

Bayesian Optimization-Driven U-Net Architecture Tuning for Brain Tumor Segmentation [†]

Shoffan Saifullah ^{1,2,*}  and Rafał Dreżewski ¹ ¹ Faculty of Computer Science, AGH University of Krakow, 30-059 Krakow, Poland; drezew@agh.edu.pl² Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Yogyakarta 55281, Indonesia

* Correspondence: saifulla@agh.edu.pl or shoffans@upnyk.ac.id

[†] Presented at the 6th International Electronic Conference on Applied Sciences, 9–11 December 2025; Available online: <https://sciforum.net/event/ASEC2025>.

Abstract

Precise brain tumor segmentation from magnetic resonance imaging (MRI) scans is critical for clinical diagnosis and treatment planning. However, determining an optimal deep learning architecture for such tasks remains a challenge due to the vast hyperparameter space and structural variations. This paper presents a novel approach that integrates Bayesian Optimization (BO) to automatically tune the U-Net architecture for effective brain tumor segmentation. The proposed BO-UNet framework searches over encoder, bottleneck, and decoder configurations using a Gaussian Process-based surrogate model, guided by a fitness function derived from Dice Similarity Coefficient (DSC) and Jaccard Index (JI). Experiments were conducted on two benchmark datasets: the Figshare Brain Tumor Segmentation (FBTS) dataset and the BraTS 2021 dataset (focused on Whole Tumor segmentation). The best-discovered architecture [64, 64, 64, 256, 64, 128, 256] achieved notable performance: on the FBTS dataset, it reached 0.9503 DSC and 0.9054 JI; on BraTS 2021, it obtained 0.9261 DSC and 0.8631 JI, outperforming several state-of-the-art methods. Convergence and segmentation-map evolution confirm that BO effectively guided the architectural search process. These findings demonstrate the potential of BO-driven deep learning in medical imaging, opening new avenues for architecture-level optimization with minimal manual intervention.

Keywords: Bayesian Optimization; U-Net architecture; brain tumor segmentation; medical image analysis; MRI

1. Introduction

Brain tumor segmentation from MRI is essential for diagnosis, surgical planning, and treatment monitoring [1]. However, tumors exhibit high variability in size, shape, and intensity across modalities, making manual delineation slow and inconsistent. Deep convolutional neural networks, particularly U-Net, have become the foundation of medical image segmentation due to their encoder–decoder architecture and skip connections that effectively integrate local and contextual information [2].

Despite its success, U-Net performance is strongly dependent on architectural hyperparameters such as filter width, network depth, and decoder configuration [3,4]. Manual trial-and-error tuning is inefficient and often suboptimal for complex multimodal MRI data, leading to limited reproducibility and architecture bias [5,6]. Metaheuristic approaches [7], including Particle Swarm Optimization (PSO) and Genetic Algorithms (GAs), have been



Academic Editor: Lucia Billeci

Published: 9 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

applied to optimize U-Net parameters; however, their population-based nature typically requires extensive evaluations and may suffer from premature convergence, restricting scalability and robustness in architecture-level optimization [8].

Bayesian Optimization (BO) offers a more principled and sample-efficient alternative by modeling the performance landscape with a Gaussian Process (GP) surrogate and exploring promising configurations via an acquisition function [9]. Although BO has shown effectiveness in hyperparameter tuning, its application to architecture-level optimization of encoder–decoder structures for medical image segmentation remains largely unexplored. To address this gap, we propose Bayesian Optimization-Driven U-Net (BO-UNet), an automated framework that integrates BO directly into the U-Net architecture search to discover optimal encoder, bottleneck, and decoder configurations. Guided by a composite fitness function that combines the Dice Similarity Coefficient (DSC) and Jaccard Index (JI), BO-UNet enables data-driven architectural discovery tailored to segmentation performance on the Figshare Brain Tumor Segmentation (FBTS) and BraTS 2021 datasets. Unlike prior studies that focus on tuning isolated training parameters, BO-UNet formulates architectural design itself as a probabilistic optimization problem, enabling the automated discovery of compact yet high-performing U-Net structures.

The major contributions of this study are as follows:

- We propose BO-UNet, a novel framework that integrates Bayesian Optimization with U-Net for automated architecture-level tuning in brain tumor segmentation.
- We design a composite fitness function that combines DSC and JI to guide the probabilistic search toward segmentation-specific objectives.
- We validate BO-UNet on two benchmark datasets (FBTS and BraTS 2021), showing consistent improvements over manually designed and state-of-the-art models through quantitative metrics and visual analysis.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 describes the proposed method and optimization strategy; Section 4 presents the experimental results and the discussion; and Section 5 concludes the paper.

2. Related Work

Deep learning has transformed medical image segmentation, particularly for brain tumor analysis. U-Net [10] remains the foundational architecture due to its encoder–decoder design and skip connections that fuse localization and context. Numerous extensions—such as ResUNet [11], UNet++ [12], and Attention U-Net [13]—introduce residual links, nested skip paths, and attention modules to enhance feature representation. Further hybrids, including YOLO-UNet [14], Residual Attention U-Net [15,16], ViT-based UNETR [17], and ASPP-integrated U-Net [18], integrate localization cues, transformer encoders, or multi-scale aggregation to improve boundary precision.

