Brain Tumor Detector using Hybrid Logistic Regression Algorithm with MRI images

Manorama Chouhan[1], Jatin Kushwaha[2] and Devendra Thakur[3]

[1,2,3]School of Computing Artificial Intelligence, VIT Bhopal University, Madhya Pradesh - 466114

Abstract. Early detection of brain tumors is vital for improving treatment outcomes. Brain tumors are typically classified through biopsy, which requires definitive brain surgery. This study utilized a dataset comprising 3,000 MRI brain images, which included cases of glioma, meningioma, pituitary gland tumors, as well as healthy brains. Various preprocessing and augmentation techniques were employed to improve the quality and diversity of the MRI images. This report presents the findings of our brain tumour detection methods using various machine learning models. As we have searched a lots of review paper and models of data we got the best accuracy in this Dataset. The balanced dataset contains 1500 tumour images and 1500 non-tumour images. We assessed the effectiveness of several models, including Logistic Regression, Support Vector Classifier (SVC), k-Nearest Neighbors (KNN), Naive Bayes, Neural Networks, Random Forest, and K-means clustering. Each model's performance was evaluated by analyzing metrics such as accuracy, precision, recall, F1-score, true positive and negative rates, false positive and negative rates, and area under the curve (AUC). Logistic Regression (96%), Neural Network (95%), and Random Forest (96%) exhibited the highest performance, particularly in accuracy, precision, recall, and F1-score, successfully distinguishing between tumor and non-tumor images, thereby yielding reliable results. Early detection of brain tumors plays a crucial role in determining patient prognosis and treatment plans. This study employed advanced image processing techniques to improve MRI scan quality. A variety of machine learning models were utilized for a comprehensive analysis of the data, with performance rigorously evaluated using multiple metrics. Future research could aim to optimize these models for clinical use, contributing to more accurate and timely tumor diagnosis.

Keywords: Naive Bayes Algorithm, support vector machine (SVM), Neural network, K- means, image classification, Brain Tumour

1. Introduction

Tumours are abnormal cell growths that can be classified as either malignant or benign. Among the more than 200 types of tumours that can affect humans [2] brain tumours represent a particularly serious condition. These tumours, which involve irregular cell proliferation within the brain, can significantly impair neurological function. Over the past three decades, the mortality rate due to brain tumours has surged by 300%. Without timely intervention, [5] brain tumours can be fatal, making early detection and treatment crucial for enhancing patient survival rates. Diagnosing brain tumours poses significant challenges due to their complex nature. Traditional methods such as biopsies require invasive surgical procedures, underscoring the need for non-invasive diagnostic alternatives. Magnetic Resonance Imaging (MRI) is widely utilized in the diagnosis of brain tumours, offering detailed images of brain structures. Symptoms of brain tumours can vary but commonly include - *Headaches, Seizures, Vision problems, Muscle twitching and shaking, Drowsiness, and nausea, and vomiting* Brain tumours are a critical medical condition characterized caused by the growth of abnormal cells in the brain, which may be benign (non-cancerous) or malignant (cancerous). [4] These tumours can significantly impact brain function, leading to severe health issues and increased mortality rates if not diagnosed and treated promptly. The detection and classification of brain tumours are vital for effective treatment planning and improved patient outcomes. Recent advancements in machine learning, especially deep learning, have transformed the analysis and classification of medical images. These techniques have been effectively used across various medical fields, including disease prognosis, diagnosis, image classification, and tissue segmentation.[3] Given the variability in pathology and the potential for human error, computer-assisted methods can provide critical support in identifying and classifying tumours. Machine learning and deep learning in particular, plays a pivotal role in the analysis, segmentation, and classification of cancer images, with significant applications in brain tumour detection. These methods offer precise and dependable tumour identification, distinguishing them from other conditions. Brain tumours are a diverse group of neoplasms that originate in the brain or its surrounding tissues. They can present significant clinical challenges due to their potential to disrupt normal brain function, leading to a wide range of neurological symptoms. Malignant brain tumours are associated with high morbidity and mortality rates, emphasizing the need for effective diagnostic tools and timely intervention. Magnetic Resonance Imaging (MRI) plays a crucial role in the non-invasive diagnosis of brain tumors [6]. This technology uses strong magnetic fields and radio waves to produce detailed brain images, enabling the detection of tumors and other abnormalities. Despite the efficacy of MRI in detecting brain tumours, the interpretation of these images can be complex, requiring considerable expertise and experience. Advanced machine learning techniques offer a promising solution to enhance the diagnostic process. These Algorithms can be trained to detect patterns in MRI data that may signal the presence of tumors, potentially enhancing the accuracy and speed of diagnosis. The Naive Bayes algorithm, a statistical classifier grounded in Bayes' theorem, is especially suitable for this purpose due to its efficiency and capability to manage high-dimensional data. In this study, we focus on employing the Naive Bayes algorithm for brain tumour classification. Our approach leverages existing knowledge while addressing previous limitations, and we compare the performance of seven different modelling methods to identify significant performance differences.

