Advances in Engineering Software 173 (2022) 103221

Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

Research paper

Performance analysis of machine learning algorithm of detection and classification of brain tumor using computer vision

Ashwini S Shinde [a,*], BM Mahendra [b], Santosh Nejakar [c], Santosh M Herur [d], Nagaraj Bhat [b]

[a] Department of CSE-AI, Nutan College of Engineering and Research, Pune, Maharashtra, India
[b] Department of ECE, RVCE, Bengaluru, Karnataka, India
[c] Department of Electrical Engineering, JIT, Davangere, Karnataka, India
[d] Department of ECE, JIT, Davangere, Karnataka, India

ARTICLE INFO

Keywords:
Artificial Intelligence
MRI
Logistic regression
OASIS
Multilayer Perceptron

ABSTRACT

Brain tumor is one of the undesirables, uncontrolled growth of cells in all age groups. Classification of tumors depends no its origin and degree of its aggressiveness, it also helps the physician for proper diagnosis and treatment plan. This research demonstrates the analysis of various state-of-art techniques in Machine Learning such as Logistic, Multilayer Perceptron, Decision Tree, Naive Bayes classifier and Support Vector Machine for classification of tumors as Benign and Malignant and the Discreet wavelet transform for feature extraction on the synthetic data that is available data on the internet source OASIS and ADNI. The research also reveals that the Logistic Regression and the Multilayer Perceptron gives the highest accuracy of 90%. It mimics the human reasoning that learns, memorizes and is capable of reasoning and performing parallel computations. In future many more AI techniques can be trained to classify the multimodal MRI Brain scan to more than two classes of tumors.

1. Introduction

Computer Aided Diagnosis (CAD) is one of the major contributions of technology implemented in the field of medical science for better precision and high accuracy. It is considered as a high throughput for the expediency to investigate the outgrowth expanse. The implementation of technology, diagnosis and treatment planning becomes easy and gives the physician a second taught of his predictions. The most dominant tool for imaging the brain is the Magnetic Resonance Imaging (MRI) which are multimodal, where in these modalities can reveal different parts in the tumor, and provides information concomitant to anatomical assemblies as well as potential anomalous tissues essential for diagnosis and treatment planning [1,2]. The different sequences like the T1 weighted, T2 weighted, Fluid Attenuation Inversion Recovery (FLAIR) and Diffusion Weighted Imaging (DWI) show the different intensity variations that help in identifying the region of interest. Extracting reckonable data from MRI helps to capture the functions of different consequence crevices in case of different types of tumors. The possibility of survival of a patent is increased if the tumor is perceived at an early stage. However precise segmentation and categorization of abnormalities are not forthright. There exist a number of semiautomatic and fully automatic methods for the classification of tumors but clinical acceptance depends on simplicity and less human intervention. Classification of tumors in the human brain is possible by implementing the Supervised Machine Learning techniques, in this research we work upon the Naïve Bayes, The Logistic, Multilayer Perceptron (MLP), The Support Vector Machine (SVM) and the Decision Tree (DT) for classification and the results are compared on the basis of classification accuracy for the data used. Classification is performed with more discriminative features initially on the OASIS and ADNI [3,4] database

Yet there are some challenges to be addressed for classification of abnormalities in medical imagining like selection of appropriate model, describing the given data, finding the errors in the data, the adequacy of data used and confidence about the results. Therefore, there is no universal recognized method for medical image classification. So, it remains a challenging problem in computer vision and Machine learning. Fully automatic systems determine the tumor part without human intervention these systems generally include human intelligence and prior knowledge about the throughput. These algorithms which are mostly developed by using soft computing and model-based techniques prove to produce accurate results. The study of automatic brain tumor classification represents interesting research in Machine Learning and

* Corresponding author.
E-mail address: ashsshinde@gmail.com (A.S. Shinde).

https://doi.org/10.1016/j.advengsoft.2022.103221
Received 9 July 2022; Accepted 27 July 2022
Available online 10 August 2022
0965-9978/

Table 1
A Feature set of thirteen features.

Sr.No	Contrast	Correlation	Energy	Homogeneity	Mean	Standard Deviation	Entropy	RMS	Variance	Smoothness	Kurtosis	Skewness	IDM	Class
1	0.2088	0.1990	0.7621	0.9352	0.0031	0.0898	3.1735	0.0898	0.0080	0.9205	7.3282	0.4690	0.0577	Benign
2	0.2717	0.0931	0.7686	0.9338	0.0024	0.0898	3.2698	0.0898	0.0081	0.8974	7.9567	0.8862	0.4926	Benign
3	0.2414	0.1065	0.7440	0.9277	0.0030	0.0898	3.5550	0.0898	0.0080	0.9178	6.3661	0.6499	0.4729	Benign
4	0.2161	0.1382	0.7548	0.9325	0.0025	0.0898	3.3156	0.0898	0.0081	0.9032	6.2320	0.3121	0.5631	Benign
5	0.2250	0.0991	0.7691	0.9365	0.0021	0.0898	3.5182	0.0898	0.0080	0.8850	6.7672	0.4413	0.5462	Benign
6	0.2333	0.1284	0.7491	0.9308	0.0019	0.0898	2.6632	0.0898	0.0081	0.8778	7.2707	0.6117	0.0366	Benign
7	0.2033	0.1126	0.7554	0.9331	0.0019	0.0898	3.6549	0.0898	0.0080	0.8783	5.8117	0.3408	1.0010	Benign
8	0.2558	0.0895	0.7557	0.9314	0.0025	0.0898	3.0756	0.0898	0.0081	0.9040	7.7971	0.5774	0.2601	Benign
9	0.2341	0.1321	0.7530	0.9315	0.0035	0.0897	3.1562	0.0898	0.0080	0.9291	7.4848	0.5212	1.0392	Benign
10	0.2689	0.0977	0.7861	0.9410	0.0007	0.0898	2.7465	0.0898	0.0081	0.7186	10.9703	0.7365	0.1190	Benign
11	0.3059	0.1421	0.7862	0.9379	0.0063	0.0896	3.2051	0.0898	0.0080	0.9591	12.2408	1.1048	1.2156	Malignant
12	0.2255	0.1345	0.7466	0.9299	0.0043	0.0897	3.6004	0.0898	0.0081	0.9407	6.0137	0.5267	0.3801	Malignant
13	0.2433	0.0933	0.7613	0.9329	0.0037	0.0897	3.3710	0.0898	0.0081	0.9314	7.3506	0.6350	0.1378	Malignant
14	0.2439	0.1072	0.7310	0.9246	0.0046	0.0897	3.5484	0.0898	0.0081	0.9446	6.5235	0.6204	0.5030	Malignant
15	0.2280	0.0769	0.7577	0.9319	0.0030	0.0898	3.6700	0.0898	0.0080	0.9185	5.6210	0.4119	1.0350	Malignant
16	0.2314	0.1072	0.7418	0.9298	0.0042	0.0897	3.5516	0.0898	0.0080	0.9403	6.0614	0.5104	0.3130	Malignant
17	0.2517	0.0734	0.7402	0.9267	0.0035	0.0897	3.5239	0.0898	0.0080	0.9284	6.5220	0.4979	1.6524	Malignant
18	0.3073	0.1337	0.7598	0.9334	0.0060	0.0896	2.6626	0.0898	0.0080	0.9571	16.2926	1.6120	1.2859	Malignant
19	0.2155	0.0951	0.7378	0.9274	0.0028	0.0898	3.6283	0.0898	0.0080	0.9132	5.3238	0.3230	1.0419	Malignant
20	0.2870	0.0918	0.7590	0.9320	0.0055	0.0896	3.1117	0.0898	0.0080	0.9530	12.1609	1.1228	0.5619	Malignant

Table 2
Result of Naïve Bayes Classifier for 10 Cross Fold Validation.

n^{th} Cross Fold	Percent correct
1	83.33333
2	66.66667
3	66.66667
4	66.66667
5	83.33333
6	83.33333
7	83.33333
8	83.33333
9	83 33,333
10	83.33333

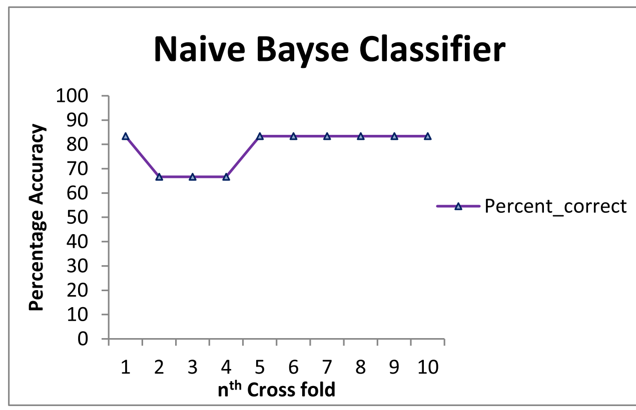

Fig. 1. Plot of Percentage Accuracy for Naïve Bayes Classifier.

