

Digital Analysis of Brain Tumor using Image Processing and Machine learning

Rajath D¹, Rajendra Prasad H S², Bhat Geetalaxmi Jairam³, Suhas N I⁴ and U Karthik⁵

¹⁻⁵Dept. of Information Science and Engineering, National Institute of Engineering, Mysuru, India

Email: {2020is_suhasni_b, bgj, 2020is_rajendraparasadhs_b, 2020is_rajathd_b, 2020is_ukarthik_b }@nie.ac.in

Abstract—The computerized examination of brain tumors is critical in the early and correct diagnosis of neurological illnesses. This study proposes a new approach that combines image processing techniques and machine learning algorithms for the automatic detection and categorization of brain tumors in medical pictures. The proposed system uses advanced image processing algorithms to enhance and preprocess magnetic resonance imaging (MRI) scans of the brain. Preprocessing processes include noise reduction, contrast enhancement, and segmentation to extract the regions of interest. Following that, a feature extraction procedure is used to extract significant features from the preprocessed pictures, collecting key properties indicative of tumor existence and type. During the machine learning phase, a robust classification model is trained on a broad collection of annotated brain images.

Index Terms— Image Processing, MRI, TensorFlow, Machine Learning.

I. INTRODUCTION

The digital technology has transformed the field of medical imaging, offering up new avenues for accurate and quick diagnosis. When it comes to brain tumors, the combination of image processing and machine learning has proven to be a promising method. Traditional techniques of manually evaluating radiological pictures can be time-consuming and have limitations. The technological evolution has led digital diagnosis of cerebral tumor. Brain tumors are a difficult health concern, and integrating digital analytics is critical to improving diagnostic capabilities. High-resolution imaging techniques, such as MRI and CT, enable the processing of massive amounts of data to extract relevant information. Image processing techniques like as segmentation and feature extraction are critical in isolating and highlighting regions of interest in the brain. these components, incorporating the applicable criteria that follows in detection.

To uncover complicated patterns in processed photographs, machine learning techniques, particularly deep learning models such as CNNs and RNNs, are applied. Using image processing skills, these strategies contribute to automated tumor classification. Using extra data to adapt machine learning models makes diagnostic solutions more resilient and scalable[7].This research seeks to provide a comprehensive framework for digital analysis, with specific goals including precise tumor segmentation, feature extraction, and the construction of machine learning models for automatic categorization. The goal is to The significance of this research stems from its ability to speed up the diagnostic process, contribute to personalized treatment regimens, and ultimately enhance patient outcomes. The successful adoption of this technique has the potential to progress neuro-oncology while also reducing the strain on the healthcare system.Accurate segmentation of tumor regions within medical images is a fundamental goal. This study attempts to use advanced image processing techniques to precisely identify and

isolate regions of interest and provide a framework for subsequent detailed investigation.

After segmentation, it is important to extract relevant features from the detected tumor regions. These features include characteristics such as shape, texture, and intensity, which can help distinguish between different types of tumors. The rigorous extraction method aims to improve the discriminatory ability of the model and thereby improve the accuracy of tumor classification. Developing machine learning algorithms to automatically classify tumors is of great importance for research. These models use techniques such as convolutional neural networks (CNN) and recurrent neural networks (RNN). These models develop intricate patterns from the collected information, allowing them to differentiate between different tumor kinds. Training entails exposing the models to a variety of datasets in order to ensure their adaptability and efficiency over a wide range of patient demographics and imaging settings. The proposed technique must be validated in order to establish the dependability and applicability of the digital analytic framework. A thorough evaluation of the models' performance on a variety of datasets, benchmarked against known neuro-oncological criteria, ensures not just accuracy but also robustness in dealing with fluctuations in imaging data and disease characteristics. Ultimately, this research promises to improve brain tumor diagnostics through the seamless integration of image processing and machine learning. This study seeks to bring valuable insights to the medical community by meeting the defined objectives and validating the proposed technique, consequently improving the efficiency, accuracy, and individuality of brain tumor diagnosis.

