where each transaction T in the set is a collection of items from the set $I = \{i_1, i_2, i_3 \dots i_n\}$. The task of mining association

rules entails identifying rules that meet the user's chosen minimal support and minimum confidence thresholds [22].

To extract relationships between attributes in a transactional database, this approach employs an enhanced version of the ARC-AC algorithm. The algorithm MARI has been elucidated in the subsequent way.

Algorithm 1: MARI Identify associations on the transactional database's training set

Input: A collection of image patches denoted as P_i ($K_1, K_2, \dots, K_m, f_1, f_2, \dots, f_n$), where k_i is a keyword associated with the patches and f_j represents the features that have been chosen for the patches, with a minimum support threshold of σ .

Output: Set of association rule of the form $f_1, f_2, \dots, f_n \rightarrow K_i$

Where K_i is the keyword and f_n is a feature and kw is a class category

Step 1: C_0 is a candidate keyword and their support

Step 2: F_0 is a frequent keyword and their support

Step 3: C_1 is a candidate keyword 1 item sets and their support

Step 4: F_1 is a frequent keyword and their support

Step 5: C_2 is candidate pairs (k,f) , such that (k,f)E P_1 and $k \in F_0$ and $f \in F_1$

Step 6: For each patches p in P_1 do

Step 7: For each $kw=(k,w)$ in C_2 do

Step 8: $kw.Support \leftarrow kw.support + count(kw,p)$

Step 9: $F_2 \leftarrow \{kw \in C_2 | kw.support > \sigma\}$

Step 10: $P_2 \leftarrow Filter\ table\ (P_1, F_2)$

Step 11: for ($i \leftarrow 3; F_{i-1} \neq \phi; i \leftarrow i+1$) do

$C_i \leftarrow (F_{i-1} F_2)$

Step 12: $C_i \leftarrow C_i - \{kw | (i-1)Item\ set\ of\ kw \in F_{i-1}\}$

Step 13: $P_i \leftarrow Filter\ table\ (P_{i-1}, F_{i-1})$

Step 14: For each patches p in P_i do

Step 15: For each kw in C_i do

Step 16: $kw.support \leftarrow kw.support + count(kw,p)$

Step 17: $F_i \leftarrow \{kw \in C_i | kw.support > \sigma\}$

Step 18: sets $\leftarrow \cup_i \{kw \in F_i | i > 1\}$

Rule = ϕ

Step 19: for each item set I in sets do

Step 20: Rule $\leftarrow Rule + \{f \rightarrow kw | f \cup kw \in I \wedge f \text{ Is an item set } \wedge kw \in C_0\}$

There are limits on the association rules so that the antecedent is always a group of traits from the brain image and the consequent is always the class name that the brain image belongs to.

3.4 Pruning Techniques

It is anticipated that the regulations formulated throughout the mining phase will be substantial in nature. This could potentially pose a challenge for applications that require immediate responses. Hence, the use of pruning procedures is crucial to eliminate specific criteria that conflict with each other, since they represent distinct groups with comparable characteristics.

This can be achieved using the following conditions:

Condition 1: If $R1 \in R2$, then the first rule $R1 \square C$ is considered a general rule. In order to achieve this, the association rules must be ordered according to condition 2.

Condition 2: If we have two rules, $R1$ and $R2$, we may say that $R1$ is considered to have a higher rank than $R2$ if

- i. $R1$ has more confidence than $R2$
- ii. If the confidences are equal, the support of $R1$ must be greater than the support of $R2$.
- iii. If both the levels of confidence and support are similar, but $R1$ has fewer qualities on its left hand side than $R2$.

Condition 3: The rules $R1 \rightarrow C1$ and $R1 \rightarrow C2$ reflect a conflict. Considering the aforementioned requirements, any duplicate instances have been removed. The collection of rules chosen after trimming becomes the definitive classifier. Utilizing these criteria, the class to which the new test image belongs has been predicted.

3.5 Classification

After completing the training phase, a classifier may be created by using a reduced set of association rules to train the brain images. The feature vector derived from the extracted attributes of the test image may be supplied to the classifier. The classifier employs association rules to construct a set of keywords that are used to generate the diagnostic of the test image.

