



Research article

Performance comparison of deep learning models for MRI-based brain tumor detection

Abdulmajeed Alsufyani*

Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

* **Correspondence:** Email: a.s.alsufyani@tu.edu.sa.

Abstract: Brain tumors pose a significant threat to human health, as they can severely affect both physical well-being and quality of life. These tumors often lead to increased intracranial pressure and neurological complications. Traditionally, brain tumors are diagnosed through manual interpretation via medical imaging techniques such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). While these methods are effective, they are time-consuming and subject to errors in consistency and accuracy. This study explored the use of several deep-learning models, including YOLOv8, YOLOv9, Faster R-CNN, and ResNet18, for the detection of brain tumors from MR images. The results demonstrate that YOLOv9 outperforms the other models in terms of accuracy, precision, and recall, highlighting its potential as the most effective deep-learning approach for brain tumor detection.

Keywords: deep learning; brain tumor detection; YOLOv9

1. Introduction

Brain tumors are considered one of the most severe health conditions because they can severely compromise one's health and general quality of life. The infiltration of normal brain tissue with brain tumors leads to increased intracranial pressure and severe neurologic consequences. This is a massive threat to people's well-being and physical safety. Brain cancers are the most common and deadliest primary tumors in people. Moreover, in children who survive cancer, medical treatments such as radiation, chemotherapy, and surgery may have lasting effects on their brains [1]. The biological characteristics of these tumors, which often impede progression, make them challenging to treat.

Central Nervous System (CNS) disorders such as stroke, infection, brain tumors, or migraines pose significant challenge in terms of diagnosis, evaluation, and the development of effective remedies. Compared with other cancers, brain tumors are less common; therefore, pharmaceutical organizations fund and pay less attention to them, challenging their research. The frequency of occurrence of brain and CNS tumors varies, ranging from extremely rare to often diagnosed. They also differ in their potential for injury, ranging from nonmalignant to cancerous [2]. Primary brain tumors and secondary or metastatic brain tumors vary in terms of their origins. The latter spreads to the brain from other body parts, whereas primary brain tumors are neoplasms that develop within the brain itself. Timely identification can significantly increase the likelihood of favorable treatment outcomes, reduce mortality rates, and enhance quality of life. On the other hand, if the condition is not discovered immediately, it often advances to more severe stages, which makes therapy harder and less effective.

Brain tumors are often identified via conventional approaches that rely on the manual interpretation of medical images obtained via CT and MRI. These methods are effective yet time-consuming and may have consistency and accuracy issues. Manual MRI-based brain tumor detection is complicated and prone to errors. Noninvasive MRI techniques produce exact and comprehensive images of various inner body structures, encompassing blood vessels, muscles, bones, and organs. An MRI scanner employs radio waves and a strong magnet to generate human body images [3]. Unlike X-rays, MRI does not produce ionizing radiation. Doctors use these images to diagnose and treat patients. MRI is preferred over CT for brain tumor detection, as it can better distinguish between normal and cancerous brain tissues because of its superior contrast resolution. MRI provides detailed brain tumor size, position, and scope data, which are essential for accurate diagnosis and therapy planning.

Owing to advances in imaging tools and the requirement for fast, accurate, and automated image interpretation, deep learning is taking over medical image processing. Deep learning is the most advanced machine learning method, using multilayer neural networks for representation learning. This system automatically converts incoming data into numerous abstractions to learn data representations. Deep learning has revolutionized medical image processing, offering new ways to detect and analyze diseases such as brain tumors [4]. Convolutional neural networks (CNNs) can identify and classify brain tumors by learning complex MRI patterns. Supervised learning trains CNNs on sizeable medical image collections. CNNs have transformed computer vision and are used in speech recognition and Natural Language Processing (NLP) [5]. CNNs mimic the human eye to quickly process and evaluate visual data such as photos and videos.

Brain tumors present a significant challenge in health care, as delayed or inaccurate detection often leads to severe neurological consequences or reduced survival rates [6]. Early and accurate detection enables timely intervention, facilitates treatment planning, and improves outcomes while also alleviating patient anxiety during the diagnostic process [7]. For example, rapid detection of high-grade glioblastomas allows for early surgical planning and the initiation of targeted therapies, which are critical for slowing disease progression [8]. Conversely, misdiagnosis of benign tumors as malignant—or vice versa—can lead to unnecessary interventions or delays in essential treatments, negatively impacting patient outcomes [9].

Advanced artificial intelligence (AI)-driven diagnostic tools, such as YOLOv9, have the potential to revolutionize clinical workflows by providing consistent and accurate results even in resource-limited settings. For example, AI models could assist in triaging cases during high workloads, identifying urgent conditions that require immediate attention [10]. Such abilities are especially

valuable in emergency neurosurgical scenarios, where rapid and accurate localization of brain tumors is critical for life-saving interventions [11].

Medical image denoising, segmentation, and classification are necessary to ensure precise analysis. Deep Learning has emerged as a powerful method for image analysis, with promising results in lowering noise in various medical imaging methods. The information generated is comprehensive and precise for identifying illnesses via MRI, Positron Emission Tomography (PET), ultrasound, and X-ray even in the presence of noise [12]. These neural networks strive to duplicate their input data, and when trained on noisy medical images, they proficiently learn to produce clearer equivalents [13]. CNNs, transfer learning, and deep residual networks have shown considerable advancements in precisely identifying various skin structures and lesions. In addition to detecting tumors, DL can improve health care, especially in areas where medical imaging is used for diagnosis, prediction, and treatment [14]. Informative and accurate data describing the present symptoms and previous medical background result in a more valuable report for the physicians and an optimal outcome for the patient.

This study aims to make the following key contributions:

1. Comprehensive model comparison: We evaluate the performance of four state-of-the-art deep-learning models (YOLOv8, YOLOv9, Faster R-CNN, and ResNet18) for brain tumor detection, considering metrics such as accuracy, precision, recall, and F1 score.
2. Focus on practicality: This work emphasizes runtime efficiency and computational resource requirements, which are crucial for real-time deployment in clinical settings.
3. Enhanced training strategies: Advanced data augmentation and transfer learning techniques were applied to address data scarcity and improve model robustness across diverse tumor types.
4. Model interpretability: By employing Grad-CAM visualizations, the study ensures that the models' predictions are interpretable, building trust for AI-assisted diagnostics among clinicians.

These contributions advance the application of deep learning in medical imaging, particularly in addressing real-world challenges of clinical deployment.

2. Literature review

Medical Image Analysis has attracted substantial attention from the research community because of its extensive applications in the healthcare sector for evaluating and diagnosing patients. A study [9] reported that machine learning can categorize brain scans and examine the structure of the brain. Deep learning was extensively employed to automate the identification and segmentation of brain tumors. They used several brain tumor images to identify brain cancers in MRI scans. Through transfer learning, the researchers increased the ability of the state-of-the-art YOLOv7 model to detect pituitary brain tumors, meningiomas, and gliomas [9]. They found that deep learning can yield several valuable applications in pattern categorization. ZainEldin et al. [10] explored a more accurate brain tumor diagnosis algorithm in 2023. They discovered that the BCM-CNN classifier performed best. This was due to CNN hyperparameter adjustments, which increased performance. A brain tumor takes time to diagnose; therefore, the radiologist's ability and understanding are vital. The growing patient population has increased the volume of data for evaluation, making outdated methods expensive and inaccurate.

