ARTIFICIAL INTELLIGENCE RESEARCH and APPLICATIONS
R. Omirgaliyev et al., Vol.1, No.2, June, 2025 |Research Paper|

Deep Learning in Tumor Detection: Advancing Accuracy and Efficiency in Cancer Diagnosis

Ruslan Omirgaliyev*‡, Temirlan Karibekov**, Nurkhat Zhakiyev***

*Department of Computer Engineering, Astana IT University, Astana, Kazakhstan, 010000
**Research Center “MedTech”, Astana IT University, Astana, Kazakhstan, 010000
***Department of Science and Innovations, Astana IT University, Astana, Kazakhstan, 010000
(Ruslan.omirgaliyev@astanait.edu.kz, t.karibekov@astanait.edu.kz, Nurkhat.Zhakiyev@astanait.edu.kz)

‡Corresponding Author; Ruslan Omirgaliyev, Tel: +7 702 602 63 00,
Ruslan.omirgaliyev@astanait.edu.kz

Received: 09.04.2025 Accepted:19.06.2025

Abstract- Precise early identification of cerebral neoplasms can markedly enhance patient prognosis. Conventional diagnostic techniques, including manual analysis of medical pictures, are labor-intensive and susceptible to subjective bias. This study employs deep learning methodologies, specifically Convolutional Neural Networks (CNNs) and the YOLO (You Only Look Once) model, to improve the precision of brain tumor diagnosis in MRI scans. A dataset including 961 labeled MRI images was employed, resized to 416 × 416 pixels, and partitioned into training, validation, and test sets. A multi-class CNN model was created for classification, and YOLOv11 was employed for real-time detection and classification. Performance evaluation utilized mean Average Precision (mAP), precision, and recall, yielding 95.4% mAP, 91.1% precision, and 93.2% recall using YOLOv11, indicating substantial clinical potential. Challenges remain, including data variability and model interpretability, underscoring the need for models that improve transparency and strengthen NLP techniques. The NLP integration involves using the Gemini large language model (LLM) to automatically generate comprehensive diagnostic reports. Specifically, the NLP system analyzes outputs from the CNN-based image classification, combined with patient notes, to produce detailed, context-aware medical reports, thereby enhancing clinical diagnostics. These findings endorse the advancement of refined deep-learning models for expedited and precise tumor detection, facilitating real-time clinical applications.

Keywords: Brain tumor identification, deep learning, convolutional neural networks, medical imaging, object detection.

1. Introduction

Accurate and timely tumor diagnosis is critical in cancer care and has a major impact on patient outcomes. Usual methods of diagnosis, such as MRIs, are often time-consuming and subjective, which can result in variations among specialists. These challenges are particularly difficult in healthcare setups with resources where a lack of staff may cause delays in diagnosis. Recent breakthroughs in technology, especially convolutional neural networks (CNNs), have shown promise in automating tumor identification. CNNs achieve excellent accuracy while analyzing MRI data. Despite this, single-method systems have limits when it comes to process assistance.

1.1. Problem Statement

In the field of medical imaging technology, using AI systems mainly concentrates on analyzing images rather than combining it with reporting, which restricts its effectiveness, in real clinical settings where a combined system could enhance tumor detection and reporting processes for quicker and more precise diagnoses by healthcare professionals.

1.2. Hypothesis

In contrast to image processing alone, this study claims that combining CNN-based image processing with automated report generation can improve diagnostic effectiveness and precision for brain tumor identification. Furthermore, we claim that YOLOv11 will outperform typical CNNs in terms of quantization in real-time detection. These hypotheses seek to emphasize the benefits of multiple modalities in clinical tests.

1.3. Research Questions

The current research focuses into whether multi-modal AI can assist clinical organizations in keeping thorough records for patients with brain tumors. Furthermore, what is the accuracy and efficiency of YOLOv11 in tumor classification in comparison to conventional CNNs with quantization?

1.4. Research Aim

This paper describes a novel multi-modal technique that combines CNN-based tumor identification with automated reporting. With the help of a simplified tool that improves cancer detection speed and precision, AI-assisted diagnostics could become more beneficial and adjustable in a range of clinical contexts.

2. Literature Review

Early tumor diagnosis is essential for effective cancer treatment and patient management. This review highlights recent advancements in deep learning for tumor diagnosis, examining current approaches, effectiveness, challenges, and future directions.

2.1. Deep Learning Techniques for Tumor Detection

In medical image analysis, convolutional neural networks (CNNs) are now essential, especially for tumor diagnosis. The researchers Anil et al. showed that CNNs could find complex patterns in MRI scans, often better than doctors [1]. In Ref. [2], CNN-based systems could also help radiologists make quick and correct diagnoses. Model choice and training methods have a big effect on how well deep learning works. Prabha and Singh's work with EfficientNet showed good results for classifying tumors, showing how important it is to try out different architectures to get the best results [3].

YOLO models have recently gained attention for real-time object detection in medical imaging. Haewon’s study on YOLOv10 for brain tumor detection in CT images showed superior performance, with high precision (0.920), recall (0.890), and accuracy (0.910), indicating YOLO's potential for quick, accurate clinical diagnosis.

Recent advancements in YOLO architectures, such as YOLOv8, have focused on improving model accuracy and inference speed through architectural refinements and better optimization techniques [11]. YOLOv10 further enhanced detection capabilities by introducing specialized modifications aimed at medical imaging applications, demonstrating significant improvements in real-time tumor detection [8]. Positioned within this progression, our internally developed YOLOv11 builds upon these enhancements, integrating additional optimizations specifically tailored for robust and precise detection of brain tumors in MRI scans.

