International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

A Comprehensive Analysis Of Machine Learning Techniques For Brain Tumor Detection And Prediction: Methods, Outcomes, And Comparisons

Teena Chinnam[1], Naga Rajeswari.K[2], Dilip Kumar.K[3]
[1]M.Tech Student, Department of Computer science Engineering, Sree vahini institute of Science & Technology, Tiruvuru-521 235, AP, India.
[2]Assistant Professor, Department of Computer science Engineering, Sree vahini institute of Science & Technology, Tiruvuru-521 235, AP, India.
[3]M.Tech Student, Department of Computer science Engineering, Sree vahini institute of Science & Technology, Tiruvuru-521 235, AP, India.

Abstract
Brain cancer diagnosis remains a complex task because tumors vary significantly in location, size, shape, and appearance across patients. These variations make precise detection and classification essential for timely treatment and improved clinical outcomes. This study presents a comprehensive review of brain tumor analysis using Magnetic Resonance Imaging (MRI) and compares the performance of multiple machine learning algorithms on a pre-processed dataset containing both tumor and healthy brain scans. The algorithms evaluated include Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, Logistic Regression, CatBoost, and Neural Networks. To enhance classification performance, meaningful features were extracted from the MRI images, including texture characteristics, pixel intensity patterns, and shape-based descriptors. These features were used to train and test the models systematically. The performance of each algorithm was measured using standard evaluation metrics such as accuracy, sensitivity, specificity, and precision to ensure a comprehensive assessment of diagnostic effectiveness. The results demonstrate that all models achieved satisfactory performance in distinguishing tumor from non-tumor cases. Among them, CatBoost and Random Forest delivered the highest accuracy and sensitivity, indicating strong predictive capability and reliability. SVM showed consistent efficiency in tumor detection tasks, while Neural Networks were particularly effective in classifying tumor subtypes. Overall, the findings highlight the strengths and limitations of different machine learning approaches and suggest that ensemble techniques offer significant potential for improving the accuracy, robustness, and reliability of automated brain tumor prediction systems.
Keywords: Brain Cancer, Machine Learning Algorithms, Magnetic Resonance Imaging, Accuracy, Texture, Precision.

1. INTRODUCTION

The rapid growth of Artificial Intelligence (AI) and Machine Learning (ML) has transformed clinical decision-making, particularly in digital pathology and medical imaging. In brain tumor analysis, where MR image interpretation is complex and time-consuming, these technologies play a crucial role [1-8]. Gliomas, the most common brain malignancies, are highly heterogeneous, making manual detection and classification challenging for radiologists. To address this, computer-aided methods using ML have been developed to improve accuracy and efficiency [9-10]. This study reviews a range of ML techniques for brain tumor segmentation and classification, aiming to identify the most effective approaches for diagnosis and treatment planning. Using a dataset of both low- and high-grade tumors, we evaluated twenty-one state-of-the-art algorithms [11,12]. Findings highlight the potential of AI and ML to enhance brain tumor analysis, support radiotherapy planning, and ultimately revolutionize patient care.

DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

This study analyzed eight machine learning techniques for brain tumor classification, focusing on accuracy. Results showed that combining algorithms, particularly with feature selection methods such as KNN, SVM, Naïve Bayes, Random Forest, and CatBoost, significantly improved performance. These findings highlight the value of algorithmic integration in refining tumor segmentation and diagnosis for clinical use. The introduction emphasizes the role of IT and e-healthcare in enhancing communication and diagnosis, noting the limitations of traditional radiology and the importance of timely, accurate brain tumor detection for effective treatment.

Brain tumors are abnormal tissue growths that increase intracranial pressure and disrupt brain function. They are classified as benign (non-cancerous) or malignant (cancerous), with malignant tumors growing rapidly, damaging healthy tissue, and sometimes spreading [13-15]. Tumors are graded I–IV: Grade I (slow-growing, often removable, e.g., pilocytic astrocytoma), Grade II (slow but infiltrative, e.g., oligodendroglioma), Grade III (aggressive, requiring chemo/radiotherapy, e.g., anaplastic astrocytoma), and Grade IV (fastest growing, highly malignant, e.g., glioblastoma multiforme). Understanding these grades is essential for accurate diagnosis and personalized treatment planning [16,17].