Passa et al. [11] employed the YOLOv8 framework and implemented data augmentation techniques to achieve precise detection of meningioma, glioma, and pituitary tumors. The investigation generated a series of T1-weighted, contrast-enhanced images. Testing, validation, and training were all

performed with the dataset. Techniques for preprocessing and augmentation were used to increase the caliber and volume of the training data. The recommended method used YOLOv8 and data augmentation to improve model accuracy and efficiency. According to their study, this technique enhances treatment planning and early diagnosis, improving patient care for brain tumor identification. YOLOv9, a groundbreaking object identification technology, is noted for its speed and precision. According to [12], all the variants of YOLO-v8 achieved better throughput while maintaining an exact number of parameters when comparing the YOLO-v8, YOLO-v5, and YOLOv6 models trained on 640 images. This implies that hardware-efficient architecture changes have been made to YOLO-v8.

A detailed investigation [13] sought to understand YOLOv9's essential components, benefits, drawbacks, and possible applications in numerous fields. In terms of input data fluctuations, YOLOv9 showed outstanding stability and generalizability. A comparison between YOLOv9, previous algorithms, and other item identification methods revealed that YOLOv9 achieves a favorable balance between speed and accuracy [14]. The authors' method outperformed current methods, with an average classification accuracy of 84.6. YOLOv9's improved architecture simplified and improved complex brain tumor patterns.

Owing to its excellent accuracy levels, Faster R-CNN has become a popular item recognition model. These detectors use a categorization network and a Region Proposal Network (RPN) to deliver likely item suggestions. In subsequent stages, these suggestions are honed and categorized [15]. Two publicly available datasets were shared to investigate the R-CNN approach, with promising results in object detection, with an average accuracy of 98.83%. This paradigm's main advantage is its capacity to function on computers with low system requirements, allowing for real-time processing. Deep-learning algorithms have been successfully applied by [16] for the automatic segmentation of several MR images. Faster R-CNN has proven to be superior in terms of speed and classification accuracy performance among deep-learning approaches, attaining a classification accuracy of 91.66%. This method also returns the object's position along with its class label.

Accurate diagnoses are essential for patients to receive the correct treatments and have a better chance of long-term survival [17]. AI approaches are essential for the use of Computer-Aided Diagnosis (CAD) systems to diagnose medical images, namely MR images [17]. Deep-learning algorithms have been studied, with an explicit focus on AlexNet and ResNet-18. Research has demonstrated that the use of deep neural network learning techniques enhances the accuracy of identifying different medical images and aids clinicians and radiologists in detecting diseases at their initial stages [17]. Multiple ResNet designs, built around deep architectures with varying layers, have been designed and proven to have high performance and accuracy. Scientists have employed DL and ML artificial intelligence methods to generate deep feature maps and classify the extracted features via deep learning. The ResNet-18+SVM model achieved an accuracy of 91.20%, according to the researchers' findings. A summary of all prior investigations is presented in Table 1.

Table 1. Summary of previous studies.

Model evaluated	Accuracy	Key findings	Ref.
YOLOv7	Higher than previous models	Higher accuracy in brain tumor detection.	[9]
BCM-CNN	99.98%	Hyperparameter optimization with adaptive dynamic sine–cosine fitness gray wolf optimizer yielded good accuracy.	[10]
YOLOv8	Improved with augmentation	Improves precision, recall, and mAP50 scores.	[11]
YOLO-v8	92.7%	YOLO-v8 improved while maintaining a similar parameter count compared to YOLO-v5 and YOLO-v6.	[12]
YOLOv9	An effective balance of accuracy and speed	Highlighted YOLOv9’s exceptional generalization ability and stability with variances in input data, balancing accuracy and speed effectively compared to prior YOLO versions and other object-detection algorithms.	[13]
YOLO (v5, v8, v9)	84.6%	Compared to previous YOLO versions and VGG16 for Alzheimer’s disease detection, YOLOv9 had the highest overall accuracy at 84.6%, while VGG16 achieved 99% accuracy for training but only 78% for testing.	[14]
Fast R-CNN	91.2%	Novel RCNN architecture with low complexity for brain tumor classification and detection, achieving high accuracy and short execution time.	[15]
Faster R-CNN	91.66%	Applied Faster R-CNN for automatic segmentation and classification of MR images, achieving high speed and classification accuracy.	[16]
AlexNet, ResNet-18	91.20%	Combining deep learning with classical machine learning, the application of AlexNet and ResNet-18 for feature extraction and SVM for classification yielded accurate and specific brain tumor diagnosis.	[17]

CNNs such as YOLOv8, YOLOv9, Faster R-CNN, and ResNet18 have consistently shown high accuracy and utility in medical image processing, as evidenced by multiple publications [9]. YOLOv7 achieved improved accuracy in detecting pituitary tumors, meningiomas, and gliomas after fine-tuning. Brain tumor image analysis is a difficult task because of the considerable variation in size, form, and location of these tumors. However, various proposed models have been demonstrated to be helpful. In [11], authors reported that data augmentation enhanced YOLOv8 accuracy and efficiency. Training data can be expanded via data augmentation to correct the imbalance between earlier YOLO versions. Owing to its resilience and adaptability, the YOLOv9 model is ideal for fast, precise tumor detection. Avşar and Salçın questioned Faster R-CNN’s trustworthiness in 2019. Their method can be used to diagnose several tumors, making it suitable for automatic segmentation and detection.

ResNet-18, a modified ResNet framework, also works well. Senan et al. [17] reported that ResNet-18 with an SVM could detect brain cancers with high sensitivity, specificity, and accuracy. The ability of ResNet models to extract and categorize deep properties helps radiologists make accurate diagnoses. Many MR images have been effectively segmented via deep-learning techniques. Specifically, CNN-based algorithms for automatic MRI segmentation of brain tumors achieved favorable outcomes by discerning unique characteristics.

Mathivanan et al. [18] explored the use of deep transfer learning models to improve brain tumor diagnosis. They evaluated four architectures—ResNet152, VGG19, DenseNet169, and MobileNetv3—using a Kaggle dataset. Through fivefold cross-validation and image enhancement, MobileNetv3 achieved the highest accuracy at 99.75%, outperforming the other methods. The results highlight the potential of transfer learning in overcoming data scarcity and enhancing the accuracy of medical imaging for brain tumor detection. This approach shows promise for advancing diagnostic methods in medical fields.

Recent advancements in nature-inspired optimization algorithms have demonstrated their potential to improve the performance of deep-learning models in medical imaging. For example, El-Kenawy et al. [19] proposed the greylag goose optimization algorithm, which mimics the migratory behavior of geese to achieve a balance between exploration and exploitation during optimization processes. This method has shown significant improvements in model generalizability and robustness, particularly in tasks requiring efficient hyperparameter tuning. The approach aligns with our efforts to evaluate and optimize models such as YOLOv9 and ResNet18 for MRI-based brain tumor detection, emphasizing the importance of efficient optimization in achieving reliable clinical performance.

