

BRAIN TISSUE SEGMENTATION USING DIFFERENT ARTIFICIAL INTELLIGENCE
ARCHITECTURES

A THESIS SUBMITTED TO
THE FACULTY OF ARCHITECTURE AND ENGINEERING
OF
EPOKA UNIVERSITY

BY

MARIO CAUSHI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR BACHELOR DEGREE
IN COMPUTER ENGINEERING

JUNE, 2025

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Mario, Caushi

Signature:

ABSTRACT

BRAIN TISSUE SEGMENTATION USING DIFFERENT ARTIFICIAL INTELLIGENCE ARCHITECTURES

Mario Caushi

B.Sc Department of Computer Engineering

Supervisor: M.Sc Stela Lila

Brain tissue segmentation in MRI scans plays a critical role in neuroimaging, aiding diagnosis, treatment planning, and research into neurological conditions. This thesis was driven by a strong academic interest in artificial intelligence and its application in medical imaging, where data-driven methods have the potential to improve diagnostic precision and reduce the burden of manual analysis. A range of classical and modern segmentation models were implemented and evaluated on the IBSR18 dataset, a publicly available collection of 3D T1-weighted brain MRIs with manually annotated labels. Models including K-Means, Fuzzy C-Means, 3D CNNs, U-Net, DenseUNet, and a custom hybrid architecture were assessed using the Dice Similarity Coefficient (DSC) as the evaluation metric. The best-performing model was a U-Net variant (D5), which achieved an average DSC of 0.7508, attributed to its patch-based training strategy, skip connections, and balanced encoder–decoder design. Despite this success, limitations such as the relatively small dataset size, class imbalance, and the absence of post-processing affected the segmentation of underrepresented structures. Future work should focus on leveraging larger multimodal datasets, applying transformer-based models, and integrating post-processing methods to enhance segmentation accuracy and generalizability.

Keywords: Brain Tissue Segmentation, Medical Imaging, Deep Learning, U-Net, MRI.

ABSTRAKT

SEGMENTIMI I INDEVE TË TRURIT DUKE PËRDORUR ARKITEKTURA TË NDRYSHME TË INTELIGJENCËS ARTIFICIALE

Mario Caushi

B.Sc Departamenti i Inxhinierisë Kompjuterike

Mbikëqyrëse: M.Sc Stela Lila

Segmentimi i indeve të trurit në imazhet MRI luan një rol thelbësor në neuroimazhëri, duke ndihmuar në diagnostikim, planifikimin e trajtimit dhe kërkimin mbi çrregullimet neurologjike. Kjo temë u motivua nga një interes i thellë akademik në inteligjencën artificiale dhe aplikimet e saj në imazhet mjekësore, ku metodat e bazuara në të dhëna kanë potencialin të përmirësojnë saktësinë diagnostike dhe të reduktojnë nevojën për analizë manuale. Një sërë modelesh klasike dhe moderne segmentimi u implementuan dhe u vlerësuan mbi dataset-in IBSR18, një koleksion publik i imazheve 3D të trurit të marra me MRI T1-weighted dhe të etiketuar manualisht nga ekspertë. Modelet si K-Means, Fