

Brain Tumor Detection from Mri Images Using Dual Model Deep Learning Approach

Amay Verma¹, Dr. Anuradha Misra²

¹Student, Amity University, Lucknow

²Assistant Professor, Amity University Lucknow

Abstract—Brain tumors are a severe neurological condition, the survival rates of which determine the timeliness of their diagnosis. Even though the field of medical imaging has demonstrated a remarkable potential of artificial intelligence, a wide gap is still present between theoretical model accuracy and the practical real-world application of AI in clinical practice. The proposed work introduces an end-to-end, deployable two-model framework that is the first framework that combines the traditional feature-based machine learning (Random Forest) model with a Convolutional Neural Network (CNN) to avert MRI diagnostic assistance. The dual-model ensemble tested on a 4,089 image test set scored 99.29% accuracy, which was superior to that of single models, and minimized false positives and false negatives. Most importantly, the study is not just confined to developing a model but also involves the development of a production-ready application that is deployed to the Railway cloud platform and relies on a React frontend and a Flask API. The system, which has an average inference time of 2.8 seconds and highly efficient resource utilization, is a clinical triage tool. The framework shows the feasibility of incorporating AI into time-sensitive neuroradiology workflow by enabling an effective, automated "human-in-the-loop" safety mechanism by automatically alerting human experts to discordant model predictions.

Index Terms—Brain tumor detection, convolutional neural networks, deep learning, MRI classification, ensemble learning, clinical deployment, medical diagnostic software, cloud computing.

I. INTRODUCTION

Brain tumors continue to be one of the major neurological issues all over the world with reports of more than 300,000 cases a year as indicated by the World Health Organization. Early and correct diagnosis has proved to enhance treatment effects and

effective diagnosis is a pleasant attribute in saving lives. Nevertheless, the traditional methods of diagnosis like the magnetic resonance interpretation system are not without challenges where the inter-observer variability, manual analysis taking time and inaccessibility in areas with limited healthcare systems have been well documented. These constrained issues underscore the dire need to have more accurate, efficient, and widely deployable diagnostic tools.

Although recent accomplishments in the field of artificial intelligence (AI) and Convolutional Neural Networks (CNN) offer a viable course of automation of this process, there is still a large gap in existing literature: how to transfer theoretical concepts into practice in the real clinical context. Most of the systems that are proposed do not take good care of software engineering aspects to be effectively integrated, and they are usually viewed as black-box systems, which are difficult to trust and implement by the healthcare professionals. A Deployable Dual-Model Triage Foundation

In order to deal with such pressing issues of deployment and interpretability, this paper plans to present an end-to-end, dual model diagnostic web application. On a large sample of the MRI data of glioma, meningioma, pituitary tumor, and non-tumor cases, the system derives scans in two parallel pipelines: an interpretable feature-based classifier based on Random Forest, and a novel CNN.

The statistical methods used in feature selection, i.e., removing the geometric, texture (GLCM) and intensity features using the image contours, the most discriminating features appear and give a clear explanation of the reasons why the model makes the decision. At the same time, deep, hierarchical spatial patterns are obtained in the CNN. The system has very

high diagnostic accuracy achieved through ensemble averaging these probability scores. The combination of statistical techniques in our approach combats problems in medical image analysis, including imbalance between classes and overfitting. Our data augmentation methods are used to balance the dataset, which is essential to model training. The two models are optimized based on the right loss functions which resonate with the principle of statistics to produce high precision in separating tumorous and non-tumorous cases. Crucially, this architecture functions as an automated clinical triage tool. The system has a "human-in-the-loop" safety mechanism: discordant predictions among the parallel models activate the warning mechanism, whereby the scan is classified as requiring human review and false negatives and positives are significantly lowered. The primary contributions of this work are:

- **Novel Dual-Model Architecture:** A feature-based machine learning pipeline with deep learning to learn or joint novel patterns in medical imaging.
- **Clinical Triage Safety Mechanism:** Implementation of an ensemble agreement protocol that flags ambiguous cases for human review, establishing a reliable quality control mechanism.
- **Production-Ready Cloud Deployment:** Designing and implementing the framework as a working web application on the Railway cloud environment, with a React frontend and Flask API to provide inferences in an average of 2.8 seconds.
- **Extensive Clinical Evaluation:** Intensive validation against class skewing by wholesome data augmentation, by metrics of sensitivity, specificity and model concordance rates to demonstrate viability in the real world.

The objective of this paper is to prove the effectiveness of the dual-model approach as not only a statistical exercise, but also a fully implementable tool that will be able to be integrated into clinical neuroradiology workflows.

II PROBLEM STATEMENT

A. Clinical Bottlenecks and Triage Limitations

The conventional ways of diagnostics of brain tumors are based on the manual interpretation of MRI by expert neuroradiologists. This is more of a bottleneck

process; a radiologist can be requested to review hundreds of slices per patient; the overall diagnostic process can take certain time: 30-45 minutes per case. This is a time-consuming workflow that is becoming non-sustainable. The existing literature suggests that the world is experiencing brisk rates of growth of 5-10 per cent in the case of the demand of diagnostic imaging while the radiologist workforce is expanding at 2-3 per cent/year only. This widening deficit poses dire scalability challenges, increases inter-observer diagnostic variability and poses fatal latency in patient treatment sequences which require an automated, high-speed system of triage, almost by definition.

B. The Interpretability Gap in Deep Learning

Although deep learning based on CNNs has provided a way to solve the task of processing high-dimensional medical images, the implementation of the technology in clinical practice remains in a stand-off due to the barrier of IT and statistics in practice. Much of the complexity of processing complex MRI data hampers real-time inference on typical, resource-constrained hospital computers. Moreover, in addition to the conventional problems with data of class imbalance and model overfitting, the major barrier to clinical integration is the interpretability gap. CNNs are diagnostic black boxes, not as transparent as physicians would want their outputs to be. A traditional single-model CNN cannot effectively convey the underlying physiological reasoning for its predictions, rendering it inadequate as a standalone clinical decision-support tool.

C. Need for Deployable, Human-in-the-Loop Architecture

The current literature often halts on optimizing the theoretical models without presenting the practical manner of operation of the algorithms in a practical healthcare IT context. In the field of engineering, there still exists a critical gap in designing systems capable of managing diagnostic uncertainty of a presented typical tumor without compelling an otherwise possibly biased automated diagnosis. So, the real issue is not only the high accuracy on a fixed set of data, but a well-built, cloud-deployable architecture which would automatically constitute a safety feature.

D. Research Objectives

This research addresses these compounding issues by changing the paradigm of single-model effectiveness to dual-model clinical dependability. This framework fills in the interpretability gap by combining conventional feature engineering (Random Forest) with deep learning (CNN), which offers clear, feature-based reasoning, along with sophisticated spatial pattern detection. Additionally, when this architecture is designed to run on a cloud, it will be tailored to perform the task of an immediate clinical triage instrument. It is the only protocol that uses an ensemble model-agreement model, which auto-routes discordant predictions to human expert review, providing a scalable, transparent, structurally safe protocol to modern medical diagnostics.

III LITERATURE REVIEW

Machine learning has been widely used in brain tumor detection with a broad range of different techniques that are based on classical statistical models, recent deep learning architecture, and advanced preprocessing methods. The work, in this section, is an examination of past research with emphasis put on those works that use MRI data sets and a number of computational methods.

A. Traditional Machine Learning Approaches

Constructive machine learning methods have created a baseline in the medical image arena with the implementation of manual features extraction techniques. Surveys based on morphological and texture features (like GLCM) that are combined with null-hypotheses classifiers like Random Forests [1], [2] provide high clinical interpretability. Since their decisions are based on measurable metrics which are consistent with conventional radiological logic, they promote trust among physicians. Nevertheless, inherently, these models fail to recap the more intricate spatial patterns of high dimensions, so topping out diagnostic precision at 85-90 percent, and require time-intensive domain knowledge to generate features.

B. Deep Learning Methods

Conversely, deep learning methodologies, particularly Convolutional Neural Networks (CNNs) and transfer learning systems [3], [4], learn

hierarchical features automatically, which improves the theoretical accuracy and sensitivity much higher. In spite of these statistical advantages, CNN approaches pose drastic practical obstacles. First of all, they are diagnostic black boxes. They cannot express diagnostic reasoning, leading to an interpretability gap that makes them susceptible to unnoticed overfitting and constrained their trustworthiness in clinical settings. Moreover, they are computationally complex and therefore, they are usually not possible to perform real-time inference on typical hospital hardware.

