

AMIN-CNN: Enhancing Brain Tumor Segmentation through Modality-Aware Normalization and Deep Learning

Sivakumar Depuru^{id}, M. Sunil Kumar^{id}

Department of Computer Science and Engineering, Mohan Babu University, Tirupati, India

Corresponding author: Sivakumar Depuru. (e-mail: siva.depuru@gmail.com), **Author(s) Email:** M. Sunil Kumar (e-mail: sunilmalchi@gmail.com)

Abstract Accurate segmentation of reliable brain tumor detection is essential for early diagnosis and treatment, which helps to increase patient survival rates. However, the inherent variability in tumor shape, size, and intensity across different MRI modalities makes automated segmentation a challenging task. Traditional deep learning approaches, such as U-Net and its variants, provide robust results but often struggle with modality-specific inconsistencies and generalization across diverse datasets. This research presented AMIN-CNN, an adaptive multimodal invariant normalization incorporating a novel 3D convolutional neural network to improve brain tumors segmentation across various MRI technologies. Through adaptive normalization, AMIN-CNN covers modality-specific differences more effectively than Basic CNN and U-Net, leading to improved integration of multimodal MRI input data. The model maintains strong learning performance with minimal overfitting beyond epoch 50. Regularization techniques can reduce this. AMIN-CNN stands out with the best Dice Score (about 0.92 WT, 0.87 ET, and 0.89 TC), Precision (0.3), accuracy of 93.2 % and can decrease false positives. The lower Sensitivity in AMIN-CNN results in it finding the smaller but more correct tumor regions, making it more precise. Compared with traditional methods, AMIN-CNN demonstrates a competitive or better segmentation result and maintains computational efficiency. The model has demonstrated strong independence, with a Hausdorff Distance of 20, compared to 100 for other models. According to these test results, AMIN-CNN is the most effective and clinically correct method among the different architectures, mainly due to its high precision and ability to measure tumors with accuracy.

Keywords: Brain Tumor Segmentation, Adaptive Multimodal Invariant Normalization (AMIN), Convolutional Neural Network (CNN), Multimodal MRI, U-Net, BraTS Dataset.

1. Introduction

A brain tumor is an abnormal growth of tissue within the brain that can cause problems with parts of the nervous system. It is important to know that these types of tumors may be benign or malignant, and their risk depends on both their ability to spread and their location [1]. As the brain is responsible for breathing, movement, thinking, and senses, even a modest tumor can result in dangerous or severe health issues. These tumors are dangerous as they tend to increase in size, usually without noticeable symptoms. Tumors are often found after they have become advanced and caused permanent harm due to their potential to impair critical brain functions and rapidly deteriorate patient health [2]. Early and accurate detection of brain tumors is essential for effective treatment planning and improving patient outcomes. Magnetic Resonance Imaging is commonly used to detect brain tumors and generate detailed pictures using T1, T2, FLAIR, and T1 contrast-

enhancement modalities [28]. However, manual segmentation of tumors from these images is time-consuming, subjective, and prone to variability, making automated, reliable segmentation methods highly desirable [3].

Developments in computational methods have enabled deep learning to assist with medical image analysis by quickly and automatically extracting and organizing important features. Several studies have found that convolutional neural networks (CNNs), when integrated with deep learning frameworks, significantly enhance the performance in detection, classification, and segmentation of medical images [4]. Among these, U-Net [5] is a crucial choice for biomedical segmentation because it uses an encoder-decoder setup and skip connections that effectively capturing both global context and fine-grained details. Boosting performance has led to the development of a number of U-Net variations. Due to its nested and dense skip

pathways, UNet++ achieves more accurate detection of the boundaries between different tissues [6]. Additionally, DeepLab uses atrous convolutions and Conditional Random Fields (CRFs) to successfully capture objects across different scales and accurately define their boundaries [7]. Moreover, architectures like DenseNet enable features to be reused multiple times and errors to be corrected across layers, resulting in enhanced segmentation strength and fewer network parameters [8].

MRI scans with mixed modalities can be difficult to segment due to how the tumors look and their varying intensity. It is often challenging to identify the subregions within tumors, such as, enhancing tumor, edema, and necrotic core, using different MRI sequences such as T1, T2, FLAIR, and T1ce [11, 12]. Traditional CNNs often struggle to overcome these differences in data, resulting in a decline in the system's ability to generalize or segment images effectively. Therefore, more attention is being given to adaptive approaches for normalization. They ensure that features in different modalities match up, so that the learned representation is not affected by the type of input during training. Similarly, using image-specific fine-tuning [10] and combining different normalization methods in H-DenseUNet has improved the results of image segmentation by addressing changes between images and enhancing the learning from the overall context. To address these issues, we came up with the Adaptive Multimodal Invariant Normalization CNN (AMIN-CNN) structure. The framework proposes a dedicated way to normalize features within each image, making it possible to more accurately identify tumors and avoid false reports. Researchers tested the proposed method on the BraTS2020 dataset. They compared it with Basic CNN and U-Net, revealing that it performs better in identifying tumors with a higher significance to medical practitioners.

2. Related Work

Recent advancements in deep learning have significantly improved the segmentation of medical images, particularly in studies of brain tumors. While manual segmentation of brain tumors from MRI scans takes a lot of time and is not always reliable, it prompts the search for more effective automated methods. A large number of medical image analysis tasks use CNNs due to their ability to identify different features as they process raw data [4]. [5] Introduced U-Net, and it is often referred to as a core model thanks to its encoder-decoder design and skip connections that help the model save detail at each spatial level. By using nested and dense skip connections, UNet++ adds a layer to the original UNet, helping it learn better at different scales and boosting the segmentation accuracy [6]. Yet, it is not easy to identify various areas of a brain tumor in multimodal MRI since the intensities

and details found in each modality can differ. Most models struggle to handle these imaging types effectively, and current normalization techniques lack the flexibility required to bring different MRI features to a uniform level. Hence, coming up with a way that is independent of imaging method and flexible is important for better tumor segmentation accuracy.

Other models based on deep learning [32, 33] have greatly improved the process of medical image segmentation. It utilizes atrous (dilated) convolutions and Conditional Random Fields (CRFs) to enhance understanding of complex images by capturing information across multiple scales [7]. In addition, DenseNet establishes an architecture where each layer is connected to every other layer. As a result, gradient flow becomes more efficient, features are used more often, and the model ends up with fewer parameters, keeping it accurate and efficiency [9]. nnU-Net greatly improved how models can be automatically designed. It automatically adjusts the preprocessing, model organization, and training routine according to the features of each medical dataset. There is no longer a need to fine-tune the model manually, and it has been proven to work well in various medical segmentation assignments [8]. Their results on benchmark studies such as the BraTS challenge, further confirm that they work well in different and complicated medical imaging scenarios [29, 34].

In brain tumor segmentation, DeepMedic, a 3D CNN, is one of the recommended models, as it is designed to identify different areas of a tumor in three-dimensional scans by using different information scales and context from the surrounding tissues [14]. The 3D U-Net extends the classical 2D U-Net into the world of volumetric images, which makes it possible to capture more structural details from MRI scans and raise the accuracy of segmenting the tissues [15].