Table 3
Mean Error for Naïve Bayes Classifier.

Mean absolute error	Root mean squared error
0.111185	0.259234
0.387215	0.585724
0.330763	0.566769
0.291527	0.506187
0.170592	0.407284
0.182931	0.409818
0.191837	0.391273
0.152055	0.366976
0.167413	0.408239
0.246425	0.416944

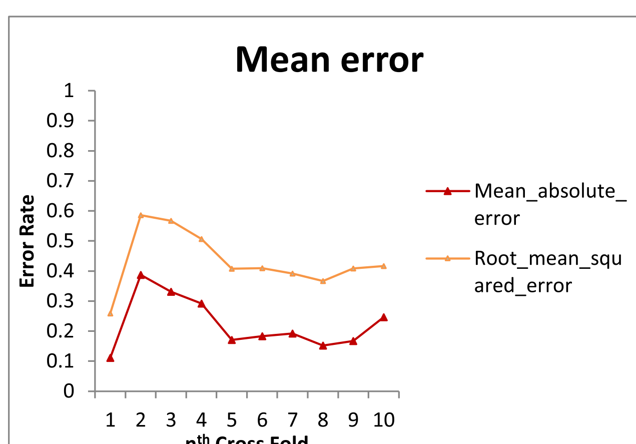

Fig. 2. Plot of mean error for Naïve Bayes Classifier.

Table 4
Detailed Accuracy Rate for Naïve Bayes Classifier.

True Positive Rate	0.900
False Positive Rate	0.300
Precision	0.75
Recall	0.900
Accuracy	80%

Artificial intelligence (AI)

The organization of this paper is as follows. In Section II, the existing scientific research in medical image classification is reviewed, along with the motivation for this research. Section III presents the materials and methods used in this work it describes the dataset implemented, it also shows the proposed work. Section IV, represents the experimental results obtained and finally the section V, elaborates the outcomes and conclusion.

2. Literature survey

The literature survey elaborates the details of research work studied in the domain to diagnose this life-threatening disease; it deals with the previous work carried out in the field of Tumor recognition and classification implementing computer vision and Machine learning. The initial papers describe the review of some of the best segmentation techniques using image processing followed by research based on automatic segmentation and classification techniques by building neural network and finally discussing about the recent techniques using Artificial Intelligence and Machine Learning.

Artificial Intelligence is an emerging trend not only in digital marketing but in almost every field, it has an extensive range of application in the area of brain tumor detection and classification. In the resent years most of the researchers made more of their inventions in the field of biomedical science through AI, this section deals with the review on some of the tumor detection and classification techniques using AI and ML.

Sergio Pereira, Adriano Pinto [5], in their paper mentioned that reliable and automatic techniques are required for tumor segmentation

Table 5
Classification Instances for Naïve Bayes Classifier.

Correctly Classified Instances	80
Incorrectly Classified Instances	20
Percentage Accuracy	80

Table 6
Confusion Matrix for Naïve Bayes Classifier.

True positive = 45	False Positive = 5
False Negative = 15	True Negative = 35

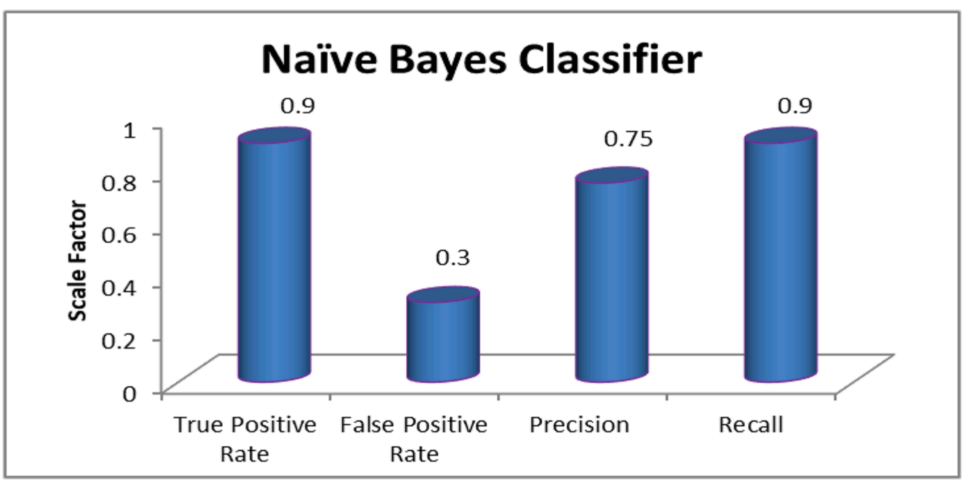

Fig. 3. Plot of Detailed Accuracy Rate for Naïve Bayes Classifier.

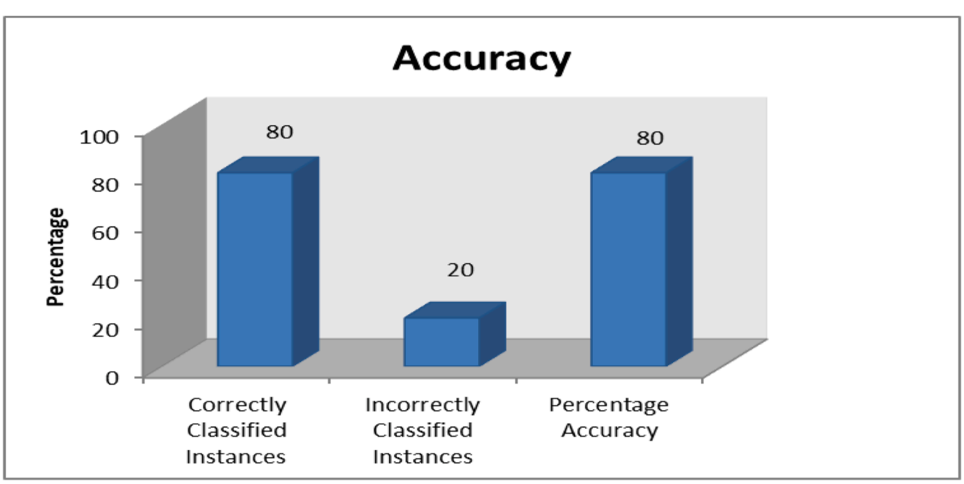

Fig. 4. Plot of Classification Instances for Naïve Bayes Classification.

Table 7
Result of Logistic Classifier for 10fold cross Validation.

n^{th} Cross Fold	Percent correct
1	75
2	50
3	100
4	75
5	75
6	100
7	100
8	100
9	100
10	100

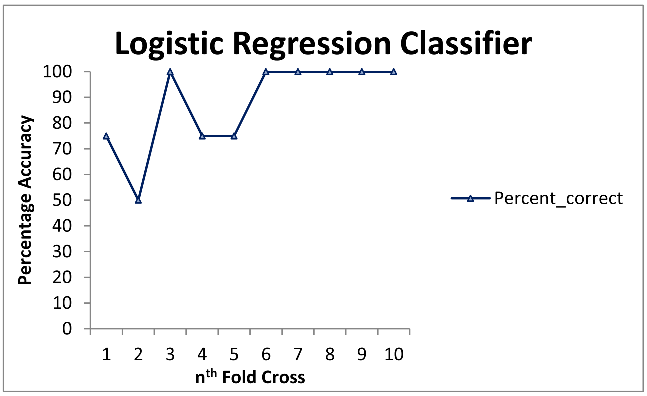

Fig. 5. Plot of Percentage Accuracy for Logistic.