2. Literature Review

The field of brain tumour detection has seen significant advancements, with researchers employing various techniques to enhance the accuracy and efficiency of diagnostic methods. [3] This literature review summarizes key studies and their contributions to brain tumour detection, focusing on challenges, segmentation techniques, and the application of One of the key challenges in detecting brain tumors using machine learning algorithms is the presence of varying intensity levels and indistinct boundaries within the tumor image.[6] These variations can arise due to differences in the magnetic field during MRI scans, which can affect image quality. To address this, researchers have implemented techniques such as intensity normalization and bias field correction. These features provide crucial information that can be leveraged by different algorithms to distinguish between tumour and non-tumour regions.

Segmentation Techniques: -Segmentation is a critical step in brain tumour detection, involving the partitioning of images into meaningful segments.

- **Spatial Clustering**: This technique differentiates between image segmentation and clustering by performing grouping in the spatial domain rather than the measurement space. Spatial clustering algorithms consider the spatial proximity of pixels to create segments that represent different tissue types or structures.
- **Split and Merge Segmentation**: This method starts by considering the entire image and recursively splitting it into smaller regions until a homogeneity criterion is met. Conversely, the merge method combines adjacent segments of the same object, improving the representation of homogeneous regions.
- **Region Growing**: In this approach, segmentation begins with a seed point and expands by connecting neighbouring pixels that meet certain criteria, such as intensity similarity. The success of region growing heavily Relies on choosing a suitable threshold value to achieve precise region expansion.

These techniques represent just a fraction of the methods developed for brain tumour segmentation. Many other various algorithms have been applied to classify tumour and non-tumour regions based on features extracted from brain images. Key machine learning techniques include. Support Vector Classifier (SVC) is A technique that identifies the best hyperplane to distinguish between different classes within the feature space. K-Nearest Neighbors (KNN) is a non-parametric approach that classifies samples by considering the majority class among their closest neighbors. Decision Trees: A model that divides data into branches, making decisions based on feature values. Random Forests: An ensemble technique that enhances classification accuracy by combining multiple decision trees. Logistic Regression: A statistical method that estimates the probability of a binary outcome from given input features. Neural Networks: Deep learning architectures that utilize layers of interconnected neurons to capture and learn intricate patterns within the data The availability of benchmark datasets, such as the MICCAI BraTS dataset, has been instrumental in advancing brain tumour detection research. These datasets provide standardized data that researchers can use to train, test, and compare their algorithms.[7] The BraTS dataset offers a comprehensive collection of MRI scans with annotated tumour region. In the table 1 Detail of Author name and used algorithm name with accuracy in their article.

Table 1. Detail of Author name and used algorithm name with accuracy in their article

Author Name	**Algorithm name**	**Algorithm (description)**	**Accuracy**
Hein Tun Zaw; Noppadol Maneerat 2017[13]	Neural network	Identifying and separating tumor regions from the surrounding brain tissues by analyzing intensity differences	95%.
Danda, Shashank Reddy2014 [1]	Decision tree Algorithm	The Decision Tree algorithm classifies brain MRI images by repeatedly dividing the data into subsets according to the most important attribute at each node	90%
Sandeep Kumar Mathivanan, Sridevi 2019 [4]	Support vector machine	MRI images by finding the optimal hyperplane that separates tumor and non-tumor classes based on feature vectors.	92%
Nitesh Kumar Singh, Geeta 2015 [12]	K-means Clustering Algorithm	The K-means algorithm partitions the image into K clusters based on. pixel intensity.	66%

3. Proposed Work

The goal of this proposed work is to create a reliable and the effective system for detecting brain tumours using the Naive Bayes algorithm. This system is designed to categorize MRI brain images as either containing tumours or being tumour-free by utilizing the unique features found in the images.

3.1 Dataset

The dataset for this project will be sourced from Kaggle and includes 3,000 MRI brain images. These images are categorized into four groups: glioma, meningioma, pituitary gland tumours, and healthy brains. The dataset is evenly distributed with 1,500 images of tumorous brains and 1,500 images of non-tumorous brains, providing a solid foundation for both training and testing the model.

