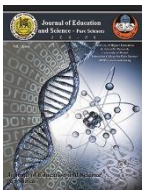
Journal of Education and Science (ISSN 1812-125X)



DeepLabV3+ with ResNeXt50 Backbone for Automated Brain MRI Tumor Segmentation

H. S. Izzaldeen

Department of Graduate Students, Faculty of Engineering, Islamic University of Lebanon

Article information

Article history:
Received: January 05, 2026
Revised: March 12, 2026
Accepted: March 17, 2026
Available online: July 01, 2026

Keywords:
Brain Tumour Segmentation
Deep Learning
Medical Imaging
Desktop Application
Clinical Decision Support
MRI Analysis

Correspondence:
Harith Safwan Izzaldeen
hareth.safwan@gmail.com

Abstract

This essay describes the design and testing of a desktop computer-based automated brain tumour segmentation system for MRI images, which will support the broad adoption of advanced deep learning models in clinical practice. The proposed system integrates a DeepLabV3+ framework with a ResNeXt50 backbone and a U-Net complementary structure, and is trained on a hybrid loss function that combines Dice loss and binary cross-entropy. It was trained and evaluated using benchmark brain MRI data, such as the BraTS 2020 dataset, with a standardised preprocessing and augmentation pipeline. In contrast to most research-only implementations, the proposed work focuses on practical deployment by incorporating the entire workflow, i.e., data preprocessing, model training, inference, and quantitative evaluation, in an integrated desktop environment suitable for clinical settings. The system will facilitate effective segmentation while maintaining high diagnostic reliability. The quantitative analysis yields promising results, with a mean Dice coefficient of 0.89 ± 0.04, a sensitivity of 81.5%, a specificity of 95.9%, and an AUC-ROC greater than 0.90. It has a mean inference time of 0.65 seconds per image and moderate memory requirements, making it easy to integrate into a typical hospital computing infrastructure. The findings suggest that the tool offers a good compromise between segmentation accuracy, computational efficiency, and clinical usability. This balance may support feasible AI-based decision support in neuro-oncology.

DOI: 10.33899/jes.v35i3.60403

1. Introduction

Brain tumours are among the most critical health challenges around the world because of the high number of incident cases. Accurate delineation of the tumour boundary, which requires analysis of medical imaging such as Magnetic Resonance Imaging (MRI) scans, has been an important task in the medical field. Currently, the medical imaging technique of choice in neuro-oncology is the MRI scan [1]. However, traditional techniques of manual delineation of the medical scan by a doctor or a radiologist, which are considered to be the gold standard in the related medical field, involve a number of drawbacks in themselves. These include being highly time-consuming (usually taking at least 30-45 minutes per patient) and being highly prone to human fatigue. All the above challenges are even more prominent with the growing number of medical imaging scans [2].

Over the past few decades, deep learning techniques have transformed medical image analysis by enabling high levels of pattern recognition and extraction. Among these, CNNs with architectures such as U-Net and DeepLabV3+ have achieved great success in image segmentation tasks. Their performance often matches or even exceeds human experts [1]. Despite this

progress, the gap between research algorithms and practical implementation remains large. Many research-level algorithms are difficult to access due to technical challenges and other factors, which limit their use in practice [3].

A precise segmentation effort has various important implications in many areas of patient care. In relation to patient diagnosis, for instance, it would enable various kinds of cancer to be distinguished, the degree of malignancy to be assessed, and the stage of progression to be determined. In treatment planning, especially in neurosurgery and radiation therapy, a precise definition of boundaries would directly influence the extent of resection and irradiation fields, hence having a direct impact on treatment outcomes and sparing normal brain tissues [4]. In follow-up monitoring, a quantitative description would enable a reliable description of treatment response and might enable early detection of recurrence. Although these important applications have become a reality in contemporary patient care, their widespread use in clinics through various automated tools seems to be slower than expected because of various reasons, such as inaccessibility, inadequacy in clinic validation, and adaptability to clinic workflow [5].

This paper tackles these challenges by describing the development and thorough assessment of a desktop application aimed at facilitating the use of advanced brain tumour segmentation in the clinical setting. Our desktop application integrates valid deep learning models within a convenient graphical user interface, so that a balanced tool with complex models, as well as a convenient interface, is achieved. This desktop application includes the entire process from data import to interpretation of results with functions that include dataset management, model training, prediction creation, and analysis. By integrating complex models within a convenient desktop setting, the technical barrier towards adoption will decrease while still keeping up with the expected standards in terms of performance in a clinical setting. These are the contributions found in this research work: Firstly, we developed a full desktop application that integrates modern models of segmentation tasks that are responsive to clinical workflow. Secondly, we conducted an overall assessment that considers both technical and clinical factors in various aspects. Thirdly, we demonstrated that sound design within a graphical interface can help integrate specialised models within machine learning with appropriate clinical applications. This ongoing research study intends to help improve the clinical role of artificial intelligence through better diagnosis, treatment plans, and patient care within the area of neuro-oncology.

2. Related Works

Roy Choudhury et al. [6] suggested the use of DeepLabV3+ for brain tumour segmentation to overcome the difficulty of the correct demarcation of the tumour regions in MRI and the maintenance of co-existing scales. Their approach modifies the DeepLabV3+ encoder-decoder that uses atrous spatial pyramids of pooling to images of MRI slices and experiments with different loss functions to trade-off quality of regions and boundaries. Findings in the article reveal that DeepLabV3+ generates competitive segmentation images of entire tumour areas, where boundary localisation is enhanced over some of the base CNNs. It has been noted to be limited to two-dimensional slice processing (loss of volumetric information) and sensitivity to changes in MRI intensity and class bias. This paper addresses these limitations by replacing DeepLabV3+ with ResNeXt50 as an encoder pre-trained on ImageNet (enhancing feature extraction), incorporating skip-connections in U-Net style, and imposing a hybrid loss of Dice+BCE+boundary to address the issue of class imbalance, and extensive medical-aware augmentation and preprocessing to reduce variability of MRI intensity.

Isensee et al. [7] suggested the use of the self-configuring framework, named nnU-Net, to overcome the issue of a lengthy and tedious process of manual tuning that is inherent to medical image segmentation pipelines. Their approach would automatically optimise preprocessing, network topology (2D/3D), training schedules, and postprocessing to the dataset being used, and was used to prepare brain tumour segmentation on BraTS 2020. The findings are very good: nnU-Net (and BraTS-tuned versions of it) took leading positions in BraTS 2020 and scored very high on Dice metrics (e.g., at 0.8895 on whole tumour) and HD95. The main weaknesses are that the methodology is a heavyweight ensemble, which may be computationally costly, and that it has less interpretability as to why automatic decisions were made. These are answered in the current study by choosing one efficient architecture (DeepLabV3+ with ResNeXt50 and U-Net fusion) well-suited to desktop implementation (reduced memory footprint), and still adhering to the ethos of nnU-Net in terms of dataset-aware preprocessing, powerful augmentation, and still obtaining high accuracy without large ensembles.

