TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.03 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.290626001

Integrating Watershed-Based Image Segmentation with Deep CNN Classification for Automated Brain Tumor Detection in MRI Scans

Aditya Pandey
Department of Computer Science and Engineering, Galgotias University, Greater Noida, India
aditya.21scse1420024@galgotiasuniversity.edu.in

KEYWORD

Brain Tumor Detection, Watershed Segmentation, Convolutional Neural Network, MRI Analysis, Hybrid Deep Learning

ABSTRACT

Intracranial neoplasms carry a disproportionate clinical burden, with survival outcomes strongly correlated to the speed and accuracy of initial identification. Radiologists rely heavily on Magnetic Resonance Imaging (MRI) for tumor localization, yet manual interpretation introduces delays and examiner-dependent inconsistencies that can jeopardize treatment planning. This paper proposes a two-stage computational framework that first isolates suspicious tissue boundaries through morphological watershed image segmentation and subsequently classifies the segmented region as malignant or benign using a purpose-built Convolutional Neural Network (CNN). Unlike single-stage approaches that operate on full MRI frames, the proposed pipeline channels the CNN's attention toward anatomically constrained candidate regions, reducing false positive detections and improving gradient quality during training. The model was evaluated on a publicly sourced, class-balanced dataset of axial, coronal, and sagittal brain MRI scans. Segmentation quality was measured via the Dice Similarity Coefficient, yielding a mean score of 0.87 across the test partition. Classification accuracy reached 94.3% on the training set and 90.1% on the held-out validation split, outperforming a baseline single-stage CNN by approximately 3.5 percentage points. The complete pipeline was encapsulated within a Flask-based web interface, enabling point-and-click diagnostic queries without any programming requirement.

1. Introduction

Central nervous system malignancies occupy a uniquely challenging position within oncology. The brain's functional density means that even a spatially limited tumor can produce disproportionate cognitive, motor, or sensory deficits, and surgical access is constrained by the imperative to preserve surrounding eloquent cortex. According to data published by the World Health Organization, primary brain tumors account for roughly 2 percent of all adult cancers, yet they contribute a far higher share of cancer-related years of life lost due to their propensity to affect working-age individuals. Secondary brain metastases from lung, breast, or melanoma primaries further compound the diagnostic workload facing neuroradiologists worldwide.

Magnetic Resonance Imaging stands as the cornerstone of neuro-oncological work-up, offering multi-contrast, high-resolution visualization of parenchymal architecture without ionizing radiation. Standard MRI protocols for tumor evaluation combine T1-weighted pre- and post-gadolinium sequences with T2 and FLAIR acquisitions, producing a rich multi-modal dataset for each patient. Interpreting this volume of data accurately and consistently

Corresponding Author: Aditya Pandey, Galgotias University, Greater Noida, India
Email: aditya.21scse1420024@galgotiasuniversity.edu.in

is cognitively demanding; studies on radiological fatigue suggest that error rates climb measurably during high-volume reporting sessions, a concern that is particularly acute in healthcare systems where specialist neuroimaging capacity is limited relative to demand.

Computer-aided detection systems have long been explored as a means of augmenting radiological throughput and consistency. Early approaches relied on hand-engineered features such as Gabor filter responses, intensity histograms, and shape descriptors fed into classical classifiers. These methods achieved modest accuracy on restricted datasets but generalized poorly when scanner models, imaging protocols, or patient demographics shifted. The advent of deep learning, and convolutional architectures in particular, fundamentally changed the calculus: networks trained end-to-end on labeled image corpora routinely match or exceed the sensitivity and specificity of board-certified specialists on well-defined radiological tasks.