Table 8
Mean Error for Logistic Classifier.

Mean absolute error	Root mean squared error
0.230653	0.458894
0.5	0.707107
0.000357	0.000646
0.249999	0.499997
0.175671	0.351273
9.43E-10	1.83E-09
1.32E-09	2.65E-09
0.048946	0.097892
7.5E-05	0.00015
5.2E-13	1.04E-12

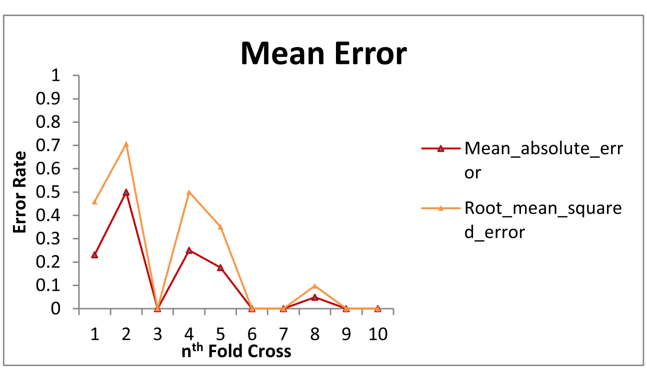

Fig. 6. Plot of mean error for Logistic Classifier.

and classification. The Convolution Neural Network (CNN), was proposed by the authors with 3 × 3 kernel that allows designing a deeper architecture, and gets an optimistic effect against over fitting by providing a smaller number of weights in the network. Though intensity

Table 9
Detailed Accuracy Rate for Logistic Classifier.

True Positive Rate	0.900
False Positive Rate	0.900
Precision	0.900
Recall	0.900
Accuracy	90%

normalization is not a pre-processing step in CNN the results of segmentation after normalization were very effective. Their algorithm was structured as follows 1. Initialization, important to achieve convergence 2. The activation function is applied which is responsible for non-linear transformation of the data. Rectifier linear unity (ReLU) gives better results than the classical sigmoid functions and also speeds up training [6]. 3. Pooling forms a feature map by combining spatially nearby features.4. Regularization to reduce over fitting. 5. Data Augmentation used to increase the number of training set. 6. Loss function it is a function to be minimized during training. Mohammad Havaei, Axel Davy et al. [7] presented a fully automatic segmentation method. This method is based on Deep Neural Network (DNN) the network is designed to classify low- and high-grade tumors. They explored an architecture based on CNN i.e. DNN adapted to image data, it exploits local as well as global features concurrently. Here the final layer i.e. the convolutional layer allows 40 fold speed up and their two phase training allows to handle the unevenness of tumor labels. Lastly, they explore the cascaded architecture here the additional source of information is provided by the CNN. High accuracy and with low computational time is the main feature of this algorithm. Heba Mohsen, El- Sayed A et al. [8] emphasized on machine learning as a powerful tool for complex problems. They developed Deep learning architecture for classification of the tumors into four classes they combined the classifier with discrete wavelet transform and principal component analyser. The flow of their system was tumor segmentation using Fuzzy C- Means, feature extraction applying DWT and feature selection by PCA and finally classification implementing the DNN. For classification 7-fold cross validation technique was built to train the data with a 7 hidden layer structure. Ali ARL and Davut Hanbay [9] proposed a three-stage method. Initially the pre-processing, secondly the Extreme Learning machine local respective fields (ELM-LRF) based on classification and finally segmenting the tumor. Initially the noise is removed by nonlocal means and local smoothening techniques, then it is classified by applying ELM-LRF and finally the tumor is segmented, they obtained the classification precession up to 97%. Elisee Illunga Mbuyamba, Juan Gabriel Avina [10]. The authors worked on the incorrect delineation of MRI where the distance between the foreground and the background is high. It was found that the region based active contour models are more sensitive to high mean intensity hence to solve the problem of high mean intensity they developed a model known as localized active counter model. This model with background intensity compensation balances the mean intensity that aims to diminish the attraction consequence of active contour to the unsolicited border lines. This approach increases the accuracy and reduces the computational time. They applied this model on the multimodal MRI data which was helpful to detect the foreground and could segment the abnormal region that can be further used for classification. A.Sankari and S.Vigneshwari [11], The authors suggested an automatic segmentation approach applying Convolutional Neural Network (CNN) with a 3 × 3 filter size. Here because of this small size the operation on the small portion permits to extract more deep design rather than having a constructive outcome with a smaller number of weights in the system. To enhance the image spatial domain methods are implemented. The CNN is trained using a class of labels that extracts features like edges lines corners etc. The system is a layer of unit here every unit in one layer is associated with the past layer i.e. every neuron is associated with the previous layer. The construction of the CNN is first is the convolution layer the second is the fully connected layer and the third is the pooling

Fig. 7. Plot of Detailed Accuracy Rate for Logistic Classifier.

Table 10
Classification Instances for Logistic Classifier.

Correctly Classified Instances	90
Incorrectly Classified Instances	10
Percentage Accuracy	90

layer. The MRI is the input where the convolution layer has learnable filters that performs a dot product among the small regions. They are also connected to the weights. Next is the fully connected layer where all the neurons are linked to the previous layer. The output layer computes the class score of the image to show the size and the location of the tumor. Muhammad Sajjad, Salman Khan et al. [12], worked on multi-grade tumor classification. They proposed a CNN based system for the classification of the tumors, the steps involved deep learning technique for segmentation of the tumor region, next data augmentation is applied to hike the training dataset. The employed augmentation techniques for the geometric transformation are flipping, rotation, skewness and shearing and the next are sharpening, Gaussian blur, emboss and edge detection used for noise invariance. Finally, CNN is implemented for multi-grade classification, it has 3 × 3 kernels for the convolution layer with 1 stride in the initial layer. Due to this it will be able to learn multiple grades of the tumor. The CNN architecture comprises of 19 weighted layers out of which 16 are convolutional and 3 are fully connected layers. The max pooling is led by the first convolutional layers. and the same is repeated for the next two layers. Next three last layers are fully connected resulting in almost 4096,4096,1000 features this form the feature vector which is fed as input for the soft max classifier to make the final decision of the grade of the tumor.

2.1. *Motivation for the research*

The outcome of the literature survey helped us to find the challenges

Table 11
Confusion Matrix of Logistic Classifier.

True positive = 45	False Positive = 5
False Negative = 5	True Negative = 45

Table 12
Result of Multi-Layer Perceptron Classifier for 10-fold cross Validation.

n^{th} Cross Fold	Percent correct
1	50
2	50
3	75
4	75
5	100
6	100
7	100
8	75
9	100
10	100

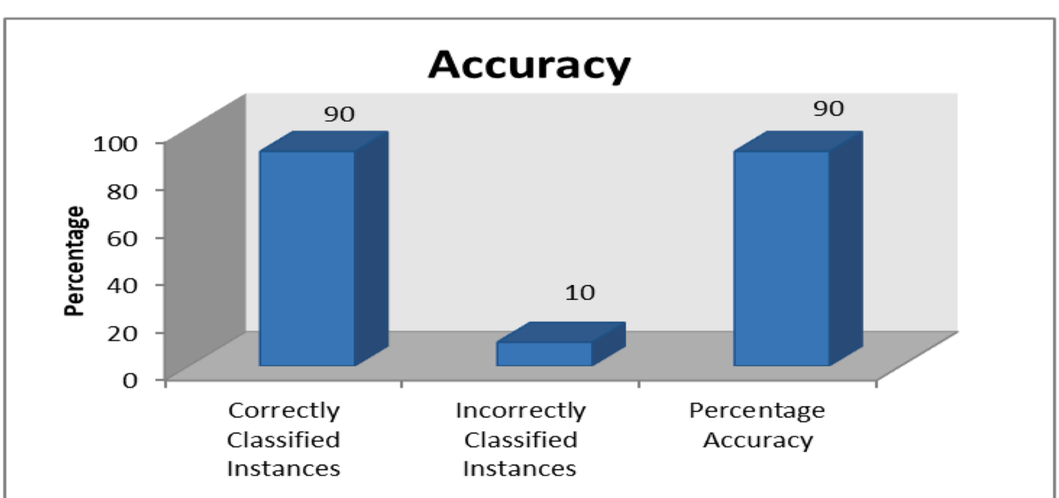

Fig. 8. Plot of Classification Instances for Logistic Classification.