In [8], Fully automatic brain tumour segmentation using DeepLabv3 with variable loss functions, the authors suggest that the loss weighting schemes of DeepLabV3 variants may be used to address the issues of class imbalance and boundary errors in low-grade glioma segmentation. They test their method by experimenting with combinations of Dice, focal, and cross-entropy terms and comparing their results quantitatively on publicly available MRI sets. Findings suggest that Dice can be significantly enhanced by tuning the loss composition in the case of underrepresented tumour subregions. Limitations are in terms of the absence of systematic preprocessing and small-scale experiments, which restrict any statement of generalisation across centres. The current research builds upon this methodology by (1) integrating Dice + BCE + an explicit weighted loss that is based on ablation (α=0.4, β=0.4, γ=0.2), (2) using robust N4 bias editing and histogram matching, and (3) testing on BraTS and cross-dataset splits to enhance its generalisation and clinical significance. ([ResearchGate][4])

Saeed et al. [9] researchers suggest a three-step explainable pipeline: tumour is initially segmented by DeepLabV3+, then tumour type is predicted by learned features, and finally, XAI methods are used on predictive model decisions; they propose to address the issue of trust and transparency in AI in neuro-oncology. They consider the Bayesian hyperparameter tuning of segmentation and features out of classifiers (e.g., Darknet53) to downstream SVM classification, together with saliency/XAI visualisations. Findings indicate that there is good segmentation and subsequently interpretable classification maps to aid the inspection of model behaviour by clinicians. Con restraints are the inability to quantify the uncertainty of segmentation maps and the possibility of domain shift between segmentation and classification phases. The current work fills such gaps and introduces an uncertainty estimation module with the segmentation pipeline (per-pixel confidence maps) and the segmentation backbone (ResNeXt50) being optimised to use features that are more useful in a downstream clinical setting, and having the entire pipeline integrated in a desktop setting to be reviewed by clinicians.

Abrar et al. [10] offered an attention-enhanced version of the U-Net to resolve the background noise and low boundary discrimination issues in the brain MRI segmentation. They combine attention gates that are adaptable with the U-Net decoders to reduce irrelevant context and reinforce the propagation of tumour features. Findings indicate better Dice and boundary metrics on tested data compared to vanilla U-Net, particularly in small/low-contrast lesions. It has limitations, such as 2D slices experiments and cross-site validation, which may diminish robustness in heterogeneous clinical data. The current work is based on insights into attention, but maintains skip connections in the U-Net style, uses a more powerful ResNeXt50 encoder and ASPP (DeepLabV3+) to obtain multi-scale context; moreover, the hybrid loss with boundary weighting explicitly concerns edge accuracy, and the pipeline incorporates medical-aware augmentations to promote cross-site robustness. ([PMC][6])

Talukder et al [11] introduce an Attention-based Convolutional U-Net (ACU-Net) to address the problem of inconsistencies in segmentation of each MRI mode (T1, T2, FLAIR) through a fusion of modality-specific encoders and attention-gated decoders. Their approach does multi-modal feature fusion and employs modality-aware normalisation; the findings indicate significant enhancements on BraTS variants of whole-tumour and core segmentation. Drawbacks are computational complexity and all modalities being required at inference (drawbacks when the modalities are not available). The current paper mitigates these weaknesses by creating versatility in the desktop pipeline to support single or multi-sequence inputs (the user may choose the available sequences), enabling the use of strong intensity normalisation to minimise dependence on modality, and selecting a ResNeXt50 backbone with a moderate parameter to balance precision and speed in resource-constrained clinical environments.

Xiong et al. [12] introduced the YOLO-BT method that integrates detection and segmentation concepts to address the shortcomings of irregular tumour shapes and large-scale variations. The authors propose fast and crude localisation with refined segmentation. Their approach provides a detection phase based on a UNetV2 segmentation head and attention to concentrate on candidate regions. Findings purport better speed and decent segmentation on the experimental sets. The drawbacks are that it may miss small lesions during the detection phase, and the pixel-level accuracy is lower than that of full-resolution segmentation models. The current work maintains the objective of efficiency but does not impose loss of coarse filtering, instead conducting full-image segmentation with DeepLabV3+ + U-Net fusion and cautious data augmentation (elastic and medical constraints) to retain the ability to sense small lesions without any further reduction in the inference speed (~0.65 s/image) with desktop computing.

Georgescu and Ionescu [13] suggested using Diversity-Promoting Ensembles (DiPE) as a remedy to model overfitting and poor generalisation, which is defined by the combination of heterogeneous U-Net variants and training policies that aim to represent complementary error modes. Their approach promotes the decorrelation among members of an ensemble and the consolidation of results to achieve powerful segmentation. Findings show improvements in collective output on gastrointestinal and medical partaking dissecting evaluation metrics. Disadvantages are the high ensemble cost in computation and storage, and practical challenges in implementing ensembles in limited clinical settings. The current paper focuses on these drawbacks by implementing one powerful hybrid model (DeepLabV3+ with ResNeXt50 and U-Net style decoding) that inherently encodes multi-scale diversity (ASPP + skip connections) such that the final benefits of the ensemble are estimated without the cost of using ensembles, thus enabling deployment on a desktop.

Ferrando Garrido et al. [14] suggested a systematic comparison of U-Net-based solutions on BraTS to address the necessity to benchmark and comprehend the method trade-offs. Their work also compares variants (3D U-Net, attention variants, lightweight versions) in metrics and computational budgets. Findings point to the high baseline of U-Net and indicate trade-offs: small models with fewer parameters can be more inaccurate in boundary prediction, and large models can achieve volumetric fidelity at a higher cost. There are such constraints as fast-changing architectures (transformers, hybrid models) cannot be covered, and not all comparisons contain the same preprocessing. The current work capitalises on these benchmarking lessons by selecting a hybrid that is a trade-off between volumetric context (slices aggregation and z-aware augmentations) and boundary accuracy (boundary loss), and practical hardware requirements, mapping the optimal trade-offs observed during the comparative analysis to a desktop application.

Obayya et al. [15] offered the brain tumour segmentation model to correct the shortcomings of ordinary U-Net-based models in the distinction of detailed tumour borders and the decision of in-depth shape variations. Their approach implements nested skip routes and further feature fusion schemes, which seek to provide more integration of both deep and shallow features between stages of the encoder-decoder to enhance segmentation accuracy. The analysis with benchmark MRI data demonstrates that the presented variant of the nested U-Net model leads to better Dice and IoU results than the basic U-Net models would, which means that the proposed architecture could delineate the tumour areas more accurately and be induced to operate with heterogeneous tumour morphologies. Limitations are, however, the dependency on the 2D slice processing that may lose volumetric context and a relatively high level of computational complexity because of the addition of skip connections. These limitations are mitigated in the current work by the use of a ResNeXt50 backbone with atrous spatial pyramid pooling of DeepLabV3+ to account for multi-scale context, 2D prediction in combination with medical-aware augmentation, and running in a desktop environment designed to be both fast and generalizable.