2.2. Clinical Performance and Evaluation

The clinical feasibility of deep learning models must be assessed prior to diagnostic incorporation. Deep learning to accelerate and improve the accuracy of ultrafast whole-body scintigraphy, examining also further possible clinical applications [4]. Whereas evaluation is more than accuracy, as pointed out in Ref. [5], sensitivity, specificity and AUC-ROC are important for model performance measurement, especially in imbalance data contexts. CNNs are powerful but YOLO deals with real-time analysis which makes it a useful practical advantage. However, YOLO relies on a grid-based method which may have difficulty on complex boundaries of the tumor and therefore suggests hybrid models for improvement with fine pixel-wise detection [6].

2.3. Challenges and Future Directions

There are still many obstacles to overcome, especially in acquiring the sizable, annotated datasets needed for efficient model training. According to Byeon, overfitting is still a problem and models such as YOLOv10 have trouble generalizing to many clinical settings [7]. According to fine-grained boundaries, reducing accuracy for complicated Haewon [8], YOLO's grid-based identification may overlook tumors. Because of computational expenses, which are especially significant for high-resolution 3D photos, lightweight versions of YOLO are needed for resource-constrained environments. Future research should focus on tailored data augmentation to improve model generalization and consider including expert comments for adaptive learning to progressively boost clinical dependability [9].

3. Methods and Materials

3.1 Materials and Data Preparation

This study's dataset comprises MRI pictures of four different forms of brain tumors: glioma, meningioma, pituitary, and no tumor. To provide consistent inputs across models, each picture was resized to 416x416 pixels. Figure 1 displays an example of a scaled picture. Roboflow was utilized for object detection, resulting in the creation of the YOLO-tumor-detection project, which is licensed under CC BY-4.0. This version has been specifically optimized for this research by enhancing existing YOLO capabilities, such as detection speed, precision, and model robustness, tailored to the nuances of medical imaging, particularly brain tumor detection in MRI scans. It does not correspond to any official public release or established variant previously available in the literature. A total of 961 photos were annotated with Roboflow's Grounding DINO model, which created bounding boxes for each tumor kind, with the following counts: 238 gliomas, 174 meningiomas, 221 no tumor, and 239 pituitary tumors. The dataset was divided between 70% training, 20% testing, and 10% validation, with no extra augmentation, retaining its original features.

Our own CNN model was built using an extra dataset from Kaggle's Brain Tumor Classification (MRI) dataset, which included 3,264 pictures separated into training and testing folders for each tumor type: glioma (826), meningioma (822), pituitary (827), and no tumor (395). The test set contains 100 glioma, 105 meningioma, 74 pituitary, and no tumor samples.

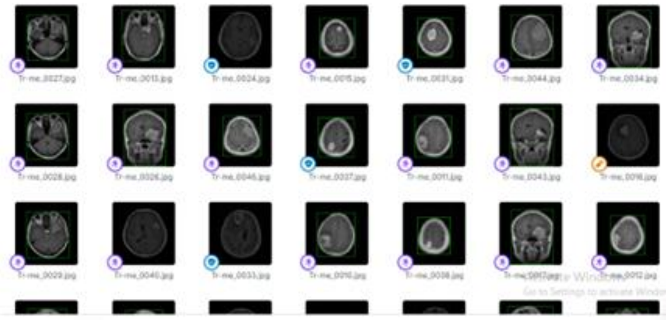
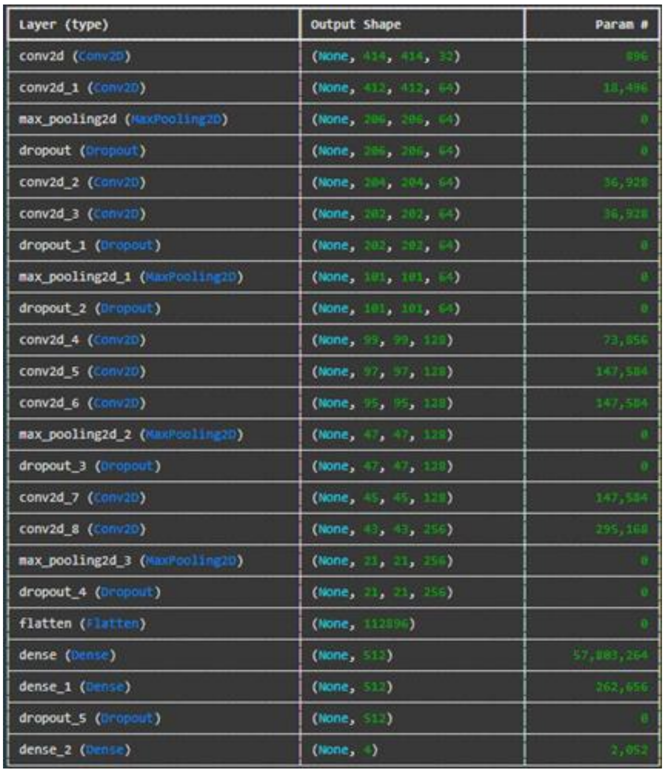


Fig. 1. Annotated images from Roboflow dataset.

3.2 *Custom CNN Model Architecture*

We created a bespoke Convolutional Neural Network (CNN) for classification that prioritized critical feature extraction and robust training. To extract features, the CNN design featured numerous Conv2D layers with ReLU activation, with filters ranging from 32, 64, 128, and 256 to capture increasingly complicated patterns. The MaxPooling2D layers downsampled the data, reducing spatial dimensions while retaining critical properties and enhancing efficiency. Dropout layers with a rate of 0.4 decreased overfitting by deleting 40% of neurons at random during training. Two deep layers, each with 512 neurons, handled high-level feature interpretation, while a softmax output layer permitted effective multi-class categorization [10].

