Benchmarking Deep Convolutional Neural Networks for Brain Tumor Detection Using Magnetic Resonance Imaging Data

Tegar Anugrah Firdaus[1], Bryan Rais[2], Marcelinus Jonathan Salim[3], Andhika Rastra Negara[4], Muhammad Qaiser Zaydan Zia[5], Yamin Thwe[6], Muhammad Nur Afnan Uda[7], Uda Hashim[8], and Yuri Pamungkas[9*]
[1−5,9]Department of Medical Technology, Faculty of Medicine and Health, Institut Teknologi Sepuluh Nopember
Jawa Timur, Indonesia 60111
[6]Department of Big Data Management and Analytics, Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi
Pathum Thani, Thailand 12120
[7,8]Department of Electrical and Electronics Engineering, Faculty of Engineering, Universiti Malaysia Sabah
Sabah, Malaysia 88400
Email: [1]5049231036@student.its.ac.id, [2]5049231037@student.its.ac.id, [3]5049231027@student.its.ac.id, [4]5049231115@student.its.ac.id, [5]5049231010@student.its.ac.id, [6]yamin_t@rmutt.ac.th, [7]nurafnan@ums.edu.my, [8]uda@ums.edu.my, [9]yuri@its.ac.id

***Abstract*—Brain tumors are associated with high mortality and severe neurological effects, making accurate and timely diagnosis essential. However, manual interpretation of Magnetic Resonance Imaging (MRI) remains subjective and time-consuming. Deep learning, particularly Convolutional Neural Networks (CNNs), offers a promising approach to improving diagnostic accuracy and efficiency. This research systematically benchmarks five CNN architectures (VGG19, DenseNet201, ResNet50, Inception-v3, and MobileNet) on balanced and naturally imbalanced MRI datasets. Its novelty lies in evaluating all models under a unified preprocessing and transfer-learning framework across both dataset distributions, with Grad-CAM used to support interpretability. Preprocessing includes resizing, adaptive masking, Gaussian blur, Contrast Limited Adaptive Histogram Equalization (CLAHE), normalization, and augmentation, followed by transfer learning using ImageNet-pretrained models. Performance is evaluated using accuracy, precision, recall, specificity, F1-score, and Area Under the Curve (AUC). Results show that VGG19 achieves the highest accuracy (0.937 for balanced and 0.962 for imbalanced data), recall (0.991 and 0.984, respectively), and F1-score, demonstrating strong sensitivity and generalization. DenseNet201 achieves the highest AUC (0.992 and 0.996) and high specificity, indicating strong overall discriminative performance. These findings suggest that VGG19 is particularly suitable for sensitivity-oriented early screening, whereas DenseNet201 is advantageous when overall discrimination and specificity are prioritized. In conclusion, CNN-based models show substantial potential for brain tumor classification, with architecture selection depending on specific clinical priorities.**

***Index Terms*—Deep Convolutional Neural Networks, Brain Tumor Classification, Magnetic Resonance Imaging (MRI), Medical Image Analysis**

I. Introduction

BRAIN tumors are among the most severe medical conditions and are associated with a high mortality rate [1]. Each year, millions of individuals worldwide are diagnosed with brain tumors, which are generally grouped into two broad categories, namely benign and malignant [2, 3]. Benign brain tumors usually grow more slowly and tend to exhibit more uniform structural characteristics. In contrast, malignant tumors, often referred to as brain cancer, grow rapidly and display heterogeneous patterns [4, 5]. Early detection and timely management are essential because even small brain tumors can cause serious disturbances in neurological functions such as motor coordination and cognitive ability. These disruptions directly affect patients' quality of life [6, 7].

Received: Sep. 18, 2025; received in revised form: Nov. 22, 2025; accepted: Nov. 28, 2025; available online: Sep. 08, 2026.
*Corresponding Author

The global incidence of brain tumors has risen significantly, particularly in developed countries. Brain and central nervous system cancers represent a substantial global health burden, with incidence varying across regions and demographic groups [8]. Although some regions, such as Africa and parts of Asia, still report lower incidence rates, the overall global trend shows a steady increase. This trend highlights a growing burden on healthcare systems and underscores the need for more accurate and efficient diagnostic solutions. Current diagnostic approaches rely heavily on Magnetic Resonance Imaging (MRI), which provides highly detailed structural images of the brain [9, 10]. However, interpreting MRI scans remains a complex task that depends on the expertise of radiologists and neurologists. Manual analysis is time-consuming and often subjective, which may result in misclassification and delayed treatment decisions [11].

Recent advancements in MRI technology have significantly improved image resolution, contrast sensitivity, and acquisition speed, enabling more detailed visualization of tumor morphology, edema, and necrotic regions [12, 13]. Advanced modalities such as Diffusion-Weighted Imaging (DWI), Perfusion-Weighted Imaging (PWI), and Functional MRI (fMRI) have further enhanced tumor characterization by providing physiological and metabolic insights beyond anatomical details [14–16]. Despite these technological developments, manual interpretation remains labor-intensive and susceptible to inter-observer variability, underscoring the necessity for automated and objective analysis methods.

Deep learning and Artificial Intelligence (AI) have emerged as promising solutions to this problem. These technologies can automatically process large volumes of medical imaging data, reducing human error and improving both the speed and accuracy of diagnosis [17–19]. Convolutional Neural Networks (CNNs), in particular, have shown strong ability to extract critical features from MRI scans, enabling highly accurate classification of tumor and non-tumor cases [20, 21], [22, 23]. However, despite this demonstrated success, a significant gap remains. The performance reported in many studies often relies on ideal balanced datasets, which do not reflect the imbalanced nature of real-world clinical data, and there is a lack of systematic benchmarking across different architectures to assess their true robustness [24–26]. This issue is compounded by other persistent challenges, such as the complexity of tumor segmentation, which is crucial for delineating tumor boundaries that directly affect therapeutic planning [27–29].

To address these challenges, the research evaluates and benchmarks five deep learning architectures (VGG19, ResNet50, DenseNet201, Inception-v3, and MobileNet) for brain tumor analysis using MRI data. The novelty of this research lies in the systematic evaluation of these models under both balanced and imbalanced dataset conditions, thereby simulating the real-world data distribution challenges commonly encountered in clinical practice. This underexplored aspect is crucial for assessing the robustness and clinical applicability of deep learning models in realistic medical imaging scenarios.

The main research contributions can be summarized as follows. First, it provides a comprehensive benchmarking of five state-of-the-art CNN architectures for binary brain tumor classification using MRI data. Second, it evaluates the models under both balanced and imbalanced dataset conditions to simulate real-world clinical variability and assess model robustness. Third, a customized preprocessing pipeline is developed by integrating adaptive masking, Gaussian blur, Contrast Limited Adaptive Histogram Equalization (CLAHE), and normalization to enhance MRI image quality prior to CNN training. Finally, the research incorporates Grad-CAM-based explainability to visualize model attention areas and validate the clinical interpretability of the classification results. Collectively, these contributions establish an extensive benchmarking framework to identify the most reliable CNN architecture for accurate brain tumor segmentation and classification. The proposed framework is expected to enhance the development of AI-assisted clinical decision-support systems, reduce diagnostic subjectivity, and improve treatment planning accuracy in clinical practice.

A. Related Works

Recent advancements in deep learning have significantly improved brain tumor classification from MRI images. A 2D CNN and a convolutional autoencoder are developed. Both models are compared with six machine-learning techniques using 3,264 MRI images. The proposed 2D CNN achieved a training accuracy of 96.47%, while the convolutional autoencoder achieved 95.63% [30]. In the same year, another previous research has developed a custom CNN architecture trained on 3264 MRI images. The model reaches 93.3% accuracy, outperforming ResNet-50, VGG16, and Inception V3 [18]. However, the long training time and the absence of tumor area visualization remained limitations that restricted its clinical applicability.

Further evidence of the potential of custom architectures has been presented by Lamrani et al. [31], who have designed a CNN with four convolutional layers and six dense layers. Using 3,000 MRI images, the model has achieved 96.33% accuracy and 97.93%

TABLE I
SUMMARY OF PREVIOUS STUDIES ON BRAIN TUMOR DETECTION USING DEEP LEARNING-BASED MAGNETIC RESONANCE IMAGING (MRI).

References	Year	Dataset	Model/Approach	Accuracy	Limitation
[30]	2023	3,264 MRI	2D CNN, convolutional autoencoder, & six ML techniques	96.47% training accuracy	Single public MRI dataset; external clinical validation not reported
[18]	2023	3,264 MRI	Custom CNN vs ResNet-50, VGG16, Inception V3	93.3%	Long training, no visualization
[31]	2022	3,000 MRI	Custom CNN (4 conv + 6 dense layers)	96.33%	Needs larger dataset
[32]	2022	253 MRI	CNN, VGG-16, Ensemble	98.15% (VGG-16)	Very small dataset
[33]	2021	3,064 MRI	Multiscale CNN	97.3%	Misclassify non-cerebral, low sensitivity
[34]	2021	3,064 & 253 MRI	Mask R-CNN + DenseNet-41 backbone	98.34% / 96.3%	Manual annotation needed
[35]	2020	253 MRI	Hybrid ResNet-50 with 10 new layers	97.01%	Need clinical validation
[17]	2020	3,064 MRI	Hybrid CNN-NADE	95%	Lower than AlexNet (96.61%)
[36]	2020	3,064 MRI	Inception-v3 + DenseNet201 fusion	99.51% (DenseNet)	No fine-tuning
[37]	2020	3,064 MRI	ResNet-50 + augmentation	99% / 97%	Need larger dataset, no grading

Note: Machine Learning (ML), Convolutional Neural Networks (CNN), Convolutional Neural Network-Neural Autoregressive Distribution Estimator (CNN-NADE), and Magnetic Resonance Imaging (MRI).

precision. Despite its strong performance, it emphasizes the necessity of evaluating the model on larger datasets to establish generalizability. Complementing these findings, Younis et al. [32] have tested CNN, VGG-16, and ensemble models on a smaller dataset of 253 MRI images. VGG-16 has outperformed the other approaches with 98.15% accuracy, although the small dataset size has posed limitations for broader clinical use.

Multiscale and hybrid approaches have also demonstrated considerable success. Díaz-Pernas et al. [33] have proposed a multiscale CNN that processes MRI scans using three independent feature streams. The model achieves 97.3% accuracy and outperforms 7 benchmark methods. Nonetheless, misclassification in non-cerebral regions and reduced sensitivity for meningioma cases are observed. Similarly, Masood et al. [34] have combined classification and segmentation using a custom Mask R-CNN with DenseNet-41 as the backbone. The system achieves 98.34% classification accuracy and 96.3% segmentation accuracy, but requires intensive manual annotation, which can limit scalability.

Hybrid deep learning architectures have also shown promise in improving efficiency. Çinar et al. [35] have modified the last five layers of ResNet-50 and added 10 new layers, leading to a 97.01% accuracy on 253 MRI images. Despite this improvement, the model requires further clinical validation. Gassenmaier et al. [17] have proposed a Convolutional Neural Network-Neural Autoregressive Distribution Estimator (CNN-NADE) hybrid architecture, which achieves 95% accuracy. However, its performance is slightly lower compared to AlexNet, which reaches 96.61% on the same dataset.

Among the most notable contributions, Noreen et al. [36] have employed a feature fusion strategy by concatenating features from Inception-v3 and DenseNet201. This approach produces an impressive accuracy of 99.51% with DenseNet201, outperforming other tested models. The limitation is the absence of fine-tuning, which can have further enhanced the model's adaptability. Finally, Ismael et al. [37] have demonstrated that ResNet-50 combines with extensive data augmentation achieving 99% accuracy at the image level and 97% at the patient level. Nevertheless, it highlights the need for larger datasets and incorporation of tumor grade classification to ensure comprehensive clinical utility. A summary of previous research relates to brain tumor detection using deep learning-based MRI is presented in Table I.

II. RESEARCH METHOD

This section outlines the overall methodology designed to enhance the performance of brain tumor classification using CNNs. The proposed framework consists of several key stages, namely dataset preparation, image preprocessing, model development with different CNN architectures, and performance evaluation. The complete research workflow is illustrated in Fig. 1. The methodology begins with identifying the research problem, followed by collecting brain MRI images from publicly available open-access repositories. Once the dataset is obtained, preprocessing steps are performed to ensure data consistency and quality. These steps include resizing images to a uniform dimension, normalizing pixel intensities to standardize values, and augmenting the data to increase variability and reduce the risk of overfitting.

After preprocessing, the dataset is systematically split into training and testing subsets. The training subset is used to develop CNN-based models for binary classification of brain tumors, while the testing subset is reserved to validate the generalization ability of the trained models. In the final stage, performance is evaluated using a set of established metrics, including accuracy, precision, recall, F1-score, and Area

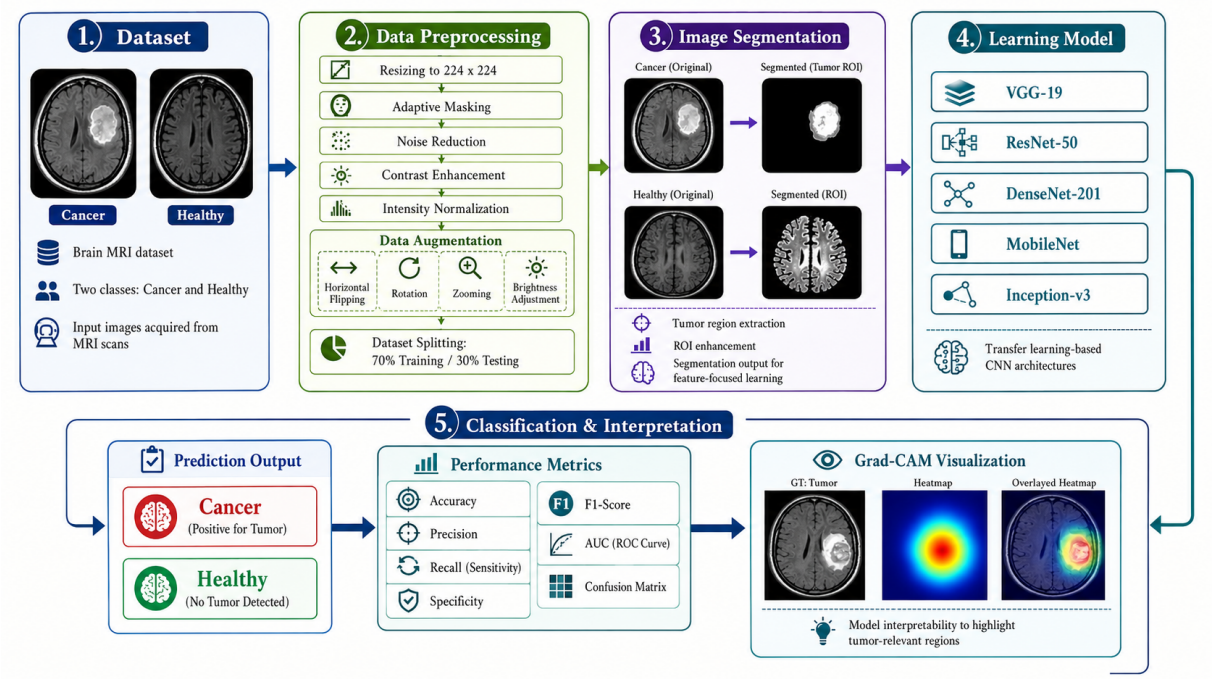


Fig. 1. The overall workflow diagram in the research. Note: Magnetic Resonance Imaging (MRI), Region of Interest (ROI), Area Under the Curve (AUC), Receiver Operating Characteristic (ROC), and Ground Truth (GT).

Under the Curve (AUC). These measures provide a comprehensive assessment of model effectiveness and robustness, thereby facilitating reliable comparisons among the implemented CNN architectures.

A. Dataset

The dataset used in the research is obtained from Kaggle's open-access repository [38]. It consists of MRI brain scans categorized into two groups, namely tumor and non-tumor. The images are available in multiple formats, including .jpg, .png, and .tif, with varying resolutions. However, the dataset does not provide segmentation annotations or tumor-type labels, limiting its scope to binary classification tasks. In the research, the dataset is used to develop CNN models to classify MRI images as healthy or tumor-bearing. Deep learning is considered appropriate because it can automatically extract relevant features from raw imaging data without manual feature engineering. This approach is further supported by prior research, which has consistently demonstrated that CNNs and their hybrid architectures achieve high levels of accuracy in brain tumor classification [30, 39]. The imbalance in the dataset originates naturally from the repository and is not artificially introduced through over- or under-sampling. The original collection contains a higher proportion of tumor MRI images compared to non-tumor scans. This natural imbalance is intentionally preserved to reflect real-world clinical data distributions and to evaluate model robustness under authentic conditions.

B. Preprocessing

The preprocessing stage comprises several sequential steps to improve image quality and prepare the dataset for deep learning analysis. The procedures include resizing images to 224 $\times$ 224 pixels, adaptive masking to focus on relevant regions, noise reduction with Gaussian blur, contrast enhancement using CLAHE, scaling pixel values to the [0-1] range, and data augmentation. These steps are essential for enhancing the visual quality of MRI images, suppressing irrelevant information, highlighting important features, and standardizing input dimensions for convolutional neural network models. Furthermore, various augmentation techniques are applied to increase dataset diversity and reduce the risk of overfitting during training. By introducing controlled variations in the dataset, these methods ensure that the models develop greater robustness and generalization ability. The complete preprocessing workflow is illustrated in Fig. 2, and a more detailed explanation of each preprocessing step is provided in the following parts.

The preprocessing pipeline begins with image resizing to ensure uniform input dimensions across all MRI samples, thereby enabling consistent processing by convolutional neural network models. This step is