Although these designs achieve strong results, most rely on manually crafted architectures or fixed hyperparameters, limiting scalability and reproducibility. AutoML and neural architecture search (NAS) methods provide partial automation, but are computationally heavy and rarely optimized for the fine boundary accuracy required in tumor delineation. Optimization-based approaches using Particle Swarm Optimization (PSO) [19,20] or Genetic Algorithms (GAs) [21] have improved U-Net configurations, but their population-based sampling demands many evaluations and often converges prematurely. Bayesian Optimization (BO) offers a more sample-efficient, probabilistic alternative by modeling the performance landscape with Gaussian Processes and balancing exploration–exploitation via acquisition functions [22]. However, its use for architecture-level tuning in medical image segmentation remains limited. The proposed BO-UNet addresses this gap by embedding BO directly into U-Net’s design to discover optimal encoder, bottleneck, and decoder

configurations guided by segmentation-specific metrics (DSC and JI), achieving robust performance on both the FBTS and BraTS 2021 datasets.

Although recent studies have explored Bayesian Optimization for tuning learning rates, loss functions, or attention parameters, its application to the full U-Net architectural configuration remains limited. Existing approaches largely focus on component-level optimization rather than holistic encoder–decoder design. This work addresses this gap by embedding Bayesian Optimization directly into the architectural search process, enabling the automated discovery of compact, yet high-performing, U-Net structures for brain tumor segmentation.

3. Method

This section describes the proposed Bayesian Optimization-Driven U-Net (BO-UNet) framework designed to automate the architectural tuning of U-Net for brain tumor segmentation from MRI scans. Instead of manually defining encoder–decoder widths, the framework employs Bayesian Optimization (BO) with a Gaussian Process (GP) surrogate to identify an optimal configuration that maximizes segmentation accuracy [9] while minimizing design effort.

3.1. Framework Overview

The BO-UNet procedure comprises five main stages. First, a search space is defined by specifying the number of filters for each encoder (E_1, E_2, E_3), bottleneck (B), and decoder (D_3, D_2, D_1) block. Unlike conventional hyperparameter tuning, these variables explicitly define the architecture-level structure of the U-Net model. Next, a Gaussian Process (GP) surrogate model approximates the performance landscape of these configurations. The Expected Improvement (EI) acquisition function is then used to select promising candidates, balancing the exploration of uncertain architectural regions and exploitation of high-performing configurations, which are subsequently trained and evaluated using the Dice Similarity Coefficient (DSC) and Jaccard Index (JI). Finally, the GP model is updated with the new results and the process repeats until convergence or the evaluation budget is reached.

This iterative optimization process enables BO-UNet to automatically discover an optimal encoder–bottleneck–decoder configuration without manual architectural design. By modeling segmentation performance as a black-box objective function, Bayesian Optimization efficiently navigates the discrete architectural search space while minimizing the number of costly network evaluations.

This process yields the architecture that maximizes a segmentation-specific fitness function. Figure 1 illustrates the workflow, where Bayesian Optimization dynamically adjusts convolutional filters at the architectural level to improve Dice and Jaccard performance while preserving the skip connection symmetry.

3.2. Search-Space Design

To maintain valid and balanced encoder–decoder structures, the discrete search domain was defined as:

$$E_1, E_2, E_3 \in \{32, 64, 128, 256\}, \quad B \in \{64, 128, 256, 512\}, \quad D_3, D_2, D_1 \in \{32, 64, 128, 256\}.$$

Here, E_1 – E_3 denote the encoder blocks, B represents the bottleneck, and D_3 – D_1 correspond to the decoder blocks of the U-Net architecture. Each variable directly controls the number of convolutional filters, thereby defining the architectural capacity at different resolution levels.

The search space was deliberately constrained to discrete filter sets to ensure architectural feasibility and encoder–decoder symmetry through skip connections. By sharing the same candidate filter values for encoder and decoder blocks, the proposed design prevents incompatible feature-map dimensions and avoids degenerate network configurations.

This bounded space constrains exploration to feasible U-Net variants and ensures compatibility between encoder and decoder depths, while still allowing sufficient flexibility for Bayesian Optimization to discover compact and high-performing architectural configurations.

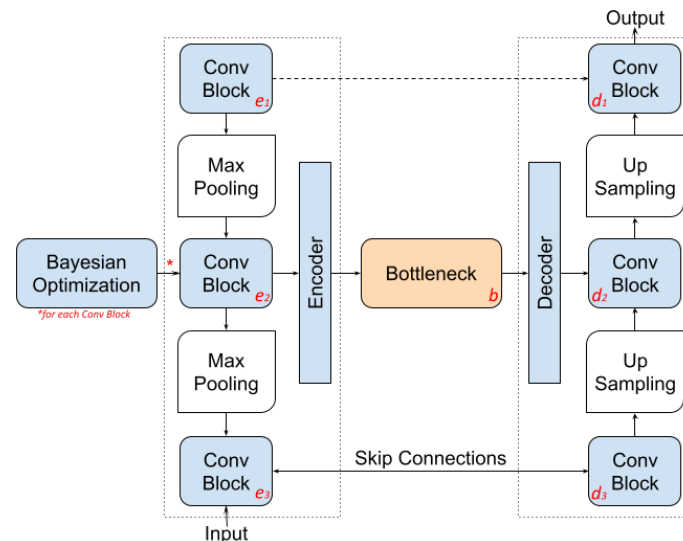


Figure 1. BO-UNet overview: Bayesian Optimization tunes encoder/bottleneck/decoder filters via a Gaussian Process surrogate and Expected Improvement to maximize Dice and Jaccard.

