Grenze International Journal of Engineering and Technology, June Issue

Brain Tumor Classification using Machine Learning Approach

Riddhi Doshi
St. Vincent Pallotti College of Engineering & Technology/Department of Computer Engineering, Nagpur, India
Email: riddhidoshi1996@gmail.com

***Abstract*—Brain tumors are a major worldwide health problem, with early and correct diagnosis critical for successful treatment and better patient outcomes. Advances in medical imaging and machine learning have created new opportunities for automated diagnostic tools to help doctors identify and categorize brain cancers with greater accuracy. This study investigates the use of machine learning techniques, notably Random Forest classifiers and 2D Convolutional Neural Networks, to classify brain tumors using MRI data. The models were trained and evaluated using a publicly available Kaggle dataset, which included a wide range of brain MRI images labeled with tumor classifications. While the Random Forest model was simple and computationally efficient, the 2D CNN surpassed it in terms of accuracy and precision, demonstrating the power of deep learning for dealing with complicated imaging data. These findings emphasize the importance of using machine learning in medical diagnostics, demonstrating how sophisticated algorithms may increase diagnostic accuracy while also providing significant insights to guide treatment decision-making. Integrating such technologies may improve early detection, streamline procedures, and eventually lead to better patient care in cancer.**

***Index Terms*—Machine Learning, Convolutional Neural Network, Random Forest, Healthcare, Brain Tumor and Biomedical.**

I. INTRODUCTION

Cerebral growths, which are differentiated by the rapid and uncontrolled development of abnormal cells, pose a major health threat. Early identification and treatment are critical, since untreated tumors can spread fast, resulting in serious health implications and a worse quality of life [1]. Despite advances in brain tumor identification and therapy, proper tumor segmentation and classification remain difficult due to the complex structure and varying appearance of tumors in medical imaging [1].

Machine learning has emerged as an effective method for addressing these difficulties. Its capacity to scan large datasets, discover detailed patterns, and make accurate predictions has transformed medical imaging. In particular, machine learning-based brain tumor categorization has showed considerable promise in terms of enhancing diagnosis accuracy and expediting treatment planning. Researchers have developed a variety of ways for classifying brain tumors from MRI images with high accuracy and speed [2]. These techniques include distinct machine learning algorithms, feature extraction methods, and pre-processing processes. These advancements illustrate the essential role of machine learning in transcending conventional restrictions in healthcare imaging and enabling more reliable and automated diagnostic solutions.

Grenze ID: 01.GIJET.11.2.101_20

II. Types of Brain Imaging Modalities

Three basic technologies as mentioned in Figure 1 are commonly utilized to study brain tumor structure.

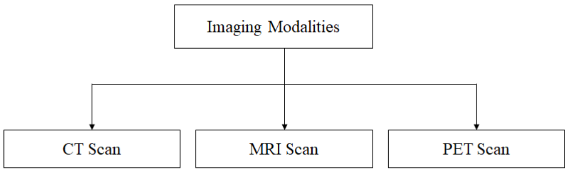

Figure 1. 2D Types of Brain Imaging Modalities

- CT Scan: Computational tomography (CT) scans are more detailed than ordinary X-ray pictures. Since its inception, the computed tomography, or CT, scan has gained widespread acceptance and usage. However, compared to routine X-rays, it exposes patients to higher radiation levels. Because of this increased radiation exposure, frequent CT scans may raise the risk of developing cancer. [1]
- MRI Scan: An MRI scanning is utilized to thoroughly determine numerous biological components alongside can discover brain abnormalities sooner than other imaging treatments. As a result, the brain's complex architecture makes tumor segmentation difficult. [1][3]
- PET Scans: The technique of positron emission tomography, or PET, employs a specific type of radioactive tracer PET is used to study the metabolic features of cerebral growths, including arterial circulation, metabolizing glucose, the generation of lipids, intake of oxygen, and metabolism of amino acids. It remains one of the most effective metabolic methods known [1]

When compared to other imaging technologies like CT scans, PET scans, and X-rays, MRI is extremely effective in detecting cancers and assessing surrounding tissues, causes no discomfort, and has no known tissue damage [4][5]. It generates detailed images of the brain and surrounding tissues using a high magnetic field, radio signals, and an electronic device rather than x-rays or ionizing radiation. [4][6][16]

III. Literature Review

In [7], MRI imaging is utilized to diagnose brain tumours. The median filter is used for picture pre-processing. The thresholding approach divides tumor regions into intensity levels. K-means clustering organizes comparable pixels for segmentation. GLCM is utilized for feature extraction. The SVM classifier detects tumours that are both malignant and benign with an accuracy of 80%.