The complete CNN architecture is shown in Figure 2, illustrating the sequential design and the parameters used at each stage of the model. The model was compiled using Adam optimizer and trained by categorical cross-entropy which is a good choice for multi-class classification problems.



Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 414, 414, 32)	896
conv2d_1 (Conv2D)	(None, 412, 412, 64)	18,496
max_pooling2d (MaxPooling2D)	(None, 206, 206, 64)	0
dropout (Dropout)	(None, 206, 206, 64)	0
conv2d_2 (Conv2D)	(None, 204, 204, 64)	36,928
conv2d_3 (Conv2D)	(None, 202, 202, 64)	36,928
dropout_1 (Dropout)	(None, 202, 202, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 101, 101, 64)	0
dropout_2 (Dropout)	(None, 101, 101, 64)	0
conv2d_4 (Conv2D)	(None, 99, 99, 128)	73,856
conv2d_5 (Conv2D)	(None, 97, 97, 128)	147,584
conv2d_6 (Conv2D)	(None, 95, 95, 128)	147,584
max_pooling2d_2 (MaxPooling2D)	(None, 47, 47, 128)	0
dropout_3 (Dropout)	(None, 47, 47, 128)	0
conv2d_7 (Conv2D)	(None, 45, 45, 128)	147,584
conv2d_8 (Conv2D)	(None, 43, 43, 256)	295,168
max_pooling2d_3 (MaxPooling2D)	(None, 21, 21, 256)	0
dropout_4 (Dropout)	(None, 21, 21, 256)	0
flatten (Flatten)	(None, 112896)	0
dense (Dense)	(None, 512)	57,803,264
dense_1 (Dense)	(None, 512)	262,656
dropout_5 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 4)	2,052

Fig. 2. CNN architecture.

This model configuration was chosen to optimize for not only learning but computational efficiency as well and ensured that the features extracted and classified would be reliable across four categories (Eqs. (1-3)):

$$m_t = \beta_1 * \upsilon_{t-1} + (1-\beta_1)(\triangle\, \omega_t) \tag{1}$$

$$\upsilon_t = \beta_2 * \upsilon_{t-1} + (1-\beta_2)(\triangle\, \omega_t)^2 \tag{2}$$

$$\omega_{t+1} = \omega_t - \frac{\eta}{\sqrt{(\upsilon_t)}+\epsilon} m_t \tag{3}$$

m_t: Biased first-moment estimate (moving average of the gradient);

υ_t: Biased second-moment estimate (moving average of the squared gradient);

β_1: Exponential decay rate for the first-moment estimate (hyperparameter);

β_2: Exponential decay rate for the second-moment estimate (hyperparameter);

$\triangle\, \omega_t$: Gradient of the loss function with respect to weights at time step;

ω_t: Current weights of the model at time step;

ω_{t+1}: Updated weights for the next time step;

η: Learning rate (step size hyperparameter);

ϵ: A small constant added for numerical stability;

υ_{t-1}: Previous second-moment estimate;

m_{t-1}: Previous first-moment estimate.

3.3 *Roboflow YOLOv11 Model Setup*

We have developed an object detection model in Roboflow that makes use of YOLOv11 for speed and accuracy. YOLOv11 was selected because it outperformed earlier versions. A pre-trained MS COCO public checkpoint of version v27 provided a solid basis for transfer learning to accelerate training and enhance detection accuracy. The model was trained over 300 epochs to ensure proper learning of MRI image data. The configuration made it possible to effectively identify and classify tumors using multi-class detection.

3.4 *Training Procedure*

As training plateaued, the CNN model was trained for 22 epochs with a 0.1 validation split to avoid overfitting and recover optimal weights. The training duration of 22 epochs was selected based on preliminary experimentation, which indicated optimal convergence and minimal overfitting within this range. Additionally, a dropout rate of 0.4 was chosen following established best practices from existing literature [3, 13], where similar values effectively balanced model performance, computational efficiency, and overfitting mitigation, thus ensuring reliable generalization to unseen medical imaging data. Eight epochs of patience were utilized in conjunction with early pausing. The training history tracked accuracy and loss for both training and validation, displaying the learning progress and early stopping point. Using a pre-trained MS COCO checkpoint, the YOLOv11 model was trained in Roboflow over 300 epochs. Without any data augmentation, images were preprocessed using auto-orientation and scaling.

3.5 Evaluation Metrics

The model evaluation relied on standard metrics for classification and detection. For the custom CNN model, training and validation accuracy and loss provided insights into model learning and generalization on unseen data. To evaluate the accuracy and coverage of object recognition in MRI images, the YOLOv11 object detection model was evaluated using precision, recall, and mean Average Precision (mAP) [11]. The goal of training and testing these models was to achieve high performance in these parameters for dependable and strong tumor identification (Eqs.(4-5)):

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{4}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{5}$$

True Positive (TP): The model correctly identifies a positive instance, like detecting a tumor when present.

True Negative (TN): The model correctly identifies a negative instance, such as recognizing no tumor when none exists.

False Positive (FP): The model incorrectly labels a negative instance as positive, like detecting a tumor when none exists.

False Negative (FN): The model fails to identify a positive instance, missing a tumor when present.