3.3. Bayesian Optimization Algorithm

The Algorithm 1 summarizes the optimization process. The initial architectures are randomly sampled, evaluated using DSC and JI, and used to fit the GP model. At each iteration, the EI function proposes the next configuration that maximizes the expected gain over the current best fitness.

Algorithm 1 Bayesian Optimization for U-Net Architecture Tuning

Require: Architecture search space S , maximum BO iterations T , initial sample size n , training dataset $\mathcal{D}_{train}$, validation dataset $\mathcal{D}_{val}$

Ensure: Optimal U-Net architecture x^*

- 1: Initialize empty evaluation history $\mathcal{H} \leftarrow \emptyset$
- 2: Initialize Gaussian Process (GP) surrogate model
- 3: Sample n initial architectures $\{x_1, \dots, x_n\} \subset S$
- 4: **for** $i = 1$ to n **do**
- 5: Train U-Net with x_i using $\mathcal{D}_{train}$
- 6: Evaluate DSC_i, JI_i on $\mathcal{D}_{val}$
- 7: Compute fitness $f_i = \frac{DSC_i + JI_i}{2}$
- 8: $\mathcal{H} \leftarrow \mathcal{H} \cup \{(x_i, f_i)\}$
- 9: **end for**
- 10: Fit GP surrogate model using evaluation history $\mathcal{H}$
- 11: **for** $t = n + 1$ to T **do**
- 12: $x_t \leftarrow \arg \max_{x \in S} EI(x | GP)$
- 13: Train U-Net with x_t using $\mathcal{D}_{train}$
- 14: Evaluate DSC_t, JI_t on $\mathcal{D}_{val}$
- 15: Compute fitness $f_t = \frac{DSC_t + JI_t}{2}$
- 16: $\mathcal{H} \leftarrow \mathcal{H} \cup \{(x_t, f_t)\}$
- 17: Update GP with (x_t, f_t)
- 18: **end for**
- 19: **return** Best architecture $x^* = \arg \max_{(x, f) \in \mathcal{H}} f$

Algorithm 1 follows the standard Bayesian Optimization paradigm with a fixed evaluation budget. The first loop corresponds to the initial design phase required to fit the Gaussian Process surrogate, while the second loop performs sequential model-based optimization guided by the Expected Improvement acquisition function. No explicit conditional branching is required, as exploration–exploitation trade-offs are handled probabilistically through EI. The termination is governed by the predefined maximum number of BO iterations T , and the final solution x^* is selected as the architecture achieving the highest observed fitness value across all evaluated configurations.

3.4. Fitness and Acquisition Functions

The fitness score guiding BO is defined as the mean of DSC and JI:

$$f = \frac{\text{DSC} + \text{JI}}{2} \quad (1)$$

This composite formulation emphasizes spatial overlap accuracy between predicted and ground truth tumor regions, while mitigating the bias of relying on a single segmentation metric. By jointly optimizing DSC and JI, the fitness function encourages balanced improvements in boundary alignment and region coverage.

The EI acquisition function selects the next candidate architecture:

$$\text{EI}(x) = \mathbb{E} [\max(0, f(x) - f^+)] \quad (2)$$

where f^+ is the best observed fitness.

The Expected Improvement (EI) criterion quantifies the expected gain over the current best-performing architecture by considering both the predicted mean performance and uncertainty of the Gaussian Process surrogate. As a result, EI enables an effective trade-off between the exploration of under-sampled architectural configurations and exploitation of architectures with high predicted segmentation performance.

This balances the exploration of uncertain regions with the exploitation of promising ones. In this study, the acquisition function operates over discrete architectural configurations, where each candidate x represents a complete encoder–bottleneck–decoder U-Net structure.

In the proposed framework, the Gaussian Process surrogate models the relationship between architectural configurations and segmentation performance. Distances are computed implicitly in the architectural parameter space (encoder, bottleneck, and decoder filter dimensions) through the kernel function of the GP. The Dice Similarity Coefficient and Jaccard Index are not used as distance measures within the GP; instead, they constitute scalar observations of the black-box objective function evaluated in each architectural configuration. Consequently, the optimization process remains fully automatic once the search space and evaluation budget are defined.

3.5. Best Discovered Architecture

After ten Bayesian-Optimization iterations, the best-performing configuration was obtained as:

$$[64, 64, 64, 256, 64, 128, 256] \Rightarrow [E_1, E_2, E_3, B, D_3, D_2, D_1] \quad (3)$$

This architecture represents the configuration that maximized the segmentation fitness defined in Equation (1) during the Bayesian Optimization process. This layout (Figure 2) offers an effective balance between model capacity and generalization. The shallower encoder layers capture essential local features, while the deeper bottleneck and decoder

blocks allow for detailed spatial reconstruction, preserving anatomical boundaries with minimal redundancy.