In [8], another approach, Cerebral growth categorization was proposed utilizing utilization of deep features combined with machine learning models. To extract features from MRI scans, the recommended technique for this reference makes use of transfer learning and several deep convolutional neural networks. The MRI scan data was then analyzed by a number of machine learning classifiers to identify between persons with malignancies and those who were healthy.

In [9], the SVM technique was used for regression and classification analysis, yielding high accuracy. However, one drawback is its lengthy execution time in tumor classification. Subsequently, one disadvantage is its long computation time for cancer categorization. A review of current computerized scanning studies for brain tumors identification was undertaken. The first study on categorizing brain cancers using magnetic resonance imaging employed SVM approach, it employs extraction of characteristics to categorize brain tumors into two types: carcinoma and normal.

In [10], the tumour classification techniques' performance in categorizing cognitive MRI scan. characteristics were evaluated. During categorizing technique, analytical properties of incoming photographs were examined, and information was carefully divided into several groups. The information have been assessed using machine learning algorithms such as KNN (k nearest neighbor), RF (random forest), SVM (support vector machines), and LDA. The SVM algorithm outperformed the others, with a 90% accuracy rate.

In [11], a deep learning network based on U-Net was created to automatically categorize tumors in preoperative MRI data. The algorithm performed well on testing and training data sets.

In [12], the approach is used to accurately segment tumor indications. The VGG-19 Transfer Learning model is fine-tuned to acquire deep features. These deep traits are then combined with manually produced form and texture properties via a sequential fusion technique. Entropy is utilized to optimize these characteristics for accurate and rapid classification, and a fused vector is fed into classifiers.

In [13], the proposed procedure involves removing the skull via brain surface extraction (BSE). The deleted skull picture is then sent into particle swarm optimization (PSO) to improve segmentation. In the following stage,

segmented pictures' local binary patterns (LBP) and deep features are retrieved, and the best features are selected using a genetic algorithm. Finally, artificial neural networks (ANNs) and other classifiers are used to categorize tumour grades.
In [14], the ROI in the brain was identified by k-nearest neighbour (K-NN) image classification. Five classes were proposed, with two indicating infection and three indicating healthy region. It was discovered that the cancer texture is substantially connected with the stroke.
To overcome time and expense constraints for multi-spectral MRI scans and other challenges, [15] proposes a single-spectral anatomical MRI technique for detecting tumor tissues. So we demonstrated a completely automated method capable of detecting tumor-containing slices and delineating the tumor region. The experimental findings using a single contrast mechanism show that our suggested approach may successfully segment brain tumor tissues with high accuracy and minimal processing complexity.

III. METHODOLOGY

To identify tumors using MRI in medical imaging, great accuracy is necessary, especially when the diagnosis has a direct influence on human lives. To do this, an effective approach for extracting and classifying tumor features is required. However, the generally utilized approaches for brain tumor classification are time-consuming and prone to overfitting. This study leverages the power of Artificial Intelligence for brain tumour classification using kaggle dataset. To achieve this, we implemented two distinct models: Random Forest and 2D Convolutional Neural Network (CNN). These models were chosen because of their shown effectiveness in dealing with complicated patterns and huge datasets.

A. 2D Convolutional Neural Network

Two Dimensional Convolutional neural networks (CNNs) have developed strong tools in a variety of disciplines, with notable performance in feature extraction and classification tasks.

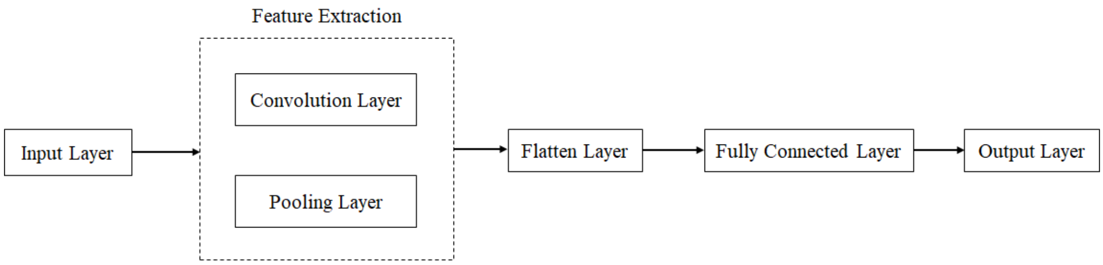

Figure 2. 2D Convolutional Neural Network

Input Layer: This layer receives primary information for input in image format.
Feature Extraction: Feature extraction incorporates the Convolution Layer and the Pooling Layer. The Convolution Layer applies filters to the input to extract crucial properties such as edges and textures. The Pooling Layer samples the feature map to decrease dimensions and computational complexity while keeping critical information.
Flatten Layer: Transforms the pooling layer's 2D feature maps into a 1D vector of features for use in the fully connected layers.
Fully Connected: A thick layer that uses all retrieved characteristics to create predictions.
Output Layer: Generated the final output.