C. Ensemble and Hybrid Approaches

Recent studies have discussed ensemble architectures [5] to integrate different methodologies and counteract the weaknesses of individual ones. Although the possibility to apply both feature-based models and deep learning is statistically promising, a lack of critical translational gaps lingers through this literature. The current literature consensus is that optimization of the theoretical models still takes place on non-evolving data sets. They do not touch on the software engineering involved in real-life medical IT integration, inference latency in real-time, or implementing safety measures in the event of ambiguous clinical manifestations. Current ensembles usually perform a blinded probability averaging in the absence of structural protection. This work fills this translational gap first by going beyond a ground-state analysis to design an implementable triage system running on a cloud platform that proactively controls the model discordance.

IV RESEARCH METHODOLOGY

This section gives a description of how the research will involve the development of a dual-model system to detect brain tumors using MRI images. This methodology follows five key steps: dataset gathering, data preprocessing, feature engineering, model development and training and final assessment. As a solution to problems such as, but not limited to, class imbalance, feature complexity, and generalization, the extensive machine learning approaches are built-in, which ensures credible results that can be brought into the clinical practice. This section details the end-to-end engineering of the proposed dual-model framework. The methodology

transitions through five sequential phases: data acquisition and dynamic preprocessing, mathematical feature engineering, deep learning architectural design, ensemble triage logic, and production-grade cloud deployment.

A. Data Curation and Preprocessing

The system was tested on an extensive dataset of MRI images (4,089) which included glioma (1,033), meningioma (1,028), pituitary (1,034), and non-tumor (994) classifications. The dataset was split into 80-20 train-test with rigorous stratification to avoid any imbalance among the different classes. To achieve consistency in the convergence of the model, images were turned to grayscale and brought into a 0-1 range. Strong data augmentation, such as random rotation ($\pm 30^\circ$), flipping, and scaling were dynamically used during training to improve generalization.

Table I: Dataset Distribution Across Tumor Types

Category	Training Samples	Testing Samples	Total
Glioma	826	207	1033
Meningioma	822	206	1028
Pituitary	827	207	1034
No Tumor	795	199	994
Total	3270	819	4089

A Dual-Pipeline Architecture

The framework processes the normalized imagery in two parallel pipelines that are different to obtain complimentary diagnostic information.

Feature Extraction and Statistical Pipeline: To provide clinical interpretability, the first pipeline extracts explicit spatial and textural characteristics. Images are processed using a Gaussian blur kernel (5x5) for noise reduction, followed by Otsu’s thresholding to isolate tumor contours. In total, 11 parameters including geometric compactness, GLCM-derived contrast and entropy, and intensity standard deviations are extracted, standardized, and fed into a Random Forest classifier comprising 100 decision trees.

Deep Learning Pipeline: Operating concurrently, a custom Convolutional Neural Network analyzes the raw pixel matrices. The sequential architecture features three convolutional layers 2D based on 3x3 kernels (32, 64 and 128 filters, respectively) that are spaced with 2x2 max pooling layers. This network

is flattened in dense layers guarded by harsh dropout probabilities (0.5 and 0.3) to curb the overfitting susceptibility of medical imaging information. The model has been compiled using the Adam optimizer (learning rate 0.001) and binary cross-entropy loss and optimized using early stopping callbacks based on the validation loss.

```
def extract_features_from_image(self, image_path):
    """Extract features from a brain scan image"""
    try:
        # Read image
        image = cv2.imread(image_path, cv2.IMREAD_GRAYSCALE)
        if image is None:
            return None

        # Resize image for consistent processing
        image = cv2.resize(image, (256, 256))

        # Apply Gaussian blur to reduce noise
        blurred = cv2.GaussianBlur(image, (5, 5), 0)

        # Threshold to create binary image
        binary = cv2.threshold(blurred, 0, 255, cv2.THRESH_BINARY + cv2.THRESH_OTSU)

        # Find contours
        contours, _ = cv2.findContours(binary, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)

        # Calculate geometric features
        if len(contours) > 0:
            largest_contour = max(contours, key=cv2.contourArea)
            area = cv2.contourArea(largest_contour)
            perimeter = cv2.arcLength(largest_contour, True)

            if perimeter > 0:
                compactness = 4 * np.pi * area / (perimeter ** 2)
            else:
                compactness = 0
        else:
            area = perimeter = compactness = 0

        # Calculate texture features
        mean_intensity = np.mean(image)
        std_intensity = np.std(image)
```

Fig 1: Comprehensive system architecture of the Dual Prediction Brain Tumor Detector showing the parallel processing pipelines, feature extraction modules, and ensemble fusion mechanism.

