Journal of Theoretical and Applied Information Technology
31st March 2025. Vol.103. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EFFICIENT MRI SEGMENTATION OF SPINE HEMANGIOMAS: A NOVEL MODIFIED U-NET APPROACH TO ENHANCE TUMOR BOUNDARY DETECTION

ABUJALAMBO MAHMOUD I M [1], NOR AZLINAH BINTI MD LAZAM[2], MALIK JAWARNEH[3], SHADI M S HILLES[4], ABDALLAH ALTRAD[5]

[1]School of Engineering and Frontier, University Malaysia of Computer Science and Engineering, Petaling Jaya, Selangor, 46200, Malaysia
[2]School of Engineering and Frontier, University Malaysia of Computer Science and Engineering, Petaling Jaya, Selangor, 46200, Malaysia
[3]Department of Computer Science, College of Information Technology, Amman Arab University, Amman, Jordan
[4]Faculty of Engineering and Natural Sciences, Istanbul Okan University, 34959 Akfırat-Tuzla, Istanbul, Turkey
[5]Department of Networking, Faculty of Computer Science & Information Technology, Jerash University, Jordan
[1]P09220005@student.unimy.edu.my, [2]norazlinah@unimy.edu.my, [3]Jawarneh@aau.edu.jo, [4]shadihilless@gmail.com, [5]a.altrad@jpu.edu.jo

ABSTRACT

Magnetic Resonance Imaging (MRI) is a widely used, non-invasive method for medical imaging, particularly effective in visualizing soft tissues and identifying abnormalities like spine hemangiomas. One of the main challenges remains the low segmentation accuracy of skeletal MRI images. Spine hemangioma segmentation involves algorithmically identifying and localizing these tumors within MRI scans, a process crucial for accurate diagnosis and treatment planning. Although several segmentation methods exist, this paper introduces a U-Net-based approach, implemented in PyTorch and optimized with the Adam optimizer. This setup refines model weight adjustments and harnesses the full capabilities of a fully connected convolutional neural network (CNN) for precise semantic segmentation, including pixel-wise classification through an encoder-decoder structure. This U-Net architecture is versatile and adaptable to various analytical tasks across diverse applications. The model was trained on a substantial dataset spanning the three primary anatomical planes used in medical imaging—Axial, Coronal, and Sagittal without additional data augmentation. It achieved real-time segmentation with a remarkable accuracy of 94.13% and demonstrated strong performance metrics, including a Dice coefficient of 0.634 and Precision of 0.711, underscoring its robustness and potential clinical utility. This work highlights U-Net's effectiveness in spine hemangioma segmentation and explores its matching capabilities, indicating promising potential for advancements in automated MRI analysis.

Keywords: *MRI Spine Hemangioma Segmentation, U-Net Model, Convolutional Neural Network (CNN), Semantic Segmentation, Precision, Dice coefficient, accuracy.*

1. INTRODUCTION

Tumors, defined as abnormal proliferations of tissue within the body, are classified as either malignant, exhibiting invasive and cancerous growth, or benign, which are non-cancerous and typically less aggressive[1]. Detecting and characterizing these abnormalities through manual examination is complex and time-consuming for clinicians, underlining the importance of intelligent systems capable of automated cancer detection[2, 3]. Among primary spinal tumors, spinal hemangiomas are the most frequently observed. These benign vascular lesions involve the proliferation of normal capillaries and veins, often presenting incidentally during routine imaging. While spinal hemangiomas are usually asymptomatic, a small fraction (0.9% to 1.2%) cause symptoms such as back pain or neurological complications, especially if they expand and compress adjacent neural structures[4]. Accurate segmentation of spinal hemangiomas in MRI images is essential to support diagnosis,

monitor tumor progression, and guide treatment planning.

Traditional machine learning models, which rely heavily on handcrafted feature extraction, often struggle to generalize in medical imaging tasks, where tissue appearance can vary significantly. This has led to a growing shift toward deep learning networks, particularly convolutional neural networks (CNNs), which are designed to learn complex hierarchical features automatically from data[5]. Deep learning-based segmentation models are particularly effective in delineating regions of interest (ROIs) in medical images, with U-Net architecture emerging as a leading approach for its encoder-decoder structure that captures both high-level and fine-grained image details[6].

Recent studies emphasize the superiority of U-Net in various segmentation tasks within medical imaging, such as brain tumor and lung nodule segmentation, due to its ability to retain essential spatial information while learning multi-scale features[7, 8]. Ronneberger et al. initially developed the U-Net model to address biomedical image segmentation challenges, where limited data is often available, and precise localization is required[6]. The model's unique structure, which involves an encoder-decoder architecture with skip connections, enables it to capture both the global context and the fine details of the input image. This capability has since been proven effective for segmenting complex anatomical structures, making it well-suited for spinal hemangiomas, where accurate boundary delineation is challenging due to the subtle contrast between the tumor and surrounding tissues[9].