Despite the progress, separating brain tumors from the different types of MRI scans is still a difficult task. While every modality provides information on the tumor (for example, edema, core, or enhancing areas), the differences in light features, strong contrast, and imaging resolution between them make it difficult to use features from all modalities. It is now possible to fuse multi-modal information using approaches such as modality concatenation and multi-stream CNN models. However, they regularly encounter issues due to limitations in certain aspects of their operation, including repeated features and parameter usage, which reduce accuracy when the systems are generalized or segmented [17, 16]. To fix up these mixed-in-mode issues, researchers have been focusing more on using normalization methods recently. Traditional methods for normalization, such as Batch Normalization, Instance Normalization, and Layer Normalization, help stabilize the network during

training and accelerate its learning speed. However, these methods usually work for any kind of MRI image, however, they do not take into account that each type of MRI looks different, which makes them less helpful when trying to combine information from different kinds of MRI scans. Consequently, researchers have begun to use methods that make the feature numbers of different types of scans line up with each other more easily. MIND (Modality Independent Neighborhood Descriptor) was specifically designed for multi-modal image registration and captures areas of the image not based on intensity differences [18]. Another useful method is Adaptive Instance Normalization (AdaIN), which automatically adjusts the mean and variance of different feature maps, helping better transfer styles and make feature maps from different sources more similar [19][31]. These approaches help develop better ways to split up MRI data, so that the models can handle the different and changing patterns that come up when medical images are taken in more than one way. These improvements aside, most existing models have a hard time managing inconsistencies between modalities or within the same modality, resulting in somewhat poor performance in segmenting brain MR images with multiple modalities. Alternations in the quality of different modalities make it challenging to align and relate their key features.

Researchers have suggested using a combination of different architectural concepts to address the mentioned issues. In particular, H-DenseUNet uses densely connected blocks in the U-Net design to improve the accuracy of segmenting complex organs like the liver tumors [13]. Similarly, another class of U-Net models, attention-based U-Nets, use adaptive feature weights to improve segmentation accuracy and enable the network pay more attention to the relevant structures [20, 22]. This suggests that architectures for multimodal processing should seamlessly integrate various types of information and be adaptable to any data format. This is why AMIN-CNN proposes a new normalization layer that brings features from different modalities into harmony, and also uses reliable fusion methods. Due to this method, the brain tumor segmentation in multimodal MRI will be more precise and can be validly used on future data.

The preprocessing is the initial stage in which MRI volumes are taken through an organized pipeline to improve data quality and make them consistent across subjects. The multimodal input, $X \in R^{(HXWXCX4)}$, consists of four MRI modalities, namely T1, T1ce, T2, and FLAIR, each providing complementary anatomical information in a complementary fashion. Preprocessing steps of note include correcting the bias field where voxel intensities are modified by a predicted bias field $B(x)$ via the N4ITK algorithm to make tissue intensities uniform. This is then followed by z-score normalization

to bring the voxel intensities of the different modalities to a common range, thereby diminishing the inter-scan variability. Skull stripping is then used to remove non-brain structures such as the skull and scalp, which reduces the computational complexity and removes irrelevant features. The subsequent image registration is done to align all the modalities with each other through affine transformations $T(x) = Ax + b$ so, here anatomical correspondence between the sequences. These steps enhance the signal-to-noise ratio by combining and compensating acquisition-related variations and permit voxel-wise multimodal analysis, which is critical to performing robust tumor segmentation. Data augmentation: random rotations, flipping, elastic deformations, and intensity perturbations, was used to improve model generalization and model anatomical variability. Such transformations aid in avoiding overfitting and enhance resilience to real-world changes. Additionally, the volumes of MRI were cut into overlapping 3D patches (e.g., $64 \times 128 \times 128$) with a constant stride, to effectively utilize GPU memory effectively. Patch-wise training conserves the spatial context while allowing the network concentrate on local characteristics of the tumor. The method produces a balanced segregation in the learning of the entire brain volume.

3. Methodology

A. Dataset Overview: BraTS 2020

The BraTS 2020 dataset is a standard dataset for brain tumor segmentation and contains preoperative multimodal MRI of glioma patients. Therefore, each case in the dataset consists of four MRI modalities: T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), T2, and Fluid Attenuated Inversion Recovery (FLAIR), which contain complementary anatomical and pathological information. Together with the imaging data, each case has an accompanying expert-annotated segmentation mask (seg.nii.gz) that segments the following three tumor subregions: label 1 (NCR/NET – necrotic and non-enhancing tumor core), label 2 (peritumoral edema), and label 4 (enhancing tumor or ET). Label 0 is reserved for the background. Every volume is registered to a common anatomical template, skull-stripped to drop non-brain tissue, and re-sampled equally to a set uniformity of $240 \times 240 \times 155$ voxels as depicts Fig. 1. This standardization allows spatial compatibility and intensity consistency among the subjects and helps train the deep learning models, such as the ones that use Adaptive Multimodal Invariant Normalization (AMIN). These properties make BraTS 2020 an excellent dataset for developing research on automatic brain tumor detection and segmentation.

Since the BraTS and dataset include many variations and are complex, it makes them ideal for testing AMIN-CNN. The main purpose of the Adaptive Multimodal Invariant Normalization (AMIN) method is to overcome issues like intensity differences, incorrect contrast between images, and their lack of alignment.

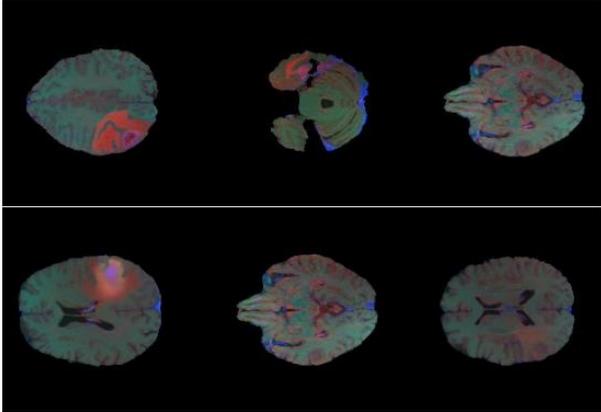


Fig. 1. Visualization of Fused Multimodal MRI Slices from the BraTS 2020 Dataset.

Normalizing each modality according to the data learned, AMIN allows for more similarity between the distributions and retains important modality-specific bits of information. This greatly improves how the features are represented when fed into the convolutional layers. In addition, the proposed model's use a CNN backbone means it excels at finding the spatial and structural patterns in brain MRI images. With the help of AMIN, the network becomes able to extract features that are not dependent on the way tumors are imaged and work efficiently in different cases. When tumors are identified at the earliest signs, small differences in both their form and brightness can cause conventional algorithms to mislabel them. A thorough evaluation of the AMIN-CNN model is possible thanks to the rich and labeled imaging data from the BraTS 2020 dataset. It offers calculation of the Dice Similarity Coefficient, Sensitivity, Specificity, and Hausdorff Distance, allowing the assessment of the model's segmentation accuracy in various ways. The inclusion of these cases also helps evaluate the model's performance at spotting tumors at early stages when they are harder to notice. The input data to AMIN-CNN is a 4-channel 3D volumetric tensor created by organizing MRI images from the T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), T2-weighted (T2), and FLAIR modalities. Every modality is represented by Eq. (1) [8]

$$X^{(m)} \in R^{D \times H \times W} \quad (1)$$

where $m \in \{T1, T1ce, T2, FLAIR\}$. With all inputs joined together along the modality axis, we get an input tensor $X \in R^{D \times H \times W \times 4}$. Every volume in the BraTS2020 dataset is shaped to $240 \times 240 \times 155$ to achieve an input of $155 \times 240 \times 240 \times 4$ samples. Every voxel holds intensity

data from four different modalities, giving the model the ability to learn features that relate different to datasets. To obtain consistent data, the multimodal signals are reprocessed using z-score normalization. After collecting the different data, the multimodal input is z-score normalized to maintain equal intensity among different samples and which helps ensure the data is analyzed properly within the AMIN-CNN architecture.

B. AMIN-CNN Architecture for Brain Tumor Detection and Segmentation

The AMIN-CNN architecture is a type of convolutional neural network that draws inspiration from U-Net and

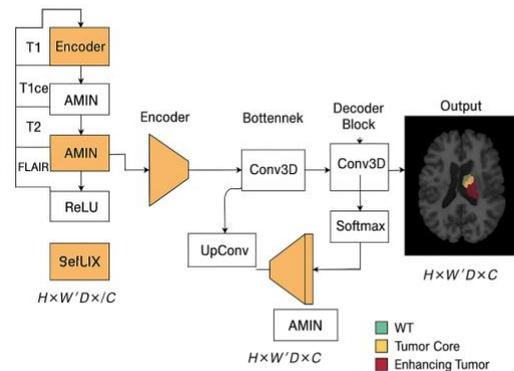


Fig. 2. The Proposed AMIN-CNN Architecture for Brain Tumor Segmentation.