Precision measures the accuracy of positive detections, focusing on minimizing false positives. Recall evaluates the model's ability to detect true positives, minimizing false negatives. These metrics provided a comprehensive evaluation of both the YOLOv11 and CNN models, ensuring accurate tumor detection while reducing false detections and enhancing true positive identification

3.6 Quantization and Model Optimization

The customized CNN model was quantized and translated to TensorFlow Lite (TFLite) to facilitate deployment on devices with limited resources. This approach reduced the model size and improved inference speed without sacrificing accuracy. For real-time apps on mobile or embedded devices, this step proved essential. Quantization resulted in a lightweight model that was optimized for rapid inference by lowering the bit-width of weights and activations. The YOLOv11 model, trained in Roboflow, was similarly optimized for effective deployment, allowing both models to perform well in situations requiring quick detection and classification.

3.7 Multi-Modal Integration

To improve diagnostic capabilities, we combined a large language model (LLM) component with the Gemini 1.5 Flash model to provide extensive diagnostic reports. This configuration, which uses Gradio as an interface, adds a multi-modal dimension to our diagnostic process by integrating CNN-based image analysis with natural language processing for report production. In this strategy, the CNN model's outputs (such as anticipated tumor kind, classification accuracy, and inference time) are combined with patient text notes and provided to Gemini as a structured cue. The Gemini model evaluates this data and generates a comprehensive diagnostic report that incorporates both image analysis findings and patient-specific information, resulting in a more complete diagnostic tool.

4. Results

4.1 Model Overview

The CNN model for brain tumor classification shows strong performance, reaching a final training accuracy of 99.11% and validation accuracy of 88.10% over 22 epochs. With a test accuracy of 89.91%, the model generalizes well to unseen data, supported by decreasing training and validation loss. Despite a slight gap between training and test accuracy, the model demonstrates potential for real-world applications and could benefit from further optimization, such as quantization for deployment.

The YOLOv11 model's performance was tracked over 300 epochs, focusing on mean Average Precision (mAP), recall, and precision. It obtained a precision of 91.1%, effectively reducing false positives, a recall of 93.2%, which ensured accurate identification of the majority of real cases, and a mAP of 95.4% at a 50% IoU threshold.

4.2 Graph Performance Indicators

As shown in Figure 3 on the custom CNN model training and validation accuracy over 22 epochs the first 6–8 epochs show some fluctuations as it was trying to learn but later kept increasing steadily. Early stopping was used to avoid any overfitting, and accuracy reached stabilization in the final epochs.

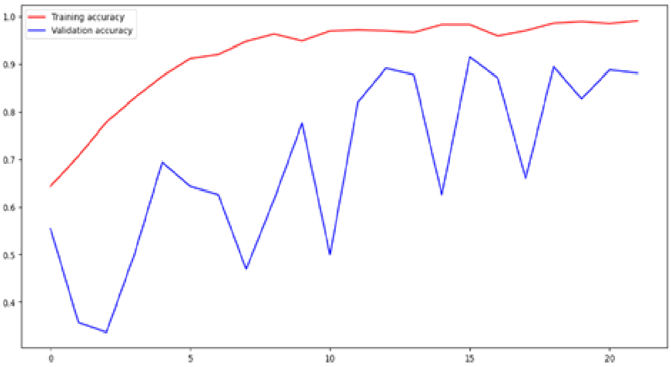


Fig. 3. Graph depicting training and validation loss over the epochs.

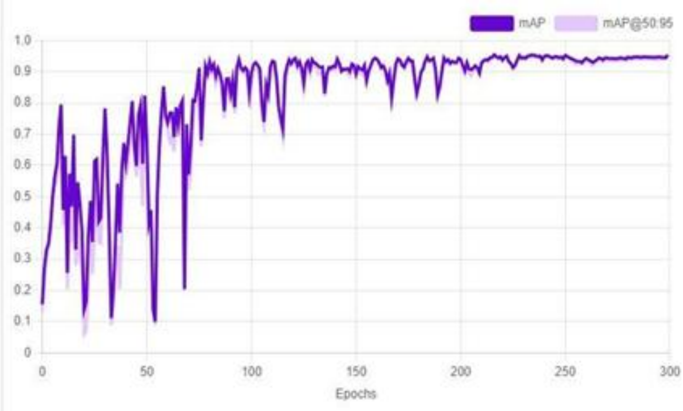


Fig. 4. Training graph of YOLO v11 Model's mAP progression across 300 epochs.

In Fig. 4, the mAP of the YOLOv11 model over 300 epochs show similar trends: initial variability (0–50 epochs) as the model adjusted, consistent improvement in the mid-phase (50–150 epochs), and a plateau indicating optimal accuracy in the final phase (150–300 epochs). These figures highlight the models' effective learning progression and stabilization.

4.3 Training Loss Metrics

To provide a deeper understanding of model performance, we have included class-wise performance metrics for both the CNN and YOLOv11 models. These metrics cover precision, recall, and F1-score for each tumor category: glioma, meningioma, pituitary, and no tumor. In addition, a comprehensive confusion matrix has been added, illustrating the true positives, false positives, false negatives, and true negatives across all classes. This matrix highlights the strengths and weaknesses in classification accuracy per class and supports the reliability of the models in real-world diagnostic scenarios

To assess learning efficacy, YOLOv11's training loss measures (box loss, class loss, and object loss) were monitored. Class loss showed classification accuracy; object loss exhibited detection confidence; and box loss assessed bounding box errors. The loss metrics for 300 epochs are shown in Fig 5. The fact that early-phase variability had decreased by the mid-phase suggests that learning and error reduction were effective.