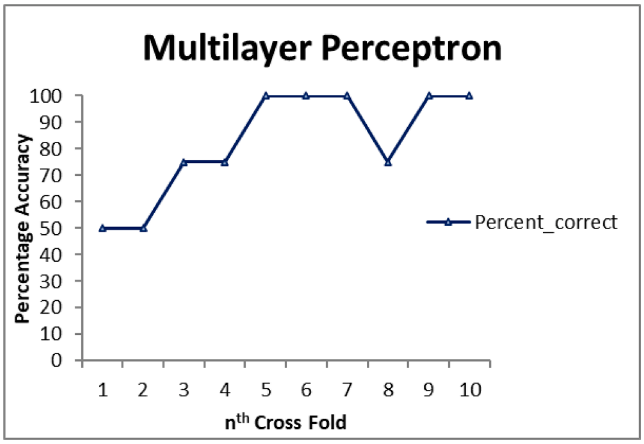

Fig. 9. Plot of Percentage Accuracy for MLP.

Table 13
Mean Error for MLP Classifier.

Mean absolute error	Root mean squared error
0.304202	0.435902
0.482741	0.679709
0.332008	0.393292
0.232097	0.429697
0.013611	0.016777
0.003172	0.003461
0.020229	0.039278
0.228384	0.452503
0.017527	0.033476
0.002333	0.002905

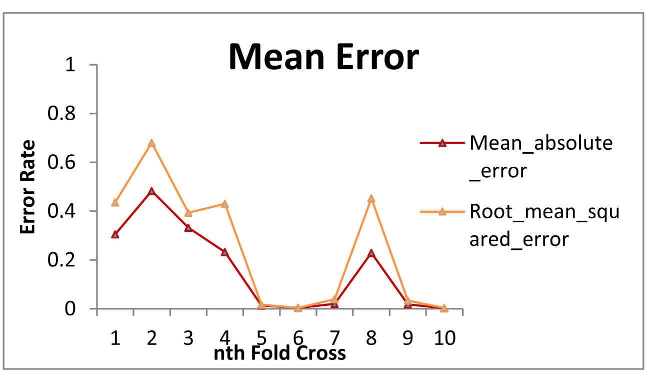

Fig. 10. Plot of mean error for MLP Classifier.

Table 14
Detailed Accuracy Rate for MLP Classifier.

True Positive Rate	0.900
False Positive Rate	0.1
Precision	0.900
Recall	0.900
Accuracy	90%

faced by researchers and to identify the gaps, it was noted through the literature survey that there exists a need to develop a Machine Learning approach that overcomes these limitations and produce better results it was found that most of the research carried out was for the classification of the image into normal and abnormal images the work can be further applied for the classification of the Brain tumors. Most of the methodologies learnt during the literature survey were implemented for data base that are available as benchmark data set which are pre-processed by the experts. The method of hybridization for multiclass classification which makes use of intensity, shape and texture features can be applied to increase accuracy. Due to the intra- and inter-intensity variability between the training data and test data, segmentation and classification is still challenging. This can be improvised by applying Machine Learning. Since MRI scans are multimodal sequences the best slice of image must be identified to get the correct class which is again done manually by the experts this can be eliminated by implementing the Machine learning techniques. The ML approach reduces the time complexity by eliminating the time required for selection of the MR sequence for segmentation which also effects the accuracy performance. Therefore, the prime objective of this research is to find an Optimal Classifier to Classify the Brain Tumor using Machine Learning Algorithm. Identifying the challenges inherent for Machine learning and applying appropriate AI technique. Comparison of different Machine Learning techniques to estimate the best classifier.

2.2. Objectives of the research work

Machine learning is evolving one of the most powerful tools in every field due to its optimal features. In the arena of medical science early diagnosis of the abnormality can save life. Brain tumor is amongst most aggressive diseases but if detected at an early stage and treated by proper plain life can be saved, with this view the prime objective of this research is to work on an Optimal Classifier to Classify the Brain Tumor using Machine Learning Algorithm so that it provides the radiologist with a second opinion for proper treatment plan for the patient.

3. Methodology followed

3.1. Pre-processing

In image analysis Pre-processing plays an important role it enhances the image to identifying the region of interest. Prior to the processing of the image for the desired task, preprocessing is an important phase where the undesired data is suppressed and the image is made ready for further analysis. The MR images are degraded by distortion during the image digitization and transmission process. The bias field distortion alters the image which causes variations in intensity when captured at different time intervals so to make the intensity and contrast range more analogous intensity normalization methods were proposed [13]. The most common noise removal technique is by applying filters. Various filters remove the different types of noise present in the image. Median filter is commonly used to eliminate the salt and pepper noise present. Gaussian filter eliminates the input MR image noise by suppressing the high frequency components that are more prone to noise resulting in a smooth image hence it is a low pass filter. The presence of extra-cranial tissue like the skin, bones, fat, skull etc. in the MRI of the brain has to be eliminated converting the heterogeneous image into homogeneous image. Anisotropic diffusion filters are well suited for preprocessing the MRI so that the undesired data is suppressed by preventing the fine details of the image for further analysis.

3.2. Feature extraction

Feature extraction techniques explore the images and the objects in the images to excerpt the maximum projecting features that represent the classes of the objects. It provides the classifier with a feature vector that characterizes the input data. Tumors are heterogeneous tissues therefore only the mean value is not sufficient to characterize the heterogeneity of various types of tumors. Discussing about the imaging features enables to understand the various modalities of the DICOM image that help in diagnostic task and the artifacts associated with the images. Three categories of features are extracted that designate the intensity, shape and texture of the image for examples

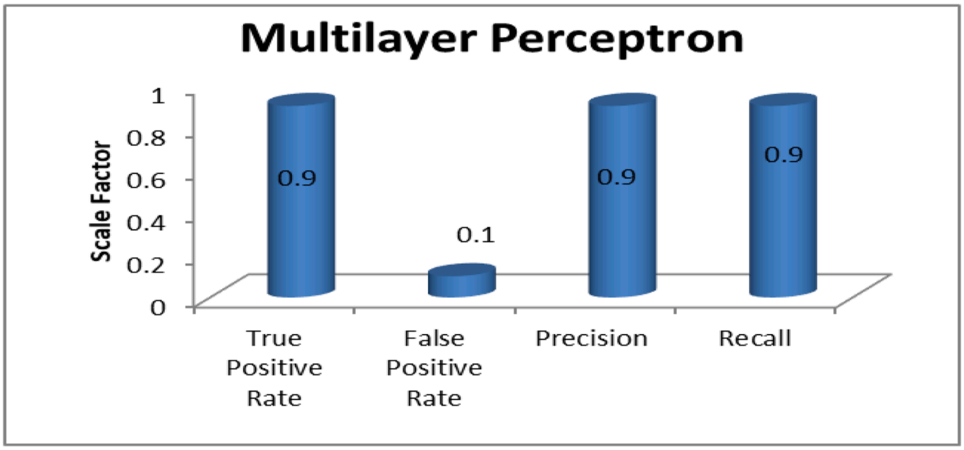

Fig. 11. Plot of Detailed Accuracy Rate for MLP Classifier.

Table 15
Classification Instances for MLP Classifier.

Correctly Classified Instances	90
Incorrectly Classified Instances	10
Percentage Accuracy	90

3.2.1. *Intensity feature*

It describes the Mean, Median intensity, Variance, Standard Variance, Skewness, Kurtosis etc.

3.2.2. *Shape feature*

Shape index, Area, Circularity, Perimeter, Irregularity etc.