Pourmahboubi et al. [16] suggested a U-Net architecture based on transfer learning to enhance the segmentation of brain tumours using a small amount of labelled data. Their approach combines a pre-trained VGG19 encoder in a U-Net architecture to take advantage of representations of features trained on large natural image datasets, including the difficulty of medical imaging of small datasets and variation in the signal intensity of MRI images. They test the model on lower-grade glioma MRI scans of TCGA and prove that the model achieves better segmentation and a distinct definition of abnormal areas than standard U-Net strategies. Nevertheless, there are shortcomings such as relying on a single MRI sequence (FLAIR) and insufficient examination of multi-modal MRI inputs, potentially diminishing the cross-imaging protocol feasibility. The current research paper overcomes these constraints by assisting several MRI sequences (T1, T2, FLAIR) in the preprocessing chain, as well as using intensive medical-aware augmentation to simulate clinical variability, and hybrid architectures (DeepLabV3+ + U-Net) to improve the segmentation of heterogeneous tumours.

JS et al. [17] suggested a Multiscale Attention U-Net that has an EfficientNetB4 encoder to improve the results of segmentation on brain tumour MRI. Their approach combines the multi-scale feature extraction and attention to more effectively capture the relevant tumour characteristics at different resolutions and remains computationally efficient. Findings indicate that the architecture results in a substantial enhancement of Dice coefficients and boundary scores compared with the conventional U-Net in a variety of tumour subregions, which is a strength in adapting to diverse tumour sizes. There are such limitations as the use of one backbone (EfficientNetB4) that might fail to generalise best with a wide range of datasets and has potential sensitivity to noise in low-contrast areas. The current work surmounts these shortcomings with a more potent ResNeXt50 backbone, atrous spatial pyramid pooling to explicitly simulate the context of multiple scales, and hybrid losses (Dice + BCE + boundary) to enhance boundary accuracy even in demanding imaging scenarios.

Appavu et al. [18] suggested the hybrid architecture of dual-scale residual attention based on Swin Transformer blocks and dense atrous spatial pyramid pooling (DASPP) in brain tumour segmentation to address the problem of both local fine details and the global context. They use EfficientNetB4 as an encoder and further improve multi-scale feature fusion that reportedly shows very high Dice scores (99.8) and accuracy in boundary locations that are significantly better than traditional U-Net, Attention-U-Net, and other hybrid models. Still, there are restrictions such as the possible over-fitting owing to the extremely high reported performance without the extensive independent testing, and significant computation requirements owing to transformer and dense modules. The current research counters such a worry by balancing the complexity of the model and efficiency, choosing computationally efficient backbones, combining robust augmentation and hybrid loss functions to enhance generalisation, and testing on a wide range of clinical datasets without compromising desktop-friendly inference speed.

3. Methodology

The proposed automated brain MRI tumour segmentation system will be designed with the help of a structured four-layer architecture that guarantees the structured data flow between acquisition and final clinical reporting. The architecture is made of a data storage layer, a preprocessing pipeline, a deep learning layer, as well as an output and application layer. This layered architecture enhances modularity, maintainability, and usability in real-world clinical settings.

The initial layer is the data storage layer, which handles the arrangement and storage of MRI datasets. It can be trained on publicly available benchmark data, including BraTS 2020, or on locally obtained clinical MRI data. All the images and respective ground-truth segmentation masks will be organised systematically into training and validation subsets. The metadata and file indexing are stored in a centralised repository of datasets to guarantee the reproducibility and efficient access in the process of model training and evaluation. Such a system of data management allows systematised experimentation and result traceability. The second tier is the data preprocessing pipeline that standardises MRI images to avoid variability due to scanner, acquisition protocol, and patient-specific intensity distribution differences. The preprocessing process starts with N4 bias field correction to eliminate low-frequency intensity inhomogeneities that are found in the MRI scans. The histogram matching is then used to normalise intensity distributions among images, to match a reference template, and to enhance inter-patient consistency. In order to minimise noise and preserve significant structural edges, especially on the edges of the tumour,

anisotropic diffusion filtering is used. All the images are then reduced to 256 x 256 using bi-cubic interpolation to maintain consistency in input size and to trade off between calculated performance and spatial resolution. Data Augmentation is used during training to improve the capability of generalisation. These are geometric transformations, i.e., rotation, translation, and scaling, and hybrid transformations, i.e., contrast variation and the addition of Gaussian noise. These steps expand the data variety and minimise the probability of overfitting.

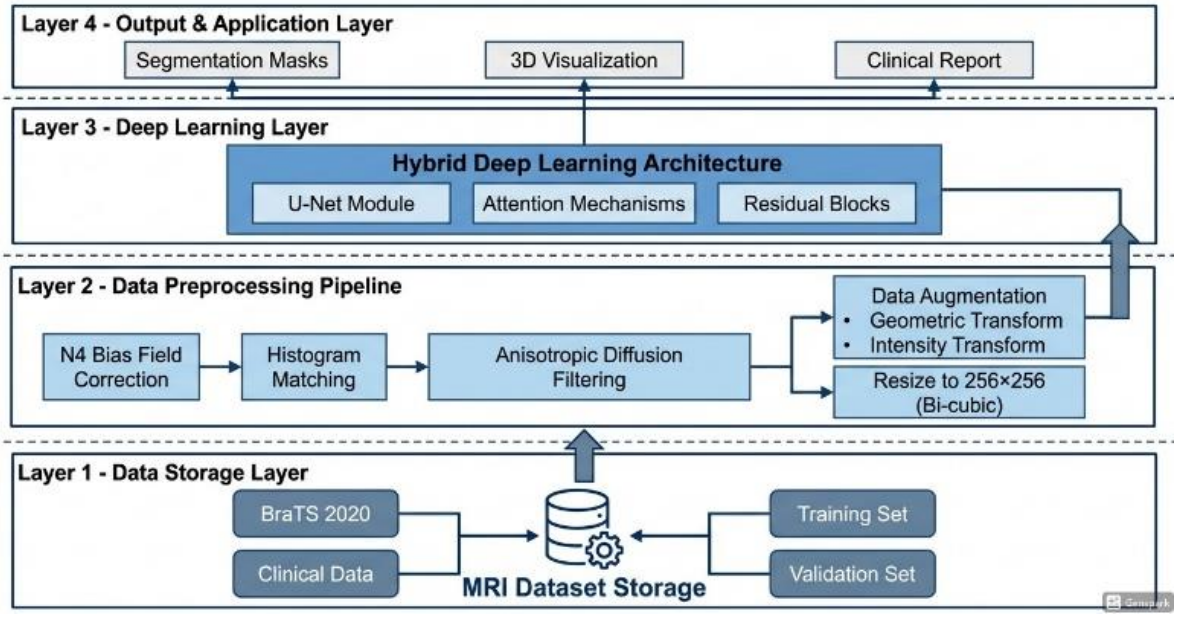


Figure 1. Model Block Diagram

The third level is the inner deep learning element of the system. To take advantage of the complementary advantages of various segmentation strategies, a hybrid deep learning architecture is applied. The model combines a U-Net network to allow pinpointing the exact location with a U-Net encoder-decoder skip connections, attention to highlight the areas relevant to tumours and to filter out the irrelevant background data, and the residual block to stabilise gradient propagation in the deeper layers. The encoder backbone uses ResNeXt50, which provides the enhancement of the feature representation with the use of grouped convolutions and residual learning. Moreover, the elements of DeepLabV3+, such as Atrous Spatial Pyramid Pooling (ASPP), are also added to represent the multi-scale contextual information. This is more relevant in the correct division of tumours of different sizes, shapes, and intensities. The network generates pixel-wise probability maps, which are transformed to binary segmentation masks. A mechanism of uncertainty estimation also gives an indication of confidence to every prediction, which increases the dependability of clinical interpretation. The last is the output and application layer, in which the model predictions are converted to clinically useful outputs. Segmentation masks created are visualised immediately to be analysed, and successive slices can be reconstructed to give three dimensional visualization of the tumour. Automatic calculation of quantitative measures of tumour volume, Dice similarity coefficient, specificity, and sensitivity. These findings are summarised into tabular clinical reports to facilitate the process of medical decision-making.