Despite these advances, a practical gap persists between laboratory benchmark performance and clinical deployment. Many published models operate on full MRI frames, requiring the classifier to simultaneously locate and characterize a lesion that may occupy fewer than five percent of total image pixels. This spatial imbalance inflates background noise during feature extraction and biases gradient updates toward non-informative background textures. The approach described in this paper addresses this limitation explicitly by introducing a watershed-based segmentation stage that isolates candidate tumor regions prior to classification, thereby focusing the CNN on the most diagnostically relevant image content.

The remainder of this paper is organized as follows. Section 2 surveys relevant prior work in deep learning neuroimaging. Section 3 describes the proposed two-stage architecture in full technical detail. Section 4 outlines the experimental configuration and evaluation protocol. Section 5 presents quantitative results and qualitative analysis. Section 6 discusses clinical implications and limitations. Section 7 identifies directions for future investigation. Section 8 concludes the paper

2. Literature Overview

2.1. Classical Segmentation Approaches

Thresholding-based segmentation, popularized by Otsu's global histogram method, was among the earliest automated tools applied to brain tumor delineation. Its reliance on bimodal intensity distributions, however, renders it fragile when imaging artifacts, partial volume effects, or heterogeneous tumor enhancement alter the expected histogram shape. Region-growing algorithms addressed some of these shortcomings by propagating from manually seeded voxels, but their sensitivity to seed placement introduced user-dependent variability that undermined reproducibility. Graph-cut methods, formulated as energy minimization problems on Markov Random Fields, improved boundary precision considerably by incorporating spatial context into the segmentation criterion, albeit at significant computational cost that limited their real-time applicability.

The watershed transform, originally borrowed from mathematical morphology, treats the image gradient magnitude as a topographic surface and identifies catchment basins corresponding to homogeneous intensity regions. Its strengths lie in its fully automatic operation, its sensitivity to weak edges, and its polynomial computational complexity. The principal vulnerability of naive watershed implementations is over-segmentation arising from minor gradient noise, which can fragment a coherent lesion into dozens of spurious micro-regions. Controlled marker-seeded variants, combined with morphological pre-smoothing, largely overcome this deficiency and form the basis of the segmentation module adopted in this work

2.2. Deep Learning in Neuroimaging

Kamnitsas et al. (2017) introduced a dual-pathway 3D CNN architecture, DeepMedic, that processed multi-scale contextual patches simultaneously and applied a fully connected conditional random field as a post-processing refinement step. Evaluated on the BRATS glioma challenge, DeepMedic demonstrated that volumetric processing and multi-scale receptive fields could capture both fine lesion boundaries and broader peritumoral characteristics that are diagnostically significant.

The encoder-decoder architecture known as U-Net, first presented by Ronneberger and colleagues (2015) for general biomedical segmentation, proved particularly influential for brain tumor delineation. Its symmetric

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.03 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.290626001

expansion path, populated with skip connections that channel high-resolution encoder feature maps directly to corresponding decoder layers, enabled precise localization even when training data were scarce. Subsequent variants incorporated residual connections, attention gates, and dilated convolutions to enhance sensitivity to irregularly shaped lesions.

Transfer learning strategies gained traction through the work of Cheng et al. (2017), who demonstrated that a VGG-style backbone pre-trained on ImageNet, when fine-tuned on a curated CE-MRI corpus, substantially outperformed networks trained from random initialization on the same target data. This finding has important practical implications for clinical deployment scenarios where annotated neuroimaging datasets are inherently smaller than their natural-image counterparts due to the expense of expert labeling.

More recent efforts have explored attention-augmented transformers for radiological image analysis. Hatamizadeh et al. (2022) proposed UNETR, a transformer-based architecture that tokenizes volumetric patches and applies self-attention across the full spatial extent of the MRI volume, capturing long-range dependencies that local convolutions cannot model. While such architectures achieve state-of-the-art benchmark scores, their training data requirements and inference latency currently limit their applicability in resource-constrained clinical environments, motivating continued investigation of more efficient hybrid approaches such as the one described here.