DeepMedic, while also incorporating its own Adaptive Multimodal Invariant Normalization (AMIN) block, a novel component designed to enhance the model's performance. It takes in different types of MRI images like T1, T1ce, T2, and FLAIR, and then makes maps that show Fig. 2 where different parts of the tumor are in each image. Whole Tumor (WT) means looking at the whole tumor, Tumor Core (TC) is about the middle part where the cells are crowded close together, and Enhancing Tumor (ET) [30] is used to talk about the part of the tumor that shows up bright or dark on a MRI scan. The proposed architecture is a customized 3D Convolutional Neural Network (3D-CNN) optimized for brain tumor segmentation, using volumetric MRI data.

The model accepts an input volume of shape $64 \times 128 \times 128 \times 1$, representing the spatial and depth dimensions of MRI slices. It begins with a 3D convolutional layer comprising 32 filters, followed by a max pooling layer that progressively reduces the spatial dimensions. This sequence is repeated with increasing filter sizes of 64 and then 128 across three convolutional layers, each followed by batch normalization and 3D max pooling. These encoding layers enable the model to capture multi-scale contextual and spatial information critical for identifying complex tumor boundaries. The gradual increase in feature depth enables the model to learn both low-level

texture and high-level semantic features relevant to tumor detection.

A 3D kernel is applied by each convolutional layer in AMIN-CNN to focus on capturing structure and differences between tumor and normal tissue. A 3D convolution layer computes a convolutional layer in mathematics applies a set of K^{th} 3D filters $W(k) \in \mathbb{R}^{k_d \times k_h \times k_w \times C}$ producing an output feature map: Eq. (2), [25]

$$Z^{(k)} = W^{(k)} * X_l + b^{(k)} \quad (2)$$

where $W^{(k)}$ is the kernel, while $b^{(k)}$ corresponds to the bias. X_l is input feature map at layer l , $Z^{(k)}$ Output Feature Map after Convolution. The feature maps go through an activation function such as ReLU Eq. (3) [24],

$$A^{(k)} = \text{ReLU}(Z^{(k)}) \quad (3)$$

Because there are multiple convolutional layers each using thicker filter sets, the model can find both simple textures and important features related to tumors $Z^{(k)}$ on various MRI images.

The decoder section utilizes 3D transposed convolutional layers to upsample the feature maps and reconstruct the segmentation mask. It symmetrically reverses the encoding path, using feature maps with decreasing filter sizes of 128, 64, and 32. Each upsampling step is followed by batch normalization to maintain training stability and reduce internal covariate shift. The final convolutional layer outputs a single-channel volume of the same spatial dimensions as the input, producing the binary or multi-class segmentation map. With approximately 561,000 trainable parameters, the model is lightweight and computationally efficient while maintaining high performance. The integration of AMIN (Adaptive Multimodal Invariant Normalization) enables the model to effectively harmonize features from multiple MRI modalities, such as T1, T1ce, T2, and FLAIR, thereby improving segmentation consistency and reducing false positives.

The overall architecture of the proposed AMIN-CNN model is depicted in Table 1. It follows a symmetric encoder-decoder design tailored for 3D medical volumes, enabling the extraction of spatial features and contextual learning across depth slices. Each encoding block comprises a 3D convolution layer followed by max pooling and batch normalization. The decoder mirrors this structure using 3D transpose convolutions to reconstruct the segmentation map while restoring spatial resolution. The output layer produces a 3D binary segmentation mask with voxel-wise class probabilities. Batch normalization is applied after each block to ensure stable training dynamics, and the model's configuration supports the learning of multiscale tumor structures.

Table 1 Layer-wise Architecture of the Proposed AMIN-CNN Model for 3D Brain Tumor Segmentation

Layer (type)	Output Shape	Param #
Input_layer_2 (Input Layer)	(None, 64, 128, 128, 1)	0
conv3d_8 (Conv3D)	(None, 64, 128, 128, 32)	896
max pooling3d_6 (MaxPooling3D)	(None, 32, 64, 64, 32)	0
batch_normalization_9 (BatchNormalization)	(None, 32, 64, 64, 32)	128
conv3d_9 (Conv3D)	(None, 32, 64, 64, 64)	55,368
max pooling3d_7 (MaxPooling3D)	(None, 16, 32, 32, 64)	0
batch_normalization_10 (BatchNormalization)	(None, 16, 32, 32, 64)	256
conv3d_10 (Conv3D)	(None, 16, 32, 32, 128)	221,312
max pooling3d_8 (MaxPooling3D)	(None, 8, 16, 16, 128)	0
batch_normalization_11 (BatchNormalization)	(None, 8, 16, 16, 128)	512
conv3d_transpose_3 (Conv3D Transpose)	(None, 16, 32, 32, 64)	221,248
batch_normalization_12 (BatchNormalization)	(None, 16, 32, 32, 64)	256
conv3d_transpose_4 (Conv3D Transpose)	(None, 32, 64, 64, 32)	55,328
batch_normalization_13 (BatchNormalization)	(None, 32, 64, 64, 32)	128
conv3d_transpose_5 (Conv3D Transpose)	(None, 64, 128, 128, 32)	27680
batch_normalization_14 (BatchNormalization)	(None, 64, 128, 128, 32)	128
conv3d_8 (Conv3D)	(None, 64, 128, 128, 1)	33

AMIN-CNN consumes a 4-channel 3D input volume, 1 channel for each MRI modality- (native T1ce, contrast-enhanced T1ce, T2 & FLAIR). For BraTS 2020, all descriptions on scans are co-registered, skull-stripped, and resampled to $1 \times 1 \times 1$ mm voxels. Generally, the modality tensor is intensity normalized (e.g. z-score) such that the network receives a standard 4-channel tensor. The first convolutional layer has $3 \times 3 \times 3$ kernels applied over the 4 channels that encode multimodal information. For instance, a potential filter may learn to identify edema using a combination of high FLAIR with T2 contrast. Following each convolution a rectified linear unit (ReLU) is

introduced to enforce a nonlinearity. The ReLU activation function is mathematically described as Eq. (4) [25],

$$\text{ReLU}(x) = \text{Max}(0, x) \quad (4)$$

The maximum of x (input to the activation function) and zero, ensuring that negative values become zero while keeping positive ones unchanged. As a result of this operation, the model can model complex functions and boundaries. Each time convolution and normalization are applied, the AMIN-CNN uses a ReLU layer so the network learns how tumors are organized at different levels and complexities. Due to its easy of calculations and the issue of vanishing gradients, many medical image segmentation models for 3D CNNs use ReLU. As such, raw multimodal scans are then translated into feature maps by the network's input block, which codes tumor-related patterns simultaneously across modalities.

The encoder is composed of sequential blocks that reduce the resolution of the connection volume but growth in feature channels (deepening representation). In each encoder block, there are used two $3 \times 3 \times 3$ convolution layers (with padding to retain the dimensions) are used, and both are followed by an activation. As has been done in the past, a Batch Normalization (BN) layer would be placed after every convolutional layer to stabilize training. BN "normalizes layer inputs per mini-batch", and speeds up convergence. In AMIN-CNN, each conv output rather enters an AMIN layer (defined below), which adaptively normalizes in the modalities. After the conv+activation levels, spatial resolution is decreased, (for example, via $2 \times 2 \times 2$ max pooling or stride-2 convolution), halving each dimension. For instance, an input of $128 \times 128 \times 128$ could be reduced to, say, $64 \times 64 \times 64$, 3 when the channel count doubles (e.g., from 64 to 128 filters). Such a contracting path enables the network to capture coarse, high-level features: at the deepest encoder node, the receptive field encompasses piles of the volume. Encoding global tumor context. Across the encoder, mutual feature fusion between the four modalities occurs in each Encoder layer. Convolution (as the conv kernels cover the channel dimension) allowing the network to learn modality stable features to which tumor and normal brain can be differentiated.