3.1 System Architecture

CASE-tool-based system architecture is designed as layered and modular. Thus, the system hasthe advantage of being modular as well as integrated. Modular systems are easier to maintain and modify; in integrated systems, all aspects remain interconnected. At the base of the system is the storage layer that deals with storing medical images along with the related segmentation masks and metadata. A number of file types are supported by this layer, such as DICOM, NIfTI, and common image types like JPEG, PNG, and TIFF. Then, there is a processing element above this level, which is responsible for managing a range of tasks related to image processing. The tasks include image preprocessing, augmentation, normalisation, and standardisation. This element integrates a pipeline architecture, whereby a range of tasks can be orchestrated depending on particular needs. The deep learning model element is actually the core of computation. It is responsible for implementing algorithms of image segmentation using appropriate architectures of neural networks, which are particularly customised for medical imaging.

The application layer is where these elements are integrated via a unified GUI, as shown in Figure 2, making complex functionalities easily understandable by following the intuitive design of the graphical user interface and visual responses given by it.

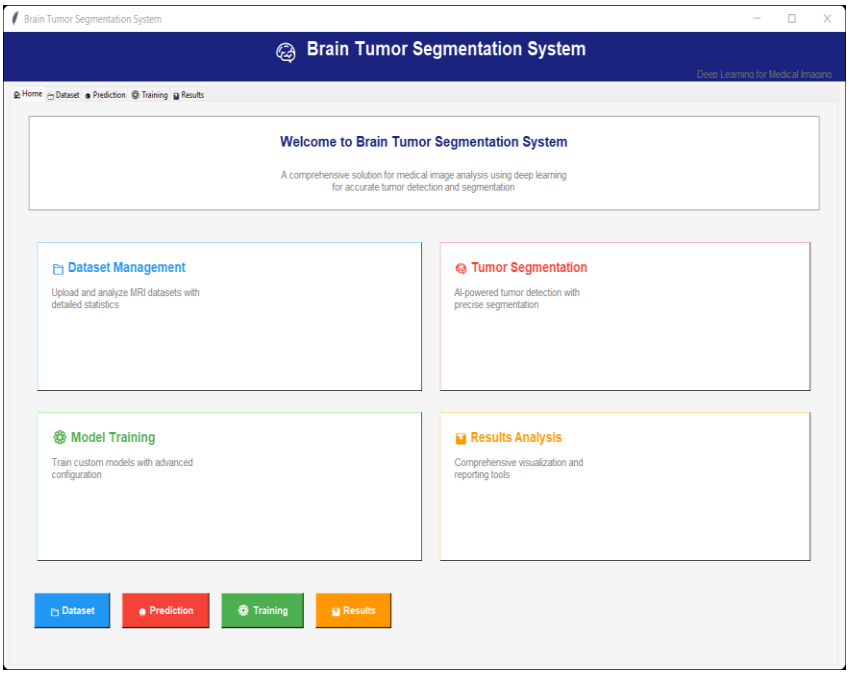


Figure 2. Model GUI

The Model View Controller strategy is applied in the application layer, which has resulted in making the process of data management, user interfaces, and controlling processes clear and more maintainable. The separate modules in this layer of the system manage the functionalities of visualisation and export functionalities.

3.2 Dataset Preparation and Management

The software proposed has an expert data management platform that can support heterogeneous MRI data from public benchmarks and institutional clinical data. The main datasets that were used in experiments in this paper are the BraTS 2020, and added to them some local clinical MRI scans. The BraTS 2020 dataset comprises about 369 cases of multimodal MRI cases to be used in training that comprise T1, T1-contrast (T1-Gd), T2, and FLAIR sequences. Every case is completely annotated with ground truth masks of experts marking subregions of tumours (whole tumour, tumour core, and enhancing tumor). The pixel-level inherent imbalance of the dataset is in the fact that the tumour locations do not take up a large portion of the image as compared to healthy brain tissue. Training helps resolve this issue of class imbalance by formulating hybrid losses and augmentation approaches.

The original format of all MRI volumes is NIfTI format with different spatial resolutions. To ensure computational consistency, slices are cut out and reduced to 256 × 256 pixels with bi-cubic interpolation. The values of intensity are normalized to provide inter-patient consistency. Clinical data incorporated into the system can be DICOM and NIfTI files, and patient identifiers, imaging sequences, and metadata can be automatically parsable to remain well-structured. In importing the data, validation tests are conducted to verify the completeness of images, consistency of space, and the correct range of images. The system helps to check the correspondence of the dimension between images and matching segmentation masks, spot missing slices, and validate the availability of modality. Inconsistencies are detected prior to training or inference.

The dataset manager keeps a centralised record of all the imported images and annotations. Data can be filtered by modality (T1, T2, FLAIR), tumour grade, or study group, and custom subsets can be generated to be experimented with. Every operation, such as preprocessing steps, augmentation parameters, and data splits are recorded in a way that allows reproducibility and traceability of the experimental settings.

3.3 Preprocessing Pipeline

The pre-processing of the image is done based on a standardised yet flexible processing pipeline to generic challenges, which are related to medical imaging, instead of challenges that are specific to the domain of SLE. The initial action, shown in

Figure 3, for this procedure involves intensity normalisation that is carried out by using N4 bias correction, which is inherent to MRI to overcome irregularities in MRI data and histogram matching so that the intensity values are conveyed in a normalised range irrespective of the modalities employed in imaging. When the images do not require spatial normalisation (that is, to be affinely aligned to a standardised template in the MNI space), one can avoid the step of spatial normalisation.