Optimization techniques have shown significant promise in improving the adaptability and efficiency of deep-learning models for medical imaging. El-Kenawy et al. [20] presented advanced strategies for optimization, including novel applications of transfer learning and data augmentation. Their work emphasized how these methods address challenges such as data scarcity and model generalization, which are critical in domain-specific tasks such as MRI-based brain tumor detection. This approach aligns closely with our use of data augmentation techniques and transfer learning to enhance the performance of the YOLOv9 and ResNet18 models.

Recent advancements in deep learning optimization have contributed significantly to improving the adaptability and performance of models in medical imaging. El-Kenawy et al. [21] introduced innovative optimization strategies in their study. Their work emphasized leveraging novel algorithms, such as the football optimization algorithm (FbOA), to address challenges in high-dimensional and nonlinear problem spaces. These strategies are particularly valuable in enhancing neural network performance, as they optimize hyperparameter tuning and model robustness in complex tasks, including MRI-based brain tumor detection. The football optimization algorithm, inspired by team dynamics in sports, balances exploration and exploitation during optimization processes, resulting in improved model generalizability. This aligns with our approach in evaluating YOLOv9 and ResNet18, where effective optimization was critical for achieving high accuracy and computational efficiency in brain tumor detection.

DL techniques for brain tumor detection in MR images have advanced, but additional research is needed to address these gaps. Most studies have used fewer samples, limiting model generalizability across populations and imaging circumstances. The diversity in MRI machine settings and patient demographics commonly causes picture quality and tumor appearance inconsistencies, which threatens model robustness. Deep learning model interpretability is another issue. CNNs have great

accuracy, but physicians need help trusting and implementing them since they are hard to interpret. Practical concerns, including real-time processing, technology restrictions, and user-friendly interfaces, must be addressed when these models are integrated into clinical procedures. The high computing costs of these methods must be addressed for resource-constrained environments to utilize deep-learning models. These limitations emphasize the need for continual research and innovation to improve the reliability, interpretability, and accessibility of clinical deep-learning models.

Deep learning has emerged as a transformative approach in brain tumor classification, leveraging MRI data for improved accuracy and clinical impact. Simo et al. [22] proposed a novel deep-learning pipeline that optimizes classification performance through advanced architectural designs and adaptive training strategies. Their method, which incorporates adaptive gradient techniques and Nesterov-based optimizers, achieved superior performance metrics in classifying brain tumor types.

This study complements our approach by emphasizing the importance of adaptive optimization and training techniques in enhancing model performance. The findings also validate the use of customized architectures for MRI-based tasks, aligning with our efforts to evaluate and optimize YOLOv9 and ResNet18 for accurate and efficient tumor detection. Additionally, the discussion of confusion matrices and training histories provides valuable insights for model convergence and reliability, which are critical for real-world deployment.

Deep learning has revolutionized medical imaging, enabling the precise detection and classification of diseases. However, several limitations in existing methods persist, which this study aims to address:

1. Limited generalization:

Many models, such as YOLOv7 and AlexNet, have demonstrated high accuracy on specific datasets but struggle to generalize across diverse imaging conditions [9,11].

Drawback: Variability in MRI machine settings and patient demographics often results in inconsistencies in tumor appearance, affecting model robustness.

2. Computational complexity:

Two-stage detectors such as Faster R-CNN offer high accuracy but require significant computational resources, making them less suitable for real-time applications [23].

Drawback: High latency and resource requirements limit deployment in resource-constrained environments.

3. Lack of interpretability:

Many CNN-based methods operate as “black boxes”, which hampers their adoption in clinical workflows [24].

Drawback: Clinicians require interpretable outputs to trust AI-based systems in critical diagnostic settings.

4. Scarcity of annotated data:

Existing studies often use small datasets, limiting the ability of models to learn robust features [17].

Drawback: Insufficient data leads to overfitting and reduced model performance in unseen cases.

By addressing these limitations, our study not only achieves the benchmark performance of state-of-the-art models but also provides practical solutions to improve their clinical applicability.

3. Materials and methods

3.1. Dataset

The dataset utilized in this research project is Kaggle's "Medical Image Dataset—Brain Tumor Detection"¹, consisting of 3903 brain MR images divided into four different groups: glioma tumors, meningioma tumors, pituitary tumors, and no tumors. The photos have an aspect ratio of 139×132 pixels and are in a PNG format. The dataset used in this study comprises MRI scans sourced from diverse institutions, ensuring a broad representation of imaging protocols. Patient demographics include ages ranging from 19 to 75, with an approximately balanced gender distribution. This diversity enhances the generalizability of the models to different clinical settings.

To prepare the input data for the models via deep learning, a few different preprocessing steps were used. The photos were scaled to a standard size of 128×128 pixels to meet the models' input standards. The values of each pixel were normalized to the 0-to-1 scale by dividing them by 255. The information being collected was divided into three sets: training, testing, and validation, with training accounting for 70% and testing accounting for 10%, while the remaining 20% was used for validation purposes [25].

3.2. Deep-learning models

This study compared the following types of deep-learning approaches:

YOLOv9: You Only Look Once (YOLO) is an immediate fashion object-detection algorithm that approaches object recognition as a regression issue with geographically separated box boundaries and likelihoods of classes. YOLOv9 is the most recent generation of the YOLO model, with better precision and speed than earlier versions [26,27].

YOLOv8: YOLOv8 is an improved version of the YOLO model that includes various enhancements over YOLOv7, including better backbone networks, more effective extraction of characteristics, and enhanced training procedures [26,27]. Faster R-CNN is a two-phase object-identification model that first creates region recommendations before classifying the items within them. It is noted for its great correctness, but its interpretation speed is lower than that of single-stage detectors such as YOLO [16,23].

ResNet18 is a convolutional neural network (CNN) design that is frequently implemented for image categorization. In this research, we fine-tune a ResNet18 model from scratch to identify brain tumors.

3.3. Training process

Learning rate selection:

- For YOLOv8 and YOLOv9, an initial learning rate of 0.001 was chosen on the basis of prior studies and preliminary experiments, as it provided stable convergence during training. A cosine annealing schedule was applied to adjust the learning rate dynamically throughout training. In addition, a learning rate of 0.001 was selected, aligning with the findings of [27], who demonstrated its

¹ <https://www.kaggle.com/datasets/pkdarabi/medical-image-dataset-brain-tumor-detection/data>

suitability for YOLO models to ensure steady convergence.

- Faster R-CNN utilized a lower learning rate of 0.0001, given its two-stage architecture, which benefits from slower updates to avoid overfitting.
- ResNet18 employs a learning rate of 0.001, with fine-tuning on pre-trained weights requiring minimal adjustments for stability.

Batch size selection:

- YOLOv8 and YOLOv9 were trained with a batch size of 16, which is a commonly used size in deep learning for balancing computational efficiency and model convergence. This aligns with the practices highlighted by [4], who emphasized using suitable batch sizes to optimize GPU utilization without compromising training stability.
- Faster R-CNN was limited to a batch size of 4 because of its computational complexity and memory requirements.
- For ResNet18, a smaller batch size of 8 was used, as its architecture and input size constraints benefit from fewer images per batch.