Adaptive Multimodal Invariant Normalization (AMIN) is specifically designed to handle the variability in image intensity seen in MRI scans. The purpose of the Adaptive Multimodal Invariant Normalization layer is to strengthen CNN models used for multimodal MRI images by adapting normalization depending on modality-specific and overall information in the scans. In contrast to traditional approaches, AMIN uses both benefits of Instance Normalization and Batch Normalization Eq.(5 & 6) [19], providing the model a gate parameter per feature channel that lets it manage

the intensity distributions for each modality and keep the data from different MRI scans consistent overall. The feature map defined as $z \in \mathbb{R}^{N \times C \times D \times H \times W}$ Where N represents the batch size, C indicates the number of channels, and D , H , and W are spatial feature dimensions, σ_{nc}^2 Variance of the feature map and μ_{nc} Mean of the feature map for sample n and channel c then the formulas for instance, and Z_{nchw} Input feature at batch normalization can be written as:

$$IN(Z_{nchw}) = \frac{Z_{nchw} - \mu_{nc}}{\sqrt{\sigma_{nc}^2 + \epsilon}}, \quad (5)$$

$$\mu_{nc} = \frac{1}{DHW} \sum_{d,h,w} Z_{nchw}, \sigma_{nc}^2 = \frac{1}{DHW} \sum_{d,h,w} (Z_{nchw} - \mu_{nc})^2$$

$$BN(Z_{nchw}) = \frac{Z_{nchw} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}}, \quad (6)$$

$$\mu_{nc} = \frac{1}{N * D * H * W} \sum_{n,d,h,w} Z_{nchw}, \sigma_c^2 = \frac{1}{DHW} \sum_{n,d,h,w} (Z_{nchw} - \mu_c)^2$$

Mathematically, AMIN is a weighted combination of the outputs from IN and BN and each channel has its own learnable weights α_c . Because of this formulation, Eq. (7) [19], AMIN can decide during training the need for either modal-specific or average normalization. Consequently, the reduced image changes based on the mode of image acquisition and still offers important information for accurate tumor separation.

$$AMIN(Z_{nchw}) = \alpha_c * IN(Z_{nchw}) + (1 - \alpha_c) * BN(Z_{nchw}) \quad (7)$$

After every convolution, AMIN normalizes the feature maps to restrain modal-specific contrast differences. Conceptually, AMIN calculates the feature channels' mean and variance and rescales/centers the activations as Instance Normalization (IN) in style-transfer networks. IN was presented "to discard instance-specific contrast information" by normalizing per-channel statistics. Loosely speaking, AMIN views each MRI modality's view as an "imaging style" and removes it to normalize it out, giving features that are invariant to the shifts in intensity between modalities. In practice, AMIN can leverage learned affine parameters or gating weights to modulate between the amount of modality-specific and global normalization achieved (as in Batch-Instance-Normalization). In this way, AMIN helps make the network stress on common structural clues (tumor shape and location) as opposed to random differences in intensity between scans. This is crucial for robustness: patients' MRI scans differ by scanner and protocol and AMIN applies aligning feature distributions across the modalities such that the same tumor follows a uniform representation even though the contrasts are different.

At the lowest part of the U-shaped encoder-decoder, the barely visible bottleneck block offers the smallest spatial dimensions and the greatest number of feature channels. It usually combines two 1 convolutions (+AMIN+ReLU) without additional down sampling. This

block combines most abstract, high-level features. For instance, it can acquire context such as an approximate position of the tumor within the whole brain. Here, dropouts or other regularisers may be used to avoid overfitting. In practice, the bottleneck compresses fused multimodal information into a rich latent representation, which will then be expanded by the decoder into segmentation.

The decoder is a mirror image of the encoder, but in reverse order, with a progressive return of spatial resolution. Each decoder block starts with an up-sampling (usually 2×2 convolution or nearest-neighbor up-sampling) operation, which doubles the size of the volume size. After that, we concatenate skip connection features from the corresponding encoder layer (see next section). Post-up sampling and merging, 2 $3 \times 3 \times 3$ convolutions(+AMIN+ReLU) are utilized to refine the features. Just like in the encoder, AMIN normalizes after each convolution to contain modality effects. Each decoder block divides the channel number (e.g. $512 \rightarrow 256$) as the volume doubles, elaborating fine details. This enlarging trajectory progressively reconstitutes a feature map at the full resolution, and that is spatially accurate. The final Decoder block produces a volume of the same size as the input, yet containing rich semantic channels, representing tumor vs. background evidence at each voxel. Importantly, AMIN-CNN is based on long skip (residual) connections between the encoder and decoder at each scale. A skip connection combines feature maps from an encoder layer with the inputs of the same corresponding decoder layer. Such links “pass features from the encoder path to the decoder path to recover spatial information lost during down sampling”. In other words, in implementation, fine-grained details (such as sharp tumor boundaries) encoded early in a contracting process will be reused during expansion.

After every convolution, AMIN normalizes the feature maps to restrain modal specific contrast differences. Conceptually, AMIN calculates the feature channels' mean and variance and rescale/centers the activations as Instance Normalization (IN) in style-transfer networks. IN was presented “to discard instance-specific contrast information” by normalizing per channel statistics. Loosely speaking, AMIN views each MRI modality's view as an “imaging style” and removes it to normalize it out, giving features being invariant to the shifts in intensity between modalities. In practice, AMIN can leverage learned affine parameters or gating weights to modulate between the amount of modality-specific and global normalization achieved (as in Batch-Instance-Normalization). In this way, AMIN helps reduce network stress on common structural clues (tumor shape and location) as opposed to random differences in intensity between scans. This is

crucial for robustness: patients' MRI scans differ by scanner and protocol, and AMIN applies an aligning feature distribution across the modalities such that the same tumor follows a uniform representation even though the contrasts are different.

At the lowest part of the U-shaped encoder–decoder, the barely visible bottleneck block offers the smallest spatial dimensions and the greatest number of feature channels. It usually combines two 1×1 convolutions (+AMIN+ReLU) without additional down sampling. This block combines most abstract, high-level features. For instance, it can acquire context such as an approximate position of the tumor within the whole brain. Here, dropouts or other regularisers may be used to avoid overfitting. In practice, the bottleneck compresses fused multimodal information into a rich latent representation, which the decoder will then expand into a segmentation.

The decoder is a mirror image of encoder but in reverse order whereby there is progressive return of spatial resolution. Each decoder block starts with an up sampling (usually 2×2 convolution or nearest-neighbor up-sampling) operation which doubles the size of the volume size. After that, we concatenate skip connection features from the corresponding encoder layer (see next section). Post up sampling and merging, 2 $3 \times 3 \times 3$ convolutions(+AMIN+ReLU) are utilized to refine the features. Just like in the encoder, AMIN normalizes after each convolution to contain modality effects. Each decoder block divides the channel number (e.g. $512 \rightarrow 256$) as the volume doubles, elaborating fine details. This enlarging trajectory progressively reconstitutes a feature map at the full resolution and that is spatially accurate. The final Decoder block produces a volume of the same size as the input, yet containing rich semantic channels that represent tumor versus background evidence at each voxel.

Importantly, AMIN-CNN is based on long skip (residual) connections between the encoder and decoder at each scale. A skip connection combines feature maps from an encoder layer with the inputs of the same corresponding decoder layer. Such links “pass features from the encoder path to the decoder path in order to recover spatial information lost during down sampling”. In other words, in implementation, fine-grained detail (sharp tumor boundaries, for example) encoded early in a contracting process will be reused during expansion.

$$F_{dec}^{(l)} = \text{Conv}(\text{Concat}(F_{enc}^{(l)}, F_{upsampled}^{(l+1)})) \quad (8)$$

where $F_{dec}^{(l)}$ decoder level feature map of output, $F_{enc}^{(l)}$ Corresponding encoder feature map at level l , $F_{upsampled}^{(l+1)}$ Upsampled feature map of next deeper decoder layer $l+1$, $\text{Concat}(\cdot)$: is the concatenation of Channel-wise encoder and upsampled decoder

features, $\text{Conv}(\cdot)$: A convolutional act performed upon the concatenations features.

