

# Brain Tumor Detection

1<sup>st</sup> Akmal

*Artificial Intelligence.CSE  
Amrita School Of Engineering  
Bengaluru,Karnataka,India*

bl.en.u4aie22150@bl.students.amrita.edu

2<sup>nd</sup> Anju Chowdary

*Artificial Intelligence.CSE  
Amrita School Of Engineering  
Bengaluru,Karnataka,India*

bl.en.u4aie22170@bl.students.amrita.edu

3<sup>rd</sup> Mahithi Reddy

*Artificial Intelligence.CSE  
Amrita School Of Engineering  
Bengaluru,Karnataka,India*

bl.en.u4aie22178@bl.students.amrita.edu

**Abstract**—This research addresses the vital task of brain tumor detection through a machine learning algorithms such as Logistic Regression and Support Vector Machines (SVM) as they are best suitable for such predictions.Principal Component Analysis(PCA) is used for the reduction of dimensions of the images.The logistic regression model and the svm model undergoes training with the datasets and then a real time prediction is done for predicting the specific image.

**Index Terms**—Medical Image Analysis, Computer-Aided Diagnosis (CAD), Feature Extraction, Training and Testing Set, Image Preprocessing, Dimensionality Reduction, Accuracy Score, OpenCV, Data Visualization, Algorithm Performance, Hyperparameter Tuning

## I. INTRODUCTION

Brain tumors pose a significant health challenge, necessitating accurate and efficient diagnostic tools. In this context, machine learning emerges as a better approach, to enhance the accuracy of brain tumor detection. This study focuses on the utilization of Logistic Regression and SVM as said in [1] and [3], two widely adopted machine learning techniques, for the purpose of brain tumor identification. This goes from preprocessing of medical imaging data to the application of dimensionality reduction techniques like PCA which was mentioned in [9].

## II. LITERATURE SURVEY

presents a survey on brain tumor segmentation and classification in MRI, employing Support Vector Machine (SVM).The segmentation process involves overcoming challenges associated with the diverse shapes and sizes of brain tumors, incorporating phases such as pre-processing, segmentation, feature extraction, feature reduction, and tumor classification.Techniques applied as given in the paper include Otsu thresholding for segmentation, Discrete Wavelet Transform (DWT) for feature extraction, Principal Component Analysis (PCA) for feature reduction, and classification using SVM, Least Squared-Support Vector Machine (LS-SVM), Proximal Support Vector Machine (PSVM), and Twin Support Vector Machine (TWSVM). Comparative analysis done in [1] reveals that TWSVM exhibits superior performance with 100 percent accuracy, positioning it as an effective tool for brain tumor classification.

paper delves into medical image processing, specifically targeting brain tumor detection using computer tomography (CT) scans.The study helps in identifying the noises and then removing them,followed by classification techniques like Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM).This [2] paper also emphasizes the broader landscape of medical image processing, highlighting the significance of noninvasive diagnostic solutions and digital image processing techniques. However, a notable gap exists in the lack of detailed discussion on comparative analysis methodology, crucial for advancing the field.

study employs a logistic regression-based machine learning technique for automatic brain abnormality detection, utilizing MRI images for training and testing. Overcoming the limitations of outdated models, the proposed method [3] achieves performance measures of 97percent accuracy, 97.9 percent precision, and 97 percent recall, outperforming existing methodologies. The work aims to contribute to the early and precise diagnosis of brain diseases through advanced machine learning technologies.

Massive datasets are generated across diverse sectors, necessitating the use of machine learning algorithms to uncover patterns and make predictions. However, not all attributes in these datasets are essential for training machine learning models, and some may even be irrelevant.[4] paper explores the impact of dimensionality reduction techniques, specifically Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA), on four widely used machine learning algorithms: Decision Tree Induction, Support Vector Machine (SVM), Naive Bayes Classifier, and Random Forest Classifier. The study employs the Cardiocography (CTG) dataset and extends the experimentation to Diabetic Retinopathy (DR) and Intrusion Detection System (IDS) datasets. Results demonstrate that PCA outperforms LDA, and the application of dimensionality reduction does not significantly degrade the performance of machine learning algorithms, particularly in scenarios with high-dimensional datasets.