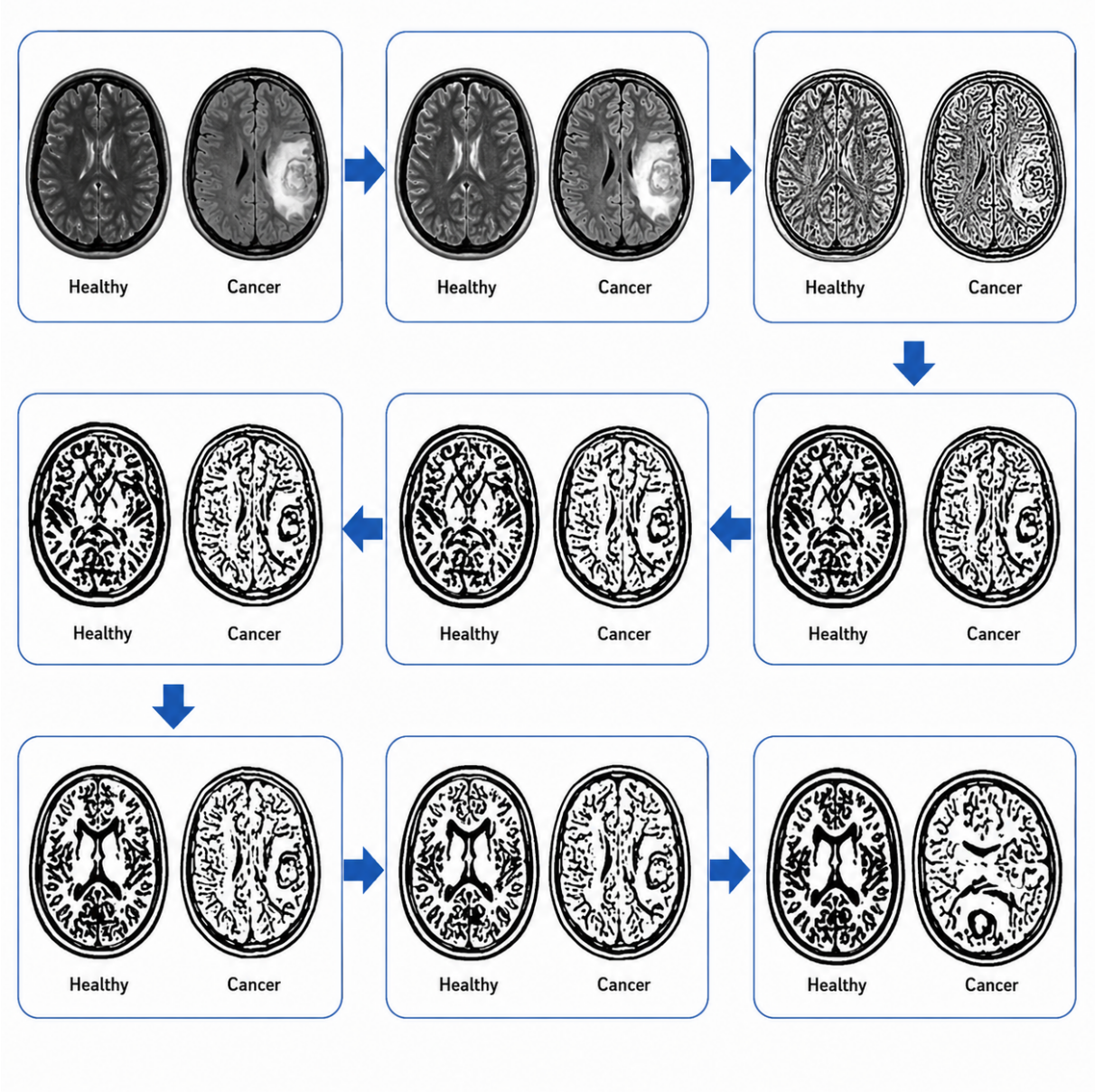


Fig. 2. Preprocessing workflow for healthy and cancerous brain Magnetic Resonance Imaging (MRI) images.

followed by adaptive masking, which isolates relevant brain regions and reduces background interference that can introduce noise. To further enhance image quality, noise reduction techniques are applied to remove artifacts, and contrast enhancement is performed to improve the visibility of both anatomical and pathological structures. After these enhancements, normalization is performed by scaling pixel intensities to a consistent range, which promotes numerical stability and enables more efficient model training. In the final stage, data augmentation strategies are applied to increase dataset variability and reduce the risk of overfitting. These strategies include vertical and horizontal flipping as well as 90-degree image rotations, all of which improve the generalization capability of the CNN models.

All MRI images are resized to 224 $\times$ 224 pixels to provide a uniform input dimension across the evaluated CNN architectures [40, 41]. Standardizing image dimensions ensures dataset consistency and helps the model to more effectively recognize and learn feature patterns [39, 40]. In addition, resizing reduces computational complexity and accelerates batch processing during training. By establishing a unified image size, this step not only optimizes model performance but

also simplifies the integration of subsequent preprocessing and analysis pipelines.

Adaptive thresholding is applied to identify relevant brain regions from MRI images based on local intensity characteristics, enabling automated region-focused feature extraction without manual tumor annotations [42]. By directing the model's attention to critical brain structures, this technique enhances the extraction of features essential for accurate classification or segmentation [39, 43]. Prior studies have shown that adaptive masking improves detection sensitivity and reduces the likelihood of misclassification. Consequently, this approach strengthens the model's ability to recognize morphological patterns within brain tissue, leading to more reliable diagnostic outcomes.

Noise reduction is performed using Gaussian blur to suppress artifacts that can interfere with image analysis [44, 45]. Gaussian blur is implemented with a 5×5 kernel with $\sigma = 1.0$. These parameters are selected empirically during preliminary experiments to balance noise suppression and preservation of anatomical boundaries. Smaller kernels (e.g., 3×3) produce insufficient smoothing, whereas larger kernels (7×7) lead to blurring of anatomical boundaries [46, 47]. The chosen parameters offer an optimal balance between artifact suppression and retention of tumor edge clarity, thereby improving model performance and stability during feature extraction. The presence of noise may obscure important structural details and negatively affect the predictive accuracy of the model [48, 49]. By mitigating these disturbances, Gaussian blur enhances overall image quality and ensures more consistent feature extraction. This process also improves the model's robustness by reducing sensitivity to unwanted variations in the input data, thereby supporting more reliable classification outcomes.

Contrast enhancement is performed using the CLAHE technique to improve local contrast and sharpen the visualization of anatomical structures.CLAHE is applied to enhance local contrast and improve visualization of anatomical structures [50, 51]. This enhancement facilitates both segmentation and classification tasks by making subtle differences between tissue regions more distinguishable [52, 53]. As a result, the model receives inputs with richer, more informative features, thereby increasing the potential for higher predictive precision.

Pixel intensity normalization is applied by scaling values to the range [0,1] to standardize the input scale and improve numerical stability during model training [54]. This adjustment improves numerical stability, reduces the risk of vanishing or exploding gradients, and accelerates convergence [55, 56]. Normalization also improves compatibility with activation functions commonly used in deep learning architectures. As a result, this step is a fundamental component of medical image preprocessing pipelines, supporting both the efficiency and reliability of subsequent model training.

Data augmentation is applied using a range of geometric and photometric transformations to artificially increase the diversity of the dataset [57, 58]. By introducing controlled variations, this technique enhances the model's generalization and reduces the risk of overfitting. Exposure to augmented data helps the model to learn more robust and adaptive feature representations that are better suited to real-world scenarios. Overall, augmentation is a critical step in improving the performance and reliability of deep learning models when applied to heterogeneous test data.

The dataset is split 70/30 for training and testing, consistent with previous studies that have identified this ratio as effective for balancing model learning and evaluation efficiency [59]. To prevent data leakage, in which images from the same volume or subject can appear in both subsets, a per-volume or per-subject split is used. This approach ensures that all images from a single case are confined to either the training or the testing set [60, 61]. Proper data partitioning is therefore essential to minimize overfitting and enhance the model's ability to generalize to unseen data.

Although the dataset did not provide tumor segmentation masks, adaptive masking is applied as an unsupervised localization technique to isolate brain regions of interest. It is achieved using adaptive thresholding based on local pixel intensity distribution, followed by Gaussian smoothing to refine the mask boundaries. This algorithm does not require manual annotations and enables region-focused feature extraction, mimicking the effect of tumor masks while maintaining automation.

C. Classification

The classification and model training stages are carried out in a technical computing environment using Visual Studio Code with Python as the programming language. In this stage, five selected CNN architectures are implemented, trained, and evaluated through a transfer learning approach. Each model is tested under two dataset scenarios, namely balanced and imbalanced distributions, to assess robustness across different conditions. Model performance is evaluated using a confusion matrix, which provides a detailed representation of classification outcomes, including true positives, false positives, true negatives, and false negatives. A comprehensive analysis of the five CNN

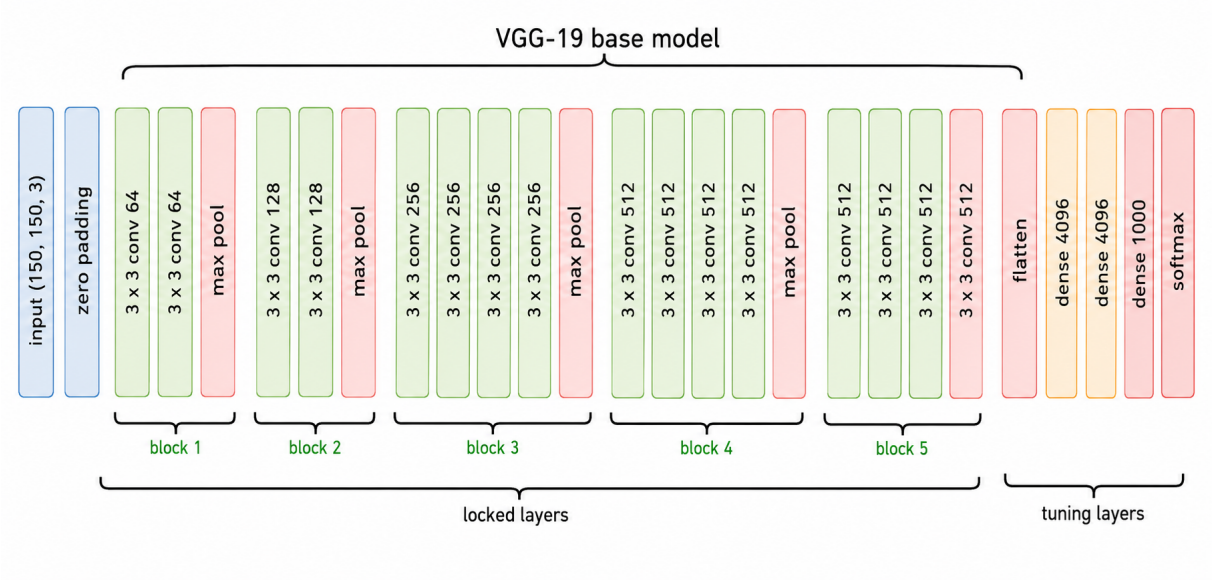


Fig. 3. Architecture of VGG19.

architectures is then conducted to determine their relative effectiveness and identify the most reliable approach for brain tumor classification.

The VGG19 architecture in Fig. 3 is selected as one of the core models because of its ability to learn hierarchical visual representations through successive convolutional layers. Recent research has further demonstrated the applicability of VGG19-based feature extraction to brain tumor MRI classification. Haque et al. [62] have developed NeuroNet19 using VGG19 as its backbone together with an Inverted Pyramid Pooling Module (iPPM), achieving an accuracy of 99.3% for four-class brain tumor classification. These findings support the suitability of VGG19-derived representations for extracting discriminative features from brain MRI images. The size of the resulting feature map in each convolutional operation is determined by Eq. (1):

$$o = \frac{W - F + 2P}{S} + 1, \tag{1}$$

where W represents the input size, F the filter size, P the padding, and S the stride. To optimize its application in the research, a transfer learning approach is adopted by utilizing the pre-trained VGG19 model on ImageNet as a feature extractor. The original fully connected layers are replaced with a newly designed classifier tailored specifically for binary brain tumor classification. Fine-tuning is then performed on MRI images, enabling the model to adapt previously learned features to medical imaging tasks. This approach allows VGG19 to achieve high accuracy even with limited data availability.