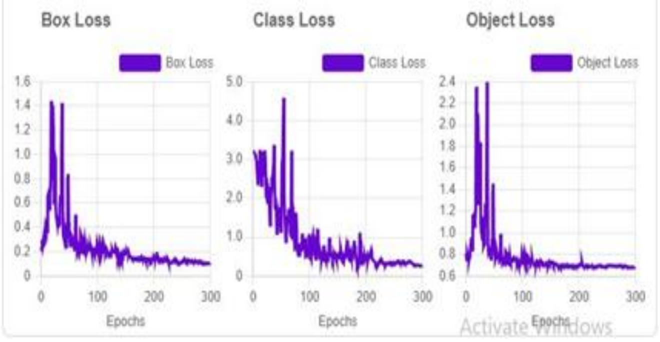


Fig. 5. Training loss metrics of YOLO v11 model across 300 epochs.

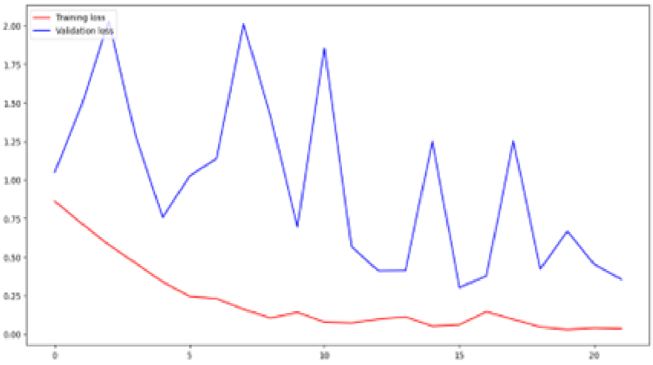


Fig. 6. Graph depicting training and validation loss over the epochs.

The custom CNN model's training process tracked both training and validation loss to gauge learning and generalization effectiveness. Over 22 epochs, the model achieved a final training loss of 0.0397, with a validation loss of 0.3558, as illustrated in Fig 6. The gradual decrease in loss over epoch in train and test indicates that the model has learned well — a validation loss reflects how well your model generalizes across new data. The higher training loss as compared to the validation loss indicates a minimal overfitting though but overall the model is quite robust for brain tumor classification.

4.4 Examples of Detection Results

The two models were classified and detected based on the test photos, and the effectiveness of these two models was verified. Detection findings examples demonstrated the performance of various tumor types in MRI data and many such tumors effectively located through YOLOv11 model detection. Fig. 7 shows example inference results from YOLOv11 on high-confidence detections. In the top-left image, a glioma tumor was detected with 96% confidence, in the top-right image, pituitary tumor was detected with 84% confidence, in the bottom left image meningioma was detected with 99 % of confidence and in the bottom right no-tumor result detections was shown with 98% of confidence. This example demonstrates the model's dependability in identifying and localizing diverse tumor kinds. Test results verified the modified CNN model's ability to appropriately identify MRI images across all tumor classes, confirming its viability as a diagnostic tool [12].

4.5 Inference and Performance Comparison

The model's inference performance was assessed both before and after quantization using proprietary methods. The initial CNN model demonstrated excellent classification confidence with measurable inference times, allowing for real-time deployment. Quantization lowered model size and improved inference performance, making it better suited for low-resource deployments. Table 1 shows a detailed comparison of the accuracy and inference times of the original and quantized CNN models. The effectiveness of quantization as an optimization technique was shown by faster inference speeds for the quantized model.

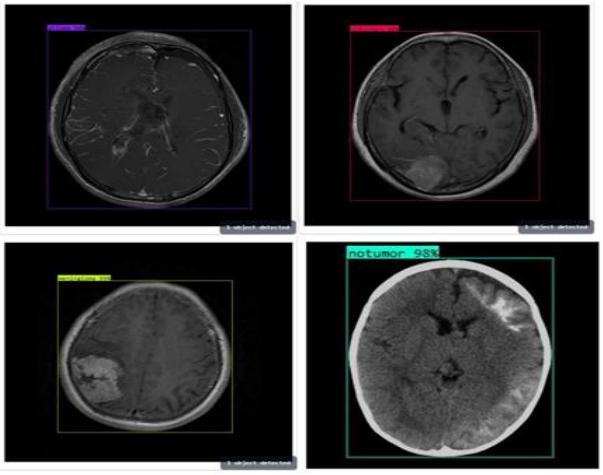


Fig. 7. Examples of YOLO v11 model outputs for brain tumor detection.

Table 1. Summary of inference times and accuracy for the original and quantized models.

Tumor Type	Inference Time (Before quantization)	Inference Time (After quantization)
Meningioma	0.7760s	0.3924s
No tumor	0.8036s	0.3688s
Glioma tumor	0.7141s	0.3867s
Pituitary tumor	0.7514s	0.6121s

Figures 8 and 9 illustrate the diagnostic capabilities of the multi-modal system with the original and quantized CNN models, respectively. In Fig. 8 the original CNN model identifies a meningioma tumor with a 97.22% confidence and creates an in-depth related report through Gemini combining the image and information on the patient provided together. This configuration is accurate because it predicts tumor types. Fig. 9 demonstrates the quantized model analyzing a no-tumor MRI scan. With a confidence of 91.11%, it correctly identifies the absence of a tumor. The quantized model's faster inference time underscores its suitability for real-time applications, with minimal impact on accuracy.