paper presents a comprehensive survey on the analyses of medical images using both traditional machine learning techniques and Convolutional Neural Networks (CNNs). The remarkable success of CNNs in diverse areas, including object detection, segmentation, and medical image analysis, is explored. The survey delves into the internal taxonomy of deep learning, various CNN models, and innovations in the

internal architecture and representational styles of CNNs. paper also traces the historical development of Computer-Aided Diagnosis (CAD) and its increasing significance in medical image interpretation. The challenges, applications, and current developments in deep learning for medical image analysis are discussed, providing a valuable overview of the field.

Brain tumors pose a critical health risk, with uncontrolled cell growth potentially leading to fatal consequences if not treated promptly. Despite notable advancements, the accurate segmentation and classification of brain tumors remain challenging due to variations in location, shape, and size. This survey comprehensively explores brain tumor detection using magnetic resonance imaging (MRI) and encompasses anatomy, publicly available datasets, enhancement techniques, segmentation, feature extraction, classification, and the application of deep learning, transfer learning, and quantum machine learning. By presenting literature on brain tumor detection as given in , the survey outlines advantages, limitations, developments, and future trends in this crucial field, providing valuable insights for researchers and practitioners.

Magnetic Resonance Imaging (MRI) serves as a valuable tool for non-invasive brain disease detection, offering three-dimensional images for precise anomaly identification. However, the manual analysis of these images from is time-consuming and labor-intensive. This paper presents a comprehensive review (covering 1998 to 2020) on brain tumor segmentation from MRI images, exploring core segmentation algorithms and methodologies. Leveraging advancements in machine learning and computational efficiency, the study delves into various approaches for brain abnormality identification, offering a thorough overview of machine learning and image segmentation techniques applied to brain tumor detection. The review highlights the effectiveness of deep learning methods in MRI-based brain tumor segmentation, providing a valuable resource for researchers and practitioners.

Brain image classification is a crucial process in the medical domain, presenting significant challenges. paper introduces a novel method, Mutual Information-Accelerated Singular Value Decomposition (MI-ASVD), for feature subset selection to enhance the efficiency of brain image classification. Experimental results demonstrate the superiority of MI-ASVD over the original feature space and two standard dimensionality reduction methods, achieving a classification accuracy of 94.91 percent. This approach outperforms existing techniques and contributes to the field of brain image classification.

MRI imaging is instrumental in brain tumor analysis, diagnosis, and treatment planning, aiding doctors in understanding the progression of brain tumors. Detecting brain tumors through MRI images is challenging due to the brain's complex structure. paper explores various preprocessing, post-processing, and image processing techniques, including filtering, contrast enhancement, edge detection, histogram analysis, thresholding, segmentation, and morphological operations. The study reviews existing literature on different filters and segmentation techniques, aiming to enhance understanding and provide insights into the diverse methodologies employed for

brain tumor detection in MR images. The comprehensive review given in covers preprocessing steps, feature extraction, segmentation, and post-processing methods, contributing to the ongoing research in this critical medical field.

The paper addresses the crucial task of brain tumor detection and classification in MRI brain images, emphasizing the distinction between normal and tumor-affected brain regions. The study focuses on various types of brain tumors, including metastatic bronchogenic carcinoma tumors, glioblastoma, and sarcoma, using magnetic resonance imaging (MRI). The proposed methodology employed in different wavelet transforms and Support Vector Machines (SVMs) for accurate detection and classification. The process involves preprocessing steps, such as denoising using diverse wavelets, feature extraction, and feature reduction. SVMs, particularly linear and kernel SVMs, are trained for classification. The effectiveness of the approach is critical for precise medical analysis and interpretation, aiding in the automated classification of MRI brain images.

## METHODOLOGY

The brain tumor classification problem is contextualized within this study, emphasizing the importance of accurate diagnosis. The dataset, comprising 'notumor' and 'positive-tumor' classes, is introduced, outlining its significance. The application of Convolutional Neural Networks (CNNs) and k-Nearest Neighbors (kNN) for classification is discussed.