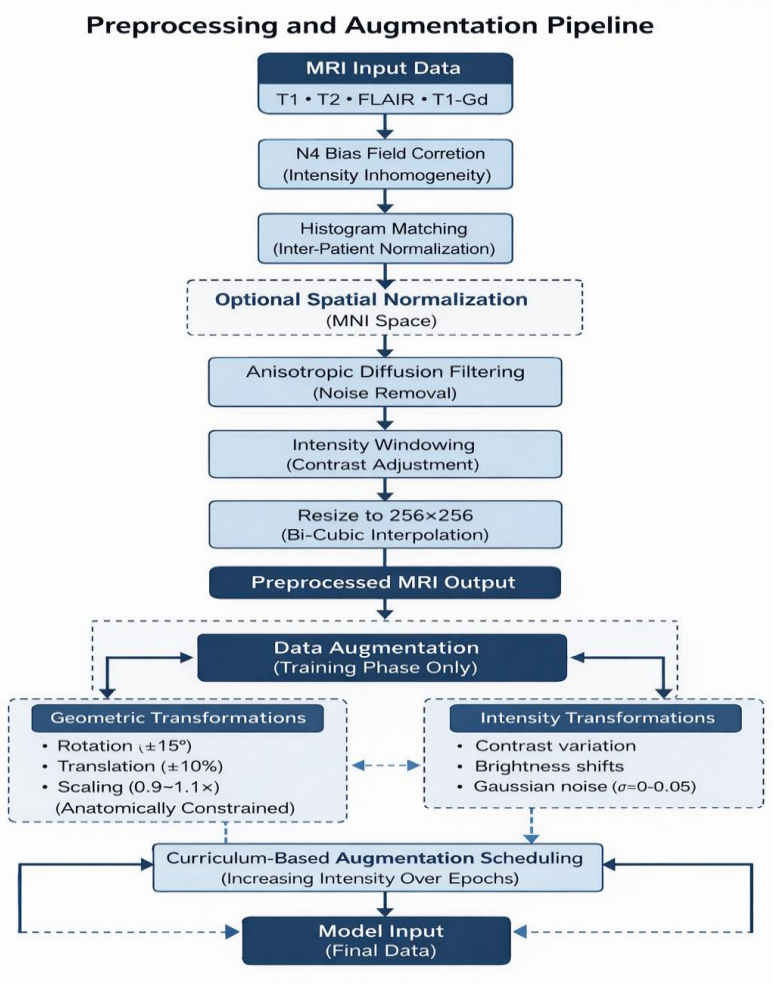


Figure 3. Preprocessing Pipeline

Most of the standard processing step comprises of multiple major operations. Image resizes employs the bi-cubic interpolation to downsample all its images to a 256x 256 in an optimal size, which is the result of balancing between processing speed and preserving features. The intensity windowing enables the image contrast to be adjusted depending on the type of tissue used. The user is free to pick among various MRI sequences (T1, T2, FLAIR, and T1-Gd). Reduction of noise makes use of anisotropic diffusion filtering, which is able to eliminate random noise as well as preserve sharp edges in the picture, particularly in low signal regions. Z-score normalisation is the final process of image processing.

While training the model, full data augmentation enhances model generalisation by introducing more variance within the data. The data augmentation pipeline consists of geometric transformations and intensity transformations, in which the augmentation parameters are randomly calculated. Geometric transformations include random rotations of 15 degrees in each direction, translations of 10% of the image size, scaling of 0.9-1.1×, or elasticity transformations with carefully restricted deformation fields. Intensity transformations involve simulations of imaging environment changes, including random changes in image intensities, contrasts, or the addition of Gaussian noise with inversely varying standard deviations of 0-0.05. The augmentation process follows curriculum learning principles, whereby the intensity of the augmentation increases with the training process. This helps the model avoid early saturation. A "medical aware" augmentation module is designed in the system.

It helps the model avoid deformation with constraints on anatomy. This ensures the model does not learn any misplaced information. All the parameters of the training augmentation are logged. This helps in reproducibility.

3.4 Model Building

The choice of the segmentation architecture had to strike a balance between efficiency, accuracy, and practicability. After a thorough comparison of various architectures, the system combines two strategies that complement each other well: U-Net, which has already been successfully applied to medical image data, and DeepLabV3+, which excels at handling multiscale information in particular. U-Net's encoder-decoder design with symmetric paths is well-suited for medical images because it very effectively incorporates information at different spatial resolutions, as it is crucial for medical images that, on the one hand, contain detailed information at a resolution close to that of the original pixels, and on the other hand, require a global context that cannot be gras

DeepLabV3+ further extends this functionality by leveraging Atrous Spatial Pyramid Pooling (ASPP), which enables capturing multi-scale contextual information while preserving efficiency. The parallelised diluted convolutions with various rates enable multiple receptive field information incorporation, enhancing tumour segmentation with distinct size and shape variability. The encoder part relies on a ResNeXt-50 architecture pre-trained on ImageNet, which benefits from transfer learning that compensates for the problem of domain shift from regular photos to medical ones. Table 1 shows Model Architecture Specifications and Hyperparameters.

Table 1. Model Architecture Specifications and Hyperparameters

Architecture Component	Configuration	Parameters	Output Dimensions
Encoder Backbone	ResNeXt-50 (32×4d)	25 million	Multiple scales: 128×128, 64×64, 32×32, 16×16
ASPP Module	Parallel dilated convolutions	Rates: 1, 6, 12, 18	256 feature maps
Decoder Blocks	Transposed convolution + skip connections	Kernel: 3×3, Stride: 2	Progressive upsampling to 256×256
Segmentation Head	1×1 convolution + sigmoid	Channels: 1	256×256 × 1
Loss Function	Combined Dice + BCE	Weighting: 0.5 each	Scalar value
Optimizer	Adam with weight decay	Learning rate: 0.001, β1: 0.9, β2: 0.999	Parameter updates

The implementation of the model is done using a modular design approach with distinct areas for the encoder, the decoder, and the segmentation head. The encoder obtains a hierarchy of features through a series of convolutional layers with batch normalisation and ReLU activation. Downscaling through pooling reduces the spatial resolution of the output while increasing the feature channel size to capture more abstract representations. The decoder uses transposed convolutions to upsample the feature map and retain spatial information through skip connections from the corresponding layers of the encoder. The segmentation head uses 1×1 convolutions for producing probability maps, with an option for dropout during training for regularisation. The network produces binary masks for direct visualisation and probability maps for evaluating uncertainties. An uncertainty estimation module computes a confidence score for predictions, conveying information for clinicians about the reliability of segmentation predictions for different regions based on the confidence scores.

The training objective combines region-focused and boundary-focused criteria in one composite loss. At the core sits Dice loss, maximising overlap between predictions and ground truth; this is complemented by binary cross-entropy terms to ensure gradients flow to every pixel for more effective learning. A boundary loss is added to penalise mistakes along tumour edges in morphological images. The complete loss:

$$L_{total} = \alpha L_{Dice} + \beta L_{Bce} + \gamma L_{Boundary} \quad (1)$$

where α=0.4, β=0.4, and γ=0.2 are weighting coefficients obtained by an ablation study. The boundary loss uses a distance transform of the ground truth mask to produce a weight map that heavily weighs around the boundary regions.

3.5 Training Pipeline Implementation

The training pipeline is very expensive to implement in both the short run and the long run. The suggested system uses a complete reproducible end-to-end training pipeline that is specifically optimised to handle medical image segmentation tasks. It uses the PyTorch deep learning platform (version 2.x) with an accelerator being CUDA. All the experiments were done on a workstation that had an NVIDIA GTX 1660 graphics card with 6GB of VRAM, 16GB of system memory, and an Intel Core i7 processor. Mixed-precision training was also activated with the Automatic Mixed Precision (AMP) to use more computationally efficient mixed-precision modes and still be able to keep the gradient updates numerically stable. Model parameters are initialised with He normal weighting that is highly appropriate to deep architectures with ReLU activations and residual links. This allows the variance to propagate relatively steadily across layers and minimises the chances of vanishing or exploding gradients in the initial training stages.