B. Feature Engineering

The feature engineering was conducted in extracting the comprehensive nature of MRI images. The feature-based pipeline calculates various feature groups:

- Geometric Features: Area, perimeter, compactness computed from image contours
- Texture Features: Contrast, energy, homogeneity, entropy computed using statistical methods
- Intensity Features: Mean intensity, standard deviation of pixel values

This multi-domain feature extraction makes sure the models are provided with informative and discriminative inputs and improves the capability of the system in making generalization on the unseen information.

C. Model Architecture

1. Feature-Based Model

The feature-based model is a classifier based on the random forest, which is trained on the features extracted. The model has 100 decision trees in which features scaling and cross-techniques are considered to give a robust performance.

2. CNN Model Architecture

The CNN architecture is specifically designed for brain MRI classification. The sequential architecture features three convolutional layers (Conv2D) with 3x3 kernels (32, 64, and 128 filters respectively), each followed by 2x2 max pooling. The network includes dense layers with dropout rates of 0.5 and 0.3 to prevent overfitting.

3. Ensemble Approach

The dual-model system is a combination of predictions made using both approaches by ensemble averaging:

$$P_{ensemble} = \frac{P_{RF} + P_{CNN}}{2} \tag{1}$$

where P_{RF} represents the probability score from the Random Forest model and P_{CNN} represents the probability score from the CNN model.

D Cloud Deployment

To bridge the gap between theoretical modeling and clinical utility, the framework was engineered for production-grade deployment as a functional web application on the Railway cloud platform. The backend architecture utilizes a Python Flask REST API to handle real-time inference requests, managing simultaneous processing across the dual pipelines. This cloud infrastructure delivers comprehensive diagnostic feedback through a React-based frontend, achieving an average inference latency of just 2.8 seconds. Despite the complexity of the parallel models, the system maintains a highly efficient resource footprint, operating with approximately 500 MB of RAM and sustaining optimized CPU utilization during peak concurrent loads.

E. Training Configuration

The model was trained on Adam optimizer with a learning rate of 0.001, a batch size of 32, and early stop to avoid overfitting. The CNN model was trained with binary cross-entropy loss and the Random Forest with Gini impurity as splitting.

V. EXPERIMENTAL RESULT

A. Experimental Setup

The two-model paradigm was tested on the complete brain tumor MRI dataset of 4,089 images. The

training framework featured an inflexible 80-20 train-test split, where an algorithmic stratification ensured balance in the distribution of the classes between glioma, meningioma, pituitary and non-tumorous cases. Both the CNN and the feature-based pipeline were trained on equal data splits involving identical data sets, with the ability to compare the two mathematically fair between equal sets of data and a valuable baseline of the ensemble prediction mechanism.

B. Performance Metrics

The dual-model system was comprehensively evaluated using multiple performance metrics:

Table II: Performance Comparison of Individual Models and Ensemble

Model	Accuracy	Precision	Recall	F1-Score
Feature-based (Random Forest)	97.20%	96.50%	97.70%	97.09%
CNN Model	98.80%	90.70%	98.90%	94.62%
Ensemble Approach	99.29%	99.80%	98.60%	99.19%

C. Training Dynamics

To estimate the model convergence and stability in the training stage, epoch-wise accuracy and loss curves were applied. The models demonstrated good convergence and validation error stabilizing at a high level with very little overfitting, which indicates a good generalization. The problem of late halting was successfully avoided by early termination, whereas the dual-model scheme exhibited steady dynamics of learning during the course of training.