The ResNet50 in Fig. 4 is employed as one of the primary architectures due to its strong performance in complex image classification tasks. The ResNet50 architecture overcomes degradation in very deep networks through shortcut connections that learn residual mappings. A residual block is expressed as Eq. (2) [63]:

$$y = F(x, \{W_i\}) + x, \tag{2}$$

where x represents the input, $F(x, \{W_i\})$ denotes the learned residual mapping, and y is the output. This formulation enables the network to approximate identity functions when necessary, prevents accuracy degradation in deeper networks, and mitigates vanishing gradient problems by facilitating smoother gradient propagation.

A ResNet50 architecture comprising 50 residual layers is implemented using a transfer learning approach. The model is initialized with pre-trained weights from ImageNet, which serves as a feature extractor. Its original fully connected classification layer is replaced with a new classifier specifically designed for binary brain tumor classification using MRI images. This strategy leverages the robust feature representations of ResNet-50, enabling high classification accuracy even with limited medical imaging data. DenseNet201 in Fig. 5 improves information flow through dense connectivity, in which each layer receives feature maps from all preceding layers. The output of the l-th layer

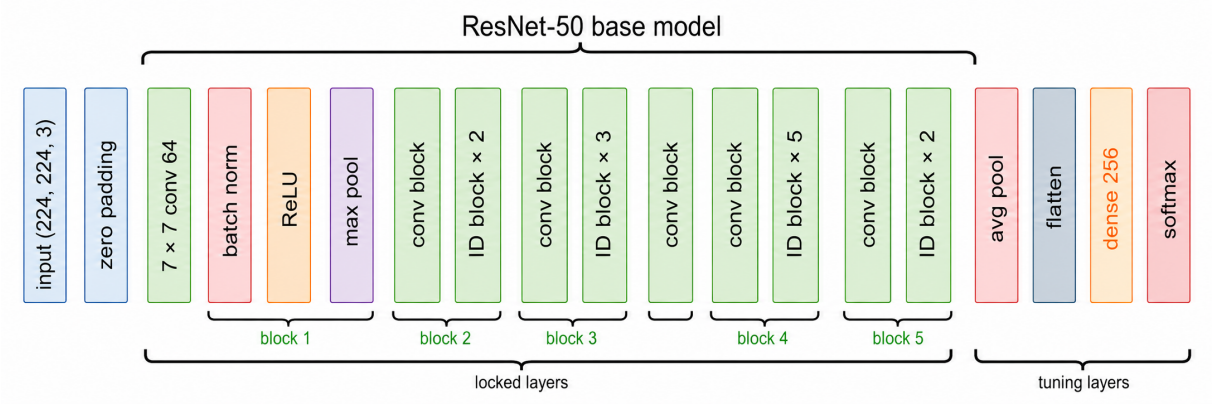


Fig. 4. Architecture of ResNet50.

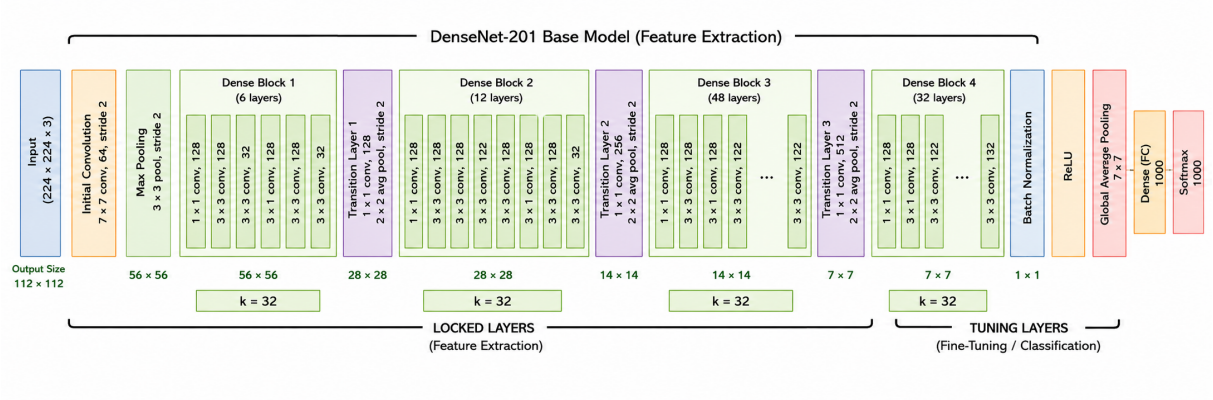


Fig. 5. Architecture of DenseNet201.

is expressed as Eq. (3) [64]:

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]), \tag{3}$$

where $[x_0, x_1, \ldots, x_{l-1}]$ denotes the concatenated feature maps from previous layers, and H_l represents a composite function consisting of batch normalization, Rectified Linear Unit (ReLU) activation, and convolution operations.

The feature of reuse mechanism in DenseNet improves parameter efficiency while preserving important low-level features such as edges and textures, even in deeper layers. This characteristic is particularly beneficial for medical imaging tasks like brain tumor classification, where fine structural details are critical. In the research, the DenseNet201 variant is implemented through transfer learning. The pre-trained model on ImageNet is adapted to MRI data, enabling accurate diagnosis by leveraging its strong representational power while maintaining efficiency in parameter usage.

MobileNet is a deep learning architecture in Fig. 6, specifically designed for computational efficiency, which makes it highly suitable for deployment on mobile devices and resource-limited platforms. Its primary innovation lies in the use of depthwise separable convolutions. This operation decomposes a standard convolution into two stages: a depthwise convolution, where one filter is applied per input channel, followed by a pointwise 1×1 convolution that combines the outputs across channels. By factorizing the convolution process in this manner, MobileNet dramatically reduces computational requirements. The cost ratio can be expressed as Eq. (4) [65]:

$$\frac{D_K \cdot D_K \cdot M + M \cdot N}{D_K \cdot D_K \cdot M \cdot N} = \frac{1}{N} + \frac{1}{D_K^2}, \tag{4}$$

where M and N represent the number of input and output channels, respectively, and D_K denotes the kernel size. For a standard 3×3 kernel, this factorization reduces computation by approximately eight to nine times, with only a minimal loss in accuracy. In the research, MobileNet is evaluated to balance computational efficiency with classification reliability in brain tumor analysis. Its lightweight structure, combined with transfer learning from ImageNet, allows the

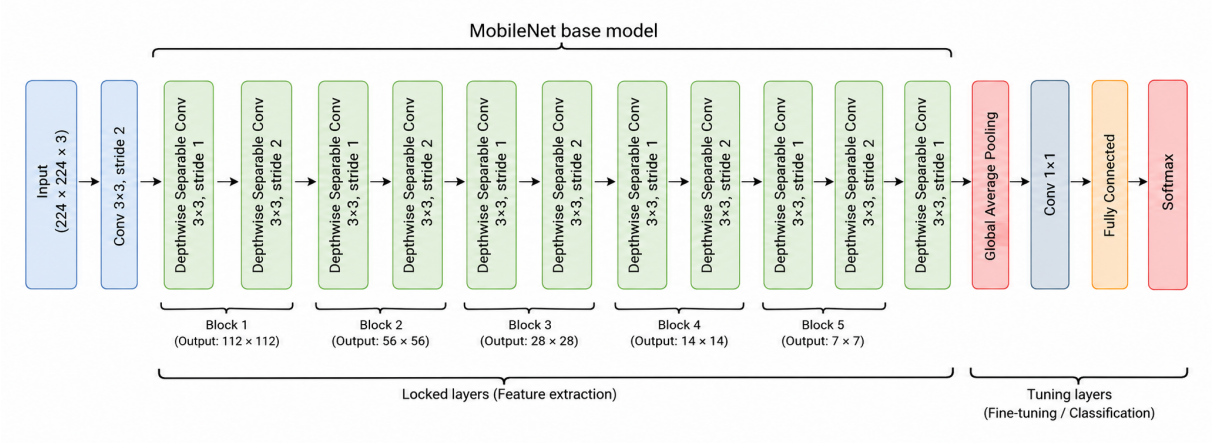


Fig. 6. Architecture MobileNet.

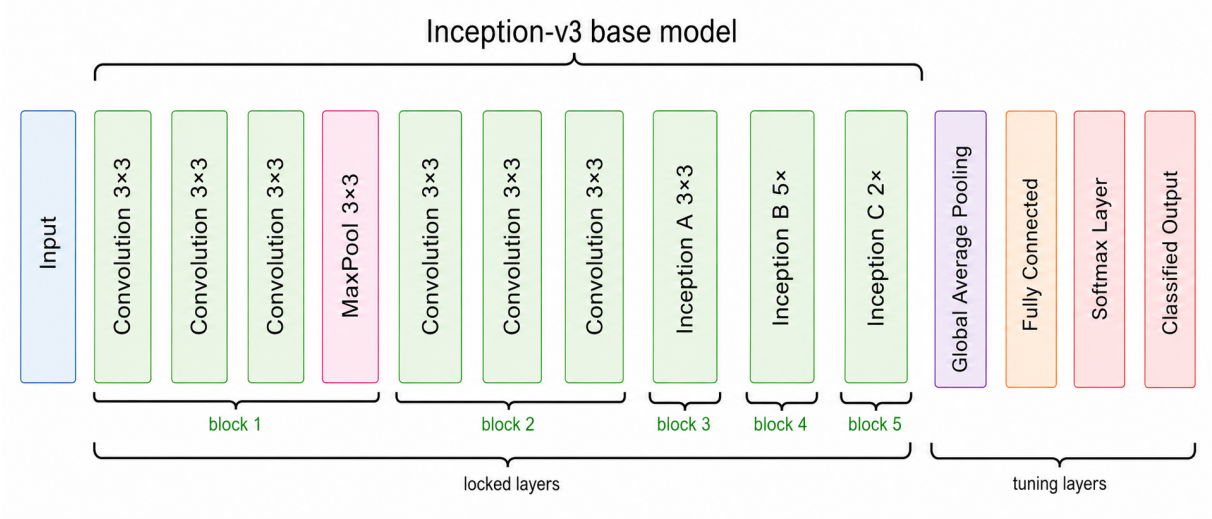


Fig. 7. Architecture of Inception-v3.

model to achieve robust performance while remaining efficient enough for use in environments with limited computational resources.