Epochs and early stopping:

- The YOLO models were trained for 50 epochs, with performance monitored on the validation set. Early stopping was implemented to terminate training if the validation accuracy did not improve for 10 consecutive epochs.
- Faster R-CNN was trained for 30 epochs, and ResNet18 was trained for 100 epochs, with similar early stopping criteria.

Optimizer and loss functions:

- The Adam optimizer was used for the YOLO models and ResNet18 to benefit from its adaptive learning rate capabilities.
- Faster R-CNN used stochastic gradient descent (SGD) with momentum, which is well suited for two-stage detectors.
- Loss functions were tailored to each model: The YOLO models used a combination of classification, objectness, and bounding box regression losses. Faster R-CNN employed region proposal loss alongside classification and regression losses. ResNet18 used categorical cross-entropy loss.

Figure 1 shows the confusion matrix for Faster R-CNN.

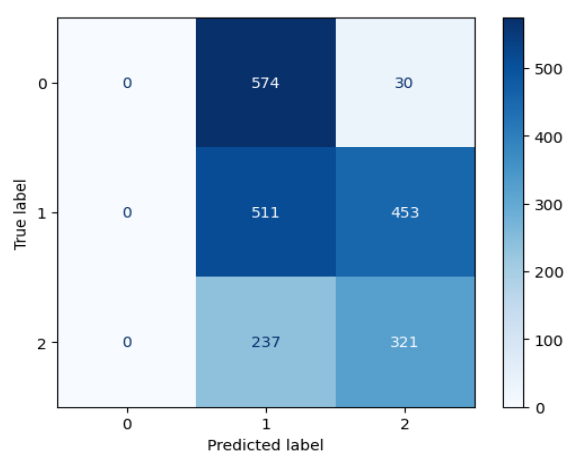


Figure 1. Confusion matrix for Faster R-CNN.

3.4. Evaluation metrics

The effectiveness of the models developed with deep learning was assessed via different metrics. Recall, which is the ratio of the number of true positives to the total number of true positives, and false negatives were used to assess the model's capacity to identify all positive cases. Precision, which is the ratio of the number of true positives to the total number of true positives, and false positives were used to indicate the capacity of the model to accurately identify cases that are positive. The F1 score, which is the harmonic average of the accuracy, and recall provided a fair assessment of the model's efficiency [28,29]. All evolution metrics are shown in Figure 2.

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 \text{ Score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

Figure 2. Performance evaluation matrices used in this study.

Confusion matrix: A matrix of data that displays the total number of false positives, true negatives, true positives, and false negatives, providing a complete picture of the efficiency of the model in every category. The confusion matrices of YOLOv8 and YOLOv9 are shown in Figures 3 and 4.

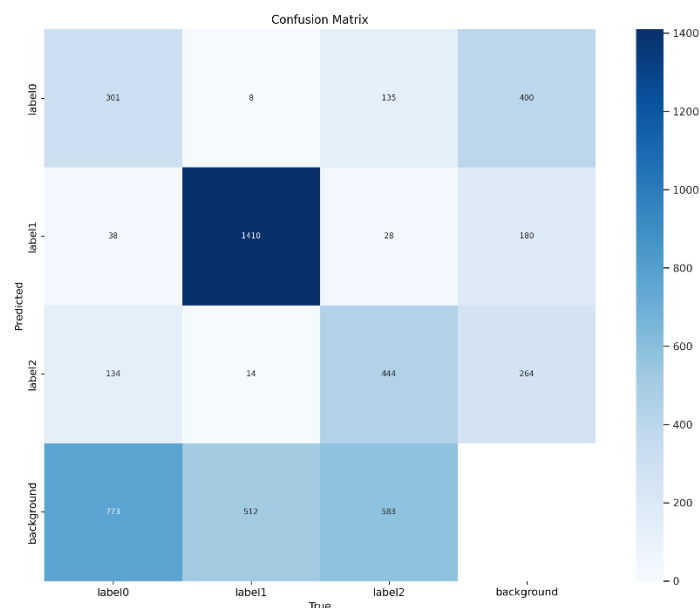


Figure 3. Confusion matrix of the YOLOv8 model.

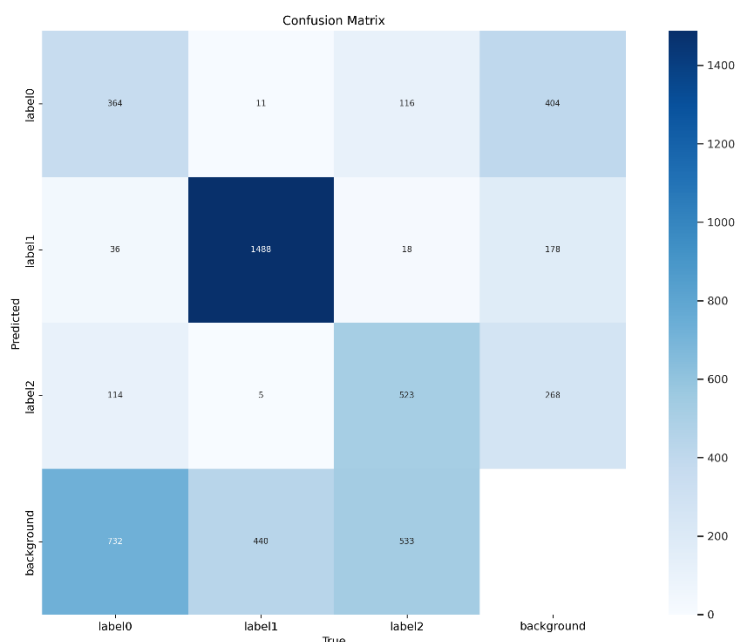


Figure 4. Confusion matrix of the YOLOv9 model on our dataset.

3.5. Image augmentation

To expand and diversify the training dataset, we used the following kinds of image enhancement approaches:

- Arbitrary horizontal rotation.
- Arbitrary vertical rotation.
- Random rotation with an angle of 20° .
- Random scaling from 0.8 to 1.2 and tearing within 0.1 radians.

These enhancement strategies were used in real-time during training with the image data generator function from the Keras model preprocessing component. This enabled us to develop new, realistic-looking brain MR images while also improving the models' generalization effectiveness [30].

3.6. Transfer learning

Using the ResNet18 algorithm, we applied transfer learning to capitalize on the information gained from large-scale dataset analysis [31]. We trained the machine learning model with pre-learned weights on the dataset provided by ImageNet before fine-tuning it on the brain tumor detection datasets. This method enables the framework to acquire key low-level properties from a broad dataset and apply them for brain tumor diagnosis [32].

Transfer learning over object detection begins with selecting an appropriate pre-trained system. Popular possibilities include Faster R-CNN, YOLO, and SSD, each with a unique trade-off between efficiency and precision. The chosen model is then adjusted to fit the new objective by substituting the final classification layers with the actual number of object categories in the target sample. This update guarantees that the model's result is suited to the precise items it is supposed to detect.