II. LITERATURE SURVEY

I. ramya [1] suggest robust segmentation algorithm using morphological operators for detection of tumor Image denoising and Image Segmentation are the main steps to be employed to detect the tumor. Noise removal is done by using fourth order partial differential equation. A preprocessing step skull removal is considered by using two main morphological operators, erosion and dilation. The initial step in the proposed algorithm involves image denoising, which is essential for enhancing the quality of medical images. Ramya [1] utilizes a fourth-order partial differential equation for noise removal, emphasizing the importance of a clean and precise image as a foundation for subsequent processing steps. Following image denoising, the algorithm incorporates a preprocessing step focused on skull removal. This step is vital in neuroimaging, where the presence of the skull can obstruct accurate tumor detection. To achieve skull removal, two main morphological operators, erosion and dilation, are employed. Morphological operations play a key role in shaping and enhancing features within images, making them valuable tools for preprocessing tasks such as skull removal. The use of erosion and dilation as morphological operators is particularly noteworthy. Erosion involves shrinking the boundaries of regions, while dilation involves expanding them. By strategically applying these operators, the algorithm effectively addresses the challenges associated with skull presence, ultimately improving the accuracy of tumor segmentation.

Xinyang Liu [2] proposed a unique method to automate the processing of patient MRI data and improve tumor progression monitoring and overall survival prediction. This study recognized the need for automated segmentation for clinical implementation and addressed the issue of data scarcity in pediatric brain tumors. To overcome this limitation, researchers investigated the effectiveness of pre-training state-of-the-art deep learning models using adult brain tumor data. The objective was to leverage knowledge transfer from adult datasets to improve segmentation performance in pediatric data. The study's rationale is based on the assumption that pre-training deep learning models on big adult brain tumor datasets can provide useful knowledge and features that can be applied to data from teenage brain tumors. Due to the rarity of such instances, this knowledge transfer strategy is especially effective in situations when obtaining a sufficiently big and diverse collection of pediatric brain tumor pictures to directly train a model is difficult. The suggested automated pipeline has the potential to revolutionize clinical practice by allowing for more efficient monitoring of tumor growth and prediction of overall survival.

T. S. Deepthi Murthy's [3] work focuses on brain tumor segmentation utilizing a mix of thresholding, morphological procedures, and tumor feature extraction. The brain, being the major control and coordinating organ of the human body, is prone to tumor growth due to unregulated cell proliferation. The paper recognizes the importance of effective brain tumor detection and segmentation and provides an algorithm for this goal that uses thresholding and morphological operations. Murthy's [3] technique is intended to efficiently segment brain tumors in scanned MRI images. Thresholding is a basic image processing technique used to differentiate between tumor and non-tumor regions based on intensity levels. Morphological processes, which involve the alteration of image structures through operations such as dilation and erosion, improve the segmentation process even further.

The program retrieves crucial features from the segmented tumor after the segmentation process. These characteristics include the tumor's centroid, perimeter, and area. These quantitative parameters provide essential information about the tumor's size, shape, and location, leading to a thorough characterization. In summary, T. S. Deepthi Murthy [3] presents an efficient algorithm for brain tumor segmentation using thresholding and morphological operations, with a subsequent focus on extracting important tumor features. The application of this algorithm in the medical field holds promise for enhancing the accuracy of tumor detection and contributing to the formulation of effective treatment strategies for patients.

Matthias W. Wagner's [4] gave a idea to predict progression-free survival (PFS) in paediatric patients with diffuse midline glioma, specifically diffuse intrinsic pontine glioma (DIPG), using radiomic characteristics collected from MRI data. The abstract gives a brief summary of the study's aims, methods, and significant findings. The retrospective investigation examined diagnostic brain MRIs from a group of children with DIPG. The dataset was divided into training and testing sets after radiomic characteristics were collected from tumor segmentations. Using the training data, a conditional survival forest model was used to predict PFS. The trained model was then validated using test data. The study's findings indicate that radiogenic characteristics, particularly those produced by FLAIR (fluid-attenuated inversion recovery) and unenhanced T1-weighted sequences, can predict PFS. This overview displays the agreement values of multiple radiomics models at critical time periods and emphasizes the models' capacity to predict survival outcomes. One of the most significant discoveries was the identification of grayscale size zones as the primary predictive radiation feature matrix. According to this summary, MRI-based radiomics can be utilized as a prognostic marker and can stratify DIPG subgroups based on predicted progression-free survival. Finally, Matthias W. Wagner's study [4] emphasizes radiomics' promising significance in predicting progression-free survival in paediatric diffuse midline gliomas, particularly DIPG. The use of MRI-based radiomic characteristics is a non-invasive approach that may provide significant prognostic information, especially when biopsy-based assessments are limited or problematic, as is frequently the case in pediatric patients.