The Inception-v3 architecture in Fig. 7 combines network depth and width through parallel computational branches that process input features using convolutional and pooling operations with different receptive fields. This multi-branch design enables the network to capture complementary local and global image characteristics. Recent studies have demonstrated the effectiveness of Inception-v3 for MRI-based brain tumor classification, including evaluations under imbalanced dataset conditions [25]. In general, the output of an Inception module can be represented as the concatenation of feature maps generated by its parallel branches, as expressed in Eq. (5).

$$y = \text{Concat}\left(f_{1\times1}(x), f_{3\times3}(x), f_{5\times5}(x), f_{\text{pool}}(x)\right), \tag{5}$$

where $f_{1\times1}(x)$, $f_{3\times3}(x)$, and $f_{5\times5}(x)$ represent convolution operations with kernel sizes of 1×1, 3×3, and 5×5, respectively, and $f_{\text{pool}}(x)$ denotes the pooling operation. The concatenation step combines these feature maps along the channel dimension to form the final output.

This multi-scale representation is particularly advantageous in brain tumor analysis, where tumors may vary significantly in size, shape, and texture. By integrating information obtained from multiple receptive fields, the architecture can represent brain tumor characteristics occurring at different spatial scales. This property is particularly relevant to MRI-based brain

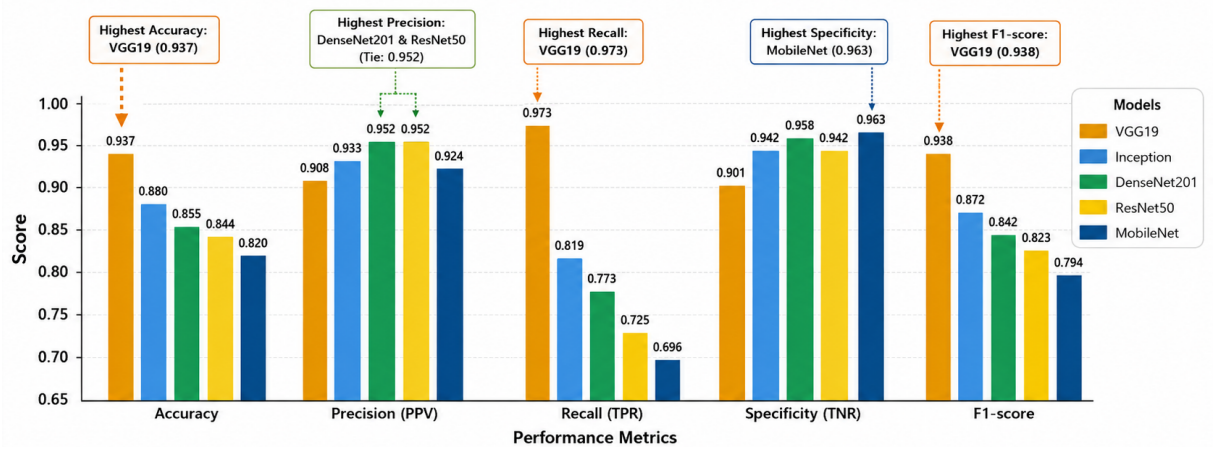


Fig. 8. Comparison of models across performance metrics (balanced data). Note: Positive Predictive Value (PPV), True Positive Rate (TPR), and True Negative Rate (TNR).

tumor classification, where lesion size, morphology, and texture can vary considerably [66]. During transfer learning, the convolutional base layers of each pretrained model (VGG19, ResNet50, DenseNet201, Inception-v3, MobileNet) are initially frozen to preserve low-level visual features such as edges and textures. Subsequently, the fully connected (dense) layers are replaced with new classifiers adapted for binary classification. Fine-tuning is then performed on the last 2–4 convolutional blocks, allowing the models to adjust high-level representations specific to brain MRI features. This selective unfreezing strategy balances computational efficiency with improved domain adaptation.

III. Results and Discussion

This section evaluates five CNN architectures, namely VGG19, Inception-v3, DenseNet201, ResNet50, and MobileNet, for brain tumor classification using MRI images. The performance of these models is measured using multiple evaluation metrics, including accuracy, F1-score, recall, True Positive Rate (TPR), and True Negative Rate (TNR). These metrics are selected to provide not only an overview of classification accuracy but also a detailed assessment of each model's ability to correctly identify both positive and negative cases, which are particularly critical in medical diagnosis. To facilitate comparison, bar charts are generated to illustrate performance differences across the five architectures. In addition, detailed analyses of the metrics are conducted to highlight the strengths and weaknesses of each model. Then, the research results are further compared with findings from prior research, allowing validation of the outcomes and providing a comprehensive interpretation of the models' effectiveness in brain tumor classification.

A. Balanced Data

Figure 8 presents a comparison of five deep learning models (VGG19, Inception-v3, DenseNet201, ResNet50, and MobileNet) evaluated on balanced brain MRI data using five key metrics: accuracy, precision (Positive Predictive Value (PPV)), recall (TPR), specificity (TNR), and F1-score. Among the models, VGG19 demonstrates the best overall performance, achieving the highest accuracy (0.937), recall (0.991), and F1-score (0.939). These results indicate that VGG19 offers strong sensitivity and balanced classification. DenseNet201 and ResNet50 achieve the highest precision (0.952), reflecting their strength in correctly identifying positive cases. Inception-v3, on the other hand, achieves the highest specificity (0.991), indicating its reliability in recognizing negative cases. Although VGG19 records slightly lower precision and specificity than these models, its superior recall and F1-score establish it as the most reliable architecture for brain tumor classification within the research scope.

Overall, the comparison shows that each CNN architecture has distinct strengths across performance metrics. VGG19 demonstrates superior accuracy, recall, and F1-score, making it highly suitable for tasks that require strong sensitivity and balanced classification. DenseNet201 and ResNet50 achieve the highest precision and specificity, which are particularly valuable for minimizing false positives during confirmatory diagnoses. Inception-v3 and MobileNet produce moderate yet consistent results, reflecting an effective balance between computational efficiency and predictive performance. These findings indicate that the

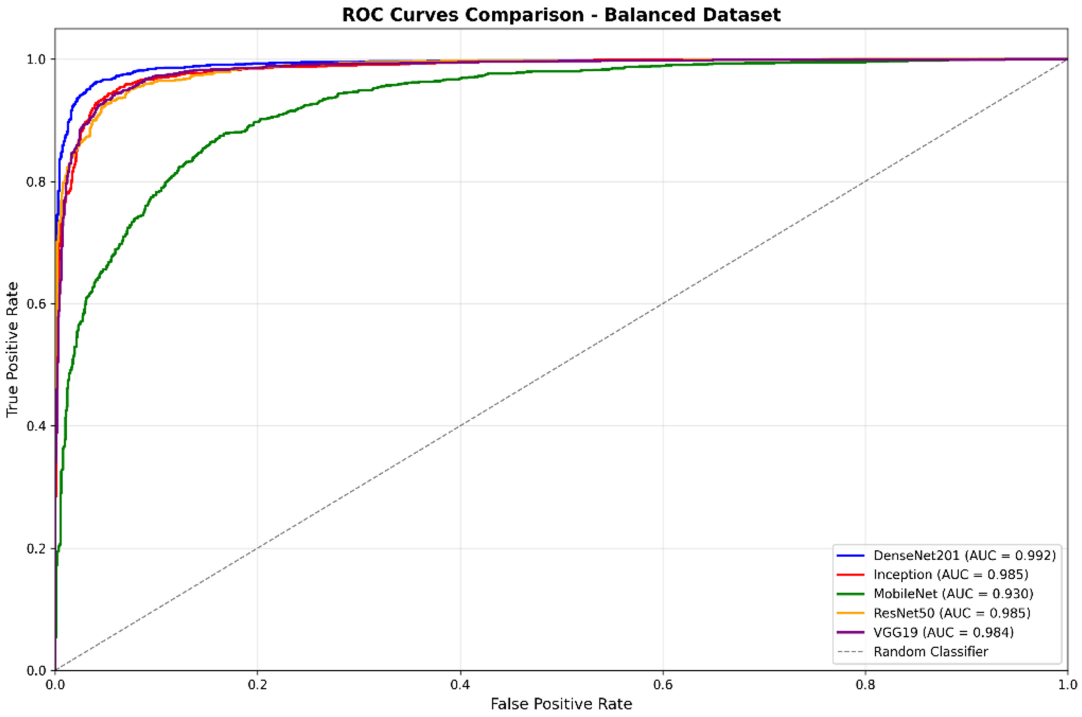


Fig. 9. Receiver Operating Characteristic (ROC) curves comparison (balanced dataset). Note: Area Under the Curve (AUC).

choice of model should align with specific clinical objectives. For example, sensitivity-oriented models such as VGG19 may be preferable for early screening, whereas precision-oriented models like DenseNet201 or ResNet50 may be more appropriate for confirmatory diagnostic settings. To strengthen the applicability of these models in real-world practice, further validation on larger and more diverse datasets is recommended to ensure robustness and generalizability. Figure 9 shows the Receiver Operating Characteristic (ROC) curves for all CNN models, with the AUC to evaluate their ability to distinguish tumor from healthy cases across varying thresholds. Among the models, DenseNet201 achieves the highest AUC (0.992), reflecting excellent discriminative power and a well-balanced trade-off between sensitivity and specificity. Inception-v3 and ResNet50 follow closely with strong AUC values of 0.985, while VGG19 obtains a slightly lower score of 0.984, which still demonstrates high classification capability. MobileNet record the lowest AUC (0.930), indicating comparatively weaker global separation between classes.