2. RELATED WORK

Extensive research has applied machine learning for brain tumor prediction, using methods such as Naïve Bayes, K-Nearest Neighbors, Gaussian Process Classification, SVM, AdaBoost, Logistic Regression, Decision Trees, and Random Forests [18,19]. Preprocessing steps like outlier removal and missing-value handling have further improved model accuracy. Recent advances also incorporate binary thresholding, PCA for feature extraction, and CNNs for classification and segmentation in medical imaging. Literature consistently highlights SVM as one of the most effective techniques for tumor classification [20-25]. The aim of this study is to classify brain tumors from MRI scans as cancerous or non-cancerous with high precision. The paper is structured as follows: Section 3 details the methods, Section 4 presents results, and Section 5 concludes with future research directions.

3. METHODOLOGY:

Description Dataset:

Model accuracy and efficacy were evaluated on a dataset with dimensions (683, 10) for independent variables and (683,) for dependent variables. Preprocessing employed a median filter, which effectively removes noise and anomalies while preserving image edges and preventing blurring [26-30]. Machine learning techniques were implemented using Python 3.7. A detailed description of the dataset attributes is provided below.

1. Clump Thickness
2. Uniformity of Cell Size
3. Uniformity of Cell Shape
4. Marginal Adhesion
5. Single Epithelial Cell Size
6. Bare Nuclei
7. Bland Chromatin
8. Normal Nucleoli
9. Mitoses
10. Class: (2 for benign, 4 for malignant)

A. Approach for Binary Classification

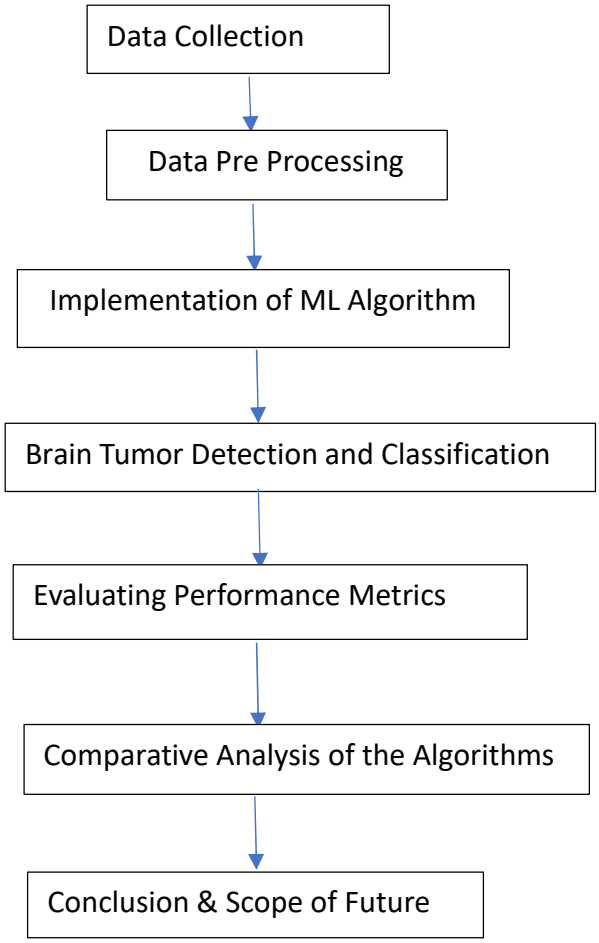
Within this segment, an elucidation of the approach employed to discern whether a patient possesses a brain tumor through brain MRI images in Dataset-A has been provided.

1) Preparation of Data Assignment of Labels to Data Brain tumor images have been categorized as '1,' while images without tumors have been denoted as '0."

2) Splitting the dataset into train and test

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

Division of Dataset The complete dataset was segregated into training and testing subsets, with a test proportion of 25%.



The depicted approach has been visually represented in FIGURE 1.

3) Applied ML algorithm for analysis

This research employs a range of machine learning methods, including Regression, KNN, SVM, SVM Kernel, Naïve Bayes, Decision Tree, Random Forest, and CatBoost, to systematically address the study objectives. Each algorithm was implemented and evaluated on the dataset, with results analyzed through confusion matrices to assess predictive performance. The findings provide insights into the strengths and applicability of each method in achieving the research goals.