### A. Data Collection and Preparation

The dataset utilized in this study consists of brain tumor images sourced from kaggle. The dataset encompasses two classes: 'notumor' and 'positivetumor.' To ensure a balanced representation, an initial analysis of the folder structure is conducted, organizing images into respective classes.

The dataset is organized into folders corresponding to different classes, as illustrated in the "Folder Structure Analysis" chart [Figure 1]. This visual representation allows for a quick understanding of the distribution of images among classes.

Images are loaded using OpenCV, resized to a standardized dimension (e.g., 100x100 pixels), and flattened into feature vectors. These image vectors are then stored in a Pandas DataFrame for subsequent analysis.

### B. Exploratory Data Analysis (EDA)

Intraclass spread, represented by standard deviation, and interclass distances between mean vectors are calculated to gain insights into the distribution of features within each class. Figures 1 depict the results of this analysis [Figure 1: "Feature Histogram"].

Additionally, a histogram of a representative feature (e.g., mean pixel intensity) is plotted, providing a visual representation of the feature distribution.

Minkowski distances are computed for different values of  $r$  (from 1 to 10), offering an understanding of the impact of distance metrics on the dataset [Figure 2: "Minkowski Distance vs.  $r$ "].

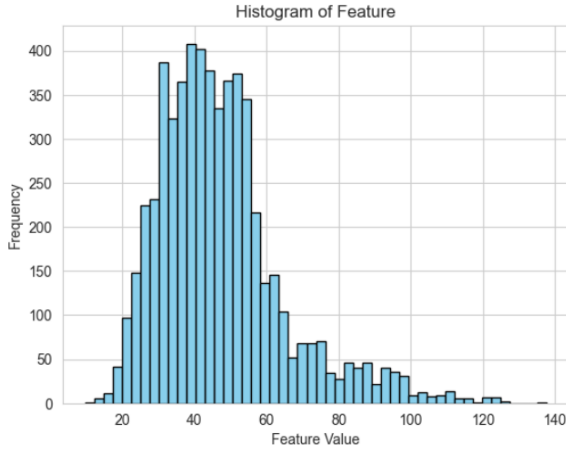


Fig. 1. Histogram

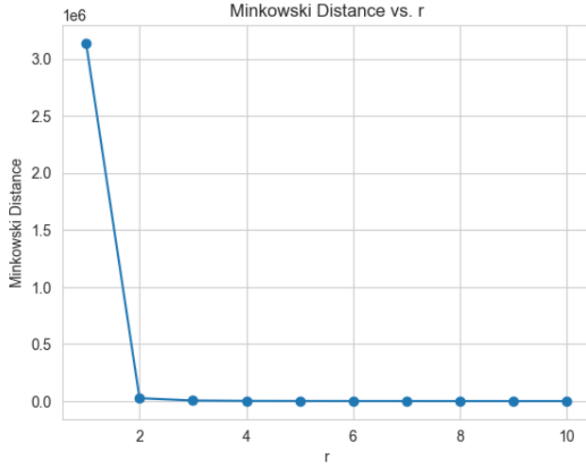


Fig. 2. Minkowski distances Graph

### C. Convolutional Neural Network (CNN) Model

The dataset is split into training and testing sets using a 70-30 ratio. Labels are converted to numerical format for binary classification. Pixel values are normalized to the range  $[0, 1]$ , and images are reshaped to be compatible with the CNN architecture.

The CNN model is designed with convolutional and pooling layers, followed by dense layers. Regularization techniques such as batch normalization and dropout are incorporated. The model is compiled using the Adam optimizer and binary crossentropy loss and is trained for 10 epochs with a 20

The trained CNN model is evaluated on the test set, providing insights into its accuracy and generalization performance.

### D. Comparative Analysis with kNN

A kNN classifier is employed on the dataset with varying  $k$  values. The results are analyzed, and a comparative study is conducted with the CNN model's performance.