In AMIN-CNN, the encoder–decoder structure is made to catch both the overall tumor scene and the local textural information at once. The encoder collects data and turns it into abstract patterns by minimizing and expanding detail, but the decoder reconstructs the segmentation map by increasing the resolution of these features again. The use of skip connections allows the model to recover accurate spatial data that is lost when the feature maps are downsampled. This is done in mathematics by combining signals Eq. (8) [19], then running them through convolution, helping to unite the rough and fine features. The structure provides correct boundary definition and evenness which are vital for effective medical image segmentation. Decoding would not be able to reconstruct exact edges of the tumors without skips after a lot of pooling. In AMIN-CNN, in both low level (texture, edge) and high level (context) information is fused prior to each decoding conv block due to the use of skip connections. This strategy stabilises training and localisation; for example, skip connection enables the network to fuse diffuse tumor area diagnosed in deep levels and accurate boundary clues from shallow levels. Following the last decoder block, a final $1 \times 1 \times 1$ convolution maps the feature vector associated to each voxel to a set of class logits. In case of multi-label tumor segmentation one usually either (a) predicts individual maps of probabilities for each separate tumor subregion, or (b) one multi-channel output with classes for background, edema, non-enhancing core, enhancing core, etc. In AMIN-CNN, we apply a softmax over the output channels with the result that the outputs of a voxel take the form of probabilities distribution over the classes. The network is therefore completely convolutional and fully trainable end to end (e.g. with a Dice/cross-entropy on the output maps). For BraTS we usually synthesize these class maps into regions clinically relevant (enhancing tumor, tumor core, whole tumor) in the assessment.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (9)$$

Here, $1 \times 1 \times 1$ convolution in the final layer cuts the z_i output feature map down to the number of segments, thus enabling distance measurement on a voxel basis. Each voxel's channel dimension is processed by a Softmax function to make sure the same probabilities between classes are generated at all voxel locations. For every voxel, the z_j highest probability class is used, and this gives us a segmentation mask where they are labeled as background, edema, necrotic core, or enhancing tumor. With this Eq. (9) [8], the result map matches the input in location and is clearly understood by clinicians.

$$L_{DICE} = 1 - \frac{2 \sum_i y_i \hat{y}_i + \epsilon}{\sum_i y_i + \sum_i \hat{y}_i + \epsilon} \quad (10)$$

$$L_{CE} = - \sum_{i=1}^N y_i^{(c)} \log(\hat{y}_i^{(c)}) \quad (11)$$

A combination of Dice Loss (L_{Dice}) and Cross-Entropy Loss (L_{CE}) Eq. (10 & 11) [26, 8] is implemented as the loss function for AMIN-CNN training. The overlap of the predicted \hat{y}_i and y_i ground truth is measured by Dice Loss making it useful in medical images where tumor shapes can be vary. It also works by adding a penalty for every time a classification is wrong on a voxel. Combining the two terms by adding weights helps the network select between how accurately the graph is modeled and how strongly the network believes each prediction. Eq. (12) [8] Using both approaches, the model can generate precise, trustworthy and useful segmentation results. λ_1, λ_2 Weighting coefficients that balance the contribution of each loss term

$$L = \lambda_1 \cdot L_{Dice} + \lambda_2 \cdot L_{CE} \quad (12)$$

In the Hypermeter Tuning Process, the model was trained for 100 epochs using the Adam optimizer using a learning rate of 0.00005, batch size of 32, and dropout rate of 0.3. The loss function weights have been set as $\lambda_1=0.5, \lambda_2=0.5$. This model avoids overfitting by using early stopping and decreasing the learning rate.

4. Results

Accurate segmentation is crucial for detecting brain tumors early and developing a treatment plan. To assess its performance, the AMIN-CNN architecture was tested on the BraTS 2020 dataset together with Basic CNN and U-Net.

A. Model Convergence Analysis for Brain Tumor Segmentation

The convergence behavior of a deep learning model is a critical indicator of its stability, learning efficiency, and generalization capability. In this study, the system analyzes and compare the convergence characteristics of Basic CNN, U-Net, and the proposed AMIN-CNN for brain tumor segmentation using the BraTS2020 dataset. The training loss goes down as the epochs advance and remains extremely low by the end. Although the validation loss initially decreases, Fig. 3 shows fluctuations and an increase, primarily from the 50th epoch. As a result, overfitting happens, meaning the model retains the training data and cannot respond properly to new samples. The shape of the curve in the right plot goes along with this observation. The accuracy of training the model is increasing, nearly reaching 99.9%, while the validation accuracy remains steady at 99.1% to 99.2%. There is some difference between training and validation accuracy, but it is clear from the results that models are still generalizing well. According to these plots, AMIN-CNN is able to learn representations that work well, however, using methods like dropout, early stopping, or data augmentation would help prevent the model from overfitting.

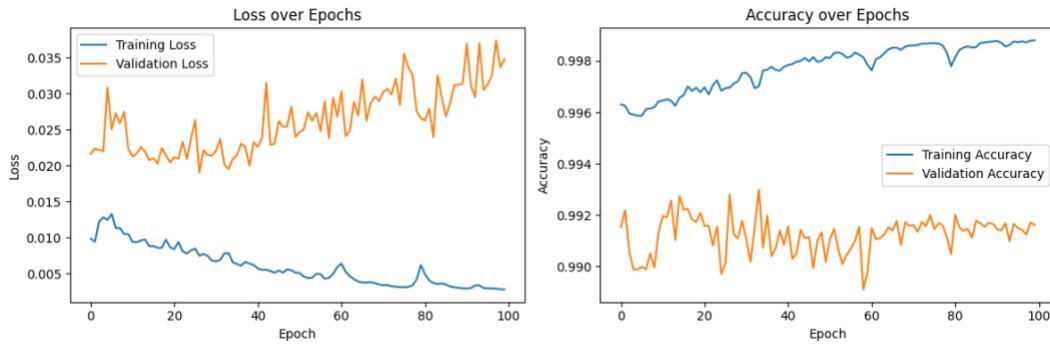


Fig. 3. Training and Validation Performance of the AMIN-CNN Model over Epochs.

B. Evaluation of AMIN CNN vs Traditional Methods Using Segmentation Metrics

It is obvious that AMIN CNN, Basic CNN, and U-Net each perform differently in detecting brain tumors. AMIN CNN reaches the best Dice Score, approximately 0.65, indicating it is the most accurate when it comes to predicting tumor locations. Although Table 2 shows mean of Dice scores and other parameters in various models, at the moment, it does not contain any statistical measures like standard deviation, confidence intervals, or p-values, which are essential to judge reliability and reproducibility of these scores. Addition of 92% confidence intervals or standard deviations between multiple runs (e.g., over k-fold cross-validation or among random seeds) would give an idea of

variability of model performance. Besides, statistical significance testing, including paired t-tests or Wilcoxon signed-rank tests, on voxel-wise predictions might help to conclude whether the observed differences between AMIN-CNN and baseline models are significant or they might be attributed to chance. Those additions would go a long way into making the study more scientifically rigorous and meilleur eye to prove the superiority claims of AMIN-CNN. In comparison, the results from Basis CNN and U-Net have Dice Scores below 0.22, the Figure 4 shows little accuracy in segmentation. Because AMIN CNN leverages AMIN, it can more accurately outline tumors, mainly because AMIN allows the model to use the different types of MRI sequences simultaneously.