III. PRELIMINARY WORK

The advancement of medical imaging technology, particularly magnetic resonance imaging (MRI) and computed tomography (CT), has been instrumental in revolutionizing diagnostic medicine. These advances have allowed for thorough, non-invasive viewing of the human body, providing clinicians with critical insights into a variety of medical disorders, including brain tumors. The move from standard two-dimensional imaging to high-resolution three-dimensional imaging enhanced anatomical visualization precision dramatically, paving the way for more advanced analysis. Because of their variety, various morphologies, and intricate placements within the brain, brain tumors represent a significant challenge. Traditional diagnostic approaches frequently entail manual interpretation of radiological images, which can be time-consuming and subjective. The limits of these standard methodologies highlight the need for more accurate and quick diagnostic tools in neuro-oncology, where early diagnosis is critical.

The importance of automation in brain tumor analysis resides in its capacity to speed up the diagnostic process while improving objectivity and reproducibility. Machine learning models, with their versatility and learning skills from varied datasets, contribute to the development of robust and effective diagnostic tools. This paradigm change corresponds with the general trend in healthcare toward personalized and data-driven treatments, emphasizing the relevance of patient-specific information for optimal treatment planning. While individual components such as image processing and machine learning show promise, there is an urgent need for comprehensive frameworks that smoothly combine these technologies. A comprehensive method that combines precise image processing for tumor localization with sophisticated machine learning models for classification is required to achieve diagnostic accuracy and reliability.

IV. ARCHITECTURE DESIGN

Dataflow diagram for MRI image analysis, with an emphasis on K-means clustering and preprocessing methods for tumor identification. The suggested approach combines many image processing phases to improve the precision and effectiveness of tumor identification in magnetic resonance imaging. The dataflow diagram shows how preprocessing, K-means clustering, and morphological filtering work together to make tumor diagnosis easier. MRI image tumor identification has been accomplished by a variety of ways in previous research, such as image enhancement techniques, segmentation strategies, and machine learning algorithms. Improving image quality and identifying pertinent features for further analysis depend heavily on preprocessing. A well-liked

unsupervised learning technique for image segmentation is K-means clustering, which divides the image into clusters according to pixel intensities. To smooth out and remove noise from the segmented sections, morphological filtering is used.

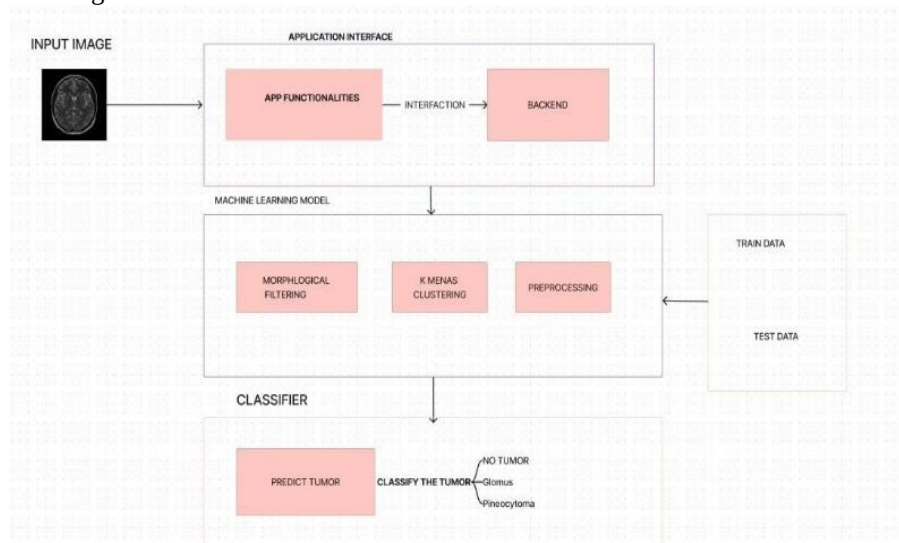


Figure 1. Architecture Design of proposed system

The following are the main elements of the suggested DFD for MRI image analysis:

A. MRI Image

We classify the datasets according to a number of important factors, such as: Imaging Modality: Collections of images from magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET) among other imaging modalities. Tumor forms: To assure representation of a range of clinical situations, datasets spanning several brain tumor forms, such as gliomas, meningiomas, and metastatic tumors, are included. Image Resolution and Quality: Details on the format, resolution, and quality of the images that are part of the datasets. Annotation Protocol: Information concerning the annotation procedure, such as the existence of ground truth labels defining the areas of tumors and the metadata that goes along with them. Data Size and Availability: The accessibility of the dataset, conditions of license, and download links to enable researchers to use it for testing and verification.