This classifier generates many classifications when analyzing a test image. The developed algorithm has been implemented to offer diagnostic

recommendations. The approach employs a data structure to store all item sets that are relevant to the head of the rules. If the equation (15) is satisfied, the suggested diagnostic will provide a collection of items called h.

$$\frac{n(M_h)}{n(M_h)+n(N_h)} \geq T \quad (15)$$

The number of instances of matches for the item set h is represented by the variable n(M_h), while the number of instances of non-matches is denoted by n(N_h). A minimum requirement for the number of matches required to generate an item set in the suggested diagnosis is imposed by the threshold T. A match is established when the attributes of the image fulfill the criteria specified for the rule's body component.

Algorithm 2: Medical Image Classification using Interest Pruning and Navie Bayes (NB)

Step 1: load and Preprocess Medical image data

Step 2: Extract Features from medical image using SGLDF

Step 3: Define Interesting Pruning Criteria C(X_i) that represents the interestingness of a feature vector X_i.

Step 4: Apply interest pruning

Prune less interesting feature

$$X_{pruned} = \{X_i | C(X_i) \geq Threshold\}$$

Step 5: Split the Dataset into Training and Testing Sets

Divide the dataset into training (X_{train}, Y_{train}) and testing set (X_{test}, Y_{train})

Step 6: Train a Naive Bayes Classifier

Train a Gaussian Naive Bayes classifier on the pruned features:

$$f(X_{train}) = GaussianNB().fit(X_{train}, Y_{train})$$

Step 7: Evaluate the model

Predict labels for the test set and calculate accuracy:

$$Y_{pred} = f(X_{test})$$

4. Result Analysis

The efficacy of the proposed methodology has been assessed by the use of a confusion matrix, as seen in Figure 2. This matrix depicts the whole spectrum of possible results that arise from a prediction, arranged in a structured table manner. The possible outcomes of a binary classification may be represented as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

		Identification by Classifier	
		Yes	No
Identification by physician	Yes	TP	FN
	No	FP	TN

Fig 2: Confusion Matrix

The normal and abnormal photos are correctly classified as TP and TN respectively. A FP arises when a negative outcome is erroneously classified as positive. A false positive occurs when an erroneous alert is generated during the classification process. A FN occurs when a good result is incorrectly predicted as negative. The accuracy and recall levels may be computed from the confusion matrix using a designated formula.

Precision

The term "relevance" refers to the proportion of the categorized picture that corresponds to the predictions. The equation (16) represents it.

Precision =

$$\frac{\text{number of instances correctly predicted as positive}}{\text{number of instances correctly classified as positive} + \text{number of instances incorrectly predicted as positive when they are actually negative}} \quad (16)$$

Recall

The term refers to the proportion of the categorized picture that consists of all the relevant predictions. The equation (17) is shown below.

$$Recall = \frac{\text{number of instances correctly predicted as positive}}{\text{number of instances correctly classified as positive} + \text{number of instances incorrectly predicted as negative when they are actually positive}} \quad (17)$$

The experiment used a collection of CT scan brain pictures, according to the procedural design shown

in Figure 1. Decisions are based on databases that have undergone pre-diagnosis by physicians. The original input image is depicted in Figure 3, whereas the outcome of histogram equalization is displayed in Figure 4. Equalization of histograms is utilized to mitigate illumination and noise fluctuations during the scanning process. Feature extraction was performed subsequent to the preparatory stage in order to eliminate superfluous and redundant data present in the input image.