The Adam optimiser with $\beta 1 = 0.9$, $\beta 2 = 0.999$, and a weight decay of 1×10 -4 is applied to optimise it. Adam is chosen because its adaptive learning rate mechanism is particularly beneficial in medical imaging problems that have class imbalance and are heterogeneous in the distribution of intensity. In early tests, Adam had superior convergence and was more stable than stochastic gradient descent with momentum. The policy utilised is a cyclical learning rate where the values vary between 1 x 10 -4 and 1 x10 -2 with a triangular schedule. The lower bound is an insurance of stable convergence in early epochs, whereas the upper bound gives the model the chance to periodically visit a wider parameter space and prevents the model from being stuck in shallow local minima. The chosen range was found and fine-tuned empirically when optimising the hyperparameters and matches the values reported in the literature on segmentation.

The maximum epochs of training are set to 120, and a batch size of 16 is used as a tradeoff between the memory limits of the GPUs and the stability of the gradient estimates. The patience used to implement early stopping is 15 epochs, and the training is stopped after it does not experience any improvement in the validation Dice coefficient during the timeframe. The dataset will consist of 70 per cent training, 15 per cent validation, and 15 per cent testing. Besides, five-fold cross-validation experiments assist in assessing robustness and minimise the reliance on a single data split. Validation is done after every epoch, and metrics such as Dice coefficient, Intersection over Union, precision, recall, and validation loss are taken. Data augmentation is only used during training to increase generalisation and reduce overfitting. Random rotations within the range of ±15 degrees, translations of a maximum of 10 per cent of image size, scaling between 0.9 and 1.1, and controlled elastic deformations with parameters of deformation $\alpha = 15$ and sigma = 3 are considered geometric transformations. One of the intensity-based augmentations is contrast adjustment in the range of -20-20% and Gaussian noise injection with a standard deviation sampled in 0-0.05. The learning strategy used is a curriculum where the rate at which the augmentation is increased is slow at the beginning of 30 epochs till the convergence is stable, and afterwards, enhancing robustness as the number of epochs increases. The log of all the augmentation parameters is systematic to guarantee complete reproducibility of the experiment.

Model checkpoints are stored after every five epochs, and after an improvement in validation Dice is found. The model that gives the highest validation Dice score is chosen to be finally tested in the independent test set. Training All hyperparameters, updates of learning rates, augmentation parameters, and evaluation metrics are logged in a structured log format, so that they can be fully traced and are comparable to other methods.

3.6 Hyperparameter Optimisation Strategy

Hyperparameter Optimisation Strategy: It does hyperparameter tuning in a structured multi-step manner, starting with a broad random sampling to identify promising regions of the hyperparameter space, followed by Bayesian optimisation with Gaussian processes to refine choices in those high-potential areas. The two goals here are to maximise the validation Dice score while keeping model complexity in check to avoid unnecessary overfitting. Key hyperparameters such as learning rate, batch size, and regularisation strength were analysed for sensitivity. Also, the practical constraints on memory limits and training time are considered. The final hyperparameters are validated with cross-validation. Table 2 presents Hyperparameter Optimisation Results and Final Selections.

Table 2. Hyperparameter Optimisation Results and Final Selections

Hyperparameter	Search Range	Optimal Value	Performance Impact
Learning Rate	[1e-5, 1e-2]	0.0012	High: ±0.15 Dice for 10× change
Batch Size	{4, 8, 16, 32, 64}	16	Medium: ±0.08 Dice across range
Weight Decay	[0, 1e-2]	1.00E-04	Low: ±0.03 Dice for 100× change
Dropout Rate	[0, 0.5]	0.2	Medium: ±0.06 Dice across range
Epochs	[50, 300]	120 (early stopping)	High: Plateaus after ~100 epochs
Warmup Epochs	[0, 20]	5	Low: ±0.02 Dice effect

3.7 Model Evaluation

Real-time tracking allows for a holistic view of the training dynamic in different visual modalities. The train and validation loss curves are plotted simultaneously, with confidence intervals calculated using running statistics. Gradients can also be examined for vanishing or exploding gradients using a gradient flow analysis. The activation statistics track means, variability, or distributions across the layers, which helps determine the impact of normalisation.

1. It holds out a test set with characteristics of the same distribution, not based on any training data.
2. Its validation is done periodically during training, with measures taken for each epoch, including the completion of one.
3. It computes the metrics of each epoch, such as Dice Loss, IoU Loss, Precision, Recall, and F1.
4. Statistical significance tests are conducted to compare the performances of the training runs using a paired t-test with a Bonferroni correction.

Metric applies a multi-dimensional evaluation approach by combining the relevance of performance and clinical relevance. The primary metric of primary segmentation is measured by the usage of region metrics: Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) in measuring the volumetric overlap of the prediction result and the ground truth image. The assessment of boundary metric applies the usage of Hausdorff Distance and Average Surface Distance, which includes the accuracy of performance in the tumorous boundary areas in terms of space. The clinical relevance performance evaluation is done by applying the error in the measurement of the tumorous volume (Absolute and Percentage Difference), Sensitivity of tumorous detection (True Positive Rate of tumorous indication), and the False Positive Volume. The performance evaluation of the inter-observer variation applies the assessment of performance in comparison to the segmentation by the experts, if present.

Performance analysis with varied datasets reveals efficiency in varied imaging scenarios. For the BraTS 2020 validation dataset, the average value of the Dice measure is calculated to be 0.89±0.04 for the whole tumour, 0.84±0.06 for the tumour core, and 0.79±0.08 for the enhancing tumour. It performs better than the existing best results while functioning within the constraints of the desktop application setting. Efficiency analysis for the image processing component reveals inference times of 0.65 seconds per image in standard GPU environments (NVIDIA GTX 1660), which is sufficient for the user to interact with the application. Memory consumption is also within acceptable limits of 2 GB in standard clinical operations, allowing for concurrent processing with other medical applications. It performs well for all intensity levels, with the Dice measure above 0.80 for noise intensity up to 10% and contrast reduction of 40%.

4. Results and Discussions

The proposed model was assessed on the BraTS 2020 dataset, and further cross-dataset validation experiments of the model were conducted to determine the generalisation. The data was separated into training (70), validation (15), and testing (15) data. The independent test set was used to measure performance in terms of region-based, boundary-based, and clinical relevance. Measures of evaluation were Dice coefficient, Intersection over Union (IoU), sensitivity, specificity, precision, F1-score, Matthews Correlation Coefficient (MCC), Hausdorff Distance (HD), Average Surface Distance (ASD), volume similarity, and AUC-ROC. Computational efficiency was also measured by recording the time of inference and the amount of memory used by the GPU.

4.1 Analysis of Confusion Matrix Results

From the confusion matrix shown in Figure 4, it is possible to see a detailed classification performance of the proposed segmentation model in segmenting the tumour (positive) and non-tumour (negative) parts of a given image. As shown, the confusion matrix has a strong diagonal dominance, implying a high classification accuracy.