2.3. Positioning of the Present Work

The framework described in this paper occupies a distinct niche relative to prior art. Rather than applying a single end-to-end learned network to raw MRI frames, it explicitly separates the localization problem from the classification problem through a deterministic segmentation stage. This separation reduces the spatial search space presented to the classifier, lowers the class imbalance ratio between foreground lesion pixels and background brain parenchyma, and provides interpretable intermediate outputs that radiologists can inspect and override. The resulting system is also considerably more data-efficient than full volumetric transformer architectures, making it trainable on moderate-sized annotated datasets typical of a university research environment.

3. Proposed Methodology

3.1. Dataset Curation and Characteristics

The experimental corpus was assembled from two publicly accessible neuroimaging repositories. Tumor-positive samples were drawn from a Kaggle brain MRI segmentation collection containing annotated T1-contrast enhanced scans labeled by board-certified radiologists. Tumor-negative samples were sourced from the IXI healthy brain MRI dataset to ensure a demographically balanced negative class. Following quality screening to remove scans with visible acquisition artifacts, motion blur exceeding a threshold of 15 percent corrupted voxels, or incomplete skull coverage, the refined dataset comprised 3,264 images: 1,532 confirmed tumor cases and 1,732 healthy controls. All identifiable metadata was stripped prior to computational processing in accordance with institutional data governance guidelines.

Class stratification was maintained across the 75:15:10 partitioning into training, validation, and test subsets, yielding 2,448 training images, 490 validation images, and 326 test images respectively. The test partition was withheld entirely from all hyperparameter tuning and architecture decisions, serving exclusively for final performance reporting. Images span axial, coronal, and sagittal acquisition planes, providing the classifier with varied anatomical perspectives that strengthen orientation invariance.

3.2. Stage One: Watershed Segmentation Module

Each input MRI slice is processed through a four-step preprocessing and segmentation pipeline before classification. First, the raw image undergoes Gaussian blur with a 5x5 kernel (sigma=1.4) to attenuate high-frequency noise while preserving macroscopic intensity transitions. Second, a morphological closing operation

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.03 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.290626001

using a circular structuring element of radius 7 pixels fills small intensity gaps that would otherwise fragment connected tumor regions during watershed flooding. Third, Otsu thresholding is applied to the smoothed image to generate a binary foreground-background mask that seeds the watershed markers: foreground seeds are obtained by erosion of the foreground mask, and background seeds are derived from a distance-transformed certainty map. Fourth, the OpenCV watershed function floods from these markers according to the original gradient magnitude image, producing a labeled region map. Regions whose area falls below an empirically determined minimum of 200 pixels are discarded as noise components, and the largest retained foreground region is extracted as the candidate lesion bounding box.

The bounding box region, padded by 12 pixels on each side to capture the peri-tumoral zone, is cropped from the original image and passed to the classification stage. On tumor-negative scans the watershed stage typically returns no candidate region meeting the size threshold, in which case the full image is downsampled and forwarded directly to the classifier to avoid false negatives from an overly conservative segmentation.

3.3. Stage Two: CNN Classification Architecture

The classification backbone consists of four convolutional blocks of increasing channel depth, followed by a global average pooling layer and a two-neuron softmax output head. Adopting global average pooling in place of a flattened dense layer reduces the parameter count by approximately 72 percent relative to a conventional fully-connected neck, while simultaneously improving spatial generalization by averaging spatial activations rather than concatenating them.