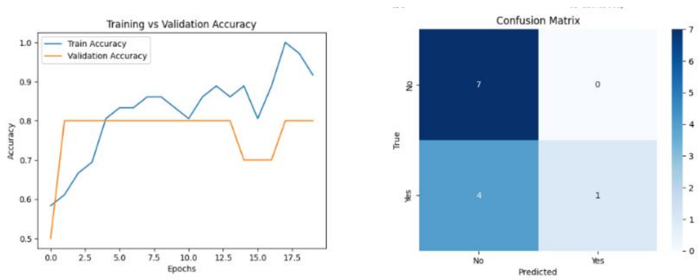

Figure 3. Results from 2D Convolutional Neural Network on Kaggle Dataset

B. Random Forest

Random Forest is a choice tree-based ensemble learning technique, was used in this work for classification tasks because to its resilience and capacity to handle high-dimensional data. During training, the method generates several decision trees using random choices of the data. using bootstrap sampling, and then combines their outputs using majority voting to get final classification. This method not only minimizes the likelihood of overfitting, but it also increases generalization. Key hyperparameters like the number of trees (n_estimators) and random state were tuned to strike a compromise between computational efficiency and model correctness. The training and evaluation procedure included preprocessing processes, model training on labeled datasets, and performance evaluation on unseen test data

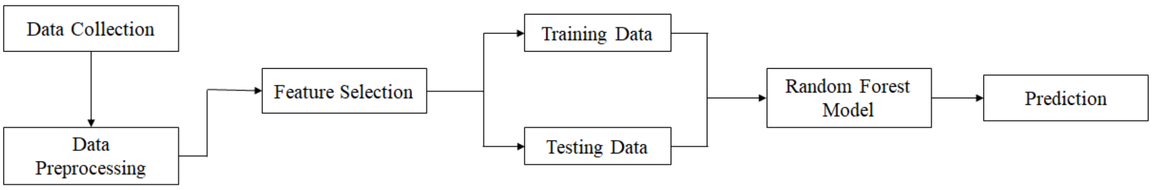

Figure 3. Random Forest Algorithm

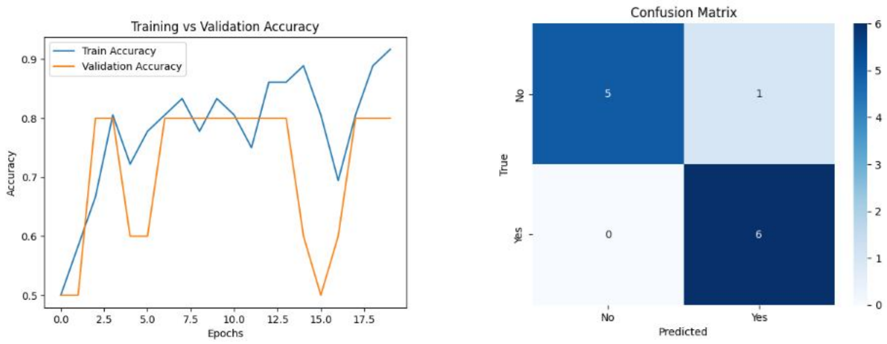

Figure 4. Results from Random Forest Model on Kaggle Dataset

IV. Result and Analysis

The performance metrics for both the 2D Convolutional Neural Network (CNN) and Random Forest classifiers indicate their effectiveness in the given classification task. The 2D CNN achieved an accuracy of 0.92, slightly outperforming the Random Forest's 0.91. Precision for the CNN was 0.93, indicating a marginally better ability to avoid false positives compared to the Random Forest, which achieved a precision of 0.92. Both models demonstrated equal F1 scores of 0.91, reflecting a balanced trade-off between precision and recall. Recall values were identical at 0.92 for both methods, showcasing their comparable ability to correctly identify relevant instances.