B. Improved Image Quality through pre-processing

Preprocessing can improve the clarity of medical images and help to eliminate artifacts and noise that impede precise tumor detection [6]. We have used gaussian blur, thresholding erosion, Dialation. counter extraction, Extreme Points Detection, Cropping. This improves machine learning algorithms by producing standardized input data for additional analysis. Feature Deletion is used to eliminate the skull region of the brain to improve detection by considering only brain cells.

The algorithm begins by converting the image to grayscale and applying Gaussian blur (5x5 kernel, sigma 0) to reduce noise. Thresholding sets pixel values above 45 to 255 (white) and below to 0 (black). Erosion (2 iterations) and dilation (2 iterations) refine the binary image, removing noise and connecting regions. Contours are extracted using `cv2.findContours()` with `RETR_EXTERNAL` mode, while `imutils.grab_contours()` ensures compatibility. The algorithm identifies the contour with maximum area, likely representing the brain. Extreme points (leftmost, rightmost, topmost, bottommost) are found using numpy. Finally, the original image is cropped using these points to isolate the brain region.

C. Clustering

The research uses machine learning arithmetics and k-means clustering techniques on MRI images to present a novel method for brain tumor identification [5]. The focus is on using cluster analysis—which includes feature extraction, data preparation, selecting the optimal cluster algorithms, setting optimization parameters, and result evaluation—to identify specific tumor locations on brain MRI scans. An actual dataset is used as an example to show how very precise results may be obtained, enabling separate tumor analysis and even suggesting the idea of

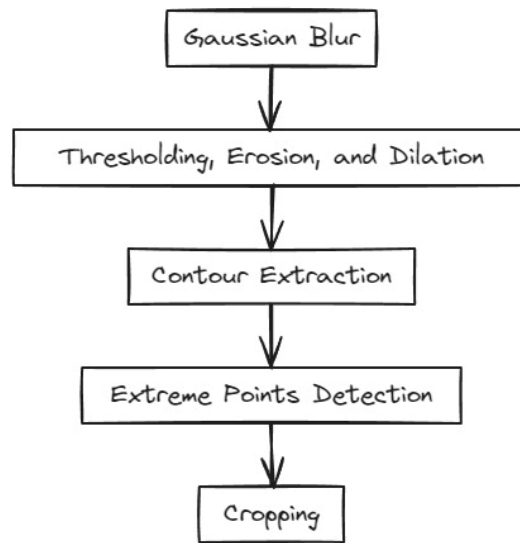


Fig 2: Image Pre-processing Dataflow

characterizing the tumors based on their geographical distribution or layout. The work being done to create better automated techniques for detecting brain tumors is probably going to help in diagnosis and treatment planning.



Figure 3. Data Flow of K means clustering algorithm

D. Filtering

To improve image quality and analysis accuracy, filtering techniques are applied to machine learning algorithms for brain tumor identification in MRI scans. In order to remove noise, get rid of artifacts, and improve features, it investigates linear, non-linear, and deep learning filters. Adaptive median and bilateral filters preserve edges and delicate details, while linear filters such as Gaussian and median filters mainly reduce noise. Large-scale datasets are used by deep learning-based filters to perform thorough transformations that greatly improve noise reduction and feature improvement.

E. ALGORITHM: Morphological Filtering

Step 1: Convert to grayscale and apply Gaussian blur

Convert the input image to grayscale

Apply Gaussian blur to the grayscale image

Step 2: Thresholding

Threshold the grayscale image to obtain a binary image.
`cv2.threshold(grey, 45, 255, cv2.THRESH_BINARY)`

Step 3: Erosion and Dilation

Perform morphological erosion with 2 iterations
Perform morphological dilation with 2 iterations

Step 4: Find Contours

Find contours in the thresholded image

Step 5: Find Extreme Points

Find extreme points (leftmost, rightmost, topmost, bottommost) of the detected contour

Step 6: Crop Brain Contour

Crop the original image using extreme points obtained

F. Diagnosis Result

The diagnostic implications of machine learning algorithms adhering to cancer tumor identification, which enhances automated medical imaging diagnostic systems. This study uses MRI scan data that has been classified as having a tumor. Machine learning algorithms are utilized for data preprocessing, feature extraction, model training, and evaluation. Results reveal that the method is accepted and that it produces a pertinent, significant diagnosis that demonstrates a high degree of precision in describing brain tumors. Machine learning therefore has the potential to increase brain tumor diagnosis efficiency and accuracy, which will provide abundant, useful data for medical practice.