3.2.3. *Texture features*

Covers a varied series of procedures built on first and second order image texture parameter such as the Correlation, Sum of square variance, Contrast, Entropy, Homogeneity, Energy, Cluster shade etc.

These feature vectors are plotted in the feature space each feature consisting of one dimension therefore the feature space is n-dimension for n feature vectors. Objects from the same class will cluster in one feature space and thus used for classification. As the feature space increases number of parameters for classification also increase.

The GLCM Features: The discrimination of the abnormal MRI from the normal MRI can be obtained through the gray level co-occurrence matrix (GLCM) [14], the selection of Haralick features that prove the applicability to analyze the irregular object outlines.

The GLCM is the measure of times a pair of pixel k_1,$k_2$with distance d and orientation θ will occur in a given matrix represented by $G_{\theta,d}(k_1, k_2)$.The dimensions of θ are horizontal, vertical, diagonal and anti-

Table 16
Confusion Matrix of MLP Classifier.

True positive = 45	False Positive = 5
False Negative = 5	True Negative = 45

Table 17
Result of Support Vector Machine Classifier for 10 fold cross Validation.

Key Run	Percent correct
1	75
2	50
3	50
4	75
5	75
6	75
7	100
8	75
9	100
10	100

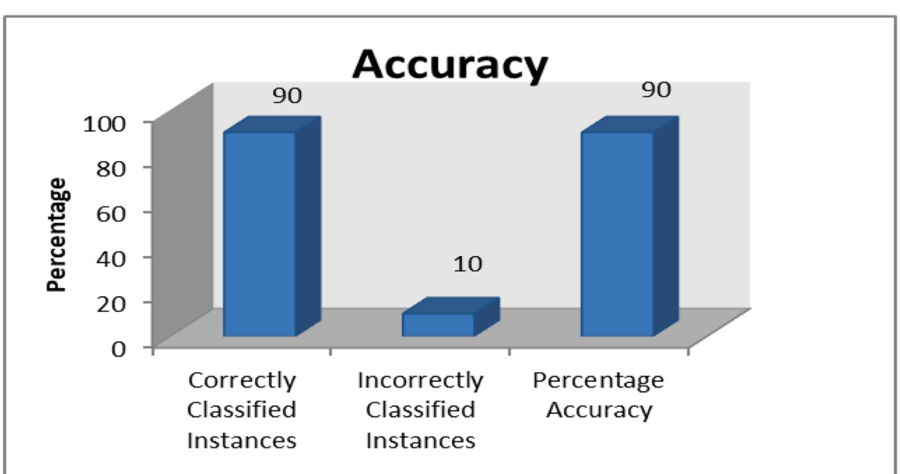

Fig. 12. Plot of Classification Instances for MLP Classification.

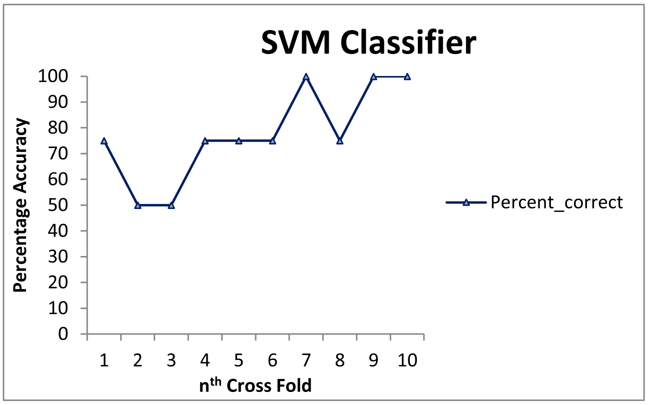

Fig. 13. Plot of Percentage Accuracy for SVM.

Table 18
Mean Error for SVM Classifier.

Mean absolute error	Root mean squared error
0.25	0.5
0.5	0.707107
0.5	0.707107
0.25	0.5
0.25	0.5
0.25	0.5
0	0
0.25	0.5
0	0
0	0

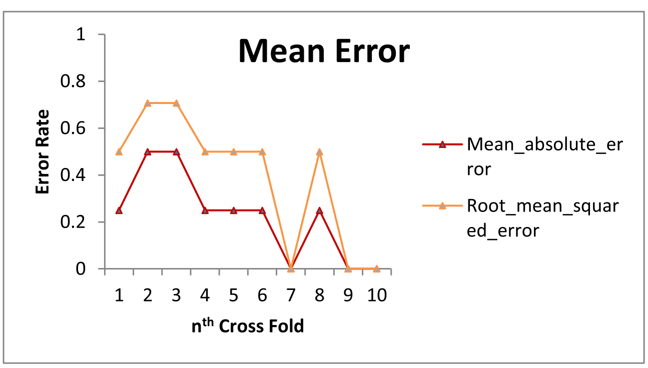

Fig. 14. Plot of mean error for SVM Classifier.

Table 19
Detailed Accuracy Rate for SVM Classifier.

True Positive Rate	1.00
False Positive Rate	0.300
Precision	0.769
Recall	1.00
Accuracy	85%

diagonal i.e.0^0,45^0,90^0,135^0.

3.3. Wavelet transform

A wavelet is a fast-decreasing wave line oscillation that has zero mean and they exist for a finite duration [15]. This enables to study every single component separately; wavelets are created from the basic wavelet Ψ(t). The two important wavelets transform concept are scaling and shifting defined by

$$\Psi_{s,\tau} = \frac{1}{\sqrt{S}}\Psi\left(\frac{t-\tau}{S}\right) \tag{1}$$

Where scaling refers to stretching and shrinking given by $\Psi(t/s)s\ > 0$ 'S' is a scaling factor the positive value resembles to the extent of the amount of a signal that is scaled in time, where S is inversely proportional to time and the constant of proportionality is the center frequency 'cf'' wavelets have band pass characteristics in frequency domain $F_{eq} = cf/S\delta t$where $\delta_t$is the sampling interval. Stretching helps in finding slow varying changes while compressing helps in capturing abrupt changes. Shifting is the delaying or the advancing the concept of wavelength along the signal.

3.4. Discrete Wavelet Transform (DWT)

DWT is based on scaling and shifting of wavelets and are used in the application of de-noising and compression of the signal as it helps in representing many signals that occur naturally with a smaller number of coefficients known as sparkle representation. The DWT base scale is set to 2 and we can obtain different scales with $2^j where j = 1, 2, 3,$ …….. Translation occurs as an integer multiple represented as $2^j*m where m = 1, 2, 3,$ …………This procedure is known as dyadic scaling and changing these results by reducing the redundancy.

The signal is filtered using high pass and low pass filters also known as Detailed (D) sub band and Approximation (A) sub band respectively, this typically filters the length of coefficient for each band which is equal to half the number of coefficients to the previous scale. The filters result in good computational performance as they have a small no of coefficients and with this method the signal of interest can be capture with some larger magnitude DWT coefficients while distortion in the signal consequences the smaller DWT coefficients. Hence the analysis of signal is done using the DWT at progressively narrow sub bands at different resolutions it likewise benefits in de-noising and compressing the signal. Table 1 gives an example of sample result of thirteen features that are extracted from the two classes of MRI scans.

3.4.1. Classification techniques

Classifying helps to understand the type of tumor depending on which the physician can plan the treatment. It is the region of interest to work upon for a given purpose therefore it is an important task in medical analysis for diagnosis of the tumor. Brain Tumors are broadly classified into Benign and Malignant. Classification categorizes each attribute to a predefined set of group or a class. This assigns the attributes to the target class; this predicts the class. Classification is performed with more discriminative features initially on the synthetic database then on the clinically database. Through the performance analysis it was analyzed that the accuracy of the synthetic images is greater than the clinical dataset hence to obtain good accuracy.