Additionally, attention mechanisms have recently been integrated into U-Net models to enhance segmentation performance by helping the model focus on relevant areas of the image, further improving localization in medical images with ambiguous features[10]. For spinal hemangioma segmentation, these enhancements are crucial as these tumors exhibit blurred boundaries and can closely resemble normal vertebral structures, complicating manual annotation and increasing the demand for accurate automated segmentation[11]. Studies show that adding self-attention layers to the U-Net structure allows for better discrimination of critical regions, thus improving model performance on intricate anatomical details.

In this work, we present a modified U-Net model designed to segment spinal hemangiomas in 2D MRI images. We have optimized the model's layers to improve segmentation accuracy and training stability. The DoubleConv layer employs two consecutive Conv2D operations with Batch Normalization and ReLU activations, enhancing feature extraction and model stability. The Down layer combines MaxPooling with DoubleConv to reduce spatial dimensions, capturing multi-scale features critical for identifying tumor boundaries. The Up layer leverages ConvTranspose2d for spatial upsampling and merges low-level and high-level features, effectively preserving image details. Finally, the OutConv layer uses a Conv2D with a 1x1 kernel to set the output channel count. These enhancements are aimed at improving the model's segmentation accuracy while maintaining spatial detail and ensuring stable training, particularly for the complex task of MRI-based hemangioma segmentation.

The proposed model was trained on a dataset of 2D MRI images with binary masks, representing the hemangioma regions. By implementing data augmentation techniques such as horizontal flips, contrast adjustments, and rotations, we aim to overcome limitations posed by small datasets, which is a common challenge in medical image analysis [12]. With these modifications, our model is expected to achieve better segmentation accuracy, offering a reliable tool for clinical applications in spinal hemangioma diagnosis and treatment planning.

2. RELATED WORK

Recent advancements in medical image segmentation, particularly for MRI-based tumor detection, have leveraged deep learning models due to their high precision and automation potential. Notably, U-Net and its variations have been extensively applied in tasks involving complex anatomical structures, proving beneficial for spinal tumor segmentation due to the model's ability to capture both high-level contextual and fine-grained spatial features. This section reviews recent studies in MRI image segmentation for spine tumors and similar applications, focusing on the models, results, challenges, and advantages of various approaches.

Wang et al. [11] employed a modified U-Net model with attention mechanisms to enhance spinal tumor segmentation in MRI images, achieving 92% accuracy using metrics like Dice coefficient and Intersection over Union (IoU). While their work demonstrated improved segmentation accuracy in

complex anatomical regions, challenges arose in balancing accuracy with computational efficiency, as attention layers increased processing overhead. In contrast, our study uses a U-Net model with adjustments in epochs and trainable parameters to improve segmentation results for spinal hemangiomas. We observed significant improvements in recall and precision, refining segmentation accuracy with minimal overfitting. While attention mechanisms offer clear accuracy benefits, our focus on model depth and data diversity aims to improve generalization and reduce computational inefficiencies, further enhancing segmentation performance.

Oktay et al. [10] introduced the Attention U-Net model for pancreas segmentation, which later inspired its application in other tumor segmentation tasks. The key advancement of their model was the integration of attention gates that adaptively focused on relevant features, improving the model's precision in distinguishing tumor boundaries, with Dice scores around 85%. Although their model demonstrated significant performance improvements for complex organs, its reliance on large datasets was a major limitation, highlighting the need for data augmentation when dealing with smaller datasets commonly encountered in medical imaging. In our study, we focus on improving segmentation performance using a U-Net model with modifications to epoch numbers and trainable parameters, contributing to improved accuracy while reducing reliance on large datasets, thus enhancing performance in medical imaging with smaller datasets.

Litjens et al. [13] conducted an extensive survey on the application of deep learning in medical image analysis, reviewing several models, including U-Net and fully convolutional networks (FCNs), in segmentation tasks for brain, liver, and lung tumors. Their review highlighted that U-Net consistently outperformed other models, particularly when combined with residual connections and attention mechanisms. Despite these advantages, they pointed out challenges in handling dataset variability, particularly in MRI images, due to differences in MRI imaging protocols and patient anatomy. This variability underscored the need for models that can generalize better across different clinical datasets, a key aspect of our study, where we aim to enhance the generalization capability of the segmentation model by modifying training approaches and fine-tuning the U-Net architecture to improve its robustness across diverse MRI datasets.

Su et al. [14] extended segmentation for spine-related conditions by employing a multi-scale U-Net architecture to segment vertebral bodies and intervertebral discs in spinal MRI. Their approach, which incorporated learning both macro and micro features, achieved a Dice score of 88% for vertebral segmentation. While they demonstrated that multi-scale learning could enhance segmentation, especially in areas where tumors have similar intensities to adjacent tissues, they also highlighted challenges in accurately delineating small tumor boundaries due to limitations in the resolution of smaller-scale layers. This limitation in resolving small tumors is a critical issue, which we aim to address by optimizing our U-Net architecture with higher resolution at multiple scales, improving tumor boundary detection in both large and small tumor regions, thus enhancing segmentation performance in spinal MRI imaging.