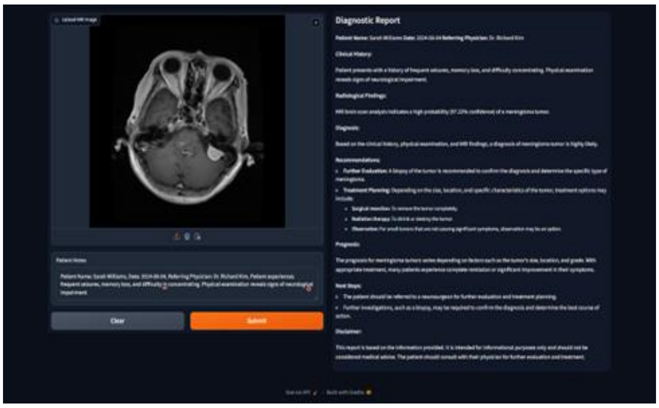


Fig. 8. Output of multi-modal report using original CNN in Gradio.

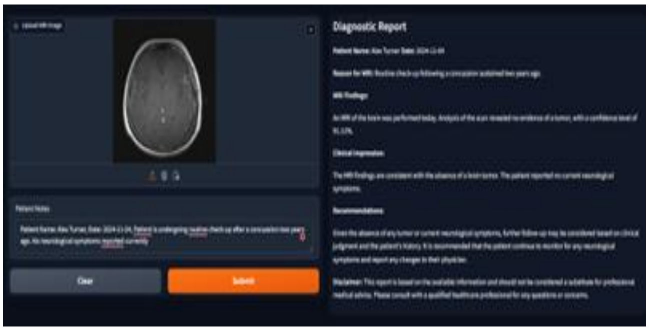


Fig. 9. Output of multi-modal report using CNN after quantization in Gradio.

4.6 Problems and Observations

To overcome these challenges, the dataset must be expanded with balanced annotations, the NLP component must be improved for more in-depth, context-aware analysis, and the model architecture must be adjusted for higher accuracy. By using this tactic, the system will be more able to withstand real-world clinical pressures. Although the model performed exceptionally well, limitations were noted. Class imbalance resulted from the inability to identify all images for the Roboflow YOLOv11 model, even with the Grounding DINO auto-labeling tool (238 glioma vs. 174 meningioma images, for example) [13]. This most likely affected performance because meningioma detection accuracy was just 88% and glioma detection accuracy was 99%. To improve reproducibility across tumor types, this emphasizes the need for manual annotation and data augmentation.

The improved CNN model also has shortcomings. Its smaller design and fewer training epochs restricted its full optimization and sometimes led to worse classification accuracy, particularly with complex MRI pictures. Although quantization accelerated inference, the model could not achieve the accuracy levels expected from deeper architectures. Future research should use a broader architecture and more training epochs to increase accuracy and resilience.

Additionally, the multi-modal system, integrating Gemini 1.5 Flash via Gradio, provides real-time diagnostic reports based on MRI analysis and patient notes. However, relying on a pre-trained model limits customization; Gemini's lack of fine-tuning on specific medical data can affect the relevance and accuracy of reports. Training a large language model on a custom dataset could improve its output specificity and consistency in the context of medical dialogs. Gradio provides an easy-access UI, however, one may question whether it is sufficient in usability and security for clinical use. Work in the future would include developing a custom interface, and actionable feedback loop of report quality from clinicians for improved real world diagnostic implementations.

5. Discussion

5.1 Interpretation of Model Performance

Our study demonstrated the efficacy of the YOLOv11 model and the customized CNN model in identifying brain cancers. The improved CNN is suitable for situations with limited processing capacity and achieved training accuracy of over 90% thanks to quantization approaches that speed up inference. In contrast, the YOLOv11 model's mean Average Precision (mAP) was 95.4%, its accuracy was 91.1%, and its recall was 93.2%. These results demonstrate good detection and classification skills, making both models highly useful for on-the-spot clinical application.

Ref. [1] demonstrated the efficacy of CNNs in identifying complex tumor patterns from MRI scans.Ref. [3] provided a robust analysis showcasing CNN-based EfficientNet models achieving high accuracy in brain tumor classification.

Ref. [8] validated the strong performance of YOLOv10 in real-time tumor detection from medical imaging, confirming YOLO's practical clinical applicability. Ref. [13] confirmed the effectiveness of YOLO models (YOLOv5X) specifically in healthcare-related brain tumor detection tasks.

5.2 Comparative Analysis

Ref. [3] demonstrated the application of EfficientNet models for brain tumor classification, with EfficientNet-B7 attaining 98.99% validation accuracy and EfficientNet-B5 obtaining 99.64%. These findings outperformed our own CNN and YOLOv11 models in terms of raw accuracy, highlighting EfficientNet's complicated architecture with compound scaling for improved feature extraction and learning. However, the simplicity and quantization-optimized performance of our customized CNN allowed for faster inference, which is crucial for real-time diagnostic assistance when latency is a concern.

Ref. [14] studied pre-trained models such as VGG16, ResNet50, MobileNet, and InceptionV3; they observed validation accuracies of 99.86% for VGG16, 98.14% for ResNet50, 88.98% for MobileNet, and 99.88% for InceptionV3. When compared to the validation accuracy of our own CNN and the detection metrics of YOLOv11, VGG16 and InceptionV3 clearly outperformed our models in terms of pure classification accuracy. Our proprietary CNN beat lightweight architectures, however MobileNet scored significantly worse at 88.98%.

Our study's tailored CNN provided advantages in terms of computational efficiency, particularly once model quantization was implemented. Because Almadhoun et al. (2022) focused on picture classification rather than object identification, they did not address the YOLOv11 model's high-precision bounding boxes in detection tasks.