During the training phase, the assessment of a validation set is critical for monitoring the model's

results and avoiding overfitting. This incremental training and validation procedure aids in the refinement of the model and achieves optimal possible outcomes. Transfer learning therefore offers a useful method for developing high-performance object-detection algorithms with minimal data and processing resources.

3.7. *Experimental setup*

We divided this dataset into training, testing, and validation sets using the most effective practices described in the documentation that was mentioned. The dataset, which included 6930 training photos, 1980 validation visuals, and 990 testing visuals, was divided via the training/validation/testing division approach. The set for training was employed to train the algorithms, the validation set was applied to tune hyperparameters and validate the model's effectiveness during training, and the experimental set was set aside for unbiased assessments of the final model that was developed.

3.8. *Training process*

3.8.1. YOLOv8 and YOLOv9

The YOLOv8 and YOLOv9 algorithms were ultralightweight and already trained. Training was performed on the training dataset for 35 epochs, using a batch number of 16 and an average learning rate of 0.001. Techniques for data enhancement, such as stochastic horizontal and vertical flipping, scaling, rotation, and splitting, were used to improve model generalizability.

3.8.2. Faster R-CNN

Using the training set, we built and trained a custom-made Faster R-CNN model. The algorithm used was trained over 4 epochs with 857 iterations per epoch, a batch size of four, and an average learning rate of 0.0001. Methods for data augmentation were used to enhance model resilience.

3.8.3. ResNet18

The ResNet18 network with torch vision was used and refined on the training dataset. Training was conducted over 100 epochs, utilizing an initial batch size of 8 and an average learning rate of 0.001. Methods for data augmentation were implemented on images used for training to improve the performance of the models.

3.9. *Evaluation and performance metrics*

As previously stated, every model's effectiveness was assessed via measures such as precision, recall, and precision, along with the F1 score. The above metrics were derived via the held-out test data to provide an unbiased evaluation of the effectiveness of the models. The method of evaluation involves contrasting the model forecasts to the test set's labeled ground truth to determine each model's accuracy in detecting brain tumors. By implementing these stringent training, testing, and validation protocols, as well as utilizing data augmentation approaches and measuring the model's effectiveness

via common metrics, this research ensures an effective and trustworthy assessment of the potential of deep-learning algorithms in identifying brain tumors.

The use of mAP50 and mAP50-95 as evaluation metrics was driven by their established relevance in object detection tasks and their applicability to clinical scenarios. mAP50 evaluates the model's ability to correctly localize and classify objects when the predicted bounding box has at least 50% overlap with the ground truth. This metric provides a clear and interpretable benchmark for detection accuracy. As highlighted by [11], mAP50 is particularly relevant in scenarios requiring high-precision localization, such as brain tumor detection, where accurate tumor boundary identification is crucial for clinical interventions like biopsy or surgical excision. mAP50-95 evaluates the model's performance across multiple intersection over union (IoU) thresholds (from 0.50 to 0.95 in 0.05 increments), offering a more comprehensive measure of detection robustness. It has been found that mAP50-90 accounts for varying detection precision required for different object sizes and shapes, such as irregular or diffuse brain tumors [12].

In conclusion, high mAP50 scores indicate the model's precision in localizing tumor boundaries, providing confidence in its immediate clinical usability for treatment planning. Robust mAP50-95 performance reflects the model's ability to generalize to diverse MRI scans, ensuring consistent results across variations in imaging protocols or noise levels. These metrics align with clinical demands, as described by [18], who demonstrated their importance in detecting brain anomalies with precision and reliability. For instance, the metrics can guide surgical interventions by accurately mapping tumor margins.

4. Results

According to the outcome of the evaluation, several important discoveries and patterns were noted: both YOLOv8 and YOLOv9 had reasonably high mAP50 values of 0.685 and 0.706, respectively, indicating good performance for identifying brain tumors. YOLOv9 has a significantly higher mAP50-95 result of 0.421 than YOLOv8's 0.394, indicating superior overall performance. The performances of YOLOv8 and YOLOv9 are plotted in Figures 5 and 6, respectively.

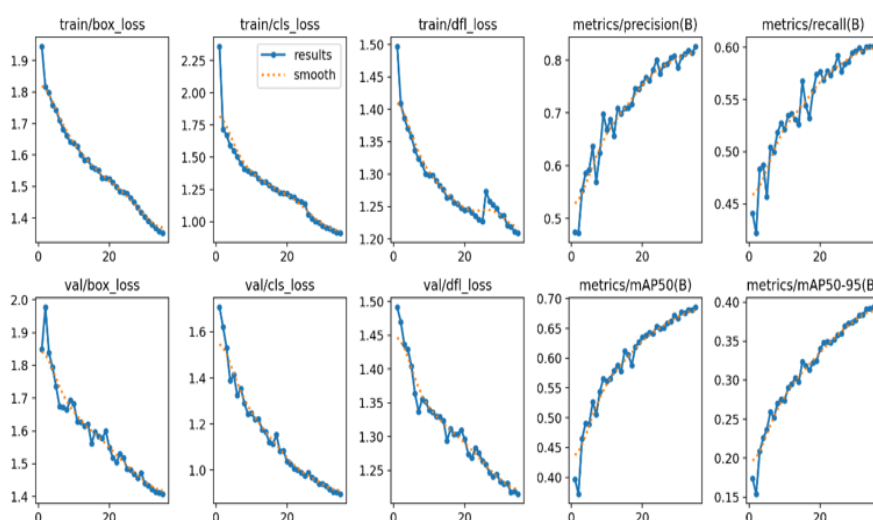


Figure 5. YOLOv8 performance metric graph.

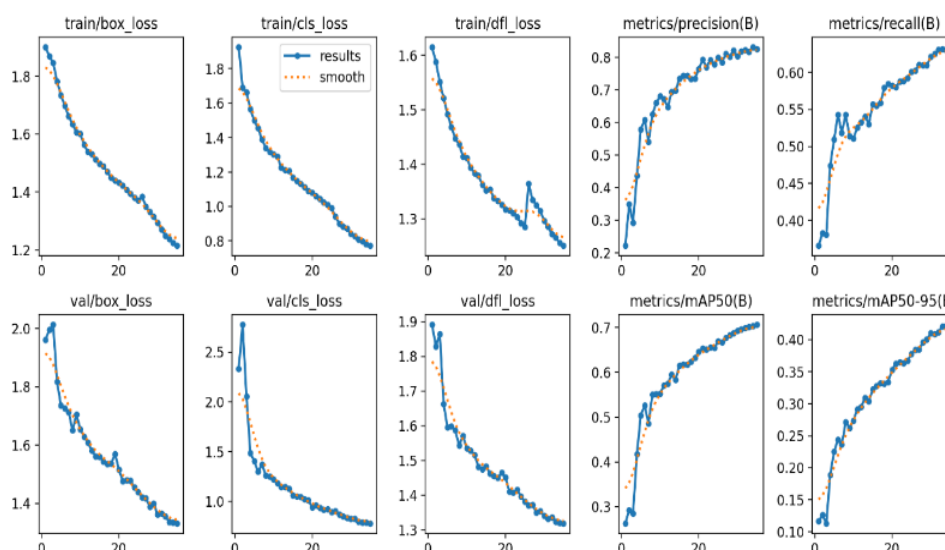


Figure 6. YOLOv9 performance metric graph.