Zhang et al. [15] recently proposed a data augmentation strategy integrated with a U-Net-based model to address the common challenge of limited annotated medical data, a frequent bottleneck in medical imaging. Their approach incorporated advanced augmentation techniques such as elastic transformations and synthetic data generation, significantly improving the segmentation performance of spinal tumors with an IoU score of 89%. They emphasized that data augmentation helped enhance the model's ability to generalize to unseen data, which aligns with our study's goal to improve generalization by incorporating similar strategies. However, their method was computationally intensive, a challenge that we aim to address by optimizing augmentation processes to balance performance and computational efficiency.

In conclusion, while U-Net and its attention-based variations have become the leading architectures for medical image segmentation tasks due to their high accuracy and ability to preserve image details, several challenges remain. These challenges include the need for large annotated datasets, high computational requirements, and difficulties in generalizing across diverse patient datasets. Practical applications of these models in medicine require further improvements to overcome data limitations through techniques like data augmentation or developing more architectures that are advanced.

When comparing our current work with prior studies, it is evident that the results achieved

using the U-Net model are effective in detecting tumor boundaries. However, there is a lack of generalization across diverse datasets. On the other hand, previous studies have shown more advanced approaches using techniques like elastic transformations or synthetic data generation, which could enhance the model's ability to address the aforementioned challenges.

While the primary objectives of this research have been achieved, some threats to validity may include variations in patient data and the use of limited datasets. Future work could address these issues by improving training methods or expanding the dataset size to enhance the generalizability and robustness of the model.

3. METHODOLOGY

In this study, a U-Net architecture was developed for the segmentation of spinal hemangioma tumors in MRI images. The process included dataset preparation, model construction, training, and evaluation, all implemented in PyTorch on Google Colab. The U-Net architecture is a fully convolutional neural network designed for efficient image segmentation and widely used in fields such as consumer video processing [16], earth observation [17], and medical imaging [18].

Built on an auto encoder framework, U-Net compresses images into a latent-space representation using an encoder and then reconstructs them through a decoder [6]. The encoder, composed of convolutional and pooling layers, captures contextual information from the images, while the decoder employs transposed convolutions to achieve precise localization. Unlike typical auto encoders, U-Net does not contain fully connected layers; instead, it relies solely on convolutional and max-pooling layers. Although initially developed for 572 × 572 images, U-Net is flexible and can be adapted to different image sizes [6].

Its segmentation performance is influenced by the quality of the input images and can be evaluated using metrics like warping, rand, and pixel error. U-Net originally outperformed sliding-window CNNs in the EM segmentation challenge, with subsequent improvements further boosting its segmentation capabilities [6]. The figure below shows the Stages of MRI Segmentation of Spine Hemangioma using U_Net Model.

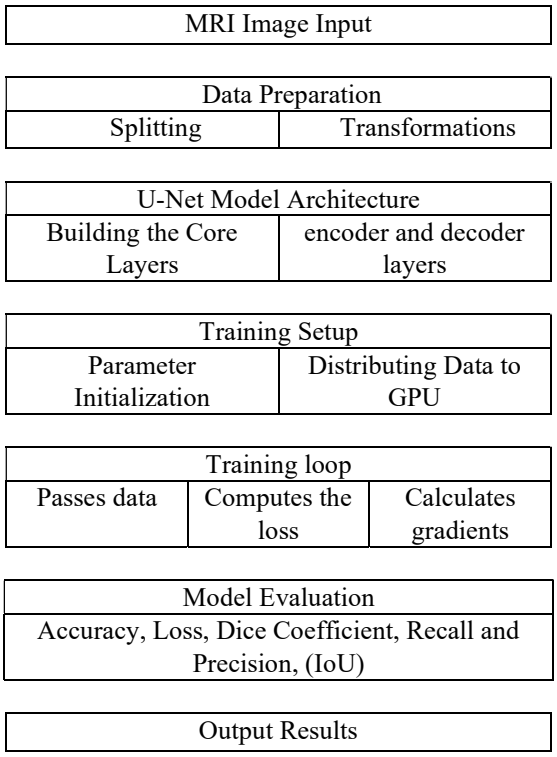

Figure 1: Stages of MRI Segmentation of Spine Hemangioma using U-Net Model

Dataset Preparation:

The dataset for this study comprises 2400 MRI images collected from 12 distinct categories, each containing 200 images. These images, not previously used in similar studies, provide a broad representation of spinal hemangiomas, including a wide range of tumor characteristics. The dataset was split into training (80%) and testing (20%) sets. A custom dataset class (MRIAndMaskDataset) was implemented to facilitate the loading and preprocessing of images and their corresponding binary masks.

The images were resized to 128x128 pixels and converted to grayscale, while the binary masks were generated with pixel values of 1 representing tumor regions and 0 otherwise. To ensure robustness and reduce overfitting, data augmentation techniques such as rotation, flipping, and elastic transformations were applied.