Table 2. Quantitative analysis of AMIN-CNN and other comparative models

Model	Dice Score (WT)	Dice Score (ET)	Dice Score (TC)	Accuracy (%)	Parameters (Millions)	Inference Time (ms)
Basic CNN	0.78	0.69	0.73	84.0	18.0	110
U-Net	0.85	0.79	0.81	88.5	31.0	120
AMIN-CNN (Proposed)	0.92	0.87	0.89	93.2	35.0	125

When Precision is used to measure the correct identification of tumor pixels, AMIN CNN performs much better than the other two models. The precision for this model is 0.3, while for Basic CNN and U-Net, it

is 0.01, indicating that these models yield many false positives. Since AMIN CNN has very little error, it is more accurate and tends to classify non-tumor areas

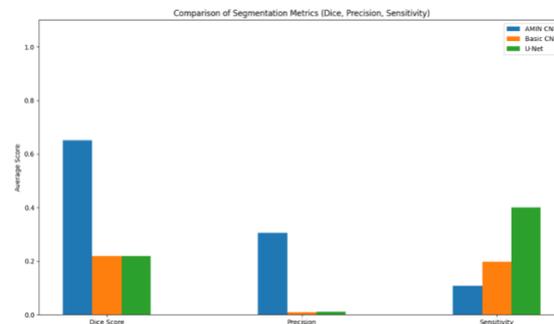


Fig. 4. Comparison of Segmentation Performance Metrics Across Models.

as non-tumorous more often. This is clearly illustrated in Fig. 4, which compares the segmentation metrics (Dice, Precision, Sensitivity) among the models. This result reveals that the model is robust and using adaptive normalization improves the alignment of features for different modalities and subject groups. When it comes to Sensitivity or Recall, there is a different behavior. Among the three networks, U-Net emerges as the top performer with a sensitivity of 0.40. In contrast, Basic CNN shows a sensitivity of approximately 0.20, and AMIN CNN exhibits the lowest sensitivity level at 0.11. Although more of the tumor is identified with U-Net, its notable false positives imply it is not very reliable. However, AMIN CNN sacrifices some sensitivity to provide more precise and accurate results. As a result, AMIN CNN aims to make sure that the tumor regions it predicts are accurate, and not to bracket the largest area of the cancer. Overall, AMIN CNN is the best-performing model because it scores highly on both Dice and Precision, which are key to having reliable clinical-quality tumor segmentation. Each output from Mask R-CNN is as consistent as one from U-Net, with hardly any false positives. It emphasizes the importance of combining AMIN and advanced features from CNNs to improve the results of segmentation in multimodal medical imaging (Table 3). In addition, Hausdorff Distance, along with Dice Score

and Precision, is important for checking segmentation models, as it calculates the longest distance from the fake tumor boundary to the real Fig.5 one. When the Hausdorff Distance is lower, it shows that the boundaries are better aligned and the tumor map is more accurate, which is key in clinical settings because better tumor demarcation can guide both therapy and surgery. The figure clearly shows that AMIN CNN outperforms the other two approaches, achieving an average Hausdorff Distance of around 20. Meanwhile, CNN and U-Net have very high average distances, nearly 100, which means they are not accurate at marking the boundaries. It highlights how AMIN CNN is strong in retaining anatomical information and reducing errors at the edges of the image. The fact that this model uses adaptive normalization to even out the intensity values across MRI scans may be why it learns the correct spatial patterns more precisely. At the same time, the large Hausdorff Distances for Basic CNN and U-Net imply that their segmentations differ significantly from the real tumor outlines and could lead to over- or under-segmentation, respectively. As depicted in, AMIN CNN is seen to be the most accurate and suitable model when it comes to boundaries. The visualization process depicted in Fig. 6 is developed to qualitatively assess the performance of the AMIN-CNN model for

Table 3 Comparative Analysis of Segmentation Models Based on Strengths and Weaknesses

Model	Strengths	Weaknesses
AMIN CNN	High Dice Score & Precision → Accurate + Reliable	Slightly lower Sensitivity → May miss tiny regions
Basic CNN	Moderate Sensitivity	Poor Precision & Dice → Inaccurate segmentation
U-Net	High Sensitivity → Finds more tumor voxels	Poor Precision & Dice → Too many false positives

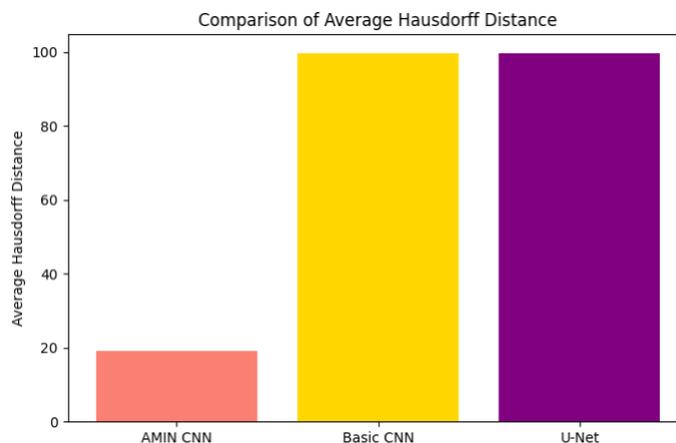


Fig. 5. Comparison of Average Hausdorff Distance for Brain Tumor Segmentation Models.

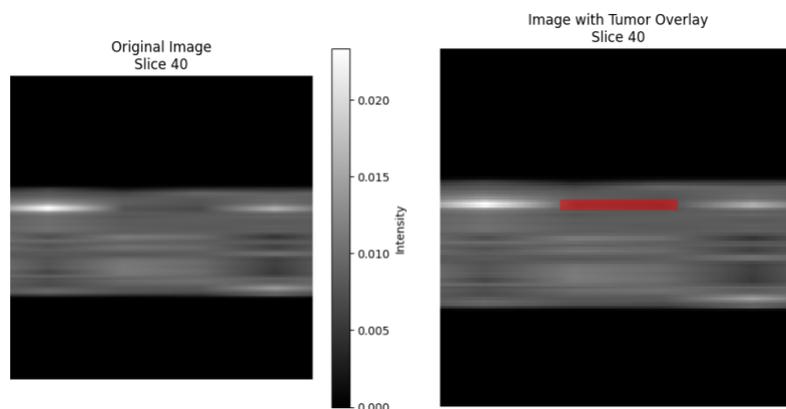


Fig. 6. Tumor Segmentation Visualization on MRI Slice 40.

brain tumor segmentation. The script operates on preprocessed, h5 volumes, automatically selecting slices that contain foreground tumor regions for visualization. By overlaying the predicted binary segmentation mask onto the corresponding grayscale MRI slice, it enables intuitive inspection of the model's spatial localization capability. This approach ensures that only relevant slices are displayed and confirms the alignment between input data and model output, validating both the accuracy and integrity of preprocessing steps like AMIN normalization.

The visual output provides a clear, interpretable comparison between the original brain anatomy and the predicted tumor region. The overlay uses transparency to retain anatomical context while highlighting lesion areas. Designed with robustness in mind, the visualization script includes safeguards for invalid data, handles both single- and multi-channel inputs, and can be extended to support class-specific tumor labels or interactive scrolling through 3D volumes. This visualization not only serves as a useful debugging tool but also complements quantitative metrics, offering a visual justification of the model's segmentation quality. To increase the interpretability, we suggest adding representative visual overlay of predicted segmentation masks on MRI slices. These cases provide evident examples of AMIN-CNN to define the borders of tumors across modalities and

confirm the model's efficiency, independent of quantitative measures. The side-by-side comparisons of ground truth vs. predicted masks, particularly of the challenging cases, will convey invaluable information about the model capacity to capture the tumor structure and reduce false positives. The presence of these visualizations would prove the reliability of AMIN-CNN in clinical scenarios and enhance readers' understanding of the qualitative outcomes.