Faster R-CNN had an estimated precision (AP) of 0.248, with an IoU of 0.50 : 0.95, which was lower than that of the YOLO algorithms. The model performed better for larger tumors, with an AP of 0.606 versus 0.090 for small tumors and 0.335 for medium tumors. The observed average recall (also known as AR) was 0.319, with an IoU of 0.50 : 0.95, which was significantly lower than that of the YOLO models. ResNet18 had a validation performance of 0.6973 after 100 epochs of learning. With the validation set of data, the model achieved a precision of 0.4573, a recall of 0.4645, and an F1 score of 0.4592. The performance of ResNet18 was lower than that of the YOLO models, implying that it might be less effective at detecting brain tumors.

In simple terms, YOLO models, particularly YOLOv9, outperformed Faster R-CNN and ResNet18 in identifying brain tumors. YOLOv9 had the greatest mAP50 and mAP50-95 scores, suggesting its ability to accurately detect the existence and positioning of brain tumors in MR images. The YOLO algorithms also had higher recalls for each meningioma tumor class, indicating that they can detect this form of tumor more correctly.

5. Discussion

The findings of the deep-learning algorithms used for brain tumor identification provide important information about how they work and how effective they are in this vital medical domain.

Table 2 shows the comparison of the deep-learning approaches used in our research.

Table 2. Comparison of the proposed models in terms of performance matrices.

Model	mAR	mAP	F1-Score/mIoU	Accuracy
YOLOv8	0.601	0.685	0.697	0.744
YOLOv9	0.635	0.826	0.718	0.784
Faster R-CNN	0.319	0.472	0.380	0.6548
ResNet18	0.4541	0.4529	0.4535	0.6973

A comparison of the outcomes of YOLOv8, YOLOv9, Faster R-CNN, and ResNet18 scenarios reveals significant differences in their efficiencies across multiple evaluation measures. YOLOv9 has the best performance score among all the methods. It yields a high mean average precision (mAP), demonstrating its ability to reliably forecast object positions across various IoU thresholds. Furthermore, YOLOv9 has a larger mean average recall (mAR), implying that its predictions include a greater number of real objects. The model's overall F1 score along with accuracy demonstrate its strong performance, integrating both precision and recall adequately while maintaining outstanding accuracy in classification.

Specifically, YOLOv8 and YOLOv9 performed well in diagnosing brain tumors, with YOLOv9 marginally outperforming YOLOv8 in most measures. YOLOv8 operates admirably; however, it falls behind YOLOv9. Although its mAP and mAR were lower than those of YOLOv9, it nevertheless demonstrated excellent precision and recall, leading to a high F1 score. This model is remarkable for its harmonious combination of detection efficiency and computing economy, making it an adaptable option for a variety of applications. The models obtained excellent mean average precision (mAP50) ratings, indicating that they can reliably locate and identify brain tumors in MR images. Additionally, the rate of recall for meningioma tumors was very high, indicating that the models had the ability to recognize this specific tumor type. The YOLOv8- and YOLOv9-predicted samples are shown in Figures 7 and 8, respectively.

**Figure 7.** YOLOv8-predicted sample.



Figure 8. YOLOv9-predicted sample.

Although Faster R-CNN produced competitive outcomes, its average precision values were lower than those of the YOLO models. In this case, Faster R-CNN is inferior to the YOLO designs but is typically a strong competitor in finding objects. It has significantly lower mAP and mAR values, indicating less precise and inferior full object detection. However, Faster R-CNN achieves a decent level of accuracy, which makes it a feasible solution in certain cases, particularly where its ability to recognize smaller components and deal with complicated environments is useful. The model performed better on larger tumors, pointing to possible difficulties in reliably recognizing small tumors. The average recall figures indicate that there is potential for enhancing the accuracy of detecting tumors of various sizes. The Faster R-CNN-predicted sample is shown in Figure 9.

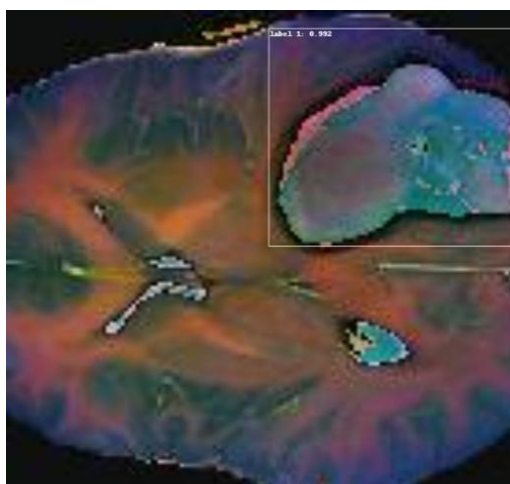


Figure 9. Faster R-CNN-predicted sample.

ResNet18, despite being a frequently used architecture, performed worse than the models built on the basis of YOLO and Faster R-CNN. ResNet18, an algorithm for classification designed for detection tasks, had the least favorable overall efficiency in the above analysis. Its mAP and mAR are not particularly impressive, resulting in a poor F1 score. Regardless, ResNet18 achieves decent

reliability, demonstrating its abilities in classification tasks. This makes it appropriate for applications that require high classification accuracy; however, its identification performance falls short of that of specialized object recognition models. The model's overall accuracy, recall, precision, and F1 score suggest that it may be less suitable for efficient brain tumor identification than models for object detection. YOLO models, particularly YOLOv9, performed best in terms of accuracy, precision, recall, and F1 score for brain tumor identification. YOLOv9's better mAP50-95 result and high recall for meningioma tumors demonstrate its ability to accurately detect and localize brain tumors.

Faster R-CNN performs competitively, especially with larger tumors, but might benefit from improved detection of smaller lesions. While having a strong design, ResNet18 might not be the best option for accurate object detection applications such as brain tumor diagnosis. The recall score, precision, F1 score, and accuracy of each model are plotted in Figure 10.