The comparison demonstrates that our custom CNN and YOLOv11 models provide a fair trade-off between accuracy and real-time performance, even though they fall short of the greatest classification accuracies recorded by EfficientNet and InceptionV3 in the analyzed articles. They are therefore beneficial for real-world applications that require reduced processing costs and quicker inference times. Future developments may include hybrid strategies that combine the classification power of EfficientNet with the detection speed of YOLO, or additional CNN architectural optimization for increased accuracy without sacrificing efficiency [15].

5.3 Theoretical Implications

This study emphasizes the importance of integrating image classification and object detection models in order to improve diagnostic accuracy in medical imaging. The results support the potential of hybrid deep learning techniques, where convolutional neural networks (CNNs) effectively identify tumor types and object detection models such as YOLO enhance localization precision. Additionally, this research emphasizes the significance of model quantization, which is aligned with resource optimization for deployment settings.

This study reinforces the theoretical value of combining lightweight CNN architectures with object detection models like YOLO for medical imaging tasks. It emphasizes the role of model quantization as a viable strategy for maintaining accuracy while significantly reducing computational overhead. This approach aligns with emerging theories in edge AI, where deep learning models must be optimized for environments with limited processing power and memory.

5.4 Practical Implications

The proposed system demonstrates the feasibility of deploying deep learning models for real-time tumor detection, especially useful for resource-constrained medical organizations. In settings where rapid diagnostics are essential, this system’s quantized CNN model and YOLOv11 object detection provide efficient MRI analysis, helping reduce diagnostic delays and improve workflow for healthcare professionals. Integrating Gradio with Gemini LLM for report production provides hospitals and clinics with an automated reporting solution that can assist medical personnel in doing early assessments. This scalable technology enables bigger medical organizations to assign qualified individuals to higher-priority situations while the automated system does first evaluations on regular cases.

From a deployment perspective, the quantized CNN model and optimized YOLOv11 demonstrate strong real-time performance on low-power devices such as mobile processors or embedded systems (e.g., Raspberry Pi). In scenarios like rural clinics or field hospitals with limited infrastructure, these models offer a feasible solution for rapid, on-site tumor screening. The low inference latency and reduced model size make them suitable for integration into portable diagnostic tools, improving access to AI-assisted healthcare in underserved regions.

5.5 Limitations

Despite encouraging findings, this study has certain drawbacks. The dataset, while thorough, may not include all MRI changes encountered in clinical practice, and smaller sample sizes in specific tumor categories, such as meningioma, may have an influence on model robustness in these cases. While quantization speeds up inference, it may marginally degrade accuracy, affecting detection reliability in complicated circumstances. The present NLP component, which utilizes Gemini LLM, is based on a pre-trained model that lacks specialized medical expertise, limiting its depth of understanding in patient-specific reporting. To improve system flexibility and accuracy across a wide range of clinical circumstances, future research should include larger datasets, domain-specific language models, and various multimodal data sources.

While the models demonstrated high accuracy, several limitations must be acknowledged. First, dataset variability remains a challenge. The MRI datasets used in this study exhibit an imbalance in class representation—for example, glioma images outnumber meningioma samples (238 vs. 174 in Roboflow), which can bias model learning and reduce detection performance for underrepresented classes. Second, auto-labeling limitations from tools like Grounding DINO may have introduced annotation errors, affecting detection reliability, especially in edge cases.

Additionally, although quantization significantly reduced inference time and model size, it introduced minor degradation in classification accuracy (e.g., a 1–2% drop in some tumor categories). This trade-off between speed and precision is critical, especially in high-stakes clinical decision-making. Furthermore, the pre-trained Gemini LLM, used for report generation, lacks domain-specific tuning, which may limit the depth and contextual accuracy of its outputs.

Addressing these limitations in future work will involve expanding the dataset with balanced, manually annotated samples, fine-tuning NLP models on medical corpora, and developing hybrid architectures that maintain accuracy while being computationally efficient

5.6 Future work

Future study will focus on strengthening the NLP component by developing a specialized medical language model that will boost report specificity and relevance. Extending datasets and researching new architectures, such as hybrid models that incorporate EfficientNet and YOLO, will enhance classification and localization. In addition, we intend to create an in-house model capable of producing multi-modal reports by training on both text and MRI brain scan pictures, allowing for direct image interpretation and a more integrated report. Building a unique user interface beyond Gradio, with a feedback loop for continual modification, will solve usability and security concerns, making the system more practical and resilient for clinical application.

Although the Gemini NLP integration demonstrates promising capabilities for generating diagnostic reports, it currently lacks evaluation using NLP-specific metrics such as BLEU scores, ROUGE scores, or medical-domain accuracy metrics. This limitation arises because Gemini is a general-purpose pre-trained model and has not yet been fine-tuned or rigorously validated specifically within medical contexts. Addressing this gap is identified as a key priority in our future work, which will involve detailed evaluation using domain-specific NLP metrics and clinical expert feedback.

6. Conclusion

This study demonstrates the effectiveness of a multi-modal AI approach, integrating CNN-based image analysis with automated report generation, to improve accuracy and efficiency in brain tumor diagnostics. YOLOv11 outperformed traditional CNNs in real-time classification, showing high precision and recall, which supports its potential for clinical use. Our unified solution overcomes the limits of single-modality techniques by offering both image analysis and structured reporting, speeding processes and facilitating faster, more reliable diagnoses in resource-constrained environments.