3.2 Preprocessing

In the preprocessing we apply noise reduction, normalization, augmentation. In noise reduction, this is Techniques like Gaussian filtering will be utilized to remove noise from MRI images and enhance their quality, ensuring that critical features for tumour detection are preserved. In normalization will be applied to standardize pixel values across different images, addressing any variations in brightness and contrast that could impact model performance and the augmentation to enhance the diversity of the training data and improve model generalization, image augmentation methods such as rotation, flipping, and scaling will be implemented. In figure 1 two image original and cropped image shown.

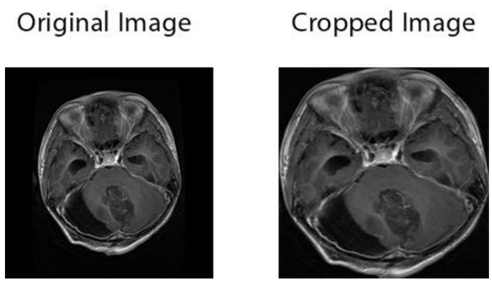

Fig .1 Original & Cropped Image

3.3 Feature Extraction

Extracting key features from MRI images is crucial for precise classification. The features to be extracted include: Texture Features, which involve using techniques like the Gray Level Co-occurrence Matrix (GLCM) to capture the textural patterns within the image. **Shape Features**: Identifying and extracting shape-related features to distinguish between tumour and non-tumour regions. **Intensity Features**: Utilizing intensity histograms to capture the distribution of pixel intensities within the images.

3.4 Model Development

The Naive Bayes algorithm, recognized for its simplicity and efficiency, will be employed for classification. The development process will involve the following steps: **Training**: The dataset will be split into training and validation sets. The Naive Bayes classifier will be trained on features derived from the training set. **Evaluation:** The model's effectiveness will be measured using the validation set. Metrics such as accuracy, precision, recall, F1-score, and area under the curve (AUC) will be used to assess performance. **Optimization**: To improve the Naive Bayes model's performance, hyperparameter tuning will be performed. Techniques like grid search or random search will be utilized to identify the optimal parameters

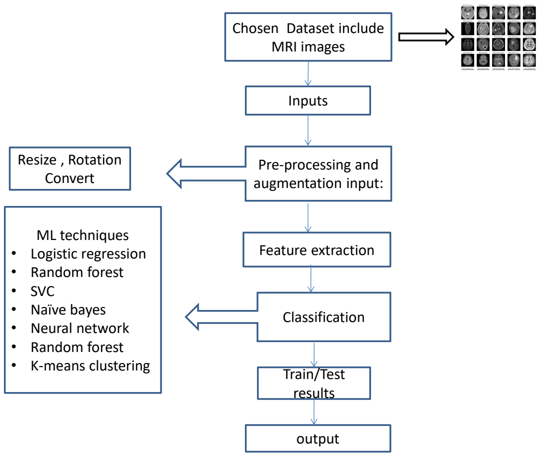

Fig.2 Architecture Diagram of Algorithms with MRI images

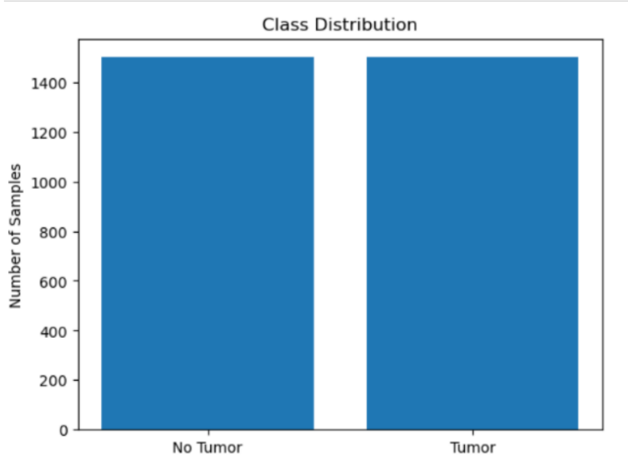

Fig.3 Class Distribution (No Tumor & Tumor)

In a binary classification scenario, outcomes can be categorized as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positives and True Negatives refer to correctly identified normal and abnormal images, respectively. A False Positive occurs when an image is incorrectly identified as positive when it is actually negative.