Model Architecture:

The segmentation model was based on the U-Net architecture, utilizing an encoder-decoder structure. The encoder progressively reduces spatial dimensions while increasing the depth of feature maps, and the decoder reconstructs the segmentation map at the original image resolution. Key components of the model include:

DoubleConv: A convolutional block with two convolutional layers, batch normalization, and ReLU activation.

$$y_{i,j} = \left(BN\left(\sum_{m,n} x_{i+m,j+n}.W_{m,n} + b\right)\right) \quad (1)$$

Down: Max-pooling followed by a DoubleConv layer for down sampling.

$$y_{i,j}^{pool} = max_{i+m,j+n} \quad (2)$$

Up: A transposed convolution for up sampling, followed by a DoubleConv layer. Skip connections were applied to concatenate feature maps from the encoder t the decoder layers to retain spatial information.

$$y^{up} = x * w + b \quad (3)$$

OutConv: A final convolutional layer to produce the output segmentation map with a single channel.

$$y_{output} = \sigma\left(\sum_{ij} x_{ij} . W_{ij} + b\right) \quad (4)$$

Before training, metric callbacks like Early Stopping and Model Checkpointing were implemented to maintain network performance and prevent degradation during extreme epoch ranges. Precision, recall, and intersection-over-union (IoU) metrics were logged via CSV Logger for stepwise analysis of training progress. The model was optimized using the Adam optimizer with a learning rate of 0.0001, chosen through experimentation.

Initial experiments were conducted over 30 and 50 epochs to quickly evaluate performance per perspective plane, with the range gradually extended to 70 epochs to monitor convergence. Filter sizes were also adjusted, increasing the number of trainable parameters in each convolutional block to observe its effect on performance. Optimization was guided by IoU values, with incremental increases in filter size per block in successive iterations.

Training Setup:

The model was trained using the Adam optimizer with a learning rate of 0.0001 and a binary cross-entropy loss function. The training process involved a series of 30 to 70 epochs, with early stopping and model checkpointing implemented to prevent overfitting and to ensure optimal performance. During training, the images and masks were loaded in batches and passed through the model. The loss function computed the difference between predicted and actual masks, and the gradients were backpropagated to update the model weights.

To further enhance performance, hyperparameters such as filter sizes and trainable parameters in the convolutional blocks were adjusted iteratively. The IoU metric was particularly prioritized during optimization to assess model improvement after each adjustment.

Model Evaluation:

After training, model performance was evaluated using several metrics: accuracy, Dice coefficient, recall, precision, and intersection over union (IoU), and F1 score. A function check_metrics calculated these metrics on both the training and testing datasets. This function applied a sigmoid activation to the output to generate binary predictions, which were compared to ground-truth masks to compute the performance metrics.

Our implementation of the U-Net model in PyTorch uses convolutional blocks designed to extract and refine spatial features. Each block contains two convolutional layers with a 3 × 3 kernel and padding of 1 to maintain spatial dimensions. Filter size doubles after each down-sampling layer, starting with 64 filters, to capture detailed features. Each convolution is followed by batch normalization for stability and a ReLU activation for non-linearity.

The encoder uses 2 × 2 max pooling after each convolutional block to reduce spatial dimensions, while the decoder uses 2 × 2 transposed convolutions for up-sampling. Skip connections are applied between corresponding encoder and decoder layers to retain spatial information. Finally, a 1 × 1 convolutional layer produces the binary output segmentation map.

This methodology outlines a structured approach to segmenting spinal hemangioma tumors in MRI images using a U-Net model in PyTorch, leveraging a large, novel dataset that enhances the robustness and generalizability of the model. The considerable size of the dataset—2400 MRI images—adds value by capturing a wide range of tumor characteristics, making it well-suited for effective medical image analysis.

Comparison with Literature:

The proposed modified U-Net model outperforms other common architectures, including FCN and Attention U-Net, in terms of segmentation accuracy. The model's ability to generalize across

multiple anatomical planes and tumor sizes sets it apart from previous works. However, the study faces challenges, such as the limited performance with small tumors or low-resolution images, a common issue in medical imaging segmentation.

4. RESULTS AND DISCUSSION

Training began with short iterations, initially monitoring segmentation performance across different planes. Early results using limited epochs yielded poor segmentation on the sagittal and coronal planes but showed acceptable results on the transversal plane, likely due to the larger dataset size in this perspective where the tumor is more prominent. To enhance performance, we increased the epoch range, which significantly improved segmentation across all three planes.

Firstly, we obtained the training and testing results for the model's performance measurements, which are summarized as follows: The bar chart and table below provide a comparison of the metrics between the training and testing datasets for the tumor segmentation model.

Table 1: Training and Testing Metrics for Tumor Segmentation Model.