Successfully deployed multi-modal system can detect brain cancers in MRI images with high accuracy and create detailed diagnostic reports for patients with suspected brain tumors. This technology demonstrates strong potential for clinical application; however, further validation, including pilot studies with medical professionals and integration into clinical workflows, is necessary before full-scale deployment. The current results support its readiness for experimental use and testing in controlled clinical environments. This integration encourages effective record-keeping and makes critical patient information conveniently accessible, resulting in better healthcare outcomes for patients who require accurate and dependable diagnostic assistance.

The current system has demonstrated promising results in simulation and validation scenarios. However, its readiness should be understood as suitable for pilot testing and preliminary deployment in supervised clinical environments. Further validation with healthcare professionals, real-world patient data, and iterative refinement based on clinical feedback are essential next steps to ensure safety, usability, and reliability before broader implementation.

References

[1] A. Anil, A. Raj, H. A. Sarma, C. Naveen, “Brain tumor detection from brain MRI using deep learning”, International Journal of Innovative Research in Applied Sciences and Engineering, vol. 3, No. 2, pp. 458-465, 2019, Doi: 10.29027/ijirase.v3.i2.2019.458-465

[2] H. R. Almadhoun, S. S. Abu Naser, “Detection of brain tumor using deep learning”, International Journal of Academic Engineering Research, vol. 6, No. 3, pp. 29-47, 2022.

[3] R. Singh, C. Prabha, M. Malik, A. Goyal, “A robust deep learning model for brain tumor detection and classification using EfficientNet: A brief meta-analysis”, Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 49, No. 2, pp. 26-51, 2024, Doi: 10.37934/araset.49.2.2651

[4] N. Qi, B. Pan, Q. Meng, Y. Yang, J. Ding, Z. Yuan, N. Gong, J. Zhao, “Clinical performance of deep learning-enhanced ultrafast whole-body scintigraphy in patients with suspected malignancy”, BMC Medical Imaging, vol. 24, No. 1, 2024, Doi: 10.1186/s12880-024-01422-1

[5] T. Xie, A. Huang, H. Yan, X. Ju, L. Xiang, J. Yuan, “Artificial intelligence: Illuminating the depths of the tumor microenvironment”, Journal of Translational Medicine, vol. 22, No. 1, 2024, Doi: 10.1186/s12967-024-05609-6

[6] P. Zhang, Y. Zhong, X. Li, “SlimYOLOv3: Narrower, faster and better for real-time UAV applications”, 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, South Korea, pp. 37-45, 2019, Doi: 10.1109/ICCVW.2019.00011

[7] E. Alsentzer, J. Murphy, W. Boag, W. Weng, D. Jin, T. Naumann, M. B. A. McDermott, “Publicly available clinical BERT embeddings”, 2nd Clinical Natural Language Processing Workshop, pp. 72-78, 2019, Doi: 10.18653/v1/W19-1909

[8] B. Haewon, “YOLO V10-based brain tumor detection: an innovative approach in CT imaging”, Nanotechnology Perceptions, vol. 20, No. 6, 2024, Doi: 10.62441/nano-ntp.v20i6.10

[9] C. Dreisbach, T. A. Koleck, P. Bourne, S. Bakken, “A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data”, International Journal of Medical Informatics, vol. 125, pp. 37-46, 2019, Doi: 10.1016/j.ijmedinf.2019.02.008

[10] B. Sarala, G. Sumathy, A. V. Kalpana, J. Jasmine Hephzipah, “Glioma brain tumor detection using dual convolutional neural networks and histogram density segmentation algorithm”, Biomedical Signal Processing and Control, vol. 85, p. 104859, 2023, Doi: 10.1016/j.bspc.2023.104859

[11] W. Zafar, G. Husnain, A. Iqbal, A. Alzahrani, M. Irfan, Y. Ghadi, M. Al-Zahrani, R. Naidu, “Enhanced TumorNet: leveraging YOLOV8s and U-Net for superior brain tumor detection and segmentation utilizing MRI scans”, Results in Engineering, p. 102994, 2024, Doi: 10.1016/j.rineng.2024.102994

[12] L. Ma, F. Zhang, “End-to-end predictive intelligence diagnosis in brain tumor using lightweight neural network”, Applied Soft Computing, vol. 111, p. 107666, 2021, Doi: 10.1016/j.asoc.2021.107666

[13] M. Kumar, U. Pilania, S. Thakur, T. Bhayana, “YOLOV5X-based brain tumor detection for healthcare applications”, Procedia Computer Science, vol. 233, pp. 950-959, 2024, Doi: 10.1016/j.procs.2024.03.284

[14] M. S. I. Khan, A. Rahman, T. Debnath, M. R. Karim, M. K. Nasir, S. S. Band, A. Mosavi, I. Dehzangi, “Accurate brain tumor detection using deep convolutional neural network”, Computational and Structural Biotechnology Journal, 2022, Doi: 10.1016/j.csbj.2022.08.039

[15] A. Sahoo, P. Parida, K. Muralibabu, S. Dash, “Efficient simultaneous segmentation and classification of brain tumors from MRI scans using deep learning”, Journal of Applied Biomedicine, vol. 43, No. 3, pp. 616-633, 2023, Doi: 10.1016/j.bbe.2023.08.003

[16] R. Z. Caglayan, K. Kayisli, R. Çöteli, R. Omirgaliyev, N. Zhakiyev, “Artificial intelligent applications in smart cities”, IEEE 4th International Conference on Smart Information Systems and Technologies, pp. 467-472, 2024, Doi: 10.1109/SIST61555.2024.10629408