Table 1: CNN Classification Backbone — Layer-wise Specification

Block / Layer	Operation	Filters / Units	Output Dimensions	Activation
Input	Crop from Seg.	—	96 × 96 × 3	—
Block 1a	Conv2D + BN	32	96 × 96 × 32	ReLU
Block 1b	Conv2D + BN	32	96 × 96 × 32	ReLU
Pool 1	MaxPool 2×2	—	48 × 48 × 32	—
Block 2a	Conv2D + BN	64	48 × 48 × 64	ReLU
Block 2b	Conv2D + BN	64	48 × 48 × 64	ReLU
Pool 2	MaxPool 2×2	—	24 × 24 × 64	—
Block 3a	Conv2D + BN	128	24 × 24 × 128	ReLU
Block 3b	Conv2D + BN	128	24 × 24 × 128	ReLU
Pool 3	MaxPool 2×2	—	12 × 12 × 128	—
Dropout	SpatialDropout2D	p=0.40	12 × 12 × 128	—
Block 4a	Conv2D + BN	256	12 × 12 × 256	ReLU
Block 4b	Conv2D + BN	256	12 × 12 × 256	ReLU
Pool 4	MaxPool 2×2	—	6 × 6 × 256	—
GAP	GlobalAvgPool2D	—	256	—
Output	Dense	2	2	Softmax

Each convolutional block follows an identical micro-architecture: two consecutive 3x3 convolutional layers with Batch Normalization and ReLU activation, succeeded by 2x2 max pooling. Batch Normalization accelerates training by reducing internal covariate shift, allowing a higher initial learning rate and enabling stable gradient flow through the four-block depth. He-normal kernel initialization is applied throughout to address the

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.03 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.290626001

variance collapse that would otherwise arise in deep ReLU networks. Channel counts progress as 32, 64, 128, and 256 across the four blocks, systematically increasing the representational capacity at each spatial scale. A spatial dropout layer with a drop probability of 0.40 is inserted between the third and fourth convolutional blocks, targeting entire feature map channels rather than individual neurons. This channel-level regularization is more effective than scalar dropout for spatially correlated convolutional activations, as demonstrated empirically by Tompson et al. (2015). The final output layer maps global average pooled features to a two-dimensional probability vector via softmax, yielding confidence scores for the tumor-present and tumor-absent hypotheses.

3.4. Training Protocol and Hyperparameter Configuration

The network was optimized using the AdamW variant, which decouples weight decay from the adaptive gradient update step, providing more effective L2 regularization than standard Adam in scenarios with heterogeneous parameter magnitudes. An initial learning rate of 1e-3 was combined with a cosine annealing schedule that reduced the rate to 1e-5 over 30 training epochs, allowing large initial weight updates to explore parameter space broadly before fine-grained refinement in later epochs. Categorical cross-entropy served as the primary loss, supplemented by a label-smoothing factor of 0.05 to penalize overconfident predictions and improve calibration of the softmax output probabilities. Online data augmentation was applied exclusively to training samples: random horizontal flipping, rotation within plus or minus 15 degrees, zoom within plus or minus 10 percent, brightness jitter of plus or minus 0.15 intensity units, and Gaussian noise injection at sigma=0.02. These transforms were selected to mimic variability introduced by different scanner orientations and acquisition parameters without introducing non-physiological distortions. Augmentation was implemented inline using TensorFlow's `tf.image` API within the data-loading pipeline, ensuring zero computational overhead during inference. Batch size was fixed at 32 across all experiments; the model was trained on a single NVIDIA GPU with mixed-precision (FP16) arithmetic.

Table 2: Training Hyperparameter Configuration

Hyperparameter	Selected Value	Justification
Optimizer	AdamW	Decoupled weight decay
Initial Learning Rate	1 × 10	Standard for AdamW
LR Schedule	Cosine Annealing	Smooth decay, avoids sharp drops
Final Learning Rate	1 × 10	Fine-grained convergence
Loss Function	Cat. Cross-Entropy	Multi-class probability target
Label Smoothing	0.05	Prevents overconfidence
Spatial Dropout Rate	0.40	Channel-level regularization
Batch Size	32	Memory-accuracy balance
Training Epochs	30	Convergence observed
Input Resolution	96 × 96 px	Post-segmentation crop size
Weight Initialization	He-Normal	Stable deep ReLU gradients