3.1.1 Logistic Regression

Logistic Regression is a statistical method primarily used for binary classification, estimating the probability of an instance belonging to a class using the logistic (sigmoid) function. Unlike linear regression, which predicts continuous outcomes, it transforms a linear combination of input features into a probability between 0 and 1 via the log-odds ratio. The model is simple, interpretable, and computationally efficient, making it widely applicable in domains such as medical diagnosis, marketing analytics, and sentiment analysis. Despite its linear nature, Logistic Regression remains a powerful baseline and foundation for more advanced classification techniques.

Table 1. Depicting Confusion Matrix for Logistic Regression model

	Brain Tumor	Non-Brain Tumor
Brain Tumor	114	1
Non-Brain Tumor	4	52

3.1.2 K-Nearest Neighbors: K-Nearest Neighbors (KNN) is a supervised algorithm used for both classification and regression. It classifies a data point based on the majority label of its ‘k’ nearest neighbors, making it simple and intuitive. As a non-parametric and lazy learner, KNN requires no

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

explicit training phase and uses all available data during classification. Its performance depends on proper data normalization and the choice of 'k.' KNN works well with datasets containing clear clusters and finds applications in image recognition, recommendation systems, and medical diagnosis.

Table 2. Depicting Confusion Matrix for KNN model

	Brain Tumor	Non-Brain Tumor
Brain Tumor	114	1
Non-Brain Tumor	4	52

3.1.3 Support Vector Machine (SVM): Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification and regression. Its goal is to identify an optimal hyperplane that separates data points of different classes with the maximum margin. By mapping inputs into higher-dimensional spaces, SVM can handle both linear and non-linear problems through kernel functions. It excels in managing high-dimensional data, creating complex decision boundaries, and achieving strong generalization. SVM represents training data as points in space, with boundaries defined by parallel lines around the hyperplane, enabling accurate prediction of new instances.

Table 3. Depicting Confusion Matrix for SVM model

	Brain Tumor	Non-Brain Tumor
Brain Tumor	114	1
Non-Brain Tumor	3	53

3.1.4 Support Vector Machine (SVM) Kernel: SVM kernels enhance the capabilities of Support Vector Machines by mapping input data into higher-dimensional spaces, enabling the handling of complex and non-linear relationships. This allows SVMs to identify non-linear decision boundaries even when the data is not linearly separable in the original feature space. Common kernels include polynomial, radial basis function (RBF), and sigmoid, each suited for different problem types. Kernels achieve computational efficiency by calculating inner products in high-dimensional space without explicitly transforming the data. The choice of kernel depends on the data structure, making it crucial for effective model performance. By leveraging kernels, SVMs can address diverse real-world challenges such as image recognition, text analysis, and bioinformatics.

Table 4. Depicting Confusion Matrix for SVM Kernel model

	Brain Tumor	Non-Brain Tumor
Brain Tumor	114	2
Non-Brain Tumor	3	53

3.1.5 Naïve Bayes: Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem, assuming conditional independence among features given the class label hence the term "naïve." It is simple, computationally efficient, and effective for both binary and multi-class classification, handling continuous and categorical data with ease. Despite its strong independence assumption, it often delivers competitive results, particularly in high-dimensional applications like text categorization. Each feature is evaluated independently for example, a fruit classified by attributes such as color, taste, and shape. Its speed and scalability make it well-suited for large datasets, and its performance can be interpreted through a confusion matrix, as shown in Table 5.

Table 5. Depicting Confusion Matrix for Naïve Bayes model

	Brain Tumor	Non-Brain Tumor
Brain Tumor	114	1
Non-Brain Tumor	3	53

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

3.1.6 Decision Tree: A Decision Tree is a graphical model for classification and regression that recursively splits data into subsets based on input features. Internal nodes represent feature-based decisions, branches denote possible outcomes, and leaf nodes provide the final prediction or class. This tree-like structure is easy to interpret, making decision-making transparent. Decision Trees handle both categorical and numerical data and can capture complex relationships, though they are prone to overfitting. To improve robustness, ensemble methods such as Random Forest and boosting are commonly applied.

Table 6. Depicting Confusion Matrix for Decision Tree model

	Brain Tumor	Non-Brain Tumor
Brain Tumor	112	3
Non-Brain Tumor	6	50

3.1.7 Random Forest: Random Forest is a robust and versatile ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. By introducing randomness in feature selection and data sampling, it lowers bias and enhances generalization. Capable of handling diverse data types and high-dimensional spaces, it also provides feature importance rankings. Known for its stability, scalability, and resilience to noisy data, Random Forest is widely applied in fields such as remote sensing, finance, and genetics, making it a reliable choice for complex classification and regression tasks.