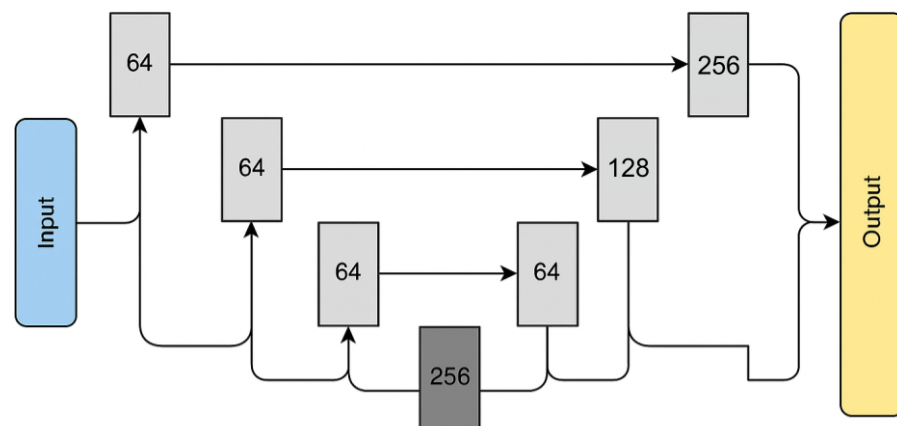


Figure 2. BO-optimized U-Net architecture ([64, 64, 64, 256, 64, 128, 256] for E_1 – E_3 , B , D_3 – D_1) using symmetric skip connections to balance feature abstraction and spatial reconstruction.

3.6. Implementation Details

The experiments were implemented in TensorFlow 2.x and executed on an HPC system with eight NVIDIA A100 (40 GB) GPUs. Input MR images were resized to 256×256 , normalized to $[0,1]$, and processed in grayscale for consistency across modalities. Training employed the Adam optimizer with a batch size of 8 and 50 epochs. Binary Cross-Entropy (BCE) was used as the loss function, and segmentation quality was assessed using accuracy, DSC, and JI metrics defined below:

- Accuracy evaluates the overall proportion of correctly classified pixels and provides a coarse measure of the voxel-wise prediction correctness. It is defined as Equation (4).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Although accuracy is reported for completeness, it may be biased in medical image segmentation tasks due to the dominance of background pixels. Therefore, overlap-based metrics are emphasized for performance interpretation.

- The Dice Similarity Coefficient (DSC) measures the overlap between the predicted segmentation (P) and the ground truth (G) (Equation (5)) and is particularly sensitive to boundary alignment and small tumor regions.

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|} = \frac{2TP}{2TP + FP + FN} \quad (5)$$

- The Jaccard Index (JI), also known as Intersection over Union (IoU), quantifies the similarity between the predicted and actual tumor regions (Equation (6)) and imposes a stricter penalty on over-segmentation compared to DSC.

$$\text{JI} = \frac{|P \cap G|}{|P \cup G|} = \frac{TP}{TP + FP + FN} \quad (6)$$

Let TP, TN, FP, and FN denote the number of true-positive, true-negative, false-positive, and false-negative pixels, respectively. Let P represent the predicted tumor mask and G the ground truth tumor mask. The operator $|\cdot|$ denotes the cardinality (i.e., the number of pixels) of a set.

- Binary Cross-Entropy (BCE) is the training loss function used to optimize voxel-wise classification probabilities and is defined as Equation (7).

$$\text{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (7)$$

where y_i and $\hat{y}_i$ represent the ground truth label and the predicted probability for pixel i , respectively.

These metrics were used both for reporting segmentation accuracy and as the foundation of the Bayesian Optimization fitness evaluation in Equation (1) [23].

4. Results and Discussion

This section presents the experimental setup, dataset details, and evaluation results of the proposed BO-UNet on two benchmark brain tumor segmentation datasets: the Figshare Brain Tumor Segmentation (FBTS) and BraTS 2021 collections.

4.1. Dataset Description and Experimental Setup

The FBTS dataset comprises T1-contrast-enhanced (T1CE) MRI slices categorized into three types of tumors: Meningioma, Glioma, and Pituitary [24]. In total, the dataset contains 3064 annotated 2D MRI slices, distributed across Meningioma (1426 slices), Glioma (708 slices), and Pituitary (930 slices). Each image and its corresponding mask were resized to 256×256 pixels and normalized to the grayscale intensity in $[0, 1]$.

The BraTS 2021 dataset contains multi-sequence MRI scans (FLAIR, T1, T1CE, and T2) with voxel-level tumor annotations [25]. The dataset includes 1251 subjects, from which axial 2D slices containing tumor regions were extracted for analysis. Following previous work, experiments targeted Whole Tumor (WT) segmentation using 2D axial slices.

For both datasets, only slices containing visible tumor regions were retained to ensure meaningful supervision and to avoid trivial background-only samples.

For both datasets, the 80:10:10 split was applied for training, validation and testing. All experiments were executed on $8 \times \text{NVIDIA A100 (40 GB)}$ GPUs using TensorFlow 2.x. Each candidate architecture was trained for 50 epochs with the Adam optimizer (learning rate = 10^{-4}), batch size = 8, and Binary Cross-Entropy loss.