4. Experimental Setup and Evaluation Protocol

4.1. Evaluation Metrics

Four complementary metrics were used to characterize system performance. Classification Accuracy measures the proportion of correctly classified scans across both classes. Sensitivity (Recall) quantifies the fraction of true tumor cases correctly identified, a clinically critical metric since missed tumors carry higher patient risk than false alarms. Specificity measures the fraction of healthy scans correctly cleared, reflecting the system's value as a

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.03 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.290626001

triage tool that avoids unnecessary follow-up imaging. The F1 Score, defined as the harmonic mean of precision and recall, provides a single balanced summary measure that is robust to class imbalance. Segmentation quality was assessed independently using the Dice Similarity Coefficient (DSC), computed as twice the intersection of the predicted tumor mask and the ground-truth annotation divided by the sum of their areas.

4.2. Baseline Comparison

To isolate the contribution of the watershed segmentation stage, the same CNN backbone was also trained on full-frame MRI inputs without prior region extraction, constituting a single-stage baseline. Identical training hyperparameters, augmentation strategies, and random seeds were used across both conditions to ensure a controlled comparison. The performance gap between the two conditions directly quantifies the benefit of structured region-of-interest preprocessing.

4.3. Software and Hardware Environment

Table 3: Software and Hardware Configuration

Component	Version / Specification	Function
Python	3.11.8	Core scripting environment
TensorFlow	2.16.1	CNN training and inference
Keras	3.3.3	High-level model API
OpenCV	4.11.0	Watershed segmentation pipeline
NumPy	1.26.4	Numerical array operations
Scikit-learn	1.5.0	Metrics and data splitting
Pillow	10.3.0	Image I/O and conversion
Flask	3.1.1	Web deployment interface
Werkzeug	3.0.3	Secure file handling
Matplotlib	3.9.0	Result visualization
GPU	NVIDIA RTX 3060	Model training acceleration
RAM	16 GB DDR4	Dataset buffering

5. Results and Analysis

5.1. Segmentation Performance

The watershed segmentation module was evaluated against radiologist-drawn ground-truth contours on the 326-image test set. The mean Dice Similarity Coefficient across all tumor-positive test scans was 0.873 (standard deviation 0.041), indicating high spatial agreement between algorithmically derived and manually annotated boundaries. Cases with DSC below 0.75, representing approximately 11 percent of positive test scans, were predominantly associated with diffusely infiltrating gliomas whose boundaries blend gradually into surrounding parenchyma, a well-recognized challenge for gradient-based segmentation algorithms. Mean processing time per slice was 38 milliseconds on the CPU-only test environment, confirming real-time compatibility.

5.2. Classification Performance

Table 4: Quantitative Performance Comparison — Proposed vs. Baseline

Metric	Two-Stage (Proposed)	Single-Stage (Baseline)	Improvement
Training Accuracy	94.3%	91.0%	+3.3 pp

Validation Accuracy	90.1%	86.4%	+3.7 pp
Test Accuracy	89.7%	85.9%	+3.8 pp
Sensitivity	91.2%	87.5%	+3.7 pp
Specificity	88.3%	84.6%	+3.7 pp
F1 Score	0.907	0.870	+0.037
Mean DSC (Seg.)	0.873	N/A	—

Across every reported metric, the two-stage pipeline demonstrated consistent advantage over the single-stage baseline, with test accuracy improving from 85.9 percent to 89.7 percent. The most clinically significant gain was in sensitivity: the proposed system correctly identified
91.2 percent of tumor-positive scans against 87.5 percent for the baseline, representing a reduction of roughly one missed tumor per 27 positive cases. This improvement is directly attributable to the segmentation stage channeling the classifier's receptive field toward the most diagnostically relevant image region, shielding convolutional filters from the distracting background skull and scalp intensities that occupied much of the feature computation budget in the single-stage model.