#### 1) Convolutional Neural Network (CNN) Implementation:

To address the brain tumor classification task, we employ a

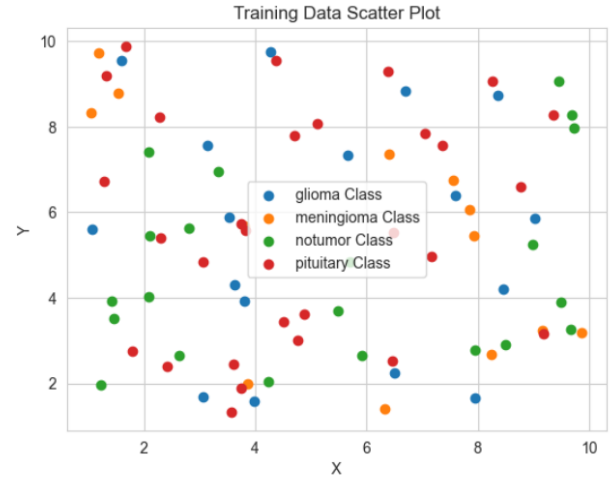


Fig. 3. Training Data plot

CNN model. The dataset is loaded from Kaggle, comprising 'notumor' and 'positivetumor' classes. The images are resized to a standardized dimension of 100x100 pixels and normalized. The CNN architecture includes convolutional and pooling layers, followed by dense layers. Regularization techniques such as batch normalization and dropout are incorporated for improved generalization.

The model is compiled using the Adam optimizer and binary crossentropy loss. It is trained for 20 epochs with a 70-30 train-test split. The evaluation metrics, including accuracy, precision, recall, and F1-score, are computed on the test set.

2) *k-Nearest Neighbors (kNN) Implementation:* For the kNN classifier, we utilize the scikit-learn library. A hyperparameter tuning process is conducted using RandomizedSearchCV to find the optimal 'k' value. The classifier is trained on the training data, and predictions are made on the test set for various 'k' values (e.g., 3, 5, 7). Additionally, the hyperparameter tuning process using RandomizedSearchCV helps identify the best model.

3) *Visualizations:* Visualizations of the training data scatter plot and test data scatter plots for different 'k' values are provided to offer an intuitive understanding of the classification results. fig 3 shows the training data scatterplot representation and fig 4 represents the best test data scatterplot representation.

## III. RESULTS AND DISCUSSION

The effectiveness of class separation is analyzed by considering both intraclass spread and interclass distances. The behavior of the kNN classifier is explored, emphasizing the impact of changing 'k' values and scenarios of over-fitting and under-fitting. Model performance metrics, including accuracy, precision, recall, and F1 score, are evaluated and presented in a comparative analysis against the CNN classifier. The learning curves of the CNN model are scrutinized to identify potential over-fitting or under-fitting situations.

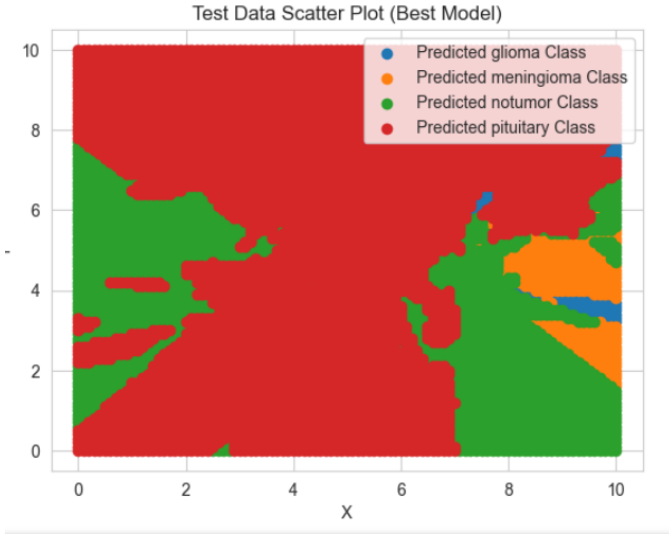


Fig. 4. Best Testing Data plot

	Precision	Recall	F1-Score	Support
Class 0	0.95	0.90	0.92	425
class 1		0.86	0.86	405
Class 2	0.95	0.96	0.96	451
Class 3	0.95	0.98	0.97	433
Accuracy	0.93			1714
Macro AVG	0.93	0.93	0.93	1714
Weighted AVG	0.93	0.93	0.93	1714