Performance was evaluated using accuracy, Dice Similarity Coefficient (DSC), and Jaccard Index (JI) defined in Equations (4)–(6). For statistical significance, the Wilcoxon signed-rank test was employed to compare BO-UNet with a baseline U-Net of fixed configuration.

4.2. Convergence Behavior of BO-UNet

Bayesian Optimization was performed for ten iterations over the defined search space. Figure 3 illustrates the convergence trend of accuracy, loss, DSC, and JI. Although accuracy remained consistently high from the early iterations, DSC and JI improved markedly after iteration 4, indicating that the Gaussian Process surrogate progressively identified better architectural configurations. Loss decreased from 0.1114 to 0.0384, confirming stable convergence of the optimization process. The best architecture, [64, 64, 64, 256, 64, 128, 256], achieved the highest fitness value, reflecting an optimal trade-off between architectural compactness and segmentation quality.

Although segmentation accuracy saturates early due to the dominance of background pixels in medical image segmentation, the continuous improvement of Dice and Jaccard metrics after the fourth iteration demonstrates that Bayesian Optimization, guided by the Gaussian Process surrogate and Expected Improvement criterion, effectively refines architectural configurations rather than converging to trivial solutions. It is important to emphasize that DSC and JI are not used as comparative metrics between different

models within Figure 3. Instead, they serve as objective functions that guide the Bayesian Optimization process. Comparative evaluation with state-of-the-art methods is performed separately in Section 4.4, following standard benchmarking protocols.

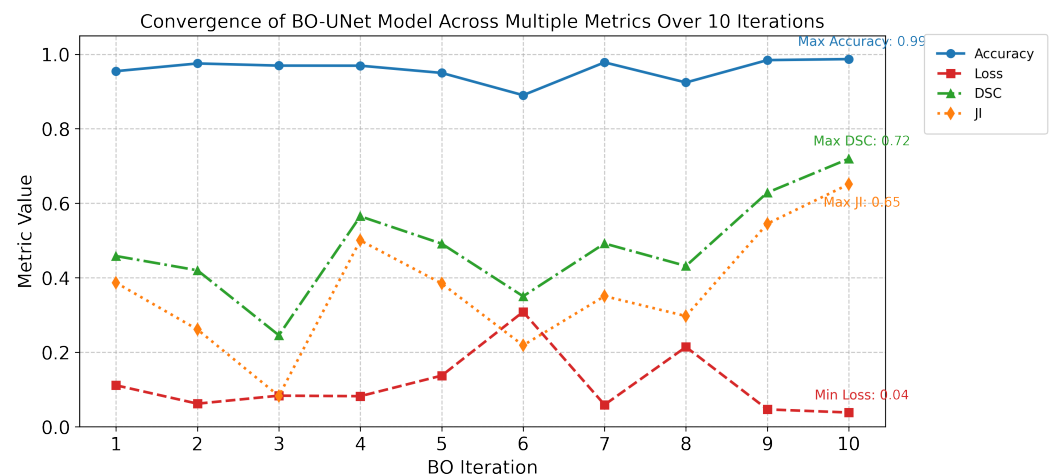


Figure 3. Convergence behavior of the Bayesian Optimization process. The evolution of the Dice Similarity Coefficient (DSC) and Jaccard Index (JI) across BO iterations reflects progressive refinement of architectural configurations guided by the Gaussian Process surrogate. Loss and accuracy are shown for training stability analysis and are not used as optimization objectives.

4.3. Quantitative Evaluation and Statistical Analysis

FBTS Dataset. Table 1 reports the results for Meningioma, Glioma, and Pituitary classes. BO-UNet consistently achieved very high accuracy (>0.998) and strong overlap metrics across all categories. The Pituitary class reached the highest performance (DSC = 0.9559, JI = 0.9156), likely due to its well-defined structural boundaries. The minimal differences between the training and testing results indicate a stable convergence and a strong generalization of the BO-optimized architecture.

BraTS 2021 Dataset. Table 1 summarizes the results for the four MRI modalities. The FLAIR modality yielded the best results (DSC = 0.9456, JI = 0.8970), reflecting its superior tumor-to-tissue contrast. T2 and T1CE followed closely, while T1 performed slightly lower due to weaker contrast in non-enhanced regions.

Statistical Validation. Wilcoxon signed-rank tests (Table 2) confirmed statistically significant improvements of BO-UNet over the baseline U-Net across all tumor types and modalities ($p < 0.01$ for DSC and JI). Median gains ranged between +0.011 and +0.021 in DSC, demonstrating that Bayesian Optimization yields consistent and non-trivial performance improvements rather than random variation.

Table 1. Segmentation results of BO-UNet on the FBTS and BraTS 2021 datasets.

Class/Modality	Training				Testing			
	Acc	Loss	DSC	JI	Acc	Loss	DSC	JI
FBTS								
Meningioma	0.9986	0.0035	0.9374	0.8826	0.9987	0.0032	0.9509	0.9066
Glioma	0.9981	0.0044	0.9375	0.8826	0.9984	0.0038	0.9441	0.8941
Pituitary	0.9995	0.0012	0.9551	0.9125	0.9995	0.0012	0.9559	0.9156
BraTS 2021								
FLAIR	0.9973	0.0065	0.9348	0.8779	0.9975	0.0060	0.9456	0.8970
T1	0.9960	0.0097	0.9020	0.8220	0.9964	0.0086	0.9143	0.8423
T2	0.9971	0.0070	0.9298	0.8691	0.9968	0.0076	0.9327	0.8743
T1CE	0.9969	0.0075	0.9228	0.8569	0.9963	0.0090	0.9119	0.8386

The consistently high accuracy values (>0.99) reflect the strong class imbalance inherent in medical image segmentation, where background voxels dominate. Therefore, overlap-based metrics such as DSC and JI provide more discriminative insight into segmentation quality.