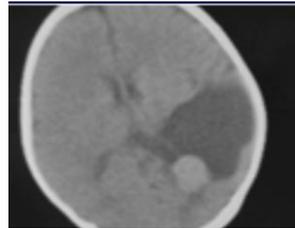


Fig 3: Input CT image

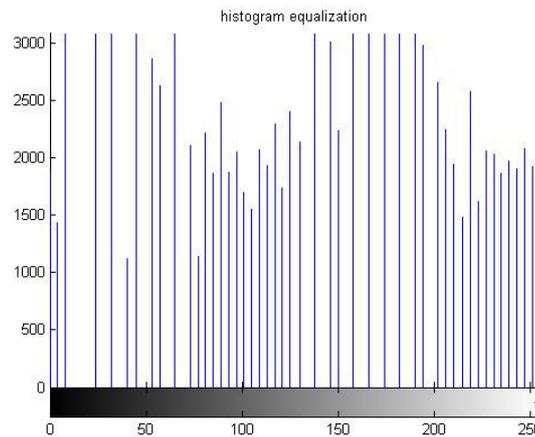


Fig 4: Histogram equalized image

The Haralick co-occurrence method was used to evaluate the distinction of alterations at the tissue level, as seen in Figure 5. Figure 6 demonstrates that pixel 1 is associated with a 45° angle, pixel 2 is associated with a 45° angle, pixel 3 is associated with a 90° angle, and pixel 4 is associated with a

135° angle relative to the center pixel. The measurement of these angles is taken at a distance of one unit. Figure 7 exhibits the pre-processed CT scan image of the brain, accompanied by an angular representation.

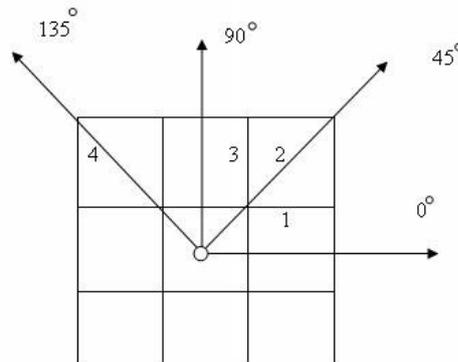


Fig 5: Matrix depiction of the central pixel and its surrounding pixels

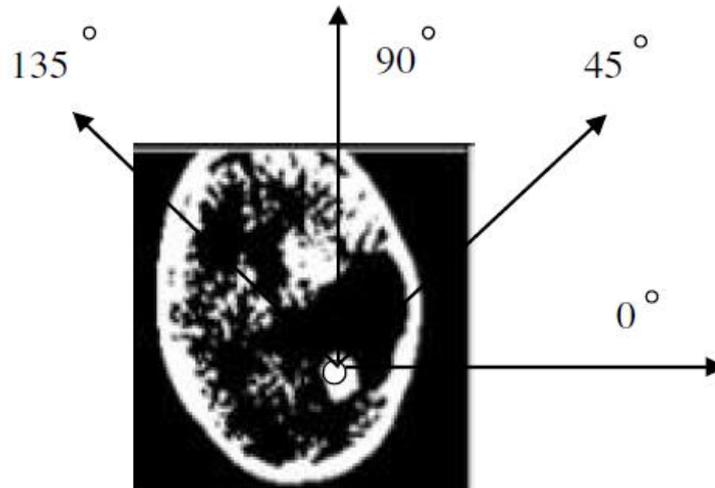


Fig 6: CT scan brain image pre-processed and combined with the angle representation.

Table 1 gives the matrix that represents the pixels at a distance of one and a degree of zero. To generate the pixel representation matrix for the additional

degrees, the same procedure must be used. Table 2 presents the co-occurrence matrices derived from the pixel representation matrix.

Table 1: Matrix depicting the pixel representation at a zero degree angle

0	0	1	3	1
3	1	1	3	1
2	1	3	0	3
3	2	1	0	3
3	3	2	1	2

Table 2: Co-occurrence matrix for a distance of one and a degree of zero

i/j	0	1	2	3
0	1	1	0	2
1	1	1	1	3
2	0	3	0	0
3	1	3	2	1

Figure 7 depicts the sequential process of extracting texture characteristics and segregating items. Co-occurrence matrices have been calculated and texture features have been extracted for each item. There are four specific orientations: 0° degrees, 45° , 90° , and 135°. The result of these orientations is the development of sixteen co-occurrence matrices.