Taken together, these results show that although VGG19 leads in accuracy, recall, and F1-score. DenseNet201 demonstrates the most comprehensive classification performance across thresholds. This finding underscores DenseNet201's potential value for clinical decision-making, where both sensitivity and specificity are equally critical to reliable diagnostic outcomes. In general, ROC analysis on balanced data confirms that all evaluated CNN models demonstrate strong discriminative performance, as indicated by their high AUC values. Among the models, DenseNet201 achieves the best overall separation, underscoring its potential for robust clinical application in scenarios where both sensitivity and specificity are equally critical. The consistently high AUC scores across all architectures further validate the effectiveness of deep learning approaches for brain tumor classification using MRI images. Nevertheless, the selection of the most appropriate model should ultimately be guided by specific clinical priorities and operational constraints in real-world healthcare settings.

The comparative analysis of balanced and imbalanced datasets highlights the distinct strengths of each evaluated model. VGG19 achieves the highest accuracy (0.937 on balanced data and 0.962 on imbalanced data) and the highest F1-score (0.939 and 0.921), confirming its robust generalization and balanced predictive performance. DenseNet201 excels in AUC (0.992 and 0.996) and specificity, demonstrating its effectiveness at minimizing false positives, which is particularly valuable in confirmatory diagnostic settings. Inception-v3 and ResNet50 deliver strong precision and high

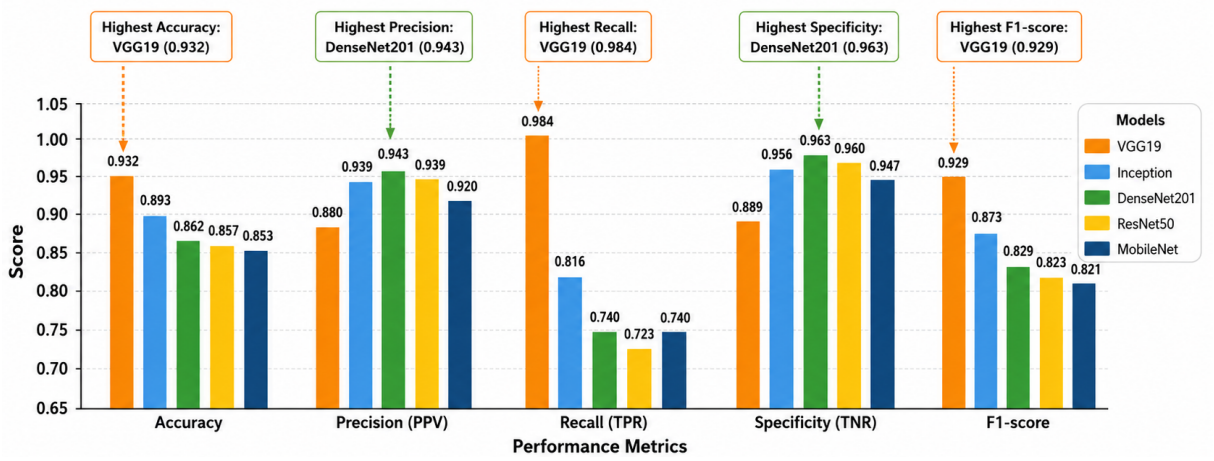


Fig. 10. Comparison of models across performance metrics (imbalanced data). Note: Positive Predictive Value (PPV), True Positive Rate (TPR), and True Negative Rate (TNR).

AUC values, indicating their suitability for applications where positive predictive value is prioritized. By contrast, MobileNet shows relatively lower performance, consistent with its lightweight, efficiency-oriented design. From a clinical perspective, models with high sensitivity, such as VGG19, are well-suited for early tumor screening, where minimizing false negatives is critical. DenseNet201, with its superior specificity, is better suited for confirmatory diagnoses that require precise exclusion of non-tumor cases. Despite these promising results, further external validation using larger, multicenter datasets is necessary to ensure robustness and generalizability. In addition, incorporating interpretability mechanisms will be essential to build clinical trust and support the safe deployment of these models in real-world medical settings.

B. Imbalanced Data

Figure 10 compares five deep learning models (VGG19, Inception-v3, DenseNet201, ResNet50, and MobileNet) evaluated on imbalanced brain MRI data using five performance metrics. VGG19 achieves the highest accuracy (0.962), recall (0.984), and F1-score (0.929), confirming its strong ability to detect true positives. However, it comes at the cost of lower precision (0.800) and specificity (0.849), indicating a higher false-positive rate. In contrast, Inception-v3 and DenseNet201 achieve the best precision values (0.943 each) and demonstrate strong specificity (0.969 and 0.960, respectively). ResNet50 and MobileNet produce moderate results across most metrics, reflecting balanced but less outstanding performance.

These findings highlight VGG19's superior generalization and high sensitivity, making it the most reliable model for diagnostic tasks in imbalanced settings. In scenarios where minimizing missed tumor cases is more critical than reducing false positives, VGG19 offers the greatest clinical utility. Overall, the comparison shows that each CNN architecture trained on imbalanced data exhibits distinct strengths across performance metrics. VGG19 stood out with the highest accuracy, recall, and F1-score, confirming its strong sensitivity and balanced predictive performance. DenseNet201 and ResNet50 perform best in precision and specificity, which are particularly important for minimizing false positives in confirmatory diagnostic tasks. Inception-v3 and MobileNet achieve moderate yet consistent results, reflecting an effective trade-off between model complexity and computational efficiency.

These findings suggest that model selection should be guided by specific clinical objectives. For early screening, where sensitivity is critical to avoid missed tumor cases, VGG19 may be the most suitable. Conversely, for confirmatory diagnosis, DenseNet201 or ResNet50 may be preferable due to their higher precision and specificity. Nevertheless, further validation on larger and more diverse datasets is recommended to ensure robustness and generalizability in real-world clinical applications. Figure 11 presents the ROC curves for all CNN models evaluated on the imbalanced dataset, illustrating their ability to distinguish tumor cases from normal tissue under skewed class distributions. DenseNet201 achieves the highest AUC (0.996), indicating outstanding discriminative power and a well-balanced trade-off between sensitivity and specificity. VGG19 (0.989), ResNet50 (0.986), and Inception-v3 (0.984) also demonstrate strong, reliable performance. Meanwhile, MobileNet obtains the low-

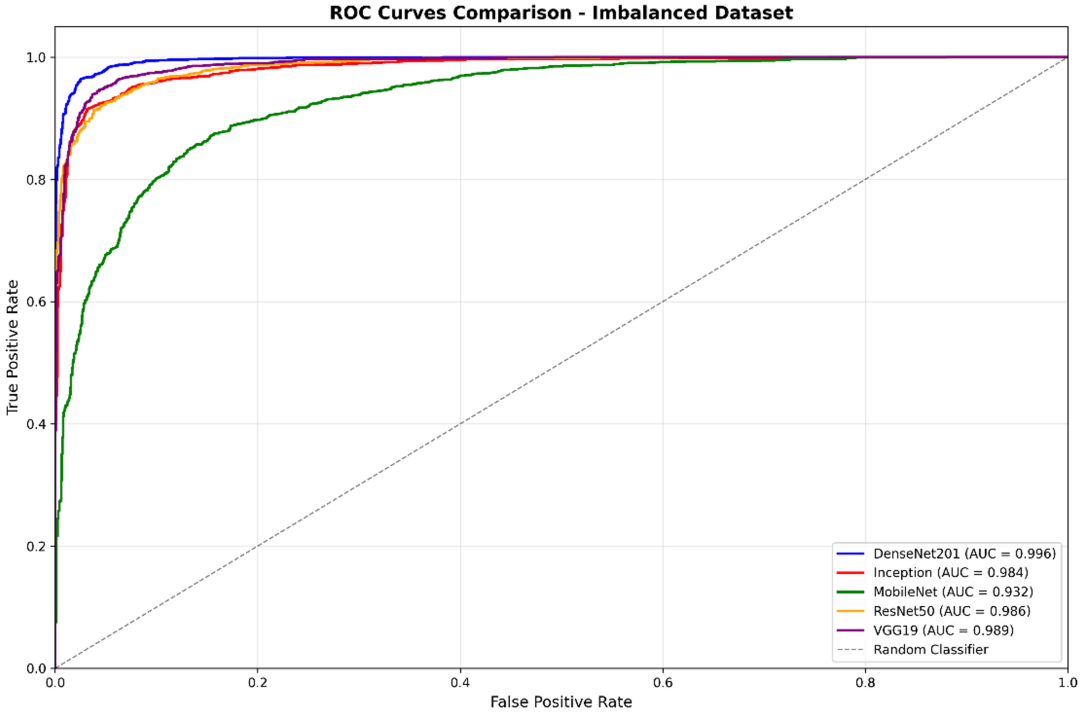


Fig. 11. Receiver Operating Characteristic (ROC) curves comparison (imbalanced dataset). Note: Area Under the Curve (AUC).

est AUC (0.932), reflecting its reduced effectiveness in imbalanced conditions.

These results highlight the clinical potential of DenseNet201, as its exceptionally high AUC provides robust diagnostic confidence by ensuring consistent detection of true positives while minimizing false positives. Such performance is particularly valuable in real-world medical applications, where imbalanced datasets are common and accurate differentiation between tumor and non-tumor cases is critical. In general, ROC analysis of the imbalanced dataset confirms that all evaluated CNN models maintain strong discriminative performance, as reflected by their high AUC values. Among them, DenseNet201 demonstrates the best overall separation capability, underscoring its potential for robust clinical application in scenarios where both sensitivity and specificity are equally important. The consistently high AUC scores across models further validate the effectiveness of deep learning methods for brain tumor classification using MRI images. Nonetheless, the final selection of a model should be guided by specific clinical priorities and operational constraints to ensure optimal implementation in real-world practice.