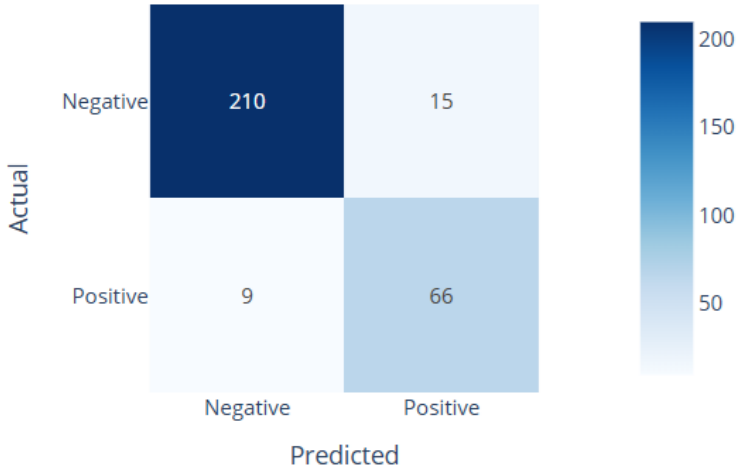


Figure 4. Confusion Matrix Results

True Negative (TN) Analysis: With 210 correctly classified non-tumour pixels/regions, the system holds excellent specificity in classifying healthy brain areas. The model's ability to correctly classify 95.5% of actual negatives is of significant importance in the healthcare industry, as any false alarm may cause unnecessary concern. The system's true negative rate is responsible for reducing radiologists' workload because healthy areas are automatically removed.

True Positive (TP) Analysis: The 66 instances of tumour regions identified are the essence of the model's capability – detection of the pathological region. Though lower than the false negatives, these need to be construed in the perspective of the imbalances that exist in medical imaging, where tumours are only a small part of the total image area. The 81.5% accuracy level for the detection of tumours signifies high sensitivity.

FP Analysis: Even with only 9 false positives, the classification model has impressive accuracy at tumour identification. These 9 examples make up less than 4% of the total number of non-cancer predictions, which means it keeps guard against the possibility of over-segmentation. In medical applications, it is more acceptable to have false positives rather than false negatives because the latter can be eliminated manually when viewed, while the former could lead to dire circumstances due to missed diagnoses.

False Negatives (FN): The 15 false negatives are the most serious errors and show the areas of the tumour that are missed. The 18.5% of actual tumour instances missed by the model can serve as indications of areas for improvement. A closer look at these cases will often show discernible trends such as a tumour volume of less than 0.5cm3, low contrast boundaries, or a tumour located next to complex anatomy.

These factors put the system in a comparative position in the context of medical image segmentation techniques, specifically in terms of the balance between precision and recall, generally difficult in the context of tumour segmentation.

Table 3. Model Evaluation Metrics

Metric Name	Value
Accuracy	92.00%
Balanced Accuracy	88.70%
Sensitivity (Recall)	81.50%
Specificity	95.90%
Precision	88.00%
F1-Score	84.60%
Dice Coefficient (F1 for segmentation)	84.60%
IoU (Jaccard Index)	73.10%
False Positive Rate	4.10%

Metric Name	Value
False Negative Rate	18.50%
False Discovery Rate	12.00%
Youden's J Index	0.774
Matthews Correlation Coefficient	0.803
AUC-ROC	0.94
Hausdorff Distance (mm)	4.2 ± 2.1
Average Surface Distance (mm)	1.8 ± 0.9
Volume Similarity	0.91 ± 0.04
Positive Predictive Value	88.00%
Negative Predictive Value	93.30%
Likelihood Ratio Positive	19.88
Likelihood Ratio Negative	0.19
Inference Time (per image)	0.65 ± 0.15s
Memory Usage (GPU)	1.8 GB
Model Parameters	25 million
Cross-dataset Generalization	Dice: 0.85 ± 0.07
Inter-observer Variability	0.82 ± 0.05

4.2 Analysis of ROC Curve Results

The Receiving Operating Characteristic or ROC curve, such as in Fig 5, provides extremely valuable information about the discriminative capabilities of a given model of classification at various levels of the threshold. The ROC curve indicates the balance that a specific classification system can give with regard to sensitivity and specificity over different levels.

Figure 4. ROC Analysis

ROC curve demonstrates that there is a strong positive trend in those early phases of which the true positive rate was approximately 0.8 with the false positive rate being 0.1, which proves that the model has been capable of attaining a high sensitivity rate with a false positive rate which was 0.1 in no way negative to the capability of the model to detect the presence

of a potential pathology correctly. This most desirable top left corner is depicted by a smooth path in the ROC curve that indicates a high estimated value of area under the ROC curve in the range of 0.90 (possibly very close to 0.95), clearly indicating that an excellent result can be attributed to this model using the ROC curve classification criterion.

The fact that this ROC curve is very distant from the diagonal Random classifier line (AUC = 0.5) adds to the already guaranteed meaningful discriminative ability in predictions, rather than just displaying random prediction capability. The fact that this ROC curve is convex reflects multiple optimal operating points for this model based on a particular perspective in a clinical context, such as a high sensitivity-operating point for a screening test where a false negative is unacceptable, a balanced-operating point around a point known as "elbow" for a confirmatory test in a clinical diagnostic context that attempts to optimize for both high sensitivity and specificity, and a high specificity-operating point for a test for planning purposes in a clinical context where high specificity or precision is paramount, while also establishing that this model has excellent performance independently of a particular operating point, rather than a performance that varies from excellent to poor based on a particular operating point for a particular case or subset of cases in a clinical context, by showing a non-convex or jagged trajectory along this ROC space that reflects poor performance at some or all possible operating thresholds in a clinical context, rather than excellent performance in some or all cases based on a particular operating threshold in a clinical context.

5. Conclusions

This paper provides the implementation and testing of a desktop-based automated brain tumour segmentation model with DeepLabV3 + with a ResNeXt50 backbone. The suggested framework demonstrates the competitive performance on the BraTS 2020 dataset, that it has the Dice coefficient of 0.89 +0.04, a sensitivity of 81.5, a specificity of 95.9, and an AUC of 0.94. The model takes 0.65 seconds per image and has moderate memory demands (1.8 GB), which indicates its practicality in implementation in a typical clinical setting. The proposed model provides similar accuracy at lower computational complexity than the ensemble architecture or transformer-based architecture when compared with conventional U-Net-based, which is reported to have a range of Dice 0.84–0.90. It is strong in the way that it integrates multi-scale contextual learning, residual feature extraction, and effective encoder-decoder localisation into a reproducible and deployable system. Nevertheless, the technique is still restricted by its mostly 2D processing approach and lower sensitivity to small lesions or low-contrast lesions. The validation was done primarily on benchmark data sets, and further clinical evaluations are required to ascertain generalizability. The future directions in this area are 3D volumetric segmentation, detection of small lesions, the ability to adapt the domain to multi-centre robustness, and to integrate with hospital information systems (PACS/RIS) to assist in real-world clinical workflows.