Following this, the texture attributes for each co-occurrence matrix are computed and entered into the database. The computation of feature vectors and subsequent storage of the resulting vectors in the transaction database are performed in accordance with the values of the co-occurrence matrix.

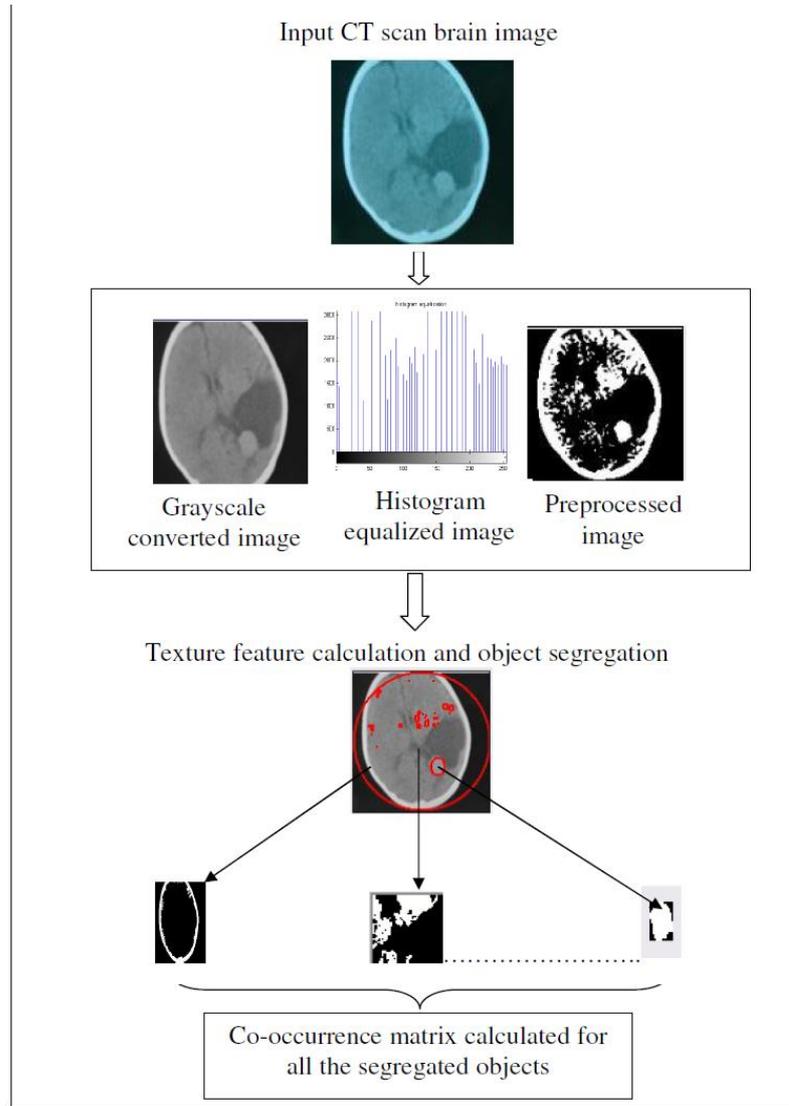


Fig 7: Texture based feature extraction and object Segregation

The MARI method has been put into motion on the transaction database, which has diagnosis data and feature vectors for the training CT scan images. Figure 8 displays the images from the sample dataset with their related diagnostic information.

The accuracy and recall values for the suggested methodology, Association Rule Mining (ARM) method, and Naïve Bayesian method are shown in Figure 8. Empirical evidence substantiates that the proposed strategy surpasses prior methodologies.

The effectiveness of the proposed method has been estimated using the following measures:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (18)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (19)$$

$$Specificity = \frac{TN}{TN+FP} \quad (20)$$

The quantities denoted as TP, TN, FP, and FN represent the proportions of correctly classified abnormal cases, correctly classified normal cases, normal cases erroneously classified as abnormal, and abnormal cases erroneously classified as normal, respectively. Accuracy is defined as the ratio of cases that are accurately diagnosed to the overall number of cases. The sensitivity of the proposed method is a metric that quantifies its capability to detect anomalous events. Specificity quantifies the capability of a method to accurately identify and classify typical instances.