Table 7. Depicting Confusion Matrix for Random Forest model

	Brain Tumor	Non-Brain Tumor
Brain Tumor	114	1
Non-Brain Tumor	3	53

3.1.8 Cat Boost: CatBoost, developed by Yandex, is a powerful gradient boosting framework designed for efficient handling of categorical features with minimal preprocessing. Using ordered boosting and an oblivious tree structure, it reduces overfitting while delivering high accuracy. With GPU support, automatic handling of missing values, and compatibility with custom loss functions, CatBoost offers strong scalability and performance. Its versatility makes it well-suited for applications such as recommendation systems, marketing analytics, and fraud detection.

Table 8. Depicting Confusion Matrix for Cat the Boost model

	Brain Tumor	Non-Brain Tumor
Brain Tumor	114	1
Non-Brain Tumor	3	53

4. RESULTS AND DISCUSSION

This study employed a range of ML algorithms, including Logistic Regression, K-Nearest

S. no.	ML Algorithm	Precision	Recall	F-Measure	Accuracy %
1	Logistic Regression	0.95	0.96	0.95	0.99
2	K-Nearest Neighbors	0.96	0.97	0.96	0.96
3	Support Vector Machine	0.97	0.97	0.97	0.96

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

4	Support Vector Machine Kernel	0.98	0.98	0.98	0.97
5	Navie Bayes	0.95	0.95	0.95	0.95
6	Decision Tree	0.95	0.98	0.96	0.95
7	Random Forest	0.98	0.99	0.98	0.98
8	Cat Boost	0.99	0.99	0.99	0.99

Table 9. Machine Learning Model Results

Neighbors, SVM, SVM Kernel, Naïve Bayes, Decision Tree, Random Forest, and CatBoost. Experiments were conducted systematically, and performance was evaluated using precision, recall, accuracy, and F-measure. These metrics enabled a thorough comparison of classification effectiveness, with detailed results and analyses presented in Table 9 to highlight performance across different scenarios.

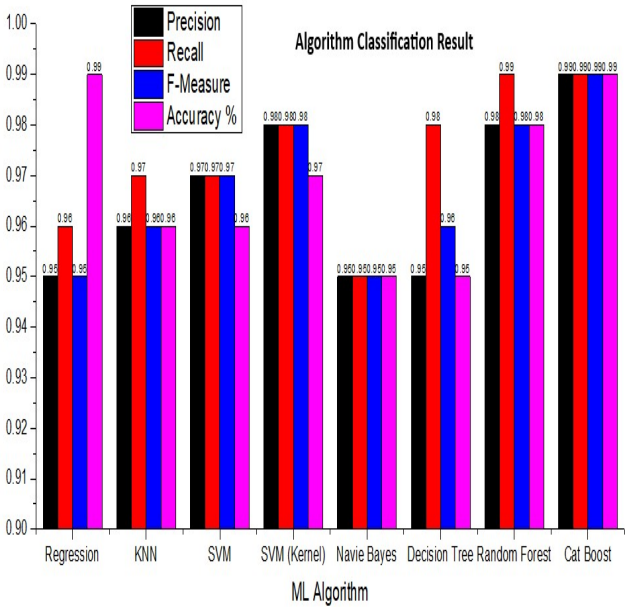


FIGURE 2: Comparison of different algorithm for precision, recall, Accuracy

Precision: Precision measures how accurately a model identifies positive instances out of all instances predicted as positive, reflecting the quality of its positive predictions. Figure 3 shows a comparative analysis of precision across different machine learning algorithms.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