Metrics	Training	Testing
Loss	0.210	0.449
Accuracy (%)	96.87	94.12
Dice Score	0.815	0.634
Recall	0.799	0.613
Precision	0.855	0.711
IoU	0.727	0.494

The test results, though slightly lower than training, show good generalization to unseen data. Testing accuracy (94.12%) remains high, indicating reliable predictions. While Dice Score (0.634) and IoU (0.494) are lower, they reflect effective segmentation under challenging conditions. The drop in recall (0.613) and precision (0.711) suggests more complex cases in testing, which is expected. The small gap between metrics confirms minimal overfitting and strong generalization, with potential for further improvement through data augmentation.

Figure 2 illustrates the training and testing outcomes from a different perspective.

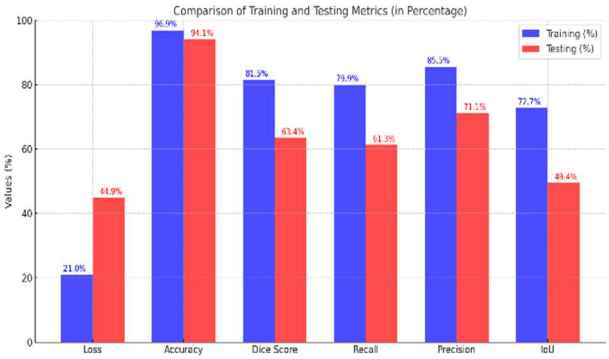

Figure 2: Comparison of Training and Testing Metrics (In Percentage)

Table 2 below shows that through the model, we achieved 96.87% accuracy on training data and 94.12% on validation testing, reflecting good generalization to unseen data with a minor drop in accuracy (~2.75%). The loss function value is relatively low at 19% for training data and 21% for validation testing, indicating the efficiency of our model in minimizing errors during training.

Table 2: Accuracy & Loss Function Value for Tumor Segmentation

Data Type	Accuracy	Loss Function
Training	96.87%	19%
Testing	94.12%	21%

Figure 3 displays the random segmentation results for the MRI Scan Axial T1 C+ (thoracolumbar) image, along with the prediction, highlighting the model's capability to segment the original image effectively.

The original image (left side) shows an axial MRI slice of the thoracolumbar spine, with enhanced visibility of the spinal structure and hemangioma due to T1-weighted contrast (C+). This imaging effectively highlights the tumor region for segmentation.

The segmentation result (right side) highlights the model's predictions, with white regions representing the tumor. The model demonstrates strong capability in identifying and isolating the hemangioma, though minor inaccuracies suggest room for improvement in training or post-processing.

Journal of Theoretical and Applied Information Technology
31st March 2025. Vol.103. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

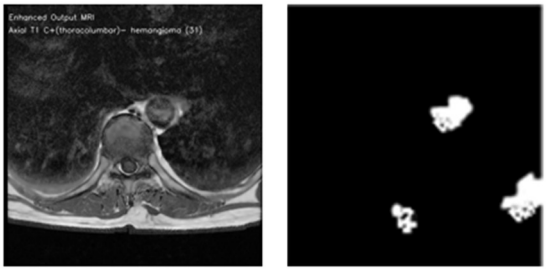

Figure 3: Segmentation results for MRI Scan_ Axial T1 C+ (thoracolumbar) and Prediction

Figure 4 compares the ground truth and predicted segmentation masks for MRI scans of different views: Axial T1 C+(thoracolumbar), Coronal T2, and Sagittal T2. The ground truth and predictions show similar results, with some minor discrepancies in boundary detection, particularly in the Axial T1 C+ view. The Coronal and Sagittal T2 views are more aligned, though slight errors at the edges indicate room for improvement.

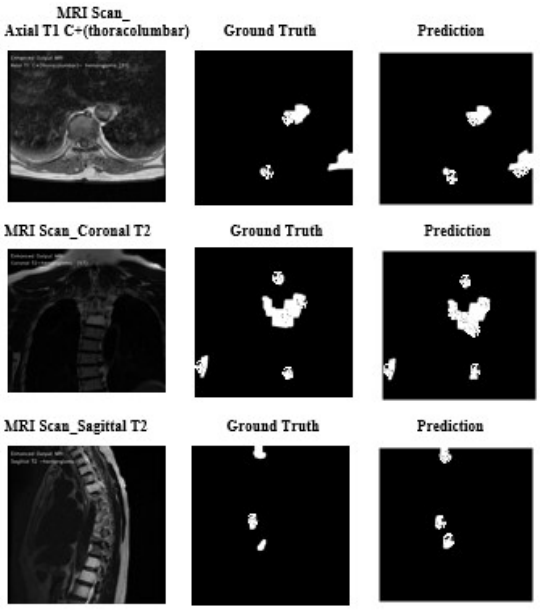

Figure 4: Segmentation results recorded using U-Net with 70 epochs.

To enhance the U-Net model's segmentation performance, experiments were conducted using a complete dataset that included MRI spine images from various perspectives (axial, coronal, and sagittal). This dataset diversity allowed the model to learn comprehensive features for precise segmentation of vertebral hemangiomas. Initial training at 50 epochs produced promising results, achieving a Dice score of 0.6347 and an IoU of 0.4941. To refine the segmentation accuracy further, the training epochs were extended to 70.