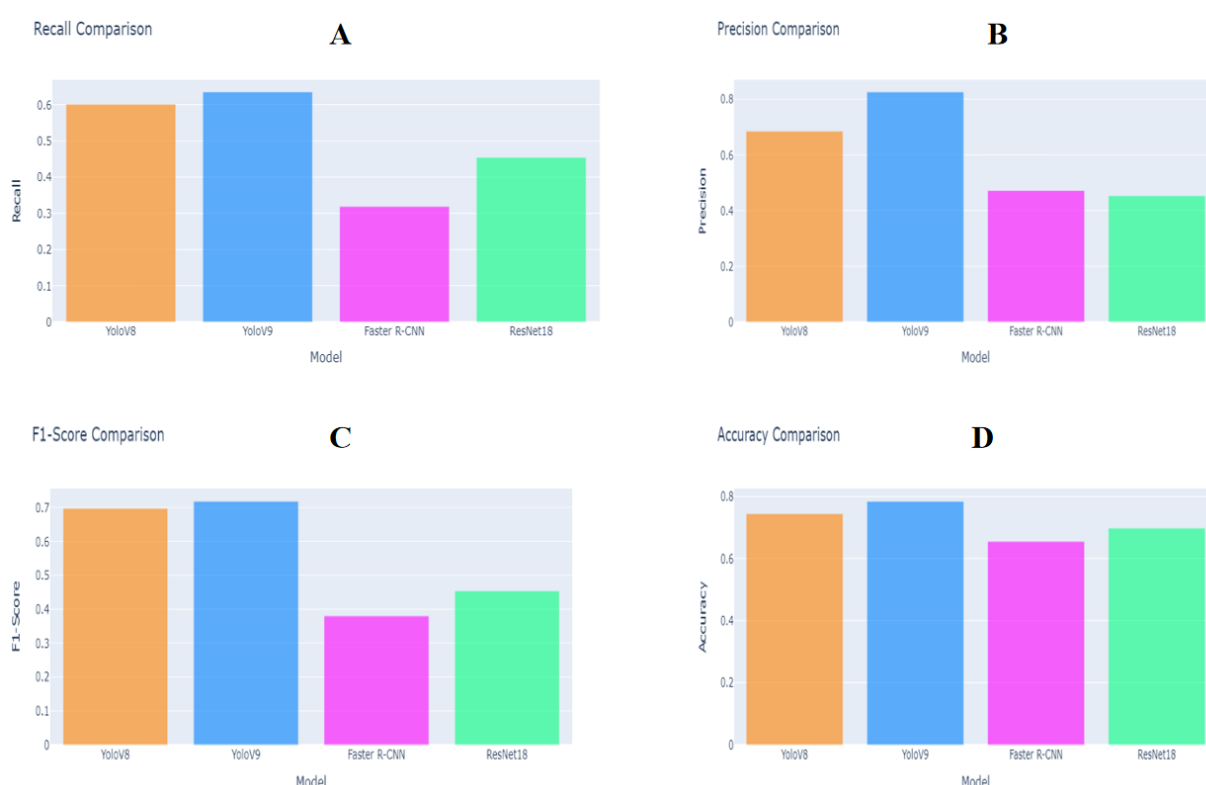


Figure 10. The recall score (A), precision (B), F1 score (C), and accuracy (D) of each model.

The evaluation findings highlight the importance of identifying the appropriate deep-learning framework for specific applications. Owing to their real-time object identification characteristics and excellent accuracy, YOLO models are potential solutions for efficiently and accurately detecting brain tumors in MR images. Additional fine-tuning and optimization of these models may improve their efficiency while contributing to more accurate and useful diagnostic tools in the discipline of healthcare imaging.

The ability to interpret AI models is essential for fostering trust among healthcare professionals and facilitating their implementation in practical medical diagnostics. In this research, interpretability was improved by using gradient-weighted class activation mapping (Grad-CAM) to visualize which areas of the MR images significantly influenced the model's predictions. Grad-CAM produces

heatmaps that are superimposed on the original images, highlighting regions that the model identifies as relevant for detecting a tumor. This method has been widely recognized as an effective way to elucidate decisions made by convolutional neural networks (CNNs) in healthcare applications [33,34]. For example, the YOLOv9 model consistently focuses on tumor edges and adjacent tissues in MR images, delivering clear visual indicators that are consistent with the insights of expert radiologists. This not only corroborates the model's precision but also provides a concrete rationale for its conclusions, enabling clinicians to verify the AI's results against their own expertise. Such methods are vital in critical medical scenarios, as they enhance transparency and decrease the chances of errors [24]. Moreover, the bounding boxes generated by the YOLO models clearly delineate the position, dimensions, and shape of the tumor, which are vital factors for both surgical planning and treatment. Although the Faster R-CNN and ResNet18 models are somewhat less interpretable due to their design, they provide probabilistic predictions for each category, assisting clinicians in assessing the confidence associated with predictions.

The following bolster trust in AI-assisted diagnostics:

- Visual feedback: Grad-CAM visualizations can be incorporated into diagnostic processes to provide clinicians with a clear understanding of the model's focus areas [35].
- Model confidence scores: Probabilistic outputs from the classification layers contribute additional context for evaluating the trustworthiness of predictions [36].
- Human-AI collaboration: These resources enable clinicians to merge their expert knowledge with AI outputs, enhancing confidence in their decision-making [37].

The interpretability techniques utilized in this research highlight the ability of deep-learning models to not only enhance diagnostic precision but also foster clinician trust, which is a crucial step toward incorporating AI into clinical practice.

6. Conclusions

The purpose of this work was to assess the effectiveness of several deep-learning models, namely YOLOv8, YOLOv9, faster R-CNN, and ResNet18, in detecting brain tumors via MRI data. The dataset included 6930 training images, 1980 validation images, and 990 testing images organized into three distinct groups: glioma tumors, meningioma tumors, and pituitary tumors. YOLOv9 outperformed all the other methods, which makes it the preferable choice for extensive object detection problems. YOLOv8 offered a balanced solution with excellent results, but Faster R-CNN and ResNet18 offered unique benefits in specific circumstances, albeit with fewer overall metrics than the YOLO alternatives.

This study highlights the potential of advanced deep-learning models, particularly YOLOv9, for accurate and efficient detection of brain tumors in MR images. Integrating such models into real-time diagnostic workflows could reduce the workload on radiologists, improve efficiency, and enhance diagnostic accuracy in clinical environments. Furthermore, the techniques developed here could be adapted for other medical imaging modalities, such as CT and ultrasound, to broaden their clinical applicability. However, challenges remain. The variability in imaging quality across institutions may impact model generalizability, necessitating further validation via diverse datasets. Regulatory hurdles must also be addressed to ensure compliance with healthcare standards and the ethical use of AI in clinical practice. Finally, reducing the computational demands of these models is crucial for deploying them effectively in resource-constrained environments, ensuring that their benefits reach a wider population.

Grad-CAM visualizations further validated the model's interpretability by highlighting regions of interest closely aligned with tumor boundaries. However, certain limitations must be noted. The models, particularly Faster R-CNN and ResNet18, struggled with smaller or irregularly shaped tumors, reflecting the need for enhanced sensitivity in such scenarios. Testing the models on an independent external dataset is a critical next step to establish their robustness and generalizability across diverse clinical settings. Addressing these limitations through improved training strategies and broader datasets could further enhance the utility of the proposed approach.

Finally, our study demonstrates the potential of deep-learning models, especially YOLOv9, which can help radiologists and doctors make rapid and precise diagnoses of brain tumors. The findings show that the use of YOLOv9 in healthcare settings might enhance brain tumor diagnostic and treatment results. However, additional studies and validations on increasingly diverse datasets are needed to accurately evaluate the models' generalizability and resilience in real-world clinical circumstances.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The author declares no conflict of interest.