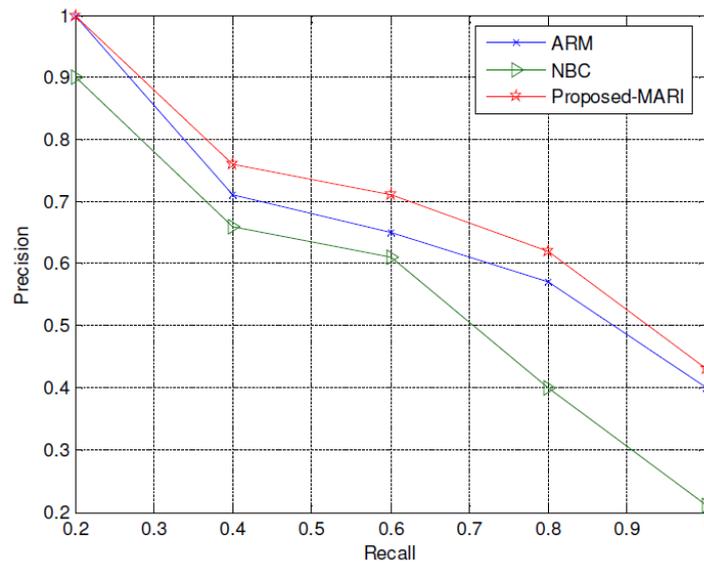


Fig 8: Precision and Recall graph using naive bayesian, association rule mining and MARI association rule mining

The minimal confidence criterion is established at 97%, whereas the minimum support threshold is established at 10%. The test images attributes and correlation rules were generated using a threshold value of 0.001. The findings suggest that the classifier that was proposed achieves superior levels of sensitivity, specificity, and accuracy, as evidenced by its respective values of 96%, 90%, and 93%. To substantiate the findings, a comparison was made between the algorithmic approach and two widely acknowledged classifiers: an associative classifier and a naïve Bayesian classifier.

Table 3 provides an overview of the association rule mining, proposed naïve Bayesian classifier, and suggested methodology. The aforementioned metrics are displayed, including processing time, area under the curve (Az), sensitivity, and specificity. The experimental findings demonstrate that the suggested methodology attains notable improvements in efficiency, sensitivity (up to 96%), accuracy (93%), and execution time (down to 0.02). Furthermore, it aids in the process of decision-making.

Table 3 Performance comparison for classifiers

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)	Az	SE	Runtime in ms
Navie Bayes Classifier	76	67	75	0.87	0.07	31.31
Association Rule based Classifier	93	83	93	0.89	0.02	9.92
Proposed Pruned Association rule with MARI Algorithm based classifier	97	93	95	0.98	0.03	3.14

The classifiers' outcomes are displayed in Table 4. A comprehensive collection of 150 photographs was utilized for training purposes, of which 95 were selected for analysis in the benign and malignant categories. Multiple classifiers were applied to the