5.3. Learning Curve Analysis

Training and validation accuracy curves converged smoothly over the 30-epoch schedule, with no evidence of overfitting beyond epoch 22. The gap between training accuracy (94.3%) and validation accuracy (90.1%) narrowed progressively under cosine annealing, indicating that the learning rate schedule successfully prevented the sharp generalization degradation sometimes observed with fixed-rate optimizers at epoch boundaries. Label smoothing visibly reduced prediction overconfidence in the final epochs, as reflected by a reduction in mean softmax output entropy from 0.93 to 0.76 on training samples.

5.4. Failure Mode Examination

Manual review of misclassified test cases revealed two dominant failure categories. The first category consisted of small peripheral lesions, predominantly meningiomas measuring less than
1.5 cm in diameter, whose post-segmentation crop contained insufficient distinctive texture to trigger confident positive classification. The second category comprised healthy scans exhibiting T2 hyperintense foci associated with small vessel disease or post-surgical encephalomalacia, which the segmentation stage occasionally extracted as candidate regions that the classifier subsequently labeled as tumors. Both failure modes identify concrete directions for architectural refinement discussed in Section 7.

6. System Deployment and Clinical Interface

6.1. Web Application Architecture

The complete two-stage pipeline was packaged within a Flask web service to enable non-technical clinical users to query the model without any programming knowledge. The application exposes three HTTP endpoints: a root GET handler that serves an HTML5 upload page, a POST endpoint at /analyze that receives an uploaded MRI image and returns a JSON response containing the predicted class, confidence score, and a Base64-encoded PNG of the segmentation overlay, and a GET endpoint at /report that renders a browser-printable diagnostic summary formatted for clinical documentation. Uploaded files are validated for format (JPEG, PNG, DICOM converted to PNG) and size (maximum 20 MB) before processing. The file path is sanitized via Werkzeug's `secure_filename` function, and the temporary upload directory is purged after each request to prevent server-side file accumulation. Response latency, measured from upload initiation to rendered result, averaged 1.2 seconds on a standard consumer laptop without GPU acceleration, confirming real-time usability in outpatient settings.

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.03 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.290626001

6.2. Output Interpretation and Clinical Safeguards

The web interface presents the classification confidence score alongside a traffic-light risk indicator: green for confidence below 50 percent (no tumor likely), amber for 50 to 75 percent (indeterminate; recommend specialist review), and red for above 75 percent (tumor suspected; urgent radiology consultation advised). This graduated output design deliberately avoids binary yes/no messaging, reinforcing the system's role as a decision-support aid rather than a diagnostic authority. A prominent disclaimer on the results page reiterates that all outputs require verification by a qualified clinical professional before any treatment decision is made.

7. Discussion, Limitations, and Future Directions

7.1. Clinical Relevance

A test accuracy of 89.7 percent and sensitivity of 91.2 percent, achieved on a heterogeneous multi-orientation dataset, positions the proposed system competitively relative to published single-modality CNN baselines for binary brain tumor detection. The explicit segmentation stage provides a clinically valuable additional output: the segmentation overlay delivered with each prediction highlights the spatial region driving the classification decision, providing radiologists with an immediate visual anchor for confirmatory review. This interpretability advantage distinguishes the proposed approach from black-box end-to-end models that produce only a class label.

7.2. Current Limitations

Four limitations warrant explicit acknowledgment. First, the dataset, while reasonably sized at 3,264 images, was drawn from two repositories and may not capture the full scanner-model and imaging-protocol diversity encountered in real-world multi-center deployments. Performance should be validated on external cohorts before clinical adoption. Second, the watershed module was parameterized empirically on the current dataset; its Gaussian kernel size and morphological structuring element dimensions may require re-tuning for scanners operating at different field strengths or slice thicknesses. Third, the current classifier produces only a binary benign/malignant distinction and does not distinguish between tumor histological subtypes, which carry critically different prognoses and treatment pathways. Fourth, the 96x96 pixel input resolution constrains fine-grained morphological characterization, particularly for sub-centimeter lesions.