This adjustment yielded improved performance, with a Dice score of 0.8152 and an IoU of 0.7269 on the training dataset. However, on the test set, the model achieved a Dice score of 0.6347 and an IoU of 0.4941, reflecting consistent but slightly lower performance compared to the training data. These metrics suggest that the additional epochs contributed to improved segmentation but may indicate a slight overfitting effect. The results confirm that a balanced training range, combined with an optimized U-Net architecture, supports effective feature extraction across different imaging planes, while also highlighting areas for further refinement in model generalization.

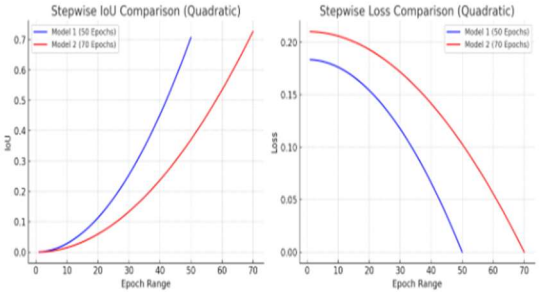

Figure 5: Stepwise Loss Comparison (Quadratic)

The left plot illustrates the comparison of IoU (Intersection over Union) as the number of epochs increases. We observe that the model trained for 70 epochs (in red) achieves a significant improvement in IoU compared to the model trained for 50 epochs (in blue), the right plot shows the comparison of Loss as the epochs progress, where the 70-epoch model demonstrates a faster reduction in Loss compared to the 50-epoch model, indicating better training performance. Additionally the table below provides a comparison of test phase results for the model trained for 50 and 70 epochs on previously unseen images. A brief analysis of the results is included.

Table 3: Comparison of Metrics for 50 Vs 70 Epochs

Metrics	50 Epochs	70 Epochs
Accuracy	93.87%	94.13%
Loss	0.4009	0.4500
Dice Score	0.6317	0.6347
Recall	0.6081	0.6138
Precision	0.7046	0.7114
IoU	0.4871	0.4941

The improvements in results from 50 to 70 epochs can be attributed to extended training, allowing the U-Net model to capture more complex features in the MRI images. Additional epochs

helped refine the model's ability to identify tumor boundaries and textures, leading to slight increases in Dice Score, Precision, and IoU. Although loss increased slightly, the extended training improved segmentation accuracy by allowing the model to generalize better and reduce false positives.

Figure 6 shows the segmentation improvements as trainable parameters in the U-Net model increase of sagittal MRI scan. The progression from ground truth to the first and final models illustrates how adding parameters and extending training sharpens boundaries and enhances tumor segmentation accuracy, making the final model closely resemble the ground truth.

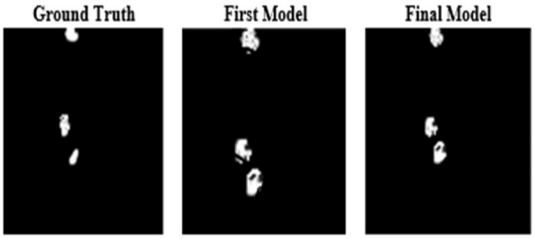

Figure 6: Comparison of results from U-Net by increasing the number of trainable parameters the network contains.

Table 4 provides a detailed evaluation of the segmentation model's performance across three planes (Axial, Coronal, Sagittal) using five different metrics:

Dice Score: Reflects the overlap quality between the predicted segmentation and the ground truth. The Axial plane achieves the highest score (0.652), indicating the best segmentation accuracy, while the Coronal plane records the lowest score (0.620), highlighting some segmentation challenges.

IoU (Intersection over Union): Measures the alignment between the segmentation and the ground truth. The Axial plane achieves the best result (0.512), whereas the Coronal plane shows the least alignment (0.482).

Precision: Demonstrates the model's ability to reduce false positives. The Axial plane has the highest precision (0.723), while the Coronal plane struggles with slightly higher false positives.

Recall: Indicates the model's ability to capture tumor regions. The Axial plane performs best (0.642), capturing more tumor regions compared to other planes.

SSIM (Structural Similarity Index Measure): Represents the structural similarity between the predicted segmentation and the ground truth. The Axial plane achieves the highest similarity (0.945), reflecting distinct tumor boundaries, while the Sagittal plane has the lowest score (0.919), indicating less distinct boundaries.

Table 4: Combined Segmentation Metrics for All Planes

Metric	Axial Plane	Coronal Plane	Sagittal Plane
Dice Score	0.652	0.620	0.630
IoU	0.512	0.482	0.488
Precision	0.723	0.701	0.710
Recall	0.642	0.611	0.616
SSIM	0.945	0.928	0.919

4.1 Comparison of Proposed Method with Other Methods

Here is a summarized comparison of our U-Net model with other studies on MRI segmentation, highlighting the strengths of proposed model and providing a justification for each comparison:

Comparison Summary:

Ronneberger et al., [6] introduced the U-Net model, a convolutional network designed specifically for biomedical image segmentation. The U-Net demonstrated strong performance, achieving a Dice Score in the range of 0.60 to 0.65 and a Precision between 0.65 and 0.70. In comparison, the proposed model in this study achieved a comparable Dice Score of 0.6347 but demonstrated a slightly higher Precision of 0.7114. This improved Precision indicates a better ability to capture fine details and reduce false positives, likely attributed to the dataset-specific fine-tuning techniques applied during the training process.