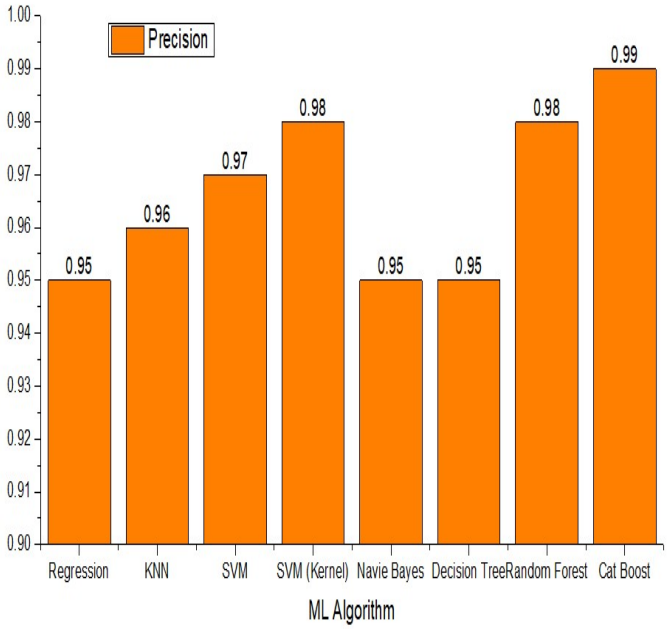


FIGURE 3. Precision comparison

Recall: Recall is a key performance metric in classification that measures a model's ability to correctly identify positive instances from all actual positives. It reflects the model's effectiveness in minimizing missed relevant cases, thereby indicating its comprehensiveness in detecting instances of interest. In essence, recall ensures accurate recognition of positive data points. Figure 4 presents a comparative analysis of recall across different machine learning algorithms.

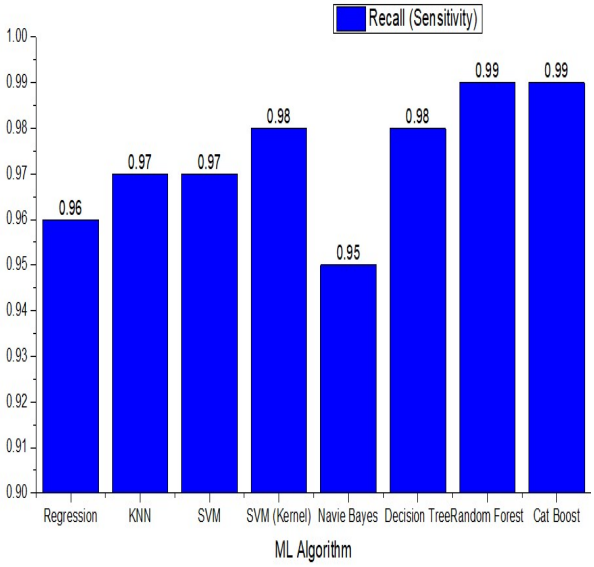


FIGURE 4. Recall comparison

F measure: The F-measure, or F1-score, is the harmonic mean of precision and recall, combining both metrics into a single value. It reflects a model's ability to correctly identify positive instances while also capturing all relevant ones, making it especially useful when balancing precision and recall is important. This metric provides a holistic view of model performance by accounting for both false positives and false negatives. Figure 5 shows a comparative analysis of F-measure across different machine learning algorithms.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

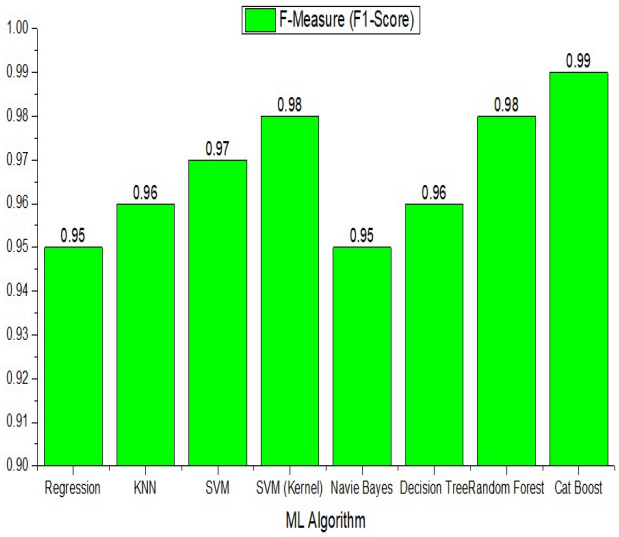


FIGURE 5. F Measure comparison

Accuracy: Accuracy is the proportion of correctly predicted instances out of all instances in a dataset, reflecting the overall correctness of a classifier in classification or regression tasks. It is a fundamental metric for evaluating model effectiveness, particularly useful with balanced datasets, though it can be misleading with imbalanced ones—where precision, recall, and F-measure provide additional insight. High accuracy indicates strong capability in distinguishing between positive and negative classes. Figure 6 presents a comparative analysis of accuracy across different machine learning algorithms.