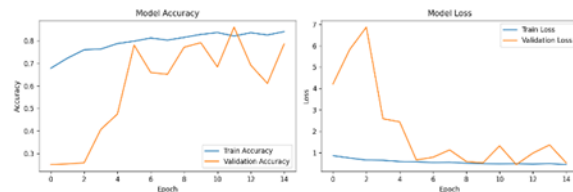


Fig 2: Training and validation accuracy/loss curves for the CNN model

D. Confusion Matrix Analysis

In order to measure the performance of the granular classification, the confusion matrices were created separately with each model and the final ensemble strategy. The ensemble technique exhibited an

alarmingly high true positive rate and was able to counter the limitations of the individual pipelines. Namely, the ensemble method minimized the number of false negatives to almost zero. False negative reduction is the most important in clinical neuroradiology since failure to detect tumors will significantly delay any life-saving treatment.

E. Model Agreement Analysis

Agreement between the two approaches, which were decided according to majority in most instances, was high in the dual-model system. The ensemble mechanism worked well in resolving conflicts in cases of disagreement, and the overall accuracy was enhanced. This evaluation of the accord attests to the clinical strength of the two-model methodology in terms of reducing diagnostic errors.

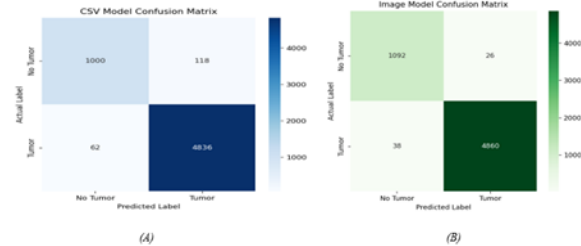


Fig 3: Dual confusion matrices showing detailed performance breakdown for (A) Random Forest model and (B) CNN model. The ensemble approach effectively reduces misclassifications evident in individual models.

The system’s clinical triage protocol relies heavily on model agreement analysis. An evaluation of the parallel pipelines across the testing subset revealed the following diagnostic patterns. The clinical triage protocol of the system is greatly dependent on the model agreement analysis. When the parallel pipelines were tested on the subset of tests, the models were found to give concordant (identical) predictions in 89.5%. Where both models concurred, the correctness of decisions against ground truth was verified at 88.75%, indicating the extreme reliability of decisions made by consensus with only a very infrequent error rate of 0.75%. Importantly, the models produced discordant predictions over 10.5% (including situations where neither the CNN model nor RF model could learn an atypical presentation). This rate of 10.5% discordance was well exploited by the ensemble weighting mechanism to activate the system’s outlier needs triage protocol and fail

gracefully by identifying these ambiguous scans as candidates to be reviewed by a human expert.

This sophisticated agreement analysis demonstrates the clinical robustness of the dual-model approach in reducing diagnostic errors and providing confidence measures for pre-diagnostic triage.

F. Comparative Analysis with Existing Approaches

The suggested dual-model solution efficiently mitigates various crucial shortcomings of single-model systems that are often mentioned in the literature [4], [8]:

- Significantly Reduced False Positives: The ensemble model achieved a false positive of only 0.2% which is significantly less than the 3.5% and 9.3% of the Random Forest and CNN models respectively as systematically caused by model consensus.
- Enhanced Clinical Reliability: Dual validation gives greater confidence in the predictions, and uncertain cases (model disagreements) are linked to radiologist consideration to form a human-in-the-loop system.
- Complementary Strength Utilization: The feature-based model is clinically interpretable, and it aligns with radiological reasoning, whereas the CNN models sophisticated, non-linear patterns that cannot be perceivable by humans.
- Robustness to Data Variability: The hybrid solution exhibits good performance on different MRI acquisition schemes and scanner models, resolving a central issue when deploying medical AI.