The comparative evaluation of deep learning models on balanced and imbalanced brain MRI datasets provides important insights into robustness and clinical applicability. VGG19 consistently achieves the highest accuracy, recall, and F1-score, with recall remaining exceptionally high in both scenarios (0.991 on balanced data and 0.984 on imbalanced data). It demonstrates strong sensitivity and resilience, making VGG19 particularly suitable for early-stage screening, where minimizing false negatives is critical. Although its precision decreases under imbalanced conditions (0.800), VGG19 still maintains the highest F1-score (0.929), reflecting an effective balance between sensitivity and specificity.

Inception-v3 excels in specificity (0.991 on balanced data and 0.969 on imbalanced data) and precision, supporting its use in confirmatory diagnosis, where minimizing false positives is essential. DenseNet201 and ResNet50 also deliver high precision but exhibit lower recall under imbalanced conditions, raising concerns about missed tumor cases. MobileNet, while optimized for computational efficiency, shows the weakest recall and F1-scores, limiting its reliability for clinical use. Overall, these findings underscore the importance of evaluating models under both balanced and real-world imbalanced conditions. VGG19's consistent robustness across both scenarios highlights its strong clinical potential, particularly for sensitive diagnostic tasks such as brain tumor detection.

In the imbalanced dataset, the ROC curves in Fig. 11 reveal how each model maintains sensitivity across

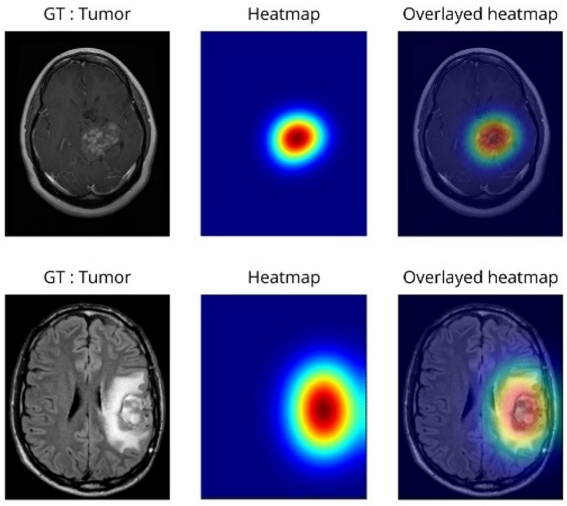


Fig. 12. Visualization results of Grad-CAM obtained from the dataset. Note: Ground Truth (GT).

varying thresholds despite class skew. The steep rise in the initial portion of the curves for DenseNet201 and VGG19 indicates strong detection of positive (tumor) cases with minimal false negatives. DenseNet201's AUC of 0.996 demonstrates superior equilibrium between TPR and FPR, while MobileNet's lower curve reflects its limited ability to separate classes under imbalance. This analysis confirms that DenseNet201's robustness is less affected by uneven data distribution compared to other architectures.

C. Grad-CAM Visualization

To complement the quantitative evaluation, visualization of the model's decision-making process is conducted using heatmap-based localization on MRI scans. Figure 12 illustrates the ground truth tumor region, the generated heatmap, and the overlaid heatmap on the original MRI. As shown, the highlighted regions produced by the model correspond closely with the actual tumor location, suggesting that the CNN is not only achieving high accuracy but also attending to clinically meaningful features.

This interpretability aspect is crucial in medical imaging. It enhances the trust of clinicians in automated systems by demonstrating alignment between algorithmic predictions and radiological findings. Furthermore, the visualization highlights the potential of such models to support early detection, treatment planning, and monitoring, thereby strengthening their clinical applicability alongside traditional diagnostic pathways.

IV. Conclusion

The research conducts a comparative evaluation of five CNN architectures (VGG19, DenseNet201, ResNet50, Inception-v3, and MobileNet) for brain tumor classification using MRI images under both balanced and imbalanced dataset conditions. The findings reveal that each model exhibits distinct strengths depending on the evaluation metrics and dataset distribution. VGG19 consistently achieves the highest accuracy, recall, and F1-score, demonstrating strong sensitivity and robust generalization across both balanced and imbalanced datasets. This performance highlights its suitability for early-stage tumor screening, where minimizing false negatives is critical. DenseNet201 achieves the highest AUC values (0.992 on balanced and 0.996 on imbalanced data) and strong specificity, underscoring its potential for confirmatory diagnostic tasks that prioritize minimizing false positives. Inception-v3 and ResNet50 also demonstrate competitive results, particularly in precision and specificity, making them valuable in clinical scenarios that require high positive predictive value. MobileNet, while computationally efficient, exhibits weaker recall and F1-scores, suggesting that its role may be better suited to low-resource applications rather than primary diagnostic use.

Overall, the results confirm the clinical potential of CNN-based deep learning for brain tumor classification. VGG19 emerges as the most reliable model for sensitive diagnostic applications, while DenseNet201 provides the strongest overall discriminative power, as indicated by its superior AUC. These findings underscore the importance of aligning model selection with specific clinical objectives, such as prioritizing sensitivity for screening or specificity for confirmatory diagnosis.

Although the results are promising, several avenues remain for future research. First, validation on larger, multi-center, and more diverse datasets is essential to ensure robustness and generalizability across populations and imaging protocols. Second, incorporating multi-class classification to distinguish among tumor subtypes and grades will yield more clinically meaningful insights beyond binary classification. Third, integrating model interpretability mechanisms, such as Grad-CAM or SHAP, is recommended to enhance transparency, facilitate clinical trust, and support decision-making. Finally, future studies should explore deployment strategies in real-world healthcare environments, including low-resource settings, where lightweight architectures or model compression techniques may play a critical role.

ACKNOWLEDGEMENT

This research was funded by Institut Teknologi Sepuluh Nopember (ITS) and managed under the Strategic Research Grant (SRG) Type D Scheme (Contract No. 1410/PKS/ITS/2026). The authors also gratefully acknowledged financial support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2025.

AUTHOR CONTRIBUTION

Conceived and designed the analysis, T. A. F., B. R., A. R. N., M. N. A. U., and Y. P.; Collected the data, T. A. F., B. R., M. J. S., M. Q. Z. Z., M. N. A. U., and U. H.; Contributed data or analysis tools, B. R., A. R. N., M. Q. Z. Z., U. H., and Y. P.; Performed the analysis, M. J. S., M. Q. Z. Z., Y. T., U. H., and Y. P.; and Wrote the paper, T. A. F., M. J. S., A. R. N., Y. T., M. N. A. U., and Y. P.

DATA AVAILABILITY

The data that support the findings of the research are openly available in Kaggle repository at https://www.kaggle.com/code/boneacrabonjac/brain-tumor-classification-with-simple-cnn.

REFERENCES

[1] B. V. S. Rao, R. Kondaveti, R. V. V. S. V. Prasad, V. S. Rao, K. B. S. Sastry, and B. Dasaradharam, "Classification of brain tumours using artificial neural networks," *ACTA IMEKO*, vol. 11, no. 1, pp. 1–7, 2022.

[2] K. Noll, A. L. King, L. Dirven, T. S. Armstrong, M. J. B. Taphoorn, and J. S. Wefel, "Neurocognition and health-related quality of life among patients with brain tumors," *Hematology/Oncology Clinics*, vol. 36, no. 1, pp. 269–282, 2022.

[3] D. P. R. Chieffo, F. Lino, D. Ferrarese, D. Belella, G. M. Della Pepa, and F. Doglietto, "Brain tumor at diagnosis: From cognition and behavior to quality of life," *Diagnostics*, vol. 13, no. 3, pp. 1–12, 2023.

[4] S. M. Vijithananda *et al.*, "Feature extraction from mri adc images for brain tumor classification using machine learning techniques," *Biomedical Engineering Online*, vol. 21, pp. 1–21, 2022.

[5] S. Lin, K. Li, and L. Qi, "Cancer stem cells in brain tumors: From origin to clinical implications," *MedComm*, vol. 4, no. 4, pp. 1–27, 2023.

[6] J. Latowska, A. Grabowska, Z. Zarebska, K. Kuczynski, B. Kuczynska, and K. Rolle, "Non-coding RNAs in brain tumors, the contribution of lncRNAs, circRNAs, and snoRNAs to cancer development–Their diagnostic and therapeutic potential," *International Journal of Molecular Sciences*, vol. 21, no. 19, pp. 1–30, 2020.

[7] F. Yamasaki, "Adolescent and young adult brain tumors: Current topics and review," *International Journal of Clinical Oncology*, vol. 27, pp. 457–464, 2022.

[8] Y. Fan *et al.*, "Burden and trends of brain and central nervous system cancer from 1990 to 2019 at the global, regional, and country levels," *Archives of Public Health*, vol. 80, pp. 1–14, 2022.

[9] T. A. Soomro *et al.*, "Image segmentation for MR brain tumor detection using machine learning: A review," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 70–90, 2022.

[10] S. Asif, W. Yi, Q. U. Ain, J. Hou, T. Yi, and J. Si, "Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from MR images," *IEEE Access*, vol. 10, pp. 34 716–34 730, 2022.

[11] M. S. I. Khan *et al.*, "Accurate brain tumor detection using deep convolutional neural network," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 4733–4745, 2022.

[12] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, and M. O. Alassafi, "Brain tumor classification based on fine-tuned models and the ensemble method," *Computers, Materials and Continua*, vol. 67, no. 3, pp. 3967–3982, 2021.

[13] A. M. D. Simo, A. T. Kouanou, V. Monthe, M. K. Nana, and B. M. Lonla, "Introducing a deep learning method for brain tumor classification using MRI data towards better performance," *Informatics in Medicine Unlocked*, vol. 44, pp. 1–24, 2024.

[14] H. Kibriya, R. Amin, J. Kim, M. Nawaz, and R. Gantassi, "A novel approach for brain tumor classification using an ensemble of deep and hand-crafted features," *Sensors*, vol. 23, no. 10, pp. 1–16, 2023.

[15] M. Agarwal, G. Rani, A. Kumar, P. Kumar, R. Manikandan, and A. H. Gandomi, "Deep learning for enhanced brain tumor detection and classification," *Results in Engineering*, vol. 22, pp. 1–12, 2024.