Table 2. Wilcoxon signed-rank test results for BO-UNet vs. baseline U-Net.

Class/Modality	DSC (<i>p</i> -Value)		JI (<i>p</i> -Value)	
	Median	<i>p</i>	Median	<i>p</i>
FBTS				
Meningioma	+0.016	0.0023	+0.014	0.0041
Glioma	+0.011	0.0065	+0.012	0.0084
Pituitary	+0.014	0.0047	+0.013	0.0053
BraTS 2021				
FLAIR	+0.021	0.0019	+0.019	0.0032
T1	+0.014	0.0096	+0.015	0.0125
T2	+0.018	0.0042	+0.017	0.0067
T1CE	+0.013	0.0078	+0.012	0.0109

These results confirm that the observed improvements are statistically robust and consistent across datasets and modalities.

4.4. Qualitative Results and Comparison with SOTA

Figure 4 shows representative visual results for both datasets. Across modalities, BO-UNet delineates tumor boundaries with high spatial consistency, particularly in FLAIR and T2 images where edema regions are better captured. The predicted masks exhibit close alignment with the ground truth annotations, especially along tumor boundaries and irregular regions. In FBTS samples, all tumor types are accurately localized with minimal false positives.

The comparison results with state-of-the-art methods are summarized in Table 3. BO-UNet achieves the highest DSC and JI values among the compared methods on both datasets, indicating consistent performance gains across different tumor types and imaging modalities. BO-UNet outperforms transformer- and attention-based models such as ViT-Self-Attention and Attention U-Net.

Unlike transformer- and attention-based models that introduce additional parameters and training complexity, BO-UNet achieves competitive or superior segmentation performance through principled, architecture-level Bayesian Optimization. These results demonstrate that efficient probabilistic exploration of U-Net architectural configurations can surpass manually engineered or heavily parameterized designs while maintaining computational efficiency.

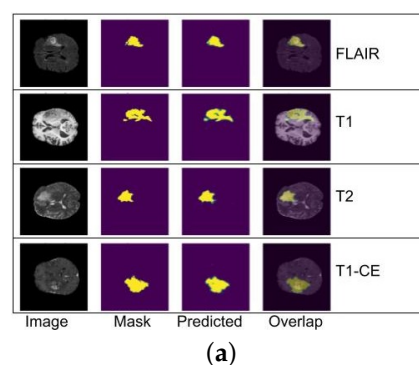


Figure 4. Cont.

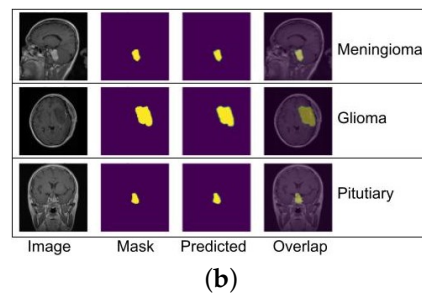


Figure 4. Qualitative segmentation results on the (a) BraTS 2021 dataset across four MRI modalities (FLAIR, T1, T2, and T1CE), and (b) FBTS dataset across Meningioma, Glioma, and Pituitary tumor classes. The columns represent: original image, ground truth mask, predicted mask, and overlap.

Table 3. Comparison of BO-UNet with state-of-the-art methods on FBTS and BraTS 2021 datasets.

Method	DSC	Jl
FBTS Dataset		
BO-UNet (Proposed)	0.9503	0.9054
EfficientNet-B4 [26]	0.9339	0.8795
Self-Attention U-Net [27]	0.9327	0.7800
YOLO-UNet [14]	0.9273	0.8915
Residual-Attention U-Net [28]	0.9110	0.8930
BraTS 2021 Dataset		
BO-UNet (Proposed)	0.9261	0.8631
U-Net-ASPP-EVO [18]	0.9251	–
ViT-Self-Attention [29]	0.9174	–
U-Net-AG [13]	0.9095	0.8323
U-Net (baseline) [10]	0.8600	0.7807

4.5. Limitations and Future Work

Although BO-UNet achieves consistent accuracy and strong generalization, each optimization iteration entails training a complete model, leading to substantial computational cost. Future extensions will focus on: (i) multi-objective Bayesian Optimization to jointly optimize accuracy and efficiency; (ii) hybrid BO–metaheuristic strategies (e.g., BO–PSO) to accelerate convergence; and (iii) cross-dataset and federated validation to evaluate robustness under domain shifts. Despite these constraints, the results indicate that BO-based architectural tuning provides a scalable and reproducible pathway toward a fully automated medical image segmentation design.