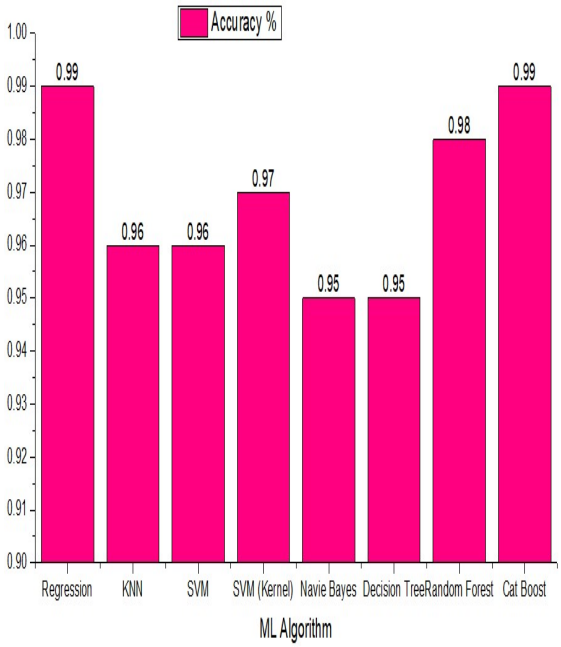


FIGURE 6: Comparison of Accuracy

CONCLUSION AND FUTURE WORK

This study aimed to identify the most effective machine learning model for early-stage brain tumor prediction. Eight classification algorithms were evaluated using multiple performance metrics on a brain tumor dataset, implemented in Python. Results indicate the model with the highest accuracy in tumor prediction. Future research will explore data augmentation and additional machine learning techniques to further enhance classification accuracy.

DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

REFERENCES

1) Lars Kunze, Nick Hawes, Tom Duckett, Marc Hanheide and Tomáš Krajník, Artificial Intelligence for Long-Term Robot Autonomy: A Survey‖, IEEE Robotics and Automation Letters, vol.3, issue.4, pp.4023-4030, 2018.
2) Li Deng, ―Artificial Intelligence in the Rising Wave of Deep Learning: The Historical Path and Future Outlook [Perspectives]‖, IEEE Signal Processing Magazine, vol.35, issue.1, pp.180-187, 2018.
3) Ruimin Ke, Yifan Zhuang, Ziyuan Pu and Yinhai Wang, ―A Smart, Efficient, and Reliable Parking Surveillance System with Edge Artificial Intelligence on IoT Devices‖, IEEE Transactions on Intelligent Transportation Systems, vol.22, issue.8, pp. 4962 – 4974, 2021.
4) P. P. Bhattacharya, Alok Kole, Tanmay Maity and Ananya Sarkar,'Neural Network Based Energy Efficiency Enhancement in Wireless Sensor Networks', International Journal of Applied Engineering Research, vol. 9, no.22, pp. 11807-11818, 2014.
5) Huimin Lu, Yujie Li, Min Chen, Hyoungseop Kim and Seiichi Serikawa, ―Brain Intelligence: Go beyond Artificial Intelligence‖, Mobile Networks and Applications, vol.23, pp.368-375, 2018.
6) Sanjeevani Bhardwaj and Alok Kole, _Review and Study of Internet of Things: It's the Future ', in Proc. IEEE International Conference on Intelligent Control, Power and Instrumentation (ICICPI-2016), Kolkata, India, 2016, pp.47-50.
7) Jiaying Liu, Xiangjie Kong, Feng Xia, Xiaomei Bai, Lei Wang, Qing Qing and Ivan Lee, Artificial Intelligence in the 21st Century‖, IEEE Access, vol.6, pp. 34403 – 34421, 2018.
8) Chinmaya Kumar Pradhan, Shariar Rahaman, Md. Abdul Alim Sheikh, Alok Kole and Tanmoy Maity, _EEG Signal Analysis Using Different Clustering Techniques ', in Proc. International Conference on Emerging Technologies in Data Mining and Information Security, Kolkata, West Bengal, 2018, pp.99-105.
9) Ankit Ghosh, Purbita Kole and Alok Kole, ―Automatic Identification of Covid-19 from Chest X-ray Images using Enhanced Machine Learning Techniques‖, International Research Journal of Engineering and Technology (IRJET), vol.8, issue.9, no.115, pp.765-772, 2021.
10)Subhamoy Mandal, Aaron B. Greenblatt and Jingzhi An, ―Imaging Intelligence: AI Is Transforming Medical Imaging Across the Imaging Spectrum‖, IEEE Pulse, vol.9, Issue.5, pp. 16 – 24, 2018.
11)Sanjay Saxena, Neeraj Sharma and Shiru Sharma, Image Processing Tasks using Parallel Computing in Multi core Architecture and its Applications in Medical Imaging, International Journal of Advanced Research in Computer and Communication Engineering, vol.2, issue.4, pp.1896-1900, 2013.
12) Chetanpal Singh, ―Medical Imaging using Deep Learning Models‖, European Journal of Engineering and Technology Research, vol.6, issue.5, pp.156-167, 2021.
13) D. Garrett, D. A. Peterson, C. W. Anderson and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 11, no. 2, pp. 141-144.
14) Webb, J. Boughton and Z. Wang, "Not So Naive Bayes: Aggregating One-Dependence Estimators", Machine Learning, vol. 58, no. 1, pp. 5-24.
15) S. Brahim-Belhouari and A. Bermak, "Gaussian process for non-stationary time series prediction," Comput. Statist. Data Anal., vol. 47, no. 4, pp. 705-712.
16) C. Cortes and V. Vapnik, "Support vector networks Mach. Learn", vol. 20, no. 3, pp. 273.
17) A. Reinhardt, T. Hubbard, Using neural networks for prediction of the sub cellular location of proteins, Nucleic Acids Research, vol. 26, Issue 9, 1 May 1998, pp. 2230–2236.