3.4.2. Naive Bayes classifier

Bayes theorem is the backbone of this classifier [16], here the event belonging to a particular class is calculated on the conditional probability for a given data given by the equation $P(C_k x) = P(C_k) + \frac{P(x|C_k)}{P(x)}$. The data $x \in$ class k and the best class is selected with the greatest probability this reduces the classification error therefore $P(C_k|x)$ is statistically independent given by $P(xC_k) = \sum_{i=0}^{n} P(x|C_k)$. Where x is a n dimensional vector of the data $= (x_1, x_2, \ldots\ldots, x_n)$. The probability of classification is determined during the training procedure where the calculation of the density function is also done. The Naïve Bayes is also called as the Maximum A Posterior (MAP) decision rule. The class is defined as given in Eq. (2)

$$Class(x) = argmax\ k\frac{1}{N}\sum_{j=1}^{N}\left(P\left(C_{i,j}\right)\prod_{i}^{n}P\left(x_i|C_{j,k}\right)\right. \tag{2}$$

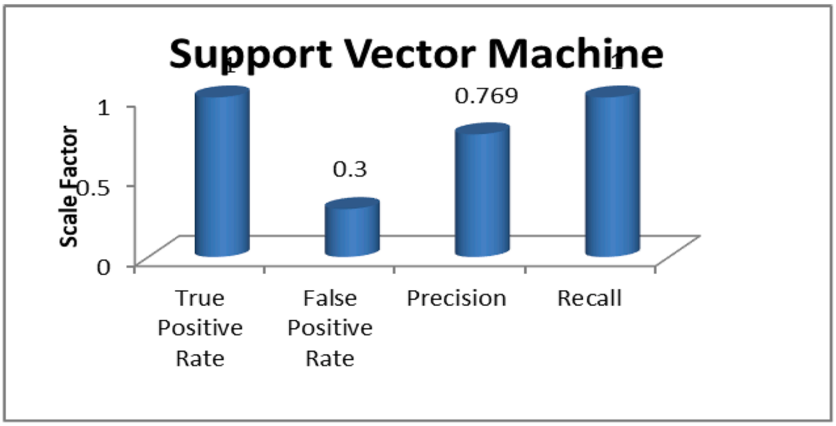

Fig. 15. Plot of Detailed Accuracy Rate for SVM Classifier.

Table 20
Classification Instances for the SVM Classifier.

Correctly Classified Instances	85
Incorrectly Classified Instances	15
Percentage Accuracy	85

Table 21
Confusion Matrix of SVM Classifier.

True positive = 50	False Positive = 0
False Negative = 15	True Negative = 35

3.4.3. *Logistic regression as a linear classifier*

A linear classifier gives an assured estimate of either 1 or 0 i.e. by passing the output of a linear function through the threshold function yet thresholding creates some problems the hypothesis hw(x) is a discontinuous function of input and is not differentiable [17]. These issues can be solved by softening the threshold function to a continuous differentiable function i.e. the integral of normal distribution and the logistic function, there are appropriate properties of the logistic functions given by

$$Logistic(Z) = \frac{1}{1 + e^{-z}} \tag{3}$$

A soft boundary is formed by the hypothesis at the center and approaches to the extremes as it moves away from the boundary. This is known as fitting the weights that is used to minimize the loss; this is called as logistic regression it is the hyper plane separating the two classes. With this function the class label can be predicted but there are limitations for the MRI data as accurate predictions becomes limited since the numbers of features are greater than the observations. Therefore, by applying regularization this limitation can be overcome i.e. by adjusting the weights to higher value for features that are sensitive to class labels and zero otherwise [18]. Gaussian prior is applied to weights leading to regularization and the maximized function is given as

Table 22
Result of Decision Tree Classifier for 10-fold cross Validation.

Key Run	Percent correct
1	50
2	75
3	50
4	75
5	100
6	75
7	75
8	75
9	75
10	75

$$L_g(\theta) = \sum_{i=1}^{N} \log\left(y^i \middle| x^i, \theta\right) - \gamma^{\theta t}\theta \tag{4}$$

Where γ Controls the degree of regularization. Hence the weights of the irrelevant features are reduced.

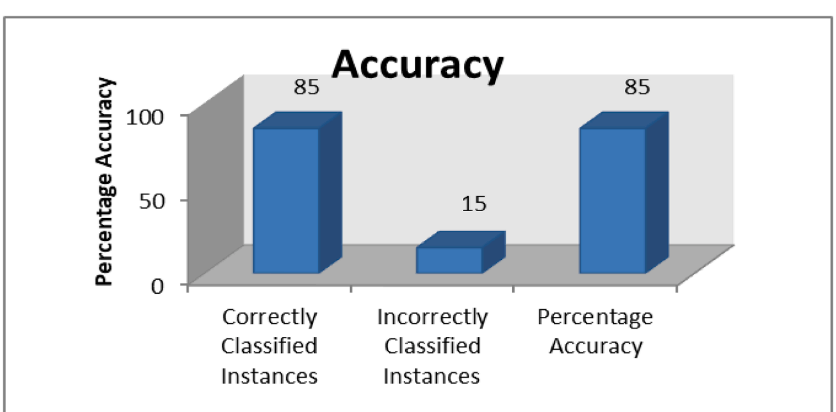

Fig. 16. Plot of Classification Instances for SVM Classifiers.

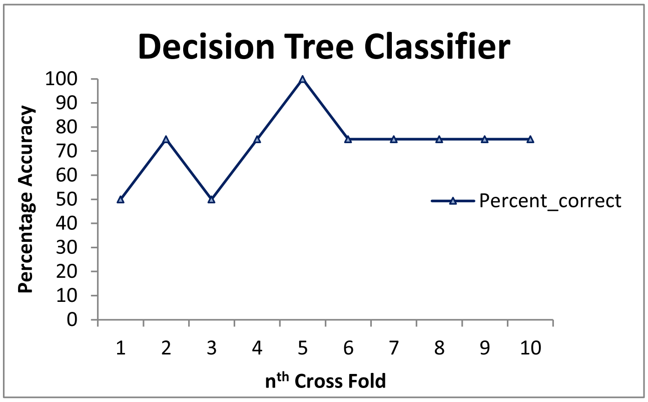

Fig. 17. Plot of Percentage Accuracy for Decision Tree.

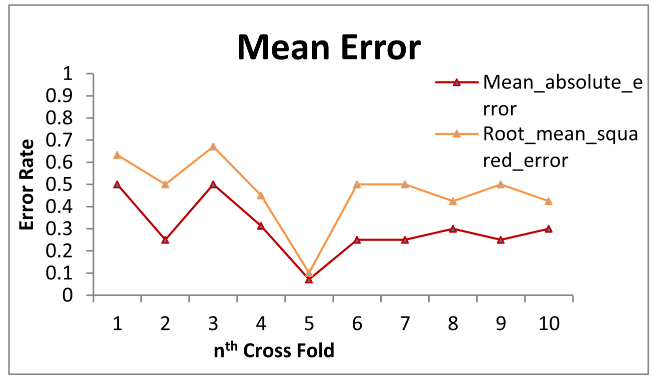

Fig. 18. Plot of mean error for Decision Tree Classifier.

Table 23
Mean Error for Decision Tree Classifier.

Mean absolute error	Root mean squared error
0.5	0.633431
0.25	0.5
0.5	0.67128
0.3125	0.450694
0.071429	0.101015
0.25	0.5
0.25	0.5
0.3	0.424264
0.25	0.5
0.3	0.424264

3.4.4. *Multilayer perceptron*

Multilayer neural network was explored during the 1980s [19] the multilayers of feature detectors were learnt from the training data. The Multilayer's are built on the back propagation algorithm [20] to compute the classification performance of the entire network that is dependent on the weight on every connection.

Multilayer is based on a unit called the Perceptron, is a single unit of a perceptron the input to a perceptron is a vector of real value x_0,x_1,x_2, ………$x_n$that computes the linear combination of these inputs. The output is unity if it is larger than the predefined threshold else the output is -1. Many of the Boolean functions like the AND, OR, NAND are represented by a single perceptron. The learning of a perceptron begins with a random weight that iteratively applies to each training samples modifying the weights until convergence to a hypothesis that classifies the data completely. According to the perceptron training rules the weights are modified at each step given by

$$w_i \leftarrow w_i + \Delta w_i \tag{5}$$

Where $\Delta w_i = \eta(t - o)x_i$ here η is the learning rate that moderates the degree to the changed weights,t is the target output and o is the output generated by the perceptron. This learning procedure converges within a finite no of iterations provided the samples are linearly separable and η

Table 24
Detailed Accuracy Rate for Decision Tree Classifier.