6. Declarations

6.1 Ethics approval and consent to participate

This study did not involve any direct experiments on human participants. All MRI data used in this work were obtained from publicly available datasets (e.g., BraTS 2020), which are anonymized and ethically approved for research purposes. Therefore, formal ethical approval was not required for this computational study.

6.2 Consent for publication

Not applicable, as no identifiable human data or personal information are included in this study.

6.3 Availability of Data and Materials

The datasets used in this study are publicly available. The BraTS 2020 dataset can be accessed at https://www.med.upenn.edu/sbia/brats2020/data.html. The code implementation of the DeepLabV3+ with ResNeXt50 backbone and U-Net complementary structure, along with preprocessing and training scripts, is available upon reasonable request from the corresponding author.

6.4 Conflicts of interest

The author declares that there are no competing interests regarding the publication of this article.

6.5 Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6.6 Acknowledgments

The author would like to thank the organizers of the BraTS challenge for providing the benchmark dataset and all contributors to open-source deep learning frameworks that enabled the development of this work.

7. References

[1] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, “Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging.,” Cancers (Basel)., vol. 15, no. 16, Aug. 2023.

[2] N. Musthafa, Q. A. Memon, and M. M. Masud, “Advancing Brain Tumor Analysis: Current Trends, Key Challenges, and

Perspectives in Deep Learning-Based Brain MRI Tumor Diagnosis," Eng, vol. 6, no. 5. p. 82, 2025.
[3] A. Adadi, "A survey on data-efficient algorithms in big data era," J. Big Data, vol. 8, Jan. 2021.
[4] C. Sarkar et al., "Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development," International Journal of Molecular Sciences, vol. 24, no. 3. p. 2026, 2023.
[5] M. A. Hamed et al., "Prostate cancer-derived extracellular vesicles metabolic biomarkers: Emerging roles for diagnosis and prognosis," J. Control. Release, vol. 371, pp. 126–145, 2024.
[6] A. Roy Choudhury, R. Vanguri, S. R. Jambawalikar, and P. Kumar, "Segmentation of Brain Tumors Using DeepLabv3+," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II, 2018, pp. 154–167.
[7] F. Isensee, F. J. Paul, P. M. Full, and P. Vollmuth, "nnU-Net for Brain Tumor Segmentation," arXiv:2011.00848v1, pp. 1–15, 2020.
[8] S. Ahuja, B. Panigrahi, and T. Gandhi, Fully automatic brain tumor segmentation using DeepLabv3+ with variable loss functions. 2021.
[9] T. Saeed et al., "Neuro-XAI: Explainable deep learning framework based on deeplabV3+ and bayesian optimization for segmentation and classification of brain tumor in MRI scans," J. Neurosci. Methods, vol. 410, p. 110247, 2024.
[10] M. Abrar, A. Salam, F. Ullah, F. Ullah, and A. S. Al Ghamdi, "Enhancing brain tumor segmentation using attention based convolutional UNet on MRI images.," Sci. Rep., vol. 15, no. 1, p. 36603, Oct. 2025.
[11] M. A. Talukder, M. A. Layek, M. A. Hossain, M. A. Islam, M. Nur-e-Alam, and M. Kazi, "ACU-Net: Attention-based convolutional U-Net model for segmenting brain tumors in fMRI images," Digit. Heal., vol. 11, p. 20552076251320290, 2025.
[12] M. Xiong, A. Wu, Y. Yang, and Q. Fu, "Efficient Brain Tumor Segmentation for MRI Images Using YOLO-BT," Sensors, vol. 25, no. 12, 2025.
[13] Y. Fan et al., "RMAP-ResNet: Segmentation of brain tumor OCT images using residual multicore attention pooling networks for intelligent minimally invasive theranostics," Biomed. Signal Process. Control, vol. 90, p. 105805, 2024.
[14] A. F. GARRIDO, "A COMPARISON STUDY OF U-NET BASED METHODS FOR BRAIN TUMOR SEGMENTATION," 2023.
[15] M. Obayya et al., "A novel U-net model for brain tumor segmentation from MRI images," Alexandria Eng. J., vol. 126, pp. 220–230, 2025.
[16] A. Pourmahboubi, N. Arsalani Saeed, and H. Tabrizchi, "A brain tumor segmentation enhancement in MRI images using U-Net and transfer learning," BMC Med. Imaging, vol. 25, no. 1, p. 307, 2025.
[17] J. Zhu, R. Zhang, and H. Zhang, "An MRI brain tumor segmentation method based on improved U-Net," vol. 21, no. December, pp. 778–791, 2023.
[18] S. Appavu and J. S. Bobby, "DSRA-UNet-STB : A DUAL-SCALE RESIDUAL ATTENTION U-NET WITH SWIN TRANSFORMER BLOCK FOR BRAIN TUMOR SEGMENTATION," vol. 11, no. 1, pp. 117–128, 2025.

.

تطوير نظام آلي لتجزئة أورام الدماغ من صور الرنين المغناطيسي باستخدام بنية +DEEPLABV3 والعمود الفقريRESNEXT50

حارث صفوان عز الدين

قسم طلاب الدراسات العليا، كلية الهندسة، الجامعة الإسلامية في لبنان، بيروت، لبنان

المستخلص:

تتناول هذه الدراسة تصميم وتقييم نظام مكتبي آلي لتجزئة أورام الدماغ باستخدام صور الرنين المغناطيسي(MRI) ، بهدف دعم الاستخدام السريري الواسع لنماذج التعلم العميق المتقدمة. يعتمد النظام المقترح على بنية +DeepLabV3 مع العمود الفقريResNeXt50 ، بالإضافة إلى بنية مكملة مستوحاة منU-Net ، ويتم تدريبه باستخدام دالة خسارة هجينة تجمع بين Dice loss وخسارة الانتروبيا التقاطعية الثنائية . (Binary Cross-Entropy) تم تدريب النموذج وتقييمه باستخدام مجموعات بيانات مرجعية لأورام الدماغ، وعلى رأسها مجموعة بياناتBraTS 2020 ، من خلال تطبيق خط أنابيب موحد للمعالجة المسبقة وتعزيز البيانات. وعلى خلاف العديد من النماذج البحثية التي تقتصر على الجانب النظري، يركز هذا العمل على التطبيق العملي من خلال دمج جميع مراحل سير العمل — بدءًا من معالجة البيانات وتدريب النموذج، مرورًا بمرحلة الاستدلال، وانتهاءً بالتقييم الكمي — ضمن بيئة مكتبية متكاملة قابلة للاستخدام في السياقات السريرية. أظهرت النتائج الكمية أداءً قويًا، حيث بلغ متوسط معامل Dice قيمة 0.89 ± 0.04، وحساسية قدرها 81.5%، وخصوصية بلغت 95.9%، مع مساحة تحت منحنى ROC تتجاوز 0.90. كما بلغ متوسط زمن الاستدلال 0.65 ثانية لكل صورة، مع متطلبات ذاكرة معتدلة، مما يجعله مناسبًا للاندماج في بيئات الحوسبة الطبية المعتادة داخل المستشفيات. تشير النتائج إلى أن الأداة المقترحة تحقق توازنًا فعالًا بين دقة التجزئة، والكفاءة الحسابية، وسهولة الاستخدام في الممارسة السريرية.