TABLE I  
CLASSIFICATION REPORT

#### A. Classification Report

Table 1 shows the classification report provides a detailed overview of the model's performance on the test set

#### B. Statistical Summary

Mean: 47.19 Variance: 311.43

#### C. Intraclass Spread (Standard Deviation)

- Glioma: [0.1027, 0.1027, 0.1027, ..., 0.0727, 0.0727, 0.0727]
- Meningioma: [9.7921, 9.7921, 9.7921, ..., 10.4363, 10.4363, 10.4363]
- Notumor: [33.9769, 33.6641, 33.5956, ..., 37.6092, 37.3328, 37.2738]
- Pituitary: [5.0374, 5.0374, 5.0374, ..., 3.9497, 3.9497, 3.9497]

#### D. Interclass Distances between Mean Vectors

- Glioma - Meningioma: 2140.87
- Glioma - Notumor: 5444.01
- Glioma - Pituitary: 3872.89
- Meningioma - Notumor: 3553.28
- Meningioma - Pituitary: 2893.87
- Notumor - Pituitary: 3949.01

## IV. CONCLUSION

The comprehensive evaluation and analysis of both the CNN and kNN models provide valuable insights into their respective performances. The CNN model exhibits remarkable accuracy and precision, particularly in classifying different tumor types. The kNN classifier, while providing competitive results, presents variations based on the choice of 'k' and demonstrates sensitivity to over-fitting and under-fitting scenarios.

Further exploration and refinement of these models could involve hyperparameter tuning, additional feature engineering, or the exploration of alternative architectures. The choice between these models should be guided by specific considerations such as computational efficiency, interpretability, and the specific characteristics of the dataset.

The detailed results and analyses presented in this section serve as a foundation for informed decision-making in the context of brain tumor classification.

## REFERENCES

- [1] Vadhvani, S., Singh, N. Brain tumor segmentation and classification in MRI using SVM and its variants: a survey. *Multimed Tools Appl* 81, 31631–31656 (2022). <https://doi.org/10.1007/s11042-022-12240-4>
- [2] Ramdas Vankodthru <https://doi.org/10.1016/j.measen.2022.100440>.
- [3] Gajula srinivasrao DO - 10.1007/s13198-022-01680-8.
- [4] G. T. Reddy et al., "Analysis of Dimensionality Reduction Techniques on Big Data," in *IEEE Access*, vol. 8, pp. 54776-54788, 2020, doi: 10.1109/ACCESS.2020.2980942
- [5] saeed eqbal DOI: 10.1109/ACCESS.2023.3301873..
- [6] Amin, J., Sharif, M., Haldorai, A. et al. Brain tumor detection and classification using machine learning: a comprehensive survey. *Complex Intell. Syst.* 8, 3161–3183 (2022). <https://doi.org/10.1007/s40747-021-00563-y>.
- [7] T. A. Soomro et al., "Image Segmentation for MR Brain Tumor Detection Using Machine Learning: A Review," in *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 70-90, 2023, doi: 10.1109/RBME.2022.3185292.
- [8] Z. A. Al-Saffar and T. Yildirim, "A Novel Approach to Improving Brain Image Classification Using Mutual Information-Accelerated Singular Value Decomposition," in *IEEE Access*, vol. 8, pp. 52575-52587, 2020, doi: 10.1109/ACCESS.2020.2980728.
- [9] Borole, Vipin Y., Sunil S. Nimbhore, and Dr Seema S. Kawthekar. "Image processing techniques for brain tumor detection: A review." *International Journal of Emerging Trends Technology in Computer Science (IJETTCS)* 4.5 (2015): 2.
- [10] M. Gurbină, M. Lascu and D. Lascu, "Tumor Detection and Classification of MRI Brain Image using Different Wavelet Transforms and Support Vector Machines," 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 2019, pp. 505-508, doi: 10.1109/TSP.2019.8769040.