[16] L. R. Carrete, J. S. Young, and S. Cha, "Advanced imaging techniques for newly diagnosed and recurrent gliomas," *Frontiers in Neuroscience*, vol. 16, pp. 1–22, 2022.

[17] S. Gassenmaier *et al.*, "Deep learning applications in magnetic resonance imaging: Has the future become present?" *Diagnostics*, vol. 11, no. 12, pp. 1–11, 2021.

[18] M. I. Mahmud, M. Mamun, and A. Abdelgawad, "A deep analysis of brain tumor detection from MR images using deep learning networks," *Algorithms*, vol. 16, no. 4, pp. 1–19, 2023.

[19] M. W. Nadeem, M. A. A. Ghamdi, M. Hussain, M. A. Khan, K. M. Khan, S. H. Almotiri, and S. A. Butt, "Brain tumor analysis empowered with deep learning: A review, taxonomy, and future challenges," *Brain Sciences*, vol. 10, no. 2, pp. 1–33, 2020.

[20] M. Aamir *et al.*, "A deep learning approach for brain tumor classification using MRI images," *Computers and Electrical Engineering*, vol. 101, 2022.

[21] D. A. Wood *et al.*, "Deep learning models for triaging hospital head MRI examinations," *Medical Image Analysis*, vol. 78, pp. 1–22, 2022.

[22] R. Zaitoon and H. Syed, "Enhanced brain tumor detection and classification in MRI scans using convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 266–275, 2023.

[23] E. Albalawi *et al.*, "Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach," *Frontiers in Computational Neuroscience*, vol. 18, pp. 1–19, 2024.

[24] S. Deepak and P. M. Ameer, "Brain tumor categorization from imbalanced MRI dataset using weighted loss and deep feature fusion," *Neurocomputing*, vol. 520, pp. 94–102, 2023.

[25] T. Agrawal, P. Choudhary, V. Kumar, P. Singh, M. Diwakar, and S. Kumar, "A comparative study of brain tumor classification on unbalanced dataset using deep neural networks," *Biomedical Signal Processing and Control*, vol. 94, 2024.

[26] K. Gasmi *et al.*, "Enhanced brain tumor diagnosis using combined deep learning models and weight selection technique," *Frontiers in Neuroinformatics*, vol. 18, pp. 1–14, 2024.

[27] P. M. S. Raja and A. V. Rani, "Brain tumor classification using a hybrid deep autoencoder with bayesian fuzzy clustering-based segmentation approach," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 440–453, 2020.

[28] S. Li, J. Liu, and Z. Song, "Brain tumor segmentation based on region of interest-aided localization and segmentation U-Net," *International Journal of Machine Learning and Cybernetics*, vol. 13, pp. 2435–2445, 2022.

[29] A. Kumar, "Study and analysis of different segmentation methods for brain tumor MRI application," *Multimedia Tools and Applications*, vol. 82, no. 5, pp. 7117–7139, 2023.

[30] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Medical Informatics and Decision Making*, vol. 23, pp. 1–17, 2023.

[31] D. Lamrani, B. Cherradi, O. El Gannour, M. A. Bouqentar, and L. Bahatti, "Brain tumor detection using MRI images and convolutional neural network," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, pp. 452–460, 2022.

[32] A. Younis, L. Qiang, C. O. Nyatega, M. J. Adamu, and H. B. Kawuwa, "Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches," *Applied Sciences*, vol. 12, no. 14, pp. 1–20, 2022.

[33] F. J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, and D. González-Ortega, "A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network," *Healthcare*, vol. 9, no. 2, pp. 1–14, 2021.

[34] M. Masood *et al.*, "A novel deep learning method for recognition and classification of brain tumors from MRI images," *Diagnostics*, vol. 11, no. 5, pp. 1–18, 2021.

[35] A. Çinar and M. Yildirim, "Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture," *Medical Hypotheses*, vol. 139, 2020.

[36] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A deep learning model based on concatenation approach for the diagnosis of brain tumor," *IEEE Access*, vol. 8, pp. 55 135–55 144, 2020.

[37] S. A. A. Ismael, A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks," *Artificial Intelligence in Medicine*, vol. 102, 2020.

[38] Kosta, "Brain tumor classification with simple CNN," 2021. [Online]. Available: https://www.kaggle.com/code/boneacrabonjac/brain-tumor-classification-with-simple-cnn

[39] J. Bleker *et al.*, "A deep learning masked segmentation alternative to manual segmentation in biparametric MRI prostate cancer radiomics," *European radiology*, vol. 32, pp. 6526–6535, 2022.

[40] A. M. Saleh and T. Hamoud, "Analysis and best parameters selection for person recognition based on gait model using CNN algorithm and image

augmentation," *Journal of Big Data*, vol. 8, pp. 1–20, 2021.

[41] X. Shen *et al.*, "Spatially-consistent feature matching and learning for heritage image analysis," *International Journal of Computer Vision*, vol. 130, pp. 1325–1339, 2022.

[42] P. Singh, "A type-2 neutrosophic-entropy-fusion based multiple thresholding method for the brain tumor tissue structures segmentation," *Applied Soft Computing*, vol. 103, 2021.

[43] V. González-Ruiz, M. R. Fernández-Fernández, and J. J. Fernández, "Structure-preserving gaussian denoising of FIB-SEM volumes," *Ultramicroscopy*, vol. 246, pp. 1–11, 2023.

[44] D. Aarthi, A. Panimalar, S. S. Kumar, and K. Anitha, "Development of adaptive Gaussian filter based denoising as an image enhancement technique," *Mathematical Modelling of Engineering Problems*, vol. 11, no. 10, pp. 2715–2725, 2024.

[45] S. S. Kolmar and C. M. Grulke, "The effect of noise on the predictive limit of QSAR models," *Journal of Cheminformatics*, vol. 13, pp. 1–19, 2021.

[46] S. Lim, M. Park, H. Kim, S. H. Kang, K. Kim, and Y. Lee, "Optimization of median modified Wiener filter for improving lung segmentation performance in low-dose computed tomography images," *Applied Sciences*, vol. 13, no. 19, pp. 1–13, 2023.

[47] R. Su, D. Zhang, J. Liu, and C. Cheng, "MSU-Net: Multi-scale U-Net for 2D medical image segmentation," *Frontiers in Genetics*, vol. 12, pp. 1–14, 2021.

[48] B. Brede *et al.*, "Assessment of workflow feature selection on forest LAI prediction with sentinel-2A MSI, landsat 7 ETM+ and Landsat 8 OLI," *Remote Sensing*, vol. 12, no. 6, pp. 1–29, 2020.

[49] I. M. Mohammed and N. A. M. Isa, "Contrast limited adaptive local histogram equalization method for poor contrast image enhancement," *IEEE Access*, vol. 13, pp. 62 600–62 632, 2025.

[50] Y. J. Ng and K. S. Sim, "A review of brain early infarct image contrast enhancement using various histogram equalization techniques," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 6, pp. 1849–1860, 2024.

[51] K. Halloum and H. Ez-Zahraouy, "Enhancing medical image classification through transfer learning and CLAHE optimization," *Current Medical Imaging*, vol. 21, no. 1, pp. 1–14, 2025.

[52] A. Zhang and A. S. Zhang, "Real-time wildfire detection and alerting with a novel machine learning approach," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, pp. 1–6, 2022.

[53] A. Crombé *et al.*, "Intensity harmonization techniques influence radiomics features and radiomics-based predictions in sarcoma patients," *Scientific Reports*, vol. 10, pp. 1–13, 2020.

[54] Z. Li *et al.*, "A systematic survey of regularization and normalization in GANs," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–37, 2023.

[55] P. Gupta, P. Jain, and O. G. Kakde, "Accelerating deep network training for radar identification using batch normalization," *Defence Science Journal*, vol. 74, no. 6, pp. 878–884, 2024.

[56] A. Kairys and V. Raudonis, "Analysis of training data augmentation for diabetic foot ulcer semantic segmentation," *Electronics*, vol. 12, no. 22, pp. 1–14, 2023.

[57] E. K. Kim, H. Lee, J. Y. Kim, and S. Kim, "Data augmentation method by applying color perturbation of inverse PSNR and geometric transformations for object recognition based on deep learning," *Applied Sciences*, vol. 10, no. 11, pp. 1–20, 2020.

[58] V. R. Joseph and A. Vakayil, "SPlit: An optimal method for data splitting," *Technometrics*, vol. 64, no. 2, pp. 166–176, 2022.

[59] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, "Data leakage inflates prediction performance in connectome-based machine learning models," *Nature Communications*, vol. 15, pp. 1–15, 2024.

[60] V. Singh, M. Pencina, A. J. Einstein, J. X. Liang, D. S. Berman, and P. Slomka, "Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging," *Scientific reports*, vol. 11, pp. 1–8, 2021.

[61] Y. Liu, D. Cheng, D. Zhang, S. Xu, and J. Han, "Capsule networks with residual pose routing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 2, pp. 2648–2661, 2024.

[62] R. Haque, M. M. Hassan, A. K. Bairagi, and S. M. Shariful Islam, "NeuroNet19: An explainable deep neural network model for the classification of brain tumors using magnetic resonance imaging data," *Scientific Reports*, vol. 14, pp. 1–22, 2024.

[63] P. K. Pandey, J. Pruthi, S. Alzahrani, A. Verma, and B. Zohra, "Enhancing healthcare recommendation: transfer learning in deep convolutional

neural networks for Alzheimer disease detection," *Frontiers in Medicine*, vol. 11, pp. 1–12, 2024.

[64] K. V. Priya and J. D. Peter, "A federated approach for detecting the chest diseases using DenseNet for multi-label classification," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3121–3129, 2022.

[65] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET image processing*, vol. 16, no. 5, pp. 1243–1267, 2022.

[66] A. M. Alhassan and N. I. Altmami, "IV3TM: Inception V3 enabled bidirectional long short-term memory network for brain tumor classification," *PLOS One*, vol. 20, no. 10, pp. 1–34, 2025.