True Positive Rate	0.800
False Positive Rate	0.300
Precision	0.727
Recall	0.800
Accuracy	75%

Table 25
Classification Instances for the Decision Tree Classifier.

Correctly Classified Instances	75
Incorrectly Classified Instances	25
Percentage Accuracy	75

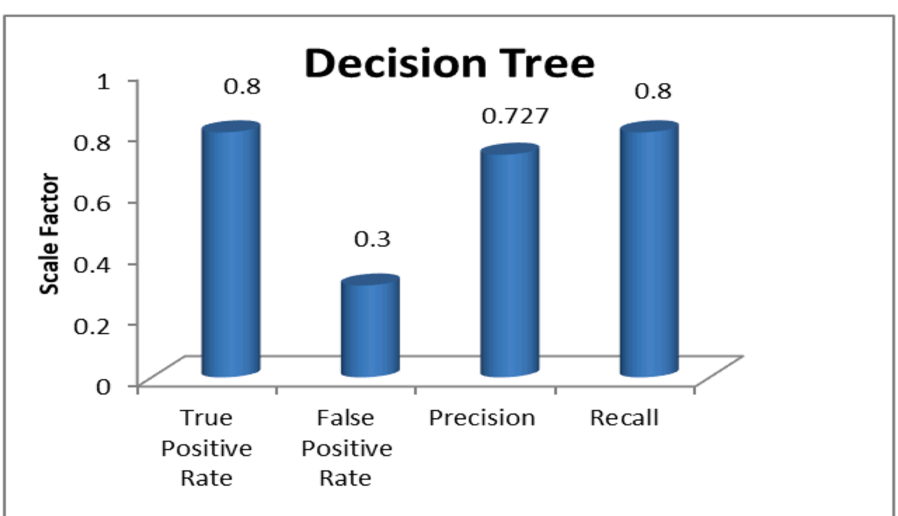

Fig. 19. Plot of Detailed Accuracy Rate for Decision Tree Classifier.

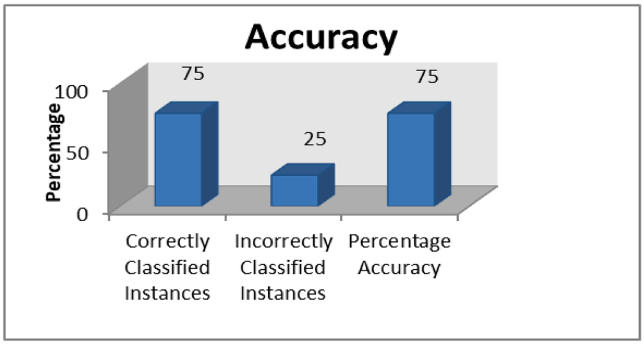

Fig. 20. Plot of Classification Instances for Decision tree Classifiers.

Table 26
Confusion Matrix of Decision Tree Classifier.

True positive = 40	False Positive = 10
False Negative = 15	True Negative = 35

Table 27
Comparison of results for different Machine Learning Techniques.

Classifiers	OASIS and ADNI Data set
Naïve Bayes	80%
Logistic Regression	90%
Multilayer Perceptron	90%
Support Vector Machine	85%
Decision Tree	75%

being sufficiently small. Convergence is not confident if the data is not linearly separable.

3.4.5. The support vector machine

SVM is a machine learning method coined from statistical theory applied for classification of images, broadly used in pattern recognition it has a good generalization performance and computational efficiently. It takes labeled data as the input from the two classes and the output is a labeled class of the two classes [21]. Training implicates feeding the known data along with prior decision values creating finite training set. Initiation of learning of the SVM classifier can be described as follows, if a vector $x \in R^n$ designate a pattern to be classified and y designate its training labels i.e. $y = \pm 1$ and if $\{(x_i,y_i), i = 1, 2, \ldots\ldots\ldots l\}$ denote the training samples then the classifier can be constructed with the function $f(x)$.

Linear classifier is the simplest case where the input patterns are linearly separable and the linear function is of the form

$$f(x) = W^T x + b \tag{6}$$

hence for each training sample x_i the function produces

$$\begin{cases} f(x_i) \geq 0 \; for \; y_i = +1 \\ f(x_i) < 0 for \; y_i = -1 \end{cases} \tag{7}$$

Hence forming the various classes that are detached by the hyper plane$f(x) = W^T x + b = 0$.

There exist numerous hyper planes for a given set however the SVM classifier considers the hyper plane that maximizes the separating margin among the two classes [22].

Nonlinear SVM classifier is the extension of the linear classifier that uses a nonlinear operator $\phi(\cdot)$ that maps input pattern x into greater dimensional space which is defined by

$$f(x) = W^T \phi(x) + b \tag{8}$$

Eq. (8) is linear in terms transformed data $f(x)$ but nonlinear for original data $x \in R^n$ the parameters of the decision function $f(x)$ is resolute by minimization criteria given by

$$Min \; J(W, \xi) = \frac{1}{2} W^2 + C \sum \xi_i \quad i = 0, 1, 2, \ldots\ldots\ldots, l \tag{9}$$

Subjected to

$$y_i \left(W^T \phi(x_i) + b \right) \geq 1 - \xi_{i,} \; \xi_i \geq 0; \quad i = 1, 2, \ldots\ldots, l \tag{10}$$

The kernel function plays a vital part of mapping the input vector into a greater dimensional feature space it defines the relaxation of the margin to admit some of the points to occupy the opposite margin. It is essential to check whether the kernel is connected with the inner product of some nonlinear mapping. The kernel function must satisfy the Mercer's theorem [23] defines that for every square integral function $g(\cdot)$ the kernel should satisfy the following condition.

$$\iint K(x, y) g(x) g(y) d_x d_y \geq 0 \tag{11}$$

The frequently used kernel function includes polynomials and radial basis function (RBF), the polynomial kernel is defined as follows

$$K(x, y) = \left(x^T + 1 \right)^p \quad where \; p > 0 \tag{12}$$

And finally, the decision function is defined as

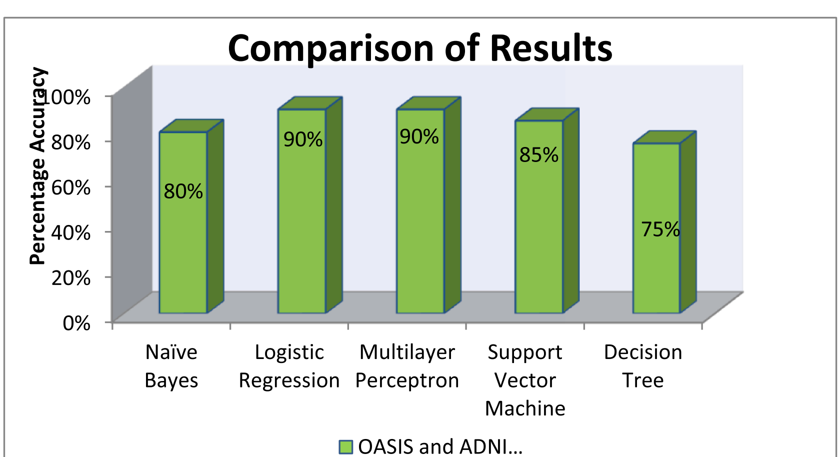

Fig. 21. Plot showing the comparison of results for different data set.

$$f(x) = \pm\left(\sum y_i \alpha_i \ k(x_i, x) - b\right. \tag{13}$$

Where x_i is a support vector, α_i Lagrange multiplier, $k(x_i,x)$ the convolutional of the inner product to get the feature space and the support vector that is equal to the decision function in the higher dimensional space.

3.4.6. Decision tree

A hierarchical data structure implementing the divide and rule method is the decision tree. It is an effectual nonparametric process applied for classification as well as regression that can be transformed to a set of rules that are easily understood [24]. Here the input space is divided into local regions that are defined by local model calculated from the training data. It is composed of internal decision nodes and terminated by leaves. Starting from the root for every input node a test is applied and one of the branches is selected depending on the outcome this process is repeated recursively till a leaf node is hit; i.e. a class is obtained at the leaf node.