A False Positive represents a false alarm in the classification process, where an image is incorrectly identified as positive when it is actually negative. Conversely, a False

Negative occurs when an image is wrongly classified as negative when it should have been positive.

TP = The count of abnormal images correctly identified.
TN = The count of normal images correctly identified.
FP = The count of normal images misclassified as abnormal.
FN = The count of abnormal images misclassified as normal.

Accuracy: It is the proportion of accurately predicted samples relative to the total number of samples.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{1}$$

Precision : Precision is a measure of how many of the samples that the model predicts as positive are actually positive.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

Recall: Recall is a measure of how many true positives (TP) are detected by the model.

$$\text{Recall} = \frac{TF}{TP+FN} \tag{3}$$

F1 Score : The F1 Score is the harmonic mean of precision and recall.

$$\text{F1 Score} = 2.\frac{Precision\ .\ Recall}{Precision+Recall} \tag{4}$$

3.4 Models

Several classification algorithms and one clustering algorithm were evaluated to assess their effectiveness in predicting class labels. Logistic Regression, which estimates the likelihood of an input belonging to a particular class using a logistic function, achieved an accuracy rate of 96%. The Support Vector Classifier (SVC), which finds the optimal hyperplane to separate data points in a high-dimensional feature space, attained an accuracy of 94%. The k-Nearest Neighbors (KNN) algorithm, which classifies inputs based on the majority vote among its k nearest neighbors, reached an accuracy of 90%. Naive Bayes, a probabilistic classifier that assumes feature independence given the class label, achieved a 64% accuracy. Neural Networks (ANNs), which consist of layers of interconnected nodes capable of learning complex data patterns, also reached a 94% accuracy rate. Finally, K-means clustering, an unsupervised learning method that groups data into k clusters based on similarity, was included in the analysis 64%.

Epochs Curves

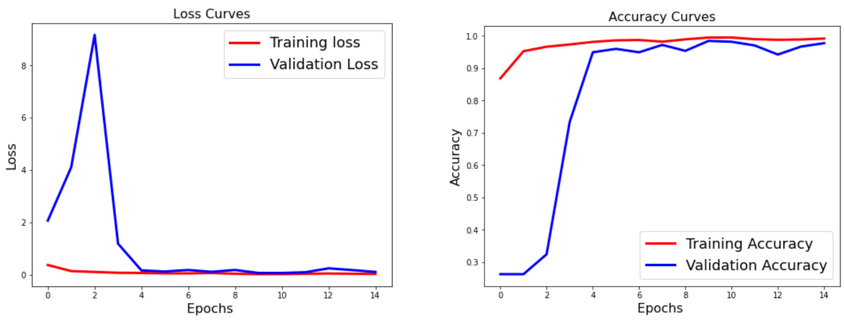

Fig .4 Loss Curve & Accuracy Curve (Epochs)

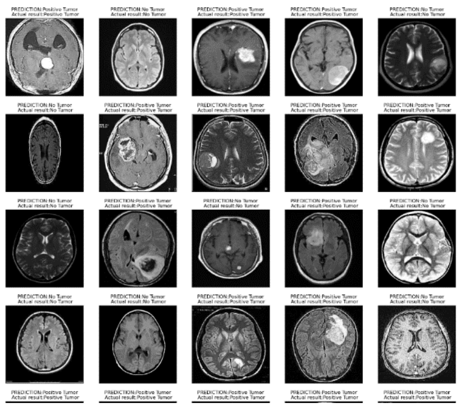

Fig .5 Sample images of Brain tumors 1,500 images of tumorous brains and 1,500 images of non-tumorous brains

ROC Curve: The Receiver Operating Characteristic (ROC) curve visually represents a classifier's performance by plotting the True Positive Rate (sensitivity) against the False Positive Rate (1-specificity) across different thresholds

Confusion Matrix: A confusion matrix is a table used to evaluate the performance of a classification model, showing the counts of true positives, true negatives, false positives, and false negatives

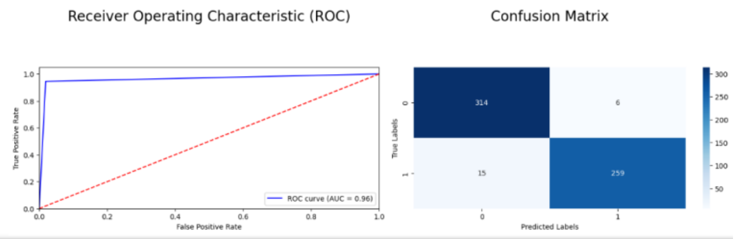

Fig .6 ROC Curves & Confusion matrix

Table 2. Result of all algorithm according to Test/Train, Precision, recall, f1-score, and support