G. Tumor Location

Using machine learning techniques, targeted MRI scans are conducted with the goal of precisely determining the tumor's location within the brain for the purpose of surgical planning and guidance. We will go over these strategies in detail, including the time-tested classical image processing techniques, the diligent traditional machine learning algorithms, and the soft competition to become an advanced deep learning architecture that locates tumors. Conventional methods include intensity normalization prior to further processing and methods for removing skulls. Support vector machines and random forests are two examples of machine learning models that rely on manually created features by researchers. In contrast, techniques like fully convolutional networks and convolutional neural networks learn hierarchical representations directly from raw pixels data, reducing localization problem to be mainly simple. In contrast, techniques like convolutional neural networks and fully convolutional networks reduce the localization problem to a primarily simple one by simply learning hierarchical representations from raw pixel data. Clinical results will ultimately improve as a result of this endeavour, which will also increase the precision of brain tumor localization.

V. EXPERIMENTAL RESULT

The tumor detection experiment makes use of various machine learning algorithms and morphological operations in a Jupyter notebook. The procedure includes several stages of image processing as shown in figure(a-h), Each step demonstrates how tumor extraction is done.

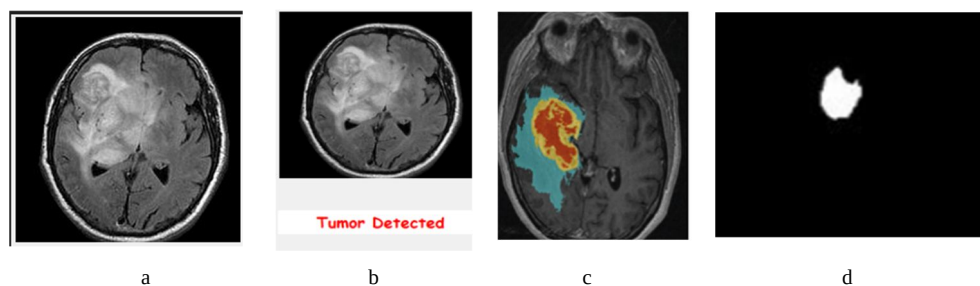


Figure 4. a: Initial MRI scan, b: final detection of tumor, c: after image enhancement, d: after morphological filter

The proposed system provides the best solution for detecting brain tumors. Finally, the procedure to calculate accuracy of model mentioned below

loss, acc = best_model.evaluate(x=X_test, y=y_test)

The evaluate method provided by Keras models is used to evaluate the model's performance on the test dataset. The method calculates the loss and accuracy of the model on the provided test data (**X_test** and **y_test**). The accuracy represents the percentage of correct predictions made by the model. The accuracy value **acc** is obtained from the evaluation, and it's calculated using the formula mentioned above.

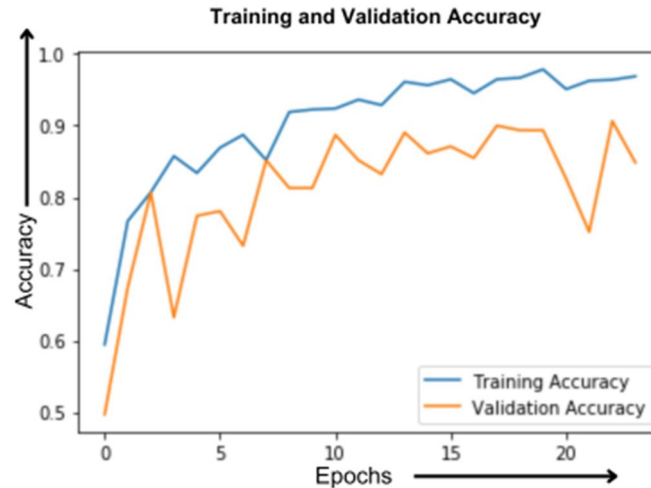


Figure 5. Training and Validation Accuracy

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

our findings demonstrate the development of a predictive model capable of detecting brain tumors with an 80% accuracy rate. We demonstrated the potential of machine learning techniques for brain tumor detection by analyzing medical imaging data using various process and a meticulous methodology. Despite encouraging results, it is critical to acknowledge our model's limitations, which include a scarcity of high-quality labeled data, difficulties in integrating diverse imaging modalities, and the need for thorough clinical validation in practical healthcare scenarios. Future research may investigate the use of additional imaging modalities or clinical data to improve predictive accuracy and effectively address these constraints.