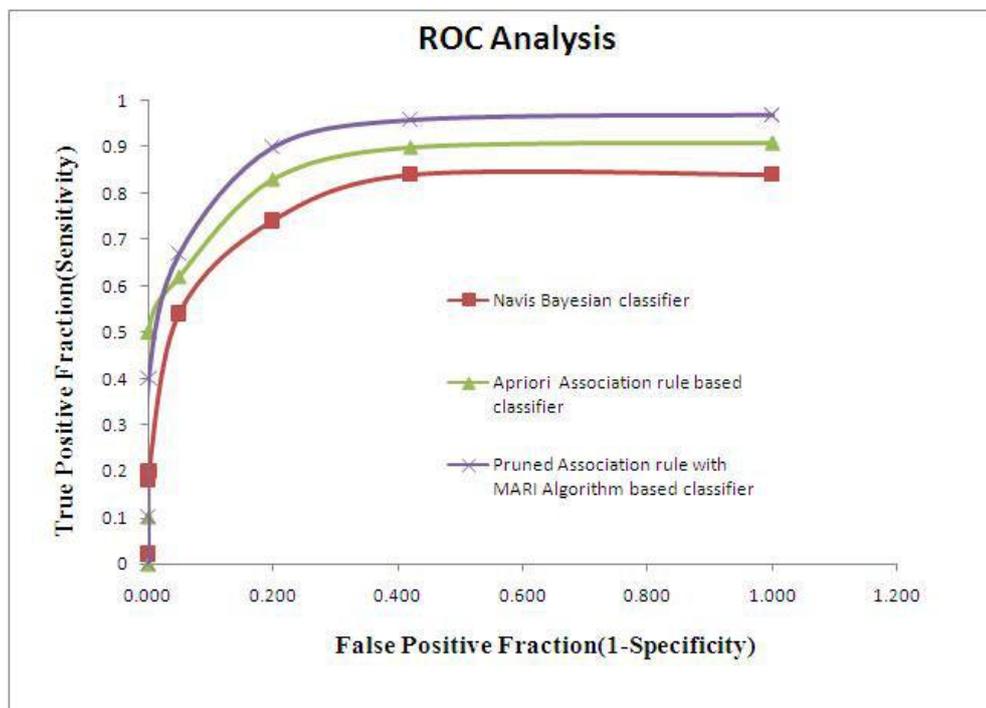
images for the purpose of classification. The findings indicate that the accuracy rate of classification achieved by the proposed system is superior to that of both the naïve Bayesian classifier and the association rule-based classifier.

Table 4: Performance Comparison of Classifier

Class	Training/ Testing	NB Classifier	Association Rule Based Classifier	Proposed Pruned Association rule with MARI Algorithm based Classifier
Benign	152/93	89.8	93.1	96.93
Malignant	152/93	84/79	93.46	99.58
Average				98

Receiver Operating Characteristic curves, the variables sensitivity and specificity are utilized. The region bounded by the ROC curve is critical because it is utilized to ascertain the classification accuracy as a whole. The ROC curve comparison among

multiple classifiers is shown in Figure 8. The findings reveal that the proposed categorization method using mining and pruned rules produces a higher value of (AZ) in comparison to other strategies.

**Fig 9:** ROC analysis of proposed work

5. Conclusion and Future work

Using pruned association rules in conjunction with the MARI algorithm, a sophisticated method for analyzing brain lesions has been developed and its performance has been assessed. The performance of the proposed methodology has been demonstrated to be superior to that of the current classifiers. The recall rate for the classification of brain tumors was 96%, while the accuracy rate was 98%. It is anticipated that the improved brain tumor classification system will furnish physicians with indispensable diagnostic methodologies.

References

[1] M. Biswas, V. Kuppili, L. Saba, D.R. Edla, H.S. Suri, A. Sharma, J.S. Suri, Deep learning fully

convolution network for lumen characterization in diabetic patients using carotid ultrasound: a tool for stroke risk, *Med. Biol. Eng. Comput.* 57 (2) (2019) 543–564, <https://doi.org/10.1007/s11517-018-1897-x>.

[2] Yu J, Yao J, Zhang J, Yu Z, Tao D (2020) SPRNet: single-pixel reconstruction for one-stage instance segmentation. *IEEE Trans Cybern.* <https://doi.org/10.1109/TCYB.2020.2969046>

[3] Ghassemi N, Shoeibi A, Rouhani M (2020) Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. *Biomed Signal Process Control* 57:101678

[4] Ayadi W, Elhamzi W, Charfi I, Atri M (2019) A hybrid feature extraction approach for brain