Algorithm Name	**Test/ Train in (Binary)**	**precision**	**recall**	**f1-score**	**Support**
Logistic Regression					
	0	0.95	0.97	0.96	306
	1	0.96	0.94	0.95	288
SVC					
	0	0.94	0.94	0.94	306
	1	0.94	0.94	0.94	288
KNN					
	0	0.85	0.97	0.91	306
	1	0.97	0.81	0.88	288
Naïve Bayes					
	0	0.63	0.70	0.66	306
	1	0.64	0.57	0.60	288
Neural Network					
	0	0.94	0.97	0.96	306
	1	0.97	0.94	0.95	288
Random Forest					
	0	0.98	0.94	0.96	306
	1	0.97	0.98	0.96	288
K-means Clustering					
	0	0.64	0.76	0.69	1482
	1	0.70	0.57	0.63	1488

4 Conclusion

After evaluating the performance of various machine learning models on the brain tumour detection dataset, several conclusions were reached. Logistic Regression, Neural Networks, and Random Forest emerged as the top performers, showcasing high accuracy, precision, recall, and F1-score. These models excelled in distinguishing between tumorous and non-tumorous MRI images, providing reliable results. Support Vector Classifier (SVC) and k-Nearest Neighbors (KNN) also demonstrated commendable performance, achieving accuracy rates above 90%, and maintaining a good balance between precision and recall, making them viable alternatives for brain tumour detection. On the other hand, the Naive Bayes model showed lower accuracy and performance metrics, indicating it may not be as effective for this particular task due to its lower precision, recall, and F1-score. K-means clustering, primarily a clustering algorithm, also showed lower performance, highlighting its unsuitability for binary classification in brain tumour detection. Overall, The Logistic Regression model proved to be the top performer, achieving 96% accuracy along with high precision, recall, and F1-score, indicating its strong ability to distinguish between classes. In contrast, Naive Bayes was the least effective, with an accuracy of 64% and higher rates of false positives and negatives, suggesting limitations in accurately classifying brain tumour images.

References

1. Janki naik 1, Sagar Patel 2 "Tumor Detection and Classification using Decision Tree in Brain MRI" IJCSNS (International journal of computer science and Network Security, Vol. 14 no.6, June 2014, pp.87-91.
2. Varun Jain Sunila Gondara, (june2017), Comparative Study of Data Mining Classification Methods in Brain Tumor Disease Detection.IJCSC, Vol.8, Issue-2 pp.12-17.
3. Kalyani A.Bhawar, Prof.Ajay S.chhajed "Brain Tumor classification using Data mining algorithms.IJESRT, November-2016, pp.239-243.
4. V.Vani, M.Kalaiselvi Geetha Automatic Tumor Classification of Brain MRI Images .IJCSC, Vol-4, Issue-10, pp.144-151.
5. Mr. Meena, E Murali Study on Various Machine Learning Algorithms for Brain Tumor Detection.Ijpam, Volume-117 No.8 2017, pp.139-143.
6. Chau, M.; Shin, D. (2009). A Comparative Study of Medical Data Classification Methods Based onDecision Tree and Bagging Algorithms. Proceedings of IEEE International Conference onDependable, Autonomic and Secure Computing, pp. 183-187.
7. Kearns M. and Mansour Y. On the boosting ability of top-down decision treelearning algorithms.Journal of Computer and Systems Sciences, 58(1): 109-128, 1999.
8. Brain Tumor ppt. Dr.Walaa Nasr Lecturer of Medical-Surgical Nursing Department Second year 2012.
9. C.Ordonez, E. Omiecinski ,"Image mining: A new approach for data mining," Technical Report GITCC-98-12, Georgia Institute of Technology, College of Computing, 1998, pp 1-21.
10. Kareem, S.M.; "Detection and determination of Brain Tumor from MRI of brain and its volume calculation", Journal of Basrah Researches ((Sciences)), Vol. (41). No. (1)A (2015)
11. Roy et al., "A Review on Automated Brain Tumor Detection and Segmentation from MRI of Brain", International Journal of Advanced Research in Computer Science and Software Engineering 3(6), June - 2013, pp. 1706-1746
12. Nameirakpam Dhanachandra∗, Khumanthem Manglem and Yambem Jina Chanu, Image Segmentation using K-means Clustering Algorithm andSubtractive Clustering Algorithm, Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science 54 (2015) 764 – 771.
13. Neeraj Battish Department of CSECGC-COE, Dapinder Kaur Department of CSECGC-COELandran, Mohali, Analysis of Pre-processing Methods using MRI Images, International Journal of Modern Computer Science (IJMCS) ISSN: 2320-7868 (Online)Volume 5, Issue 2, April 2017.
14. Poonam1, Jyotika Pruthi Computer Science and Engineering, Itm University, Gurgaon, India, Review of Image Processing Techniques for Automatic Detection of Tumor in Human Brain, Poonam et al, International Journal of Computer Science and Mobile Computing Vol.2 Issue. 11, November- 2013, pg. 117-122.
15. Liangxiao Jiang et al, "Deep feature weighting for naive Bayes and its application to text classification", Engineering Applications of Artificial Intelligence, Volume 52, June 2016, Pages 26-39.
16. S. M. Ali, Loay Kadom Abood and Rabab Saadoon Abdoon, "Brain tumor Extraction in MRI images using Clustering and Morphological Operations Techniques".
17. Rohit S. Kabade and M. S. Gaikwad,; "Segmentation of Brain Tumor and its Area Calculation in Brain MR Images using K-Mean Clustering and Fuzzy CMean Algorithm", vol. 4, issue 5, pp. 2229–3345, May (2013).