Overall, while the 2D CNN exhibited a marginally better accuracy and precision, the Random Forest achieved competitive performance with simpler implementation and reduced computational requirements. These findings underscore the suitability of both methods for the classification problem, with the choice of model depending on specific use-case constraints such as computational resources and model interpretability.

TABLE I. COMPARISON

Method	Accuracy	Precision	F1 Score	Recall
2D CNN	0.92	0.93	0.91	0.92
Random Forest	0.91	0.92	0.91	0.92

IV. Conclusion

This study illustrates the effectiveness of machine learning approaches, notably Random Forest and 2D CNN, in identifying brain cancers from MRI data. While the 2D CNN outperforms the Random Forest in terms of performance, it consumes significant processing resources and may be difficult to comprehend. The model should be carefully selected depending on individual clinical requirements and available computing resources. Future research paths may investigate the use of improved deep learning algorithms to improve classification accuracy and resilience. Furthermore, adding domain knowledge and explainable AI approaches might enhance the clinical interpretability and trustworthiness of machine learning-based diagnostic systems.

REFERENCES

[1] Javaria, Amin., Javaria, Amin., Muhammad, Sharif., Anandakumar, Haldorai., Mussarat, Yasmin., Ramesh, Sundar, Nayak. (2021). Brain tumor detection and classification using machine learning: a comprehensive survey. Complex & Intelligent Systems, 1-23. doi: 10.1007/S40747-021-00563-Y

[2] Sharmin, Akter., Md., Simul, Hasan, Talukder., Sohag, Kumar, Mondal., Mohammad, Aljaidi., Rejwan, Bin, Sulaiman., Ahmad, Alshammari. (2024). Brain tumor classification utilizing pixel distribution and spatial dependencies higher-order statistical measurements through explainable ML models. Dental science reports, 14(1) doi: 10.1038/s41598-024-74731-8

[3] Saad NM, Bakar SARSA, Muda AS, Mokji MM (2015) Review of brain lesion detection and classification using neuroimaging analysis techniques. J Teknol 74:1–13

[4] Tang,M.;Gao,L.;Bian,Y.;Xiang,S.;Zhang,K. (2023).Brain tumor MRI images classification based on machine learning.Applied and Computational Engineering,29,19-29.

[5] Radiological Society of North America (RSNA) and American College of Radiology (ACR). (2022, April 15). Brain tumors. Radiologyinfo.org. Retrieved February 15, 2023

[6] Brain MRI: What it is, purpose, procedure & results. Cleveland Clinic. (2022, September 5). Retrieved February 15, 2023, from https://my.clevelandclinic.org/health/diagnostics/22966- brain-mri.

[7] A B Malarvizhi et al 2022 J. Phys.: Conf. Ser. 2318 012042

[8] Kang J, Ullah Z, Gwak J. Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors. 2021;21:2222.

[9] Protima Khan, Md. Fazlul Kader, S. M. Riazul Islam, et al., Machine Learning and Deep Learning Approaches for Brain Disease Diagnosis: Principles and Recent Advances, Access IEEE, vol. 9, pp. 37622-37655, 2021.

[10] Çınarer G, Emirog˘lu BG, Classificatin of Brain Tumors by Machine Learning Algorithms, 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2019, pp. 1-4, doi: 10.1109/ISMSIT.2019.8932878

[11] Arora A, Jayal A, Gupta M, Mittal P, Satapathy SC. Brain tumor segmentation of mri images using processed image driven u-net architecture. Computers. 2021;10:139.

[12] Saba T, Mohamed AS, El-Affendi M, Amin J, Sharif M (2020) Brain tumor detection using fusion of hand crafted and deep learning features. Cogn Syst Res 59:221–230

[13] Muhammad Sharif, Javaria Amin, Mudassar Raza, Mussarat Yasmin, Suresh Chandra Satapathy, An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor, Pattern Recognition Letters,Volume 129,2020,Pages 150-157,ISSN 0167- 655,https://doi.org/10.1016/j.patrec.2019.11.017.

[14] Ali AH, Al-hadi SA, Naeemah MR, Mazher AN (2018) Classification of brain lesion using K-nearest neighbor technique and texture analysis. J Phys Conf Ser:012018

[15] Nabizadeh N, Kubat M (2015) Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features. Comput Electr Eng 45:286–301

[16] Wajgi, R., Yenurkar, G., Nyangaresi, V. O., Wanjari, B., Verma, S., Deshmukh, A., & Mallewar, S. (2024). Optimized tuberculosis classification system for chest X-ray images: Fusing hyperparameter tuning with transfer learning approaches. Engineering Reports, 6(11), e12906.