- MRI classification based on Bag-of-words. *Biomed Signal Process Control* 48:144–152
- [5] Shang R, He J, Wang J, Xu K, Jiao L, Stolkin R (2020) Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification. *Knowl. Based Syst* 105542
- [6] Sushma M C, Bharathi D, “Brain Tumor Identification using Bilateral Filtering and Adaptive K-Means Clustering”, *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 5, No. 5, May 2016.
- [7] Raj C.P,S., Shreeja R. “Automatic brain tumor tissue detection in T-1 weighted MRI”, *Proceedings of International Conference on Innovations in Information, Embedded and Communication Systems*, pp. 1-4, Coimbatore, India, 17-18 March 2017.
- [8] Mathew R.A, Anto B.P., “Tumor Detection and Classification of MRI Brain Image Using Wavelet Transform And SVM”, *Proceedings of the International Conference on Signal Processing and Communication*, pp. 75-78, Coimbatore, India, 28-29 July 2017.
- [9] Gamage P.T, Ranathunga L., “Identification of Brain Tumor using Image Processing Techniques”, *Independent Study, Final Report, Faculty of Information Technology University of Moratuwa*, 2017.
- [10] Gupta M., Rao B.V.V.S.N.P., Rajagopalan V., “Brain Tumor Detection in conventional MR Images based on Statistical Texture and Morphological Features”, *Proceedings of the International Conference on Information Technology*, pp. 129-133, Bhubaneswar, India, 22-24 December 2016.
- [11] Singh G., Ansari M.A., “Efficient Detection of Brain Tumor from MRIs Using K-Means Segmentation and Normalized Histogram”, *1st India International Conference on Information Processing (IICIP)*, Delhi, India, pp. 1-6, 12-14 August 2016.
- [12] Sharma K., Gujral S., Kaur A., “Brain Tumor Detection based on Machine Learning Algorithms”, *International Journal of Computer Applications*, Vol. 103 – No.1, October 2014.
- [13] Akram M.U., Usman A., “Computer Aided System for Brain Tumor Detection and Segmentation”, *Proceedings of the International Conference on Computer Networks and Information Technology*, Abbottabad, Pakistan, pp. 299-302, 11-13 July 2011.
- [14] Hemanth G., Janardhan M., Sujihelen L., “Design and Implementing Brain Tumor Detection Using Machine Learning Approach”, *Proceedings of the Third International Conference on Trends in Electronics and Informatics*, Tirunelveli, India, pp. 1289-1294, 23-25 April 2019.
- [15] S. Chaplot, L.M. Patnaik and N.R. Jagannathan, “Classification of Magnetic Resonance Brain Images using Wavelets as Input to Support Vector Machine and Neural Network”, *Biomedical Signal Processing and Control*, vol. 1(1), 2006, pp. 86–92, doi: 10.1016/j.bspc.2006.05.002.
- [16] H. Selvaraj, S. Thamarai Selvi, D. Selvathi and L. Gewali, “Brain MRI Slices Classification using Least Squares Support Vector Machine”, *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, vol. 1(1), 2007, pp. 21-33, doi: 10.1080/1931308X.2007.10644134
- [17] M. Maitra and A. Chatterjee, “Hybrid Multiresolution Slantlet Transform and Fuzzy c-means Clustering Approach for Normal-Pathological Brain MR Image Segregation”, *Medical Engineering & Physics*, vol. 30(5), 2008, pp. 615–623, doi: 10.1016/j.medengphy.2007.06.009.
- [18] Grunauer and M. Vincze, “Using Dimension Reduction to Improve the Classification of High-dimensional Data”, *OAGM Workshop*, 2015.
- [19] M. Zhou, J. Scott, B. Chaudhury, L. Hall, D. Goldgof, K.W. Yeom, R. Gillies, *Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches*, *Am. J. Neuroradiol.* 39 (2) (2018) 208–216, <https://doi.org/10.3174/ajnr.A5391>.
- [20] A. El-Baz, J.S. Suri (Eds.), *Level Set Method in Medical Imaging Segmentation*, CRC Press, 2019, <https://doi.org/10.1201/b22435>
- [21] M. Biswas, V. Kuppili, L. Saba, D.R. Edla, H.S. Suri, E. Cuadrado-Godia, J.S. Suri, *State-of-the-art review on deep learning in medical imaging*, *Front. Biosci.* 24 (2019) 392–426, <https://doi.org/10.2741/4725>.
- [22] Tandel GS, Balestrieri A, Jujaray T, Khanna NN, Saba L, Suri JS (2020) Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. *Comput Biol Med* 122:103804