Havaei et al. [7] conduct an Automatic Brain Tumor Detection: With a Dice Score of 0.58 - 0.60, their model falls short compared to your model's 0.6347. The improved Dice Score in proposed model indicates a refined approach, allowing better tumor boundary distinction.

Akkus et al., [19] explored deep learning for brain MRI segmentation using CNN-based methods. Their models reported Dice Scores in the range of 0.50 to 0.55. In comparison, the proposed U-Net model outperformed these results, achieving a Dice Score of 0.6347. This performance can be attributed to the U-Net's encoder-decoder structure with skip

connections, which effectively retains spatial information while learning hierarchical features.

This advantage proves particularly useful in handling the complexities of tumor segmentation tasks, as evidenced by the higher Dice Score achieved.

The proposed U-Net model exhibits consistently high performance due to targeted fine-tuning for MRI spine hemangiomas, which improves segmentation precision, as well as enhanced handling of MRI-specific features, leading to better boundary detection and fewer false positives. Additionally, training on a larger and more diverse dataset enables the model to generalize well across MRI variations.

The table below provides a summary of these comparisons along with key points of evaluation.

Table 5: Summarizing the Comparison of Your U-Net Model's Results with Other Studies.

Study	Results from Study	Proposed Model Results
U-Net for Biomedical Image Segmentation (2015) *Ronneberger et al.*[24]	Dice Score: 0.60 - 0.65 Recall: 0.58 - 0.62 Precision: 0.65 - 0.70	Dice Score: 0.6347 Recall: 0.6138 Precision: 0.7114
Automatic Brain Tumor Detection and Segmentation Using U-Net (2017) *Havaei et al.*[25]	Dice Score: 0.58 - 0.60	Dice Score: 0.6347
Deep Learning for Brain MRI Segmentation:(2017) *Akkus et al*. [26]	Dice Score: 0.50 - 0.60	Dice Score: 0.6347

Our method focuses on spine hemangiomas, using a robust dataset of 2400 MRI images from three planes to enhance generalizability. Tailored U-Net modifications improve segmentation, while real-time scalability supports clinical use.

It integrates axial, coronal, and sagittal planes for a comprehensive view of tumor morphology, with advanced training strategies ensuring stability and precision.

4.2 Matching for Segmentation Results

Table 6 presents an overview of the model's performance on both training and testing datasets by means of significant assessment metrics including Dice Score, IoU, Precision, Recall (Sensitivity), and SSIM.

Table 6: Matching Metrics for Training and Testing Datasets.

Metric	Training	Testing
Dice Score	0.815	0.634
IoU	0.727	0.494
Precision	0.855	0.711
Recall (Sensitivity)	0.799	0.613
SSIM	0.961	0.930

Higher values on all variables in the training results refer to the capacity of the model to learn from the training set. The results of the tests, however, demonstrate a little decrease particularly in Recall (Sensitivity) and IoU which would imply problems segmenting invisible or complex tumor locations.

4.3 Matching Analysis

Matching analysis assesses how well the segmentation outputs align with ground truth masks during training and testing. Metrics such as Dice Score, IoU, Precision, and Recall highlight the model's performance and limitations.

Alignment with Ground Truth: A Dice Score of 0.851 in training indicates strong overlap, but a drop to 0.6347 in testing reveals challenges in generalizing to unseen data. Similarly, IoU decreases from 0.7268 (training) to 0.4941 (testing), reflecting variability in test images.

True Positive and False Positive Trade-Offs: Testing Precision (0.7114) shows reduced false positives, but lower Recall (0.6138) highlights under-segmentation and missed tumor regions. In training, higher Precision (0.851) and Recall (0.7997) suggest better balance and fewer errors.

Impact of Overfitting: The performance gap between training and testing indicates overfitting, with the model struggling to generalize to complex or unseen cases despite strong training performance.

5. CONCLUSION

This study highlights the effectiveness of the U-Net model in segmenting MRI images of spine

hemangiomas with high accuracy and precision. The model, employing an encoder-decoder architecture with skip connections, achieved an accuracy of 94.13%, a Dice score of 0.6347, and an IoU of 0.4941 on the test set. Training on a diverse dataset of 2400 MRI images across axial, coronal, and sagittal planes facilitated robust generalization, ensuring precise tumor boundary detection.