7.3. Planned Future Enhancements

Multi-class Grading: Extending the classifier output layer to distinguish WHO Grade I through IV gliomas, meningiomas, and pituitary adenomas, using multi-label annotations from the TCIA TCGA-GBM collection.
Attention Mechanism Integration: Incorporating a Squeeze-and-Excitation recalibration block after each convolutional stage to enable channel-wise feature reweighting, which has been shown to improve localization of small, low-contrast lesions.
Multi-Modal Fusion: Expanding the input pipeline to accept and fuse four MRI contrasts (T1, T1c, T2, FLAIR) via an early concatenation strategy, providing the classifier with the full complement of diagnostic information used in clinical reporting.
3D Volumetric Processing: Replacing 2D slice processing with a 3D CNN operating on full volumetric patches, capturing inter-slice continuity that is routinely exploited in radiological scrolling workflows.
Grad-CAM Explainability: Integrating Gradient-weighted Class Activation Mapping overlays into the web interface to visually highlight which pixels most influenced the classification decision, supporting radiologist trust calibration.
DICOM Native Support: Implementing native DICOM file parsing to eliminate the preprocessing step currently required to convert scanner output to PNG, reducing workflow friction in radiology department integrations.
Mobile Edge Deployment: Quantizing and converting the trained model to TensorFlow Lite format for on-device inference on Android/iOS tablets, supporting field clinicians in areas with unreliable network connectivity.

8. Conclusion

This paper has presented a hybrid two-stage pipeline for automated brain tumor detection that couples marker-seeded watershed image segmentation with a four-block Convolutional Neural Network classifier. The segmentation stage isolates candidate lesion regions before classification, reducing spatial background noise and improving the quality of features available to the downstream CNN. Systematic evaluation on a 3,264-image multi-orientation MRI dataset demonstrated that this preprocessing strategy improved test classification accuracy from 85.9 percent (single-stage baseline) to 89.7 percent, with corresponding gains in sensitivity, specificity, and F1 score. Watershed segmentation quality, measured by Dice Similarity Coefficient against radiologist annotations, reached a mean of 0.873.

The entire computational pipeline was deployed as a browser-accessible Flask application that provides clinical users with segmentation overlays, confidence-calibrated predictions, and graduated risk indicators designed to reinforce appropriate deference to specialist expertise. The system's modular structure—clean separation of segmentation, classification, and deployment layers—facilitates targeted future upgrades, including multi-class tumor grading, multi-modal MRI fusion, and volumetric 3D processing. Looking forward, the convergence of interpretable segmentation-guided deep learning, efficient edge deployment, and electronic health record integration offers a credible pathway toward AI-assisted neuro-oncology screening at population scale. The goal is not to displace the specialized judgment of trained radiologists but to equip them with faster, more consistent, and geographically flexible computational tools that extend expert-quality care to patients irrespective of their proximity to a specialist center. The authors believe that responsible deployment of such systems, governed by rigorous clinical validation and transparent uncertainty communication, represents a meaningful contribution to narrowing the global brain tumor diagnostic gap.

References

[1] Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis, 36, 61–78.

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234–241.

[3] Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z., & Feng, Q. (2017). Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLOS ONE, 10(10), e0140381.

[4] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., & Xu, D. (2022). UNETR: Transformers for 3D medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 574–584.

[5] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). Efficient object localization using convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 648–656.

[6] Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. In Proceedings of the International Conference on Learning Representations (ICLR).

[7] Müller, R., Kornblith, S., & Hinton, G. (2019). When does label smoothing help? In Advances in Neural Information Processing Systems (NeurIPS), 32, 4694–4703.

[8] Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., et al. (2021). The RSNA-ASNR-MICCAI BRATS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314.

[9] Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66.

[10] Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(6), 583–598.

[11] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

[12] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.