5. Conclusions

We presented BO-UNet, a Bayesian Optimization-driven framework that automatically performs architecture-level tuning of U-Net models for brain tumor segmentation by modeling performance with a Gaussian Process and selecting candidate architectures via Expected Improvement. Guided by a DSC + Jl fitness, BO-UNet discovered compact encoder–bottleneck–decoder configurations that achieved up to DSC = 0.9559/Jl = 0.9156 on FBTS and DSC = 0.9456/Jl = 0.8970 on BraTS 2021, with median DSC gains of +0.01–0.02 over a fixed U-Net (Wilcoxon $p < 0.01$), demonstrating statistically significant and consistent improvements across datasets and imaging modalities. Unlike conventional approaches that rely on manual architectural design or population-based metaheuristics, BO-UNet formulates U-Net architecture selection as a probabilistic optimization problem, enabling efficient exploration of the architectural search space with fewer costly evaluations.

while preserving segmentation accuracy and generalization capability. These findings underscore the value of probabilistic, data-driven architecture search for medical image segmentation and its ability to reduce manual design effort, while future work will explore multi-objective BO strategies that jointly consider accuracy and computational efficiency, hybrid BO–metaheuristic schemes for faster convergence, and broader validation across modalities and clinical cohorts.

Author Contributions: Conceptualization, S.S.; Data Curation, S.S.; Formal analysis, S.S. and R.D.; Funding Acquisition, S.S. and R.D.; Investigation, S.S.; Methodology, S.S.; Project Administration, S.S. and R.D.; Resources, S.S.; Software, S.S.; Supervision, R.D.; Validation, S.S. and R.D.; Visualization, S.S.; Writing—Original Draft, S.S. and R.D.; Writing—Review and Editing, S.S. and R.D. All authors have read and agreed to the published version of the manuscript.

Funding: Research funding was provided by AGH University of Krakow (Program “Excellence initiative—research university”), Polish Ministry of Science and Higher Education funds assigned to AGH University of Krakow, and ACK Cyfronet AGH (Grant no. PLG/2024/017503 and PLG/2025/018784).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Datasets are available at (1) https://figshare.com/articles/dataset/brain_tumor_dataset/1512427 (accessed on 2 July 2025) and (2) <https://www.cancerimagingarchive.net/analysis-result/rsna-asnr-miccai-brats-2021/> (accessed on 2 July 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rasool, N.; Bhat, J.I. A Critical Review on Segmentation of Glioma Brain Tumor and Prediction of Overall Survival. *Arch. Comput. Methods Eng.* **2025**, *32*, 1525–1569. [\[CrossRef\]](#)
2. Jiangtao, W.; Ruhaiyem, N.I.R.; Panpan, F. A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation. *IET Image Proc.* **2025**, *19*, e70019. [\[CrossRef\]](#)
3. Aljohani, A. Enhancing medical image segmentation through stacked u-net architectures with interconnected convolution layers. *Egypt. Inform. J.* **2025**, *31*, 100753. [\[CrossRef\]](#)
4. Abueed, O.; Wang, Y.; Khasawneh, M. A Systematic Review of U-Net Optimizations: Advancing Tumour Segmentation in Medical Imaging. *IET Image Proc.* **2025**, *19*, e70203. [\[CrossRef\]](#)
5. Kadhim, K.A. Integrating Spatial Pyramid Pooling for Multi-scale Brain Tumor Classification in Deep Learning. *J. Image Graph.* **2025**, *13*, 315–324. [\[CrossRef\]](#)
6. Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312. [\[CrossRef\]](#)
7. Saifullah, S.; Dreżewski, R.; Yudhana, A.; Caesarendra, W.; Huda, N. Bio-Inspired Metaheuristics in Deep Learning for Brain Tumor Segmentation: A Decade of Advances and Future Directions. *Information* **2025**, *16*, 456. [\[CrossRef\]](#)
8. Thapliyal, S.; Kumar, N. Comprehensive performance metric for bio-inspired optimizers: A generation-wise population convergence approach toward global optimality. *Iran J. Comput. Sci.* **2025**, *8*, 843–891. [\[CrossRef\]](#)
9. Ramalakshmi, K.; Krishna Kumari, L. U-Net-based architecture with attention mechanisms and Bayesian Optimization for brain tumor segmentation using MR images. *Comput. Biol. Med.* **2025**, *195*, 110677. [\[CrossRef\]](#)
10. Hernandez-Gutierrez, F.D.; Avina-Bravo, E.G.; Zambrano-Gutierrez, D.F.; Almanza-Conejo, O.; Ibarra-Manzano, M.A.; Ruiz-Pinales, J.; Ovalle-Magallanes, E.; Avina-Cervantes, J.G. Brain Tumor Segmentation from Optimal MRI Slices Using a Lightweight U-Net. *Technologies* **2024**, *12*, 183. [\[CrossRef\]](#)
11. Murmu, A.; Kumar, P. A novel Gateaux derivatives with efficient DCNN-Resunet method for segmenting multi-class brain tumor. *Med Biol. Eng. Comput.* **2023**, *61*, 2115–2138. [\[CrossRef\]](#)
12. Wisaeng, K. U-Net++DSM: Improved U-Net++ for Brain Tumor Segmentation with Deep Supervision Mechanism. *IEEE Access* **2023**, *11*, 132268–132285. [\[CrossRef\]](#)
13. Saifullah, S.; Dreżewski, R.; Yudhana, A.; Wielgosz, M.; Caesarendra, W. Modified U-Net with attention gate for enhanced automated brain tumor segmentation. *Neural Comput. Appl.* **2025**, *37*, 5521–5558. [\[CrossRef\]](#)