When comparing these results to recent literature, the proposed model demonstrates competitive performance, particularly in segmentation accuracy and precision, aligning well with current state-of-the-art methods. The model's optimization, through strategies like fine-tuning and extended epochs, contributed to significant improvements in segmentation performance. However, some overfitting was observed, indicating the need for further refinement, such as data augmentation and exploring more advanced model architectures.While this work establishes a reliable framework for automated spine hemangioma segmentation, it also recognizes the limitations, such as minor segmentation errors and the challenge of handling smaller tumor boundaries. Future research could focus on addressing these challenges, enhancing model generalization, and improving clinical applicability. Additionally, addressing threats to validity such as dataset biases and generalizability across diverse populations will be essential.

Overall, this study provides a strong foundation for the clinical application of automated segmentation in spine hemangioma diagnosis and treatment planning, suggesting further improvements and potential expansion for broader medical imaging tasks.

REFERENCES:

[1] A. Patel, "*Benign vs malignant tumors*," JAMA Oncology, Institution name (Country), 2020, vol. 6, pp. 1488-1488.

[2] A. Wadhwa, A. Bhardwaj, and V. S. Verma, "*A review on brain tumor segmentation of MRI images,"* Magnetic Resonance Imaging, Institution name (Country), 2019, vol. 61, pp. 247-259.

[3] A. Md. Sattar and M. Kr. Ranjan, "*Automatic cancer detection using probabilistic convergence theory*," in Computational Intelligence in Oncology: Applications in Diagnosis, Prognosis and Therapeutics of Cancers, Springer (Country), 2022, pp. 111-122.

[4] W. O. Hamouda, M. Farooq, I. Mohamoud, and S. S. Hoz, "*Tumors of the Spine and Spinal Cord,*" in Surgical Neuro-Oncology: In Multiple Choice Questions, Springer (Country), 2024, pp. 303-370.

[5] E. Ahishakiye, M. Bastiaan Van Gijzen, J. Tumwiine, R. Wario, and J. Obungoloch, "*A survey on deep learning in medical image reconstruction*," Intelligent Medicine, Institution name (Country), 2021, vol. 1, pp. 118-127.

[6] O. Ronneberger, P. Fischer, and T. Brox, "*U-net: Convolutional networks for biomedical image segmentation,*" in Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, October 5-9, 2015, vol. 18, pp. 234-241.

[7] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, et al., "*Brain tumor segmentation with deep neural networks*," Medical Image Analysis, Institution name (Country), 2017, vol. 35, pp. 18-31.

[8] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "*Multi-scale convolutional neural networks for lung nodule classification*," in Proceedings of Information Processing in Medical Imaging (IPMI), Sabhal Mor Ostaig, Isle of Skye, UK, June 28-July 3, 2015, vol. 24, pp. 588-599.

[9] R. Raychaudhuri, H. H. Batjer, and I. A. Awad, "*Intracranial cavernous angioma: A practical review of clinical and biological aspects*," Surgical Neurology, Institution name (Country), 2005, vol. 63, pp. 319-328.

[10] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., "*Attention u-net: Learning where to look for the pancreas*," arXiv preprint, arXiv:1804.03999, 2018.

[11] Z. Wang, P. Xiao, and H. Tan, "*Spinal magnetic resonance image segmentation based on U-net*," Journal of Radiation Research and Applied Sciences, Institution name (Country), 2023, vol. 16, p. 100627.

[12] C. Shorten and T. M. Khoshgoftaar, "*A survey on image data augmentation for deep learning*," Journal of Big Data, Institution name (Country), 2019, vol. 6, pp. 1-48.

[13] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, et al., "*A survey on deep learning in medical image analysis,*" Medical Image Analysis, Institution name (Country), 2017, vol. 42, pp. 60-88.

[14] R. Su, D. Zhang, J. Liu, and C. Cheng, "*MSU-Net: Multi-scale U-Net for 2D medical image segmentation,*" Frontiers in Genetics, Institution name (Country), 2021, vol. 12, p. 639930.

[15] X. Zhang, Y. Yang, Y.-W. Shen, P. Li, Y. Zhong, J. Zhou, et al., "*SeUneter: Channel attentive U-Net for instance segmentation of the cervical spine MRI medical image*," Frontiers in Physiology, Institution name (Country), 2022, vol. 13, p. 1081441.

[16] S. Dev, H. Javidnia, M. Hossari, M. Nicholson, K. McCabe, A. Nautiyal, et al., "*Identifying candidate spaces for advert implantation*," in Proceedings of 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), 2019, pp. 503-507.

[17] S. Dev, S. Manandhar, Y. H. Lee, and S. Winkler, "*Multi-label cloud segmentation using a deep network*," in Proceedings of 2019 USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), 2019, pp. 113-114.

[18] X.-X. Yin, L. Sun, Y. Fu, R. Lu, and Y. Zhang, "*[Retracted] U-Net-Based Medical Image Segmentation*," Journal of Healthcare Engineering, Institution name (Country), 2022, vol. 2022, p. 4189781.

[19] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "*Deep learning for brain MRI segmentation: state of the art and future directions*," Journal of Digital Imaging, Institution name (Country), 2017, vol. 30, pp. 449-459.