VI. CONCLUSIONS

The goal is to extract tumor information accurately and with minimal error. Pre-processing, tumor isolation (segmentation), feature extraction, and tumor classification are all steps in the detection of brain tumors using Magnetic Resonance Imaging (MRI). This work focuses on segmentation techniques and improves performance. Compared to edge-based methods, region growing with PSO yields higher accuracy. Accurate tumor diagnosis is critical to patient outcomes. This method prioritizes accuracy, which is critical when determining the best treatment course. It not only improves precision, but also the efficacy of brain tumor detection. In future work of this study we will improve system to detect tumor size and early detection of tumors. The code sets the groundwork for brain tumor detection using a CNN but requires enhancements. The model architecture needs depth and regularization for improved pattern capture and prevention of overfitting, possibly leveraging pre-trained models like VGG or ResNet. Augmentation should expand to diverse transformations for richer training data, while evaluation metrics such as F1-score offer better performance insights, especially for imbalanced datasets. Hyperparameter tuning and robust monitoring via TensorBoard optimize model performance. Error analysis aids in understanding misclassification patterns, and deployment considerations ensure computational efficiency. Comprehensive documentation enhances code readability and maintainability, while data exploration uncovers class distributions and quality issues. Techniques to address class imbalance ensure balanced representation. By integrating these enhancements, the model becomes a powerful tool for real-world brain tumor detection.

REFERENCES

- [1] Ramya and n. Sasirekha, "a robust segmentation algorithm using morphological operators for detection of tumor in mri," 2015 international conference on innovations in information, embedded and communication systems (iciiecs), coimbatore, india, 2015, pp. 1-4, doi: 10.1109/iciiecs.2015.7192927
- [2] Liu, Xinyang & Bonner, Erin & Jiang, Zhifan & Roth, Holger & Anwar, Syed & Packer, Roger & Bornhorst, Miriam & Linguraru, Marius George. (2023). Automatic Segmentation of Rare Pediatric Brain Tumors Using Knowledge Transfer From Adult Data. 1-4. 10.1109/ISBI53787.2023.10230757
- [3] T. S. Deepthi murthy and g. Sadashivappa, "brain tumor segmentation using thresholding, morphological operations and extraction of features of tumor," 2014 international conference on advances in electronics computers and communications, bangalore, india, 2014, pp. 1-6, doi: 10.1109/icaecc.2014.7002427
- [4] Wagner MW, Namdar K, Napoleone M, Hainc N, Amirabadi A, Fonseca A, Laughlin S, Shroff MM, Bouffet E, Hawkins C, Khalvati F, Bartels U, Ertl-Wagner BB. Radiomic Features Based on MRI Predict Progression-Free Survival in Pediatric Diffuse Midline Glioma/Diffuse Intrinsic Pontine Glioma. *Can Assoc Radiol J.* 2023 Feb;74(1):119-126. doi: 10.1177/08465371221109921. Epub 2022 Jun 29. PMID: 35768942.
- [5] J. Vijay and J. Subhashini, "An efficient brain tumor detection methodology using K-means clustering algorithm," 2013 International Conference on Communication and Signal Processing, Melmaruvathur, India, 2013, pp. 653-657, doi: 10.1109/iccsp.2013.6577136.
- [6] C. S. Pawar, A. Patel, J. Bharvad and S. Ribadiya, "A Research Exploration of Features and Techniques for Brain Tumor Detection and Segmentation," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1650-1655, doi: 10.1109/ICECA49313.2020.9297619.
- [7] K. Neamah et al., "Brain Tumor Classification and Detection Based DL Models: A Systematic Review," in *IEEE Access*, vol. 12, pp. 2517-2542, 2024, Doi: 10.1109/ACCESS.2023.3347545.