14. Davar, S.; Fevens, T. Enhanced U-Net Architecture for Brain Tumour Localization and Segmentation in T1-Weighted MRI. *IEEE Trans. Circuits Syst. II Express Briefs* **2025**, *72*, 993–997. [[CrossRef](#)]
15. Yadav, A.C.; Kolekar, M.H.; Zope, M.K. Modified Recurrent Residual Attention U-Net model for MRI-based brain tumor segmentation. *Biomed. Signal Process. Control* **2025**, *102*, 107220. [[CrossRef](#)]
16. Saifullah, S.; Dreżewski, R.; Yudhana, A.; Suryotomo, A.P. Automatic Brain Tumor Segmentation: Advancing U-Net with ResNet50 Encoder for Precise Medical Image Analysis. *IEEE Access* **2025**, *13*, 43473–43489. [[CrossRef](#)]
17. Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H.R.; Xu, D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Lecture Notes in Computer Science*; Crimi, A., Bakas, S., Eds.; Springer: Cham, Switzerland, 2022; Volume 12962, pp. 272–284. [[CrossRef](#)]
18. Yousef, R.; Khan, S.; Gupta, G.; Albahlal, B.M.; Alajlan, S.A.; Ali, A. Bridged-U-Net-ASPP-EVO and Deep Learning Optimization for Brain Tumor Segmentation. *Diagnostics* **2023**, *13*, 2633. [[CrossRef](#)]
19. Saifullah, S.; Dreżewski, R. Automatic Brain Tumor Segmentation Using Convolutional Neural Networks: U-Net Framework with PSO-Tuned Hyperparameters. In *Parallel Problem Solving from Nature—PPSN XVIII. PPSN 2024. Lecture Notes in Computer Science*; Affenzeller, M., Winkler, S.M., Kononova, A.V., Trautmann, H., Tusar, T., Machado, P., Back, T., Eds.; Springer: Cham, Switzerland, 2024; Volume 15150; pp. 333–351. [[CrossRef](#)]
20. Saifullah, S.; Dreżewski, R. Particle Swarm-Optimized U-Net Framework for Precise Multimodal Brain Tumor Segmentation. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '25 Companion)*, NH Malaga Hotel, Malaga, Spain 14–18 July 2025; p. 4. [[CrossRef](#)]
21. Saifullah, S.; Dreżewski, R. GA-UNet: Genetic Algorithm-Optimized Lightweight U-Net Architecture for Multi-Sequence Brain Tumor MRI Segmentation. *IEEE Access* **2025**, *13*, 175010–175024. [[CrossRef](#)]
22. Greenhill, S.; Rana, S.; Gupta, S.; Vellanki, P.; Venkatesh, S. Bayesian Optimization for Adaptive Experimental Design: A Review. *IEEE Access* **2020**, *8*, 13937–13948. [[CrossRef](#)]
23. Saifullah, S.; Dreżewski, R. Optimizing U-Net Architecture Using Differential Evolution for Brain Tumor Segmentation. In *Computational Science—ICCS 2025. Lecture Notes in Computer Science*; Lees, M., Cai, W., Cheong, S.A., Su, Y., Abramson, D., Dongarra, J.J., Sloat, P.M.A., Eds.; Springer: Cham, Switzerland, 2025; Volume 15906, pp. 403–411. [[CrossRef](#)]
24. Cheng, J.; Huang, W.; Cao, S.; Yang, R.; Yang, W.; Yun, Z.; Wang, Z.; Feng, Q. Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition. *PLoS ONE* **2015**, *10*, e0140381. [[CrossRef](#)]
25. Baid, U.; Ghodasara, S.; Mohan, S.; Bilello, M.; Calabrese, E.; Colak, E.; Farahani, K.; Kalpathy-Cramer, J.; Kitamura, F.C.; Pati, S.; et al. RSNA-ASNR-MICCAI-BraTS-2021 Dataset. 2023. Available online: <https://www.cancerimagingarchive.net/analysis-result/rsna-asnr-miccai-brats-2021/> (accessed on 2 July 2025).
26. Preetha, P.; Priyadarsini, J.P.M.; Nisha, J.S. Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis. *Sci. Rep.* **2025**, *15*, 9914. [[CrossRef](#)]
27. Alkhalid, F.F.; Salih, N.Z. Implementation of biomedical segmentation for brain tumor utilizing an adapted U-net model. *Comput. Biol. Med.* **2025**, *194*, 110531. [[CrossRef](#)]
28. Sajid Hussain, S.; Wani, N.A.; Kaur, J.; Ahmad, N.; Ahmad, S. Next-Generation Automation in Neuro-Oncology: Advanced Neural Networks for MRI-Based Brain Tumor Segmentation and Classification. *IEEE Access* **2025**, *13*, 41141–41158. [[CrossRef](#)]
29. Ghazouani, F.; Vera, P.; Ruan, S. Efficient brain tumor segmentation using Swin transformer and enhanced local self-attention. *Int. J. Comput. Assist. Radiol. Surg.* **2023**, *19*, 273–281. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.