DOI:10.46647/ijetms.2026.v10i02.003 ISSN: 2581-4621

18) Jena, B.; Nayak, G.K.; Saxena, S. An empirical study of different machine learning techniques for brain tumor classification and subsequent segmentation using hybrid texture feature. Mach. Vis. Appl. 2021, 33, 6.
19) Deepak, S.; Ameer, P.M. Automated Categorization of Brain Tumor from MRI Using CNN features and SVM. J. Ambient. Intell. Humaniz. Comput. 2020, 12, 8357–8369.
20) Garg, G., & Garg, R. Brain tumor detection and classification based on hybrid ensemble classifier. 2101.00216, 2021.
21) Bojaraj Leena, and Annamalai Jayanthi, "Brain tumor segmentation and classification via adaptive CLFAHE with hybrid classification," in Int J Imaging Syst Technol. Wiley, 2020.
22) Prabha, S.; Raghav, R.; Moulya, C.; Preethi, K.G.; Sankaran, K. Fusion based Brain Tumor Classification using Multiscale Transform Methods. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 28–30 July 2020, pp. 1390–1393.
23) Zöllner FG, Emblem KE, Schad LR (2012) SVM-based glioma grading: optimization by feature reduction analysis. Z Med Phys 22:205–214.
24) Arakeri MP, Reddy GRM (2015) Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images. SIViP 9:409–425.
25) Yang G, Zhang Y, Yang J, Ji G, Dong Z, Wang S et al (2016) Automated classification of brain images using wavelet-energy and biogeography-based optimization. Multimedia Tools Appl 75:15601–15617.
26)Kibriya, H.; Amin, R.; Alshehri, A.H.; Masood, M.; Alshamrani, S.S.; Alshehri, A. A Novel and Effective Brain Tumor Classification Model Using Deep Feature Fusion and Famous Machine Learning Classifiers. Comput. Intell. Neurosci. 2022, 2022, 7897669.
27) Kégl B (2013) the return of AdaBoost.MH: multi-class Hamming trees. arXiv:1312.6086.
28) A. Saffari, C. Leistner, J. Santner, M. Godec and H. Bischof, "On-line Random Forests," 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, 2009, pp. 1393-1400.
29)H. Mohsen, E. El-Dahshan, E. El-Horbaty, A. Salem, Brain tumor type classification based on support vector machine in magnetic resonance images, Annals Of Dunarea De Jos" University Of Galati, Mathematics, Physics, Theoretical mechanics, Fascicle II,Year IX (XL).
30) Karl Pearson F.R.S. (1901) LIII. On lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2:11, 559-572, DOI: 10.1080/14786440109462720.