Decision tree approximates the discrete value functions with noisy data and is capable of learning disjunctive expressions. It classifies the instances by categorizing them down the tree from the root to the leaf node [25] where the leaf node provides the classification of the instances. Node represents the test of some attribute of the instance and every branch downward from the node resembles to one of the possible values of this attribute. The top-down, greedy search are the core algorithms of the decision tree illustrated by ID3 (Quinlan 1986) succeeded by C4.5 (Quinlan 1993).

Selecting the best attribute at every step of the tree for achieving accurate classification is done by the ID3 algorithm that is defined by the statistical property called the information gain; which is the measure of how well an attribute separates the training samples with respect to the target classification.

The information gain is defined by

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{14}$$

Where *values*$(A)$the values of attribute A, and S are is a subset for which A has the value $vS_v = \{S \in S|A(S) = v\}$. ID3 is characterized as searching a space of hypothesis which is set of possible decision trees. It executes a simple to complex hill climbing search through the hypothesis space starting with the empty tree.

4. Results of various implemented machine learning algorithms

4.1. Result of Naïve Bayes classifier

The results of Naïve Bayes classifier with respect to 10-fold cross validation on a set of 100 MRI scans is given in Table 2 and the same is plotted in Fig. 1. The mean absolute error and the root mean square error are tabulated in Table 3 and plotted in Fig. 2. The details of accuracy rate are tabulated in Table 4 and plotted in Fig. 3 these indicates the true positive rate, false positive, precision and recall, Table 5 gives the classification instances and the same are illustrates in Fig. 4, the confusion matrix is given in Table 6.

4.2. The result for the logistic classifier

With respect to percentage accuracy is tabulated in Table 7 and the same is plotted in Fig. 5, the mean error rates with respect to mean absolute error and root mean square error are tabulated in Table 8 and plotted in Fig. 6. The details of accuracy are given in Table 9 and illustrated in Fig. 7; Table 10 gives the classification instances applied to the logistic regression which are plotted in Fig. 8 and the confusion matrix is given in Table 11.

4.3. Result of the MLP

The results of percentage correct classification applied for multilayer Perceptron is given in Table 12 and plotted in Fig. 9, the mean error is given in Table 13 and exemplified in Fig. 10, the details of classification accuracy is tabulated in Table 14 which is illustrated in Fig. 11, the Table 15 gives the classification instances and plotted in Fig. 12 and the confusion matrix is given in Table 16.

4.4. Result of the SVM

The result of the SVM classifier with respect to accuracy in terms of percentage correct classification is particularized in Table 17 and plotted in Fig. 13 and the mean error are tabulated in Table 18 and the same is plotted in Fig. 14, the Table 19 gives the details of accuracy classification and the same is plotted in Fig. 15 and Table 20 gives the classification instances which is plotted in Fig. 16 the confusion matrix with respect to the SVM classifier is given in Table 21.

4.5. Result of the decision tree

The result of the decision tree with respect to correctly classified instances is given in Table 22 which is also illustrated in Fig. 17 the error with respect to mean absolute error and RMS error is tabulated in Table 23 which is again elaborated in Fig. 18, the detailed accuracy rate is given in Table 24 and the values are plotted in Fig. 19 the classification instances with respect to correct classification and accuracy is given in Table 25 and plotted in Fig. 20 and finally the confusion matrix is given in Table 26. Table 27 gives the comparison of results for different Machine Learning Techniques, the graph of the same is Plotted in Fig. 21.

5. Conclusion of the research

This research shows the state-of-art methods for classification of the tumor as cancerous and non-cancerous (Benign and Malignant). The advantage here is that it does not need any prior information about the probability distribution of different classes. Logistic and MLP are the branches of Machine learning that learns, and performing parallel computations. The most advantageous part in applying ML model is that the result of the system does not depends on the dataset and the structure of the network. Hence the Logistic and MLP may support the radiologist for classification of the tumors and for treatment planning which is the main objective of this research. Machine Learning for tumor classification without much human interference gives promising results, the Performance of the Logistic and MLP gives more accurate and promising results of about 90% on. Compared to the other traditional machine learning technique MLP has more powerful learning and expressing abilities that make remarkable achievements in classification task in the field of computer vision.

Declaration of Competing Interest

The authors declare that we have no conflict of interest.

References

[1] Devos A, Lukas L, Simonetti AW, Suykens JAK, Vanhamme L, van der Graaf M, Buydens LMC, Heerschap A, Van Huffel S. Does the combination of magnetic resonance imaging and spectroscopic imaging improve the classification of brain tumours?. In: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2004. p. 407–10. Vol. 1.
[2] Liu L, Zheng G, Bastian JD, Keel MJB, Nolte LP, Siebenrock KA, Ecker TM. Periacetabular osteotomy through the pararectus approach: technical feasibility and control of fragment mobility by a validated surgical navigation system in a cadaver experiment. Int Orthop 2016;40(7):1389–96.
[3] Open access series of imaging studies Available on http://www.oasis-brains.org/.
[4] http://adni.loni.usc.edu/data-samples/data-types/.

[5] Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imag 2016;35(5):1240–51.
[6] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012:1097–105.
[7] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H. Brain tumor segmentation with deep neural networks. Med Image Anal 2017;35:18–31.
[8] Mohsen H, El-Dahshan ESA, El-Horbaty ESM, Salem ABM. Classification using deep learning neural networks for brain tumors. Future Comput Inf J 2018;3(1):68–71.
[9] Ari A, Hanbay D. Deep learning based brain tumor classification and detection system. Turkish J Electric Eng Comput Sci 2018;26(5):2275–86.
[10] Ilunga-Mbuyamba E, Avina–Cervantes JG, Garcia–Perez A, de Jesus Romero–Troncoso R, Aguirre–Ramos H, Cruz–Aceves I, Chalopin C. Localized active contour model with background intensity compensation applied on automatic MR brain tumor segmentation. Neurocomputing 2017;220:84–97.
[11] Sankari A, Vigneshwari S. Automatic tumor segmentation using convolutional neural networks. In: 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM). IEEE; 2017. p. 268–72.
[12] Sajjad M, Khan S, Muhammad K, Wu W, Ullah A, Baik SW. Multi-grade brain tumor classification using deep CNN with extensive data augmentation. J Comput Sci 2019;30:174–82.
[13] http://www.madehow.com/Volume-3/Magnetic-Resonance-Imaging-MRI.html.
[14] Alpaydin E. Introduction to machine learning. MIT press; 2014.
[15] Selvaraj H, Selvi ST, Selvathi D, Gewali L. Brain MRI slices classification using least squares support vector machine. Inte Intell Comput Med Sci Image Process 2007;1(1):21–33.
[16] Jang M, Park D. Stochastic classifier integration model. Int J Appl. Eng Res 2016;11(2):809–14.
[17] Russell SJ, Norvig P. Artificial intelligence: a modern approach. Malaysia: Pearson Education Limited; 2016.
[18] Ryali S, Supekar K, Abrams DA, Menon V. Sparse logistic regression for whole-brain classification of fMRI data. Neuroimage 2010;51(2):752–64.
[19] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012:1097–105.
[20] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation (No. ICS-8506). California Univ San Diego La Jolla Inst for Cognitive Science; 1985.
[21] Chaplot S, Patnaik LM, Jagannathan NR. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomed Signal Process Control 2006;1(1):86–92.
[22] Burges CJ. A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 1998;2(2):121–67.
[23] Burges CJ, Scholkopf B, Smola AJ, editors. Advances in kernel methods: support vector learning. Cambridge, MA, USA: MIT press; 1999. p. 89–116.
[24] Lemort M, Canizares-Perez AC, Van der Stappen A, Kampouridis S. Progress in magnetic resonance imaging of brain tumours. Curr Opin Oncol 2007;19(6):616–22.
[25] Egmont-Petersen M, de Ridder D, Handels H. Image processing with neural networks—a review. Pattern Recognit 2002;35(10):2279–301.