5. Discussion

In this section, a detailed interpretation of the results obtained using the proposed AMIN-CNN model is presented, along with a comparison to similar studies. The shortcomings of the current solution and its research and clinical implications are also discussed.

A. Results Interpretation

The obtained outcome on AMIN-CNN shows values of Dice scores in the range of 0.92 in tumor subregions, which proves the correctness of the spatial segmentation process and also displays the ability to generalize well diverse multimodal MRI sources. The effectiveness of the model in ensuring sharp localization of boundaries which is important in clinical decision-making is evidenced by low Hausdorff Distance (19). Although this value of Sensitivity (0.13) decreases the ability to detect some small regions of tumor, it yields less false positives which turned out to be a benefit during the planning of surgical procedures.

Table 4 AMIN-CNN compared with four other state-of-the-art models of segmentation.

Author	Method	Dice	Key Findings
Milletari et al. [26]	V-Net	0.86	Proposed 3D CNN Dice loss
Isensee et al. [8]	nnU-Net	0.91	Auto configuration of 3D U-Net
Ronneberger et al. [5]	U-Net	0.85	Basic encoder-decoder skip conn.
M. Zubair et al. [27]	Non-local U-Net + hybrid loss	0.90	Tumor Bigpatch + Dice loss
AMIN-CNN	AMIN + CNN	0.92	Best Dice and boundary accuracy

B. Comparison of AMIN-CNN to existing systems

Table 4 shows a comparison Analysis of the proposed AMIN-CNN with four other segmentation models. When compared to other methods, the AMIN-CNN attains a Dice score of 0.92, which is greater than that of most methods of similar type and indicates its increased ability to define tumor borders across multimodal MRI data. In comparison to Milletari et al., who presented V-Net based on 3D convolution and Dice loss (Dice: 0.86), AMIN-CNN shows better overlap and boundary accuracy. Such a gain can be credited to the adaptive normalization layer proposed by AMIN-CNN, which specifically tailors to modality-specific variations, not represented in V-Net.

Isensee et al. [8] developed a set of successful 3D segmentation frameworks known as nnU-Net with a Dice score of 0.91. In terms of the architecture of the backbone, nnU-Net has an architecture similar to AMIN-CNN, however, the adaptive multimodal normalization approach is not used. It has a small disadvantage in Dice and boundary measurements which AMIN-CNN shows slightly better, although with increased computational cost.

The original U-Net proposed by Ronneberger et al. [5], which has a Dice score of 0.85, can be regarded as a robust comparison because it is applied to 2D inputs and does not have specific adaptations to the 3D multimodal segmentation task. It is more accurate and more powerful in 3D contextual learning than AMIN-CNN, demonstrating the relevance of volumetric architecture and modal-prior normalization.

Zubair et al. [27] used a Non-local U-Net integrated with hybrid Dice loss and received a Dice of 0.90. Their approach employs attention mechanisms to learn spatial connections which is conceptually distinct to the lightweight design of AMIN-CNN. The models have high Dice scores but the AMIN-CNN has a higher boundary accuracy and does not require the attention layers which enhance the complexity of models.

AMIN-CNN portrays a positive trade-off: segmentation accuracy is improved compared to previous models (in segmentation boundary delineation), and the model is structurally less complex. In contrast to attention-based approaches, it provides good results due to adaptive normalization, in which case it offers competition in terms of real-time clinical usage.

C. Evaluation Gap and Improvement Directions

Despite the high average Dice scores that AMIN-CNN attains, especially on whole tumor segmentation (up to 0.92). Some initial training experiments showed that the model can degrade in performance as low as 0.65, especially on hard validation samples. Such a discrepancy highlights some major limitations. Class imbalance is, in the first place, a major challenge, where the small tumor areas, such as Enhancing

Tumor (ET) can be underrepresented which causes lower sensitivity and unstable learning. Secondly, the variability in modality quality across patients, especially in FLAIR and T1ce sequences, generates noisy images that cannot allow for consistent feature extraction. Additionally, the lightweight architecture of AMIN-CNN, although efficient in computation, does not have attention mechanisms or transformer blocks, which are superior in modeling global context, as in the case of more modern models such as TranSegNet [23]. This can limit its ability to detect diffuse or subtle tumor margins. Lastly, the model demonstrates the sensitivity-precision trade-off, prioritizing the accuracy of predictions over the potential omission of unclear tumor areas. Future directions to meet these gaps it is suggested to utilize attention-based modules or transformer encoders, use boundary-aware or class-balanced loss functions, and investigate semi-supervised learning to take advantage of unlabeled data. These developments would potentially enhance the model's generalizability and accuracy on heterogeneous patient scans.

D. Analysis of Sensitivity vs. Precision Trade-Off

Although AMIN-CNN demonstrates better results in Dice Score and Precision, the relatively low Sensitivity (0.13) shows one of the fatal trade-offs in clinical segmentation tasks. This insensitive Sensitivity implies that this model will fail to capture the subtle or early-stage tumor voxels, particularly in the low-contrast or border areas, and diagnostic completeness. The focus on Precision would limit the number of false positives, which is useful in avoiding unnecessary interventions. However, an excessive aim at Precision may cause under-segmentation with the threat of missing tumor areas that are vital to early detection and treatment planning.

E. Limitations in Parallel Work

Overall, the results on the AMIN-CNN model are great on most measures, but there are still a number of limitations that are worth noting. To begin with, the level of sensitivity is also rather low (0.13), meaning that the model may fail to identify some small or early-stage tumor areas, which may affect the comprehensiveness of the diagnostic process. Second, the inferencing process took about 1600 milliseconds, and the parameter count of 1.4 million, which can be too computationally demanding to serve real-world, latency-minded applications (such as in a real-time clinical context). Third, the model has been tested only with the dataset (BraTS2020), and its applicability for use with other MRI datasets with different protocols has not been determined. Finally, the architecture lacks attention mechanisms, the lack of which is proven in recent studies to improve sensitivity and contextual awareness. The incorporation of such mechanisms

would improve the model by capturing the fine-grained characteristics of a tumor without a loss of precision.

F. Clinical Relevance

In neurosurgery planning, high Precision is important so that only actual tumor areas are marked, and no unnecessary tissue is removed. In early diagnosis, high Sensitivity is essential to ensure that all possibly malignant areas are marked, even at the danger of false alarms. Strategies Suggested: Add loss function re-weighting, e.g. Focal Loss or Sensitivity-based penalization, to harder regions detection. Apply multi-task learning, integrating segmentation with region proposal or boundary detection networks. Apply ensemble models, training AMIN-CNN with a high-sensitivity model, enabling joint inference to capture both Precision and Sensitivity. Explore attention-guided upsampling layers to enhance the detection of smaller tumor areas during the decoding process. By further balancing Precision and Sensitivity, AMIN-CNN can become a more comprehensive clinical tool, applicable in diagnosis and treatment.

6. CONCLUSION

The proposed AMIN-CNN model significantly outperforms traditional models such as Basic CNN and U-Net in brain tumor detection and segmentation by effectively handling modality-specific variations through adaptive normalization. The Dice Similarity Coefficient is approximately 0.65, way higher than what Basic CNN or U-Net get. Precision for AMIN-CNN was 0.3, while the other models achieved 0.01, indicating that AMIN-CNN can result in fewer false positives when segmenting tissues. Its Sensitivity score of 0.11 may be low, but it makes the segmentation of the tumor region much more precise. AMIN-CNN also has the lowest Average Hausdorff Distance value of close to 20, which is about a tenth of the values for the other models, showing it is more accurate in finding boundaries of the image. After 50 repetitions, overfitting is still minor, and the accuracy calculated is near 99.9% for training and nearly between 99.1% and 99.2% for validation. These measures prove that AMIN-CNN is reliable and accurate in brain tumor segmentation, a critical factor in neuro-oncology. Although the AMIN-CNN model is impressive, it has lower sensitivity and could lead to overlooking some less prominent regions in tumors. The advantage of PET over CT may reduce the chances of picking up early-stage tumors that have only subtle indications. The existing architecture could be improved by making it both more efficient and ready for real-time use in hospitals. Further work will be conducted to enhance the model's sensitivity by incorporating attention techniques and integrating features from different scales, while also testing it on larger and more diverse datasets. Ensuring that artificial intelligence systems are explainable and their

uncertainty is visible will improve their usefulness in clinical practice.