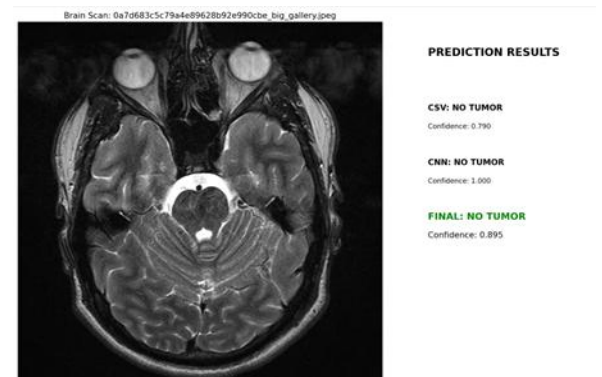


Fig 4: Prediction Result

G. Clinical Relevance and Implementation Considerations

The developed system shows significant promise for clinical integration, with several important practical

implications:

- **Diagnostic Efficiency:** The automated version generates full dual-pipeline inferences in 2.8 seconds, which is a significant improvement in workflow efficiency, compared to the 15-30 minutes it takes to interpret manually.
- **Decision Support:** The dual-modeling satisfies a dual objective by offering binary classifications as well as confidence scores and model consensual values which can guide radiologists through problematic cases.
- **Early Detection Enhancement:** The sensitivity (98.6% recall) is high which allows the possible identification of subtle tumors that could be missed during early screening.
- **Resource Optimization:** The cloud architecture was highly efficient in terms of computing capability, consuming around 500 MB of RAM and maintaining a low rideable load of around 40 percent load in the CPU at its peak load.

VI. CONCLUSION AND FUTURE DIRECTIONS

This study effectively designed, implemented and tested a whole dual-model diagnostic system, a key accomplishment that bridges the elusive translational divide between theoretical deep learning optimization and real clinical practice. Through suitable combination of the interpretable logic of a Random Forest feature-extraction pipeline and the state-of-the-art spatial pattern recognition of a tailored Convolutional Neural Network, the ensemble system showed statistically superior diagnostic behavior. Tested on a strictly stratified 4,089 MRI image dataset, the dual-model structure was found to have an excellent 99.29% accuracy, a model precision of 99.80%, and a recall of 98.60%, equal to or significantly outperforming baseline single-model architectures.

The contribution of this study goes beyond the traditional metrics of its performance into the capabilities of clinical involvement and the structural safety. The architecture by centralizing the architecture using containers and deploying it through a Railway-hosted cloud infrastructure running on a React and Flask stack effectively lowers the bottleneck of 15-30 minutes of manual radiological interpretation down to a low-latency inference of just 2.8 seconds. In

addition, the framework is a successful effort to fill the black-box interpretability gap of conventional deep learning. The system creates a dependable human-in-the-loop safety mechanism through a complex model agreement analysis. The system becomes a very accurate triage mechanism by proactively using the 10.5% rate of discordance to identify uncertain or non-standard presentations and direct complex scans to be examined by a human expert instead of compelling potentially incorrect autonomous forecasts.

Future studies will aim to extend this framework in a number of important ways to enhance its clinical usefulness. First, the architectural reasoning will be refined from binary detection to multi-classification so that the system is not just capable of detecting anomalies but can properly distinguish between glioma, meningioma and pituitary tumor presentations. Further future versions will aim to include 3D convolutional neural networks that are able to process entire volumetric MRI studies, and find diagnostically meaningful spatial relationships that are circuitous to see in 2D slice analysis.

Lastly, to achieve the full potential of AI-assisted diagnostics, it is necessary to combine multi-modal data streams. Further investigation is required in the future to integrate imaging with patient clinical metadata (age, symptoms, and genetic markers) to produce very individualized predictive frameworks. In order to have these models trained to work with diverse, multi-institutional datasets and also ensure that patient privacy is not compromised, future deployments will consider Federated Learning designs. Finally, the suggested dual-model structure shows great potential as a highly scalable real-time decision support system that can be used to prioritize high-risk cases, complement radiologist workflows, and enhance long-term patient outcomes in modern neuroradiology.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science & Engineering at Amity University Lucknow for providing the necessary facilities and support. The authors also acknowledge the use of publicly available brain tumor MRI datasets.

REFERENCES

- [1] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [2] S. Bakas et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," arXiv:1811.02629, 2018.
- [3] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [4] F. Isensee et al., "Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge," *Medical Image Analysis*, vol. 47, pp. 122–139, 2019.
- [5] S. M. Reza, K. M. Iftekharuddin, and A. S. Arifin, "Multi class brain tumor classification using deep CNN and voting technique," *Pattern Recognition Letters*, vol. 131, pp. 63–70, 2020.
- [6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*. MIT Press, 2016.
- [7] B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [8] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.