References

1. Aldape K, Brindle KM, Chesler L, et al. (2019) Challenges to curing primary brain tumours. *Nat Rev Clin Oncol* 16: 509–520. <https://doi.org/10.1038/s41571-019-0177-5>
2. Franceschi E, Frappaz D, Rudà R, et al. (2020) Rare primary central nervous system tumors in adults: an overview. *Front Oncol* 10: 996. <https://doi.org/10.3389/fonc.2020.00996>
3. McRobbie DW, Moore EA, Graves MJ, et al. (2017). *MRI: From Picture to Proton* (3rd ed.). Cambridge University Press. <https://doi.org/10.3174/ajnr.A0980>
4. Batı CT, Ser G (2023) Effects of data augmentation methods on YOLO v5s: application of deep learning with pytorch for individual cattle identification. *Yüz Yıl Üniv Tarım Bilim Derg* 33: 363–376. <https://doi.org/10.29133/yyutbd.1246901>
5. Krichen M (2023) Convolutional neural networks: a survey. *Computers* 12: 151. <https://doi.org/10.3390/computers12080151>
6. Muksimova S, Umirzakova S, Mardieva S, et al. (2023) Enhancing medical image denoising with innovative teacher-student model-based approaches for precision diagnostics. *Sensors* 23: 9502. <https://doi.org/10.3390/s23239502>
7. Rayed ME, Islam SMS, Niha SI, et al. (2024) Deep learning for medical image segmentation: state-of-the-art advancements and challenges. *Inform Med Unlocked* 47: 101504. <https://doi.org/10.1016/j.imu.2024.101504>
8. Huang SC, Pareek A, Seyyedi S, et al. (2020) Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med* 3: 136. <https://doi.org/10.1038/s41746-020-00341-z>

9. Abdusalomov AB, Mukhiddinov M, Whangbo TK (2023) Brain tumor detection based on deep learning approaches and magnetic resonance imaging. *Cancers* 15: 4172. <https://doi.org/10.3390/cancers15164172>
10. ZainEldin H, Gamel SA, El-Kenawy E-SM, et al. (2022) Brain tumor detection and classification using deep learning and sine-cosine fitness grey wolf optimization. *Bioengineering* 10: 18. <https://doi.org/10.3390/bioengineering10010018>
11. Passa RS, Nurmaini S, Rini DP (2023) YOLOv8 based on data augmentation for MRI brain tumor detection. *Sci J Inform* 10: 8. <https://doi.org/10.15294/sji.v10i3.45361>
12. Hussain M (2023) YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* 11: 677. <https://doi.org/10.3390/machines11070677>
13. Patil S, Waghule S, Waje S, et al. (2024) Efficient object detection with YOLO: a comprehensive guide. *Int J Adv Res Sci Commun Technol* 4: 519–531. <https://doi.org/10.48175/ijarsct-18483>
14. Alshahrani A, Mustafa J, Almatrafi M, et al. (2024) A comparative study of deep learning techniques for Alzheimer's disease detection in medical radiography. *IJCSNS* 24: 53–63. <https://doi.org/10.22937/IJCSNS.2024.24.5.6>
15. Kesav N, Jibukumar MG (2022) Efficient and low complex architecture for detection and classification of brain tumor using RCNN with two channel CNN. *J King Saud Univ Comput Inf Sci* 34: 6229–6242. <https://doi.org/10.1016/j.jksuci.2021.05.008>
16. Avşar E, Salçin K (2019) Detection and classification of brain tumours from MRI images using faster R-CNN. *Teh Glas* 13: 337–342. <https://doi.org/10.31803/tg-20190712095507>
17. Senan EM, Jadhav ME, Rassem TH, et al. (2022) Early diagnosis of brain tumour MRI images using hybrid techniques between deep and machine learning. *Comput Math Methods Med* 2022: 8330833. <https://doi.org/10.1155/2022/8330833>
18. Mathivanan SK, Sonaimuthu S, Murugesan S, et al. (2024) Employing deep learning and transfer learning for accurate brain tumor detection. *Sci Rep* 14: 7232. <https://doi.org/10.1038/s41598-024-57970-7>
19. El-Kenawy ESM, Khodadadi N, Mirjalili S, et al. (2024) Greylag goose optimization: nature-inspired optimization algorithm. *Expert Syst Appl* 238: 122147. <https://doi.org/10.1016/j.eswa.2023.122147>
20. El-Kenawy E, El-Sayed M, Rizk FH, et al. (2024) iHow optimization algorithm: a human-inspired metaheuristic approach for complex problem solving and feature selection. *J Artif Intell Eng Pract* 1: 36–53. <https://doi.org/10.21608/jaiep.2024.386694>
21. El-Kenawy ESM, Rizk FH, Zaki AM, et al. (2024) Football Optimization Algorithm (FbOA): a novel metaheuristic inspired by team strategy dynamics. *J Artif Intell Eng Product* 8: 21–38. <https://doi.org/10.54216/JAIM.080103>
22. Simo AMD, Kouanou AT, Monthe V, et al. (2024) Introducing a deep learning method for brain tumor classification using MRI data towards better performance. *Inform Med Unlocked* 44: 101423. <https://doi.org/10.1016/j.imu.2023.101423>
23. Ren S, He K, Girshick R, et al. (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 39: 1137–1149. <https://doi.org/10.1109/tpami.2016.2577031>

24. Holzinger A, Langs G, Denk H, et al. (2019) Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 9: e1312. <https://doi.org/10.1002/widm.1312>
25. Edalatifar M, Bote-Curiel L, Sánchez-Cano A, et al. (2021) A hybrid neuro-fuzzy algorithm for prediction of reference impact in scientific publications. *IEEE Access* 9: 35784–33579. https://doi.org/10.1007/978-3-319-99834-3_31
26. Redmon J, Farhadi A (2018) YOLOv3: an incremental improvement. *arXiv: 180402767*. <https://doi.org/10.48550/arXiv.1804.02767>
27. Bochkovskiy A, Wang CY, Liao HM (2020) YOLOv4: optimal speed and accuracy of object detection. *arXiv: 200410934*. <https://doi.org/10.48550/arXiv.2004.10934>
28. Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom* 21: 6. <https://doi.org/10.1186/s12864-019-6413-7>
29. Powers DM (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol* 2: 37–63. <https://doi.org/10.48550/arXiv.2010.16061>
30. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. *J Big Data* 6: 1–48. <https://doi.org/10.1186/s40537-019-0197-0>
31. Yosinski J, Clune J, Bengio Y, et al. (2014) How transferable are features in deep neural networks? *Adv Neural Inf Process Syst* 27: 3320–3328. <https://doi.org/10.48550/arXiv.1411.1792>
32. Tan C, Sun F, Kong T, et al. (2018) A survey on deep transfer learning, In: Kůrková V, Manolopoulos Y, Hammer B, et al., *ICANN 2018: Artificial Neural Networks and Machine Learning*, Cham, Switzerland: Springer International Publishing, 270–279.
33. Selvaraju RR, Cogswell M, Das A, et al. (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization, *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy: IEEE, 618–626.
34. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. *arXiv: 170507874*. <https://doi.org/10.48550/arXiv.1705.07874>
35. Samek W, Wiegand T, Müller KR (2017) Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. *ITU J: ICT Discov* 1: 39–48. <https://doi.org/10.48550/arXiv.1708.08296>
36. Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al. (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 58: 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
37. Tjoa E, Guan C (2020) A survey on explainable artificial intelligence (XAI): towards medical XAI. *Artif Intell Rev* 53: 591–641. <https://doi.org/10.1007/s10462-020-09825-y>



AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)