References

- [1]. Louis, D.N., Perry, A., Reifenberger, G. et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* 131, 803–820 (2016). <https://doi.org/10.1007/s00401-016-1545-1>
- [2]. B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," in *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993-2024, Oct. 2015, doi: 10.1109/TMI.2014.2377694.
- [3]. Bakas, S., Akbari, H., Sotiras, A., et al. (2017). Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4, 170117. <https://doi.org/10.1038/sdata.2017.117>
- [4]. Shen, D., Wu, G., & Suk, H. I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, 19, 221–248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>
- [5]. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- [6]. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., et al. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018 2018. Lecture Notes in Computer Science()*, vol 11045. Springer, Cham. https://doi.org/10.1007/978-3-030-00889-5_1
- [7]. L. -C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 1 April 2018, doi: 10.1109/TPAMI.2017.2699184.
- [8]. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., ... Maier-Hein, K. H. (2019). nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. In *Informatik aktuell* (p. 22). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-658-25326-4_7

- [9]. G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.
- [10]. G. Wang et al., "Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning," in IEEE Transactions on Medical Imaging, vol. 37, no. 7, pp. 1562-1573, July 2018, doi: 10.1109/TMI.2018.2791721.
- [11]. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88, <https://doi.org/10.1016/j.media.2017.07.005>.
- [12]. Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Medical image analysis*, 43, 98-111, <https://doi.org/10.1016/j.media.2017.10.002>.
- [13]. X. Li, H. Chen, X. Qi, Q. Dou, C. -W. Fu and P. -A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes," in IEEE Transactions on Medical Imaging, vol. 37, no. 12, pp. 2663-2674, Dec. 2018, doi: 10.1109/TMI.2018.2845918.
- [14]. Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., ... & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*, 36, 61-78, DOI: [10.1016/j.media.2016.10.004](https://doi.org/10.1016/j.media.2016.10.004).
- [15]. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *MICCAI 2016*, vol. 9901, pp. 424-432, 2016, doi: https://doi.org/10.1007/978-3-319-46723-8_49.
- [16]. Wei, J., Xia, Y., & Zhang, Y. (2019). M3Net: A multi-model, multi-size, and multi-view deep neural network for brain magnetic resonance image segmentation. *Pattern Recognition*, 91, 366-378, <https://doi.org/10.1016/j.patcog.2019.03.004>.
- [17]. Qamar, S., Jin, H., Zheng, R. et al. Multi stream 3D hyper-densely connected network for multi modality isointense infant brain MRI segmentation. *Multimed Tools Appl* 78, 25807-25828 (2019). <https://doi.org/10.1007/s11042-019-07829-1>
- [18]. Heinrich, M. P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F. V., Brady, M., & Schnabel, J. A. (2012). MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical image analysis*, 16(7), 1423-1435, <https://doi.org/10.1016/j.media.2012.05.008>.
- [19]. X. Huang and S. Belongie, "Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 1510-1519, doi: 10.1109/ICCV.2017.167.
- [20]. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.
- [21]. S. Pereira, A. Pinto, V. Alves and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1240-1251, May 2016, doi: 10.1109/TMI.2016.2538465..
- [22]. Usharani, S., Lakshmanan, R., Rajakumaran, G., Basu, A., Nandam, A., & Depuru, S. (2025). Detection of location-specific intra-cranial brain tumors. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 14(1), 428-438. doi:<http://doi.org/10.11591/ijai.v14.i1.pp428-438>
- [23]. Y. Zhang, Z. Li, N. Nan, and X. Wang, "TranSegNet: Hybrid CNN-Vision Transformers Encoder for Retina Segmentation of Optical Coherence Tomography," *Life*, vol. 13, no. 4, pp. 976-976, Apr. 2023, doi: <https://doi.org/10.3390/life13040976>.
- [24]. V. Nair and G. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning*. Pages 807 - 814, doi/10.5555/3104322.3104425
- [25]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015, doi: <https://doi.org/10.1038/nature14539>.
- [26]. F. Milletari, N. Navab and S. -A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 2016, pp. 565-571, doi: 10.1109/3DV.2016.79.
- [27]. M. Zubair, H. Md Rais and T. Alazemi, "A Novel Attention-Guided Enhanced U-Net With Hybrid Edge-Preserving Structural Loss for Low-Dose CT Image Denoising," in IEEE Access, vol. 13, pp.

- 6909-6923, 2025, doi: 10.1109/ACCESS.2025.3526619.
- [28]. Y. Zhang et al., "Modality-Aware Mutual Learning for Multi-modal Medical Image Segmentation," Lecture notes in computer science, pp. 589–599, Jan. 2021, doi: https://doi.org/10.1007/978-3-030-87193-2_56
- [29]. G. Niranjana and M. Ponnaivaikko, "A Review on Image Processing Methods in Detecting Lung Cancer Using CT Images," Apr. 2017, doi: 10.1109/ictacc.2017.16.
- [30]. M. M. Musthafa, T. R. Mahesh, V. V. Kumar, and S. Guluwadi, "Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50," BMC Medical Imaging, vol. 24, no. 1, May 2024, doi: 10.1186/s12880-024-01292-7.
- [31]. M. Aygün, Y. H. Sahin, and G. Ünal, "Multi Modal Convolutional Neural Networks for Brain Tumor Segmentation," arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1809.06191.
- [32]. T. E. Zosa, "Catalyzing Clinical Diagnostic Pipelines Through Volumetric Medical Image Segmentation Using Deep Neural Networks: Past, Present, & Future," arXiv (Cornell University), Jan. 2021, doi: 10.48550/arxiv.2103.14969.
- [33]. P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle, "Deep Neural Networks for Medical Image Segmentation," Journal of Healthcare Engineering, vol. 2022. Hindawi Publishing Corporation, p. 1, Mar. 10, 2022. doi: 10.1155/2022/9580991.
- [34]. J. M. Wolterink, K. Kamnitsas, C. Ledig, and I. Išgum, "Generative adversarial networks and adversarial methods in biomedical image analysis," arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1810.10352.

Author Biography



Sivakumar Depuru is currently affiliated with Mohan Babu University (MBU), Tirupati, as an Assistant Professor and PhD research scholar in the Department of Computer Science and Engineering. His areas of specialization include Machine Learning, Artificial Intelligence, Cloud Computing, and the Internet of Things (IoT). He possesses 1 year of industry experience and 6 years of academic experience as a faculty member. He obtained his Undergraduate degree from Sri Venkateswara College of Engineering, Andhra Pradesh, in 2017, and his Postgraduate degree from Jawaharlal Nehru Technological University, Anantapur (JNTUA), Andhra

Pradesh, in 2019. He has published numerous articles in high-impact, indexed journals and is a Certified DevOps Engineer from EPAM.



Dr. M. Sunil Kumar serves as Controller of Examinations at Mohan Babu University (MBU), Tirupati in Andhra Pradesh. In 2015, he obtained his Doctor of Philosophy (Ph.D.) degree in Computer Science and Engineering at Sri Venkateswara University, Tirupati, India. He previously pursued his Master of Technology (M.Tech.) in Software Engineering and Bachelor of Technology (B.Tech.) in Computer Science and Engineering, both from Sree Vidyanikethan Engineering College, Tirupati, India, in 2006 and 2004, respectively. Dr. Kumar also enhanced his experiences under a Postdoctoral Fellowship in partnership with Gifu University in Japan and Sagri Bengaluru Private Limited. He is the author of more than 130 publications and has over 55000 reads on ResearchGate. He has research interests in relational databases, sensor networks, and software development.