18. Das A.J et al., “Automatic Detection of Brain Tumor from MR Images using Morphological Operations and K-Means Based Segmentation”, Proceedings of the 2nd international conference on “Emerging research in computing, information, communication and application” ERCICA 2014.
19. Sindhushree. K. S, Mrs. Manjula. T. R, K. Ramesha, “Detection and 3d Reconstruction Of Brain Tumor from Brain Mri Images”, International Journal of Engineering Research & Technology (IJERT), vol. 2, no. 8, pp 528-534, 2013
20. Wu. T et al, “A prior feature SVM-MRF based method for mouse brain segmentation”, NeuroImage, Volume 59, Issue 3, 1 February 2012, Pages 2298-2306
21. Nameirakpam Dhanachandra∗, Khumanthem Manglem and Yambem Jina Chanu, Image Segmentation using K-means Clustering Algorithm andSubtractive Clustering Algorithm, Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science 54 (2015) 764 – 771.
22. V. Velusamy, Dr. M. Karnan, Dr. R. Sivakumar, Dr. N. Nandhagopal, Enhancement Techniques and Methods for MRIA Review, V. Velusamy et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 397-403.
23. Poonam1, Jyotika Pruthi Computer Science and Engineering, Itm University, Gurgaon, India, Review of Image Processing Techniques for Automatic Detection of Tumor in Human Brain, Poonam et al, International Journal of Computer Science and Mobile Computing Vol.2 Issue. 11, November- 2013, pg. 117-122.
24. Poonam1SivaSankari.S, Sindhu.M, Sangeetha.R, ShenbagaRajan.AU.G. Student, Department of Computer Engineering, PSRR Engineering College, Sivakasi, Tamilnadu, India, Associate Professor, Department of Computer Engineering, PSRR Engineering College, Sivakasi, Tamilnadu, India, Feature Extraction of Brain Tumor Using MRI, International Journal of Innovative Research in Science, Engineering and Technology, 2018: Volume 7, Issue 6.
25. Jay Gholap Dept. of Computer Engineering College of Engineering, Pune, Maharashtra, India, Performance Tuning of j48 Algorithm For Prediction of Soil Fertility, Asian Journal of Computer Science And Information Technology 2: 8 (2012) 251– 252,2013: ISSN 2249-5126
26. Parmeshwarappa, V., Nandish, S.: A segmented morphological approach to detect tumor in brain images. IJARCSSE 4(1), 408–412 (2014)
27. Karuna, M., Joshi, A.: Automatic detection and severity analysis of brain tumors using GUI in matlab. IJRET Int. J. Res. Eng. Technol. **2**(10), 586–594 (2013)
28. Chew, K.M., Yong, C.Y., Sudirman, R., Wei, S.T.C.: Bio-signal processing and 2D representation for brain tumor detection using microwave signal analysis. In: IEEE (2018)
29. Chauhan, S., More, A., Uikey, R., Malviya, P., Moghe, A.: Brain tumor detection and classification in MRI images using image and data mining. In: IEEE (2017)
30. Dhanwani, D.C., Bartere, M.M.: Survey on various techniques of brain tumour detection from MRI images. IJCER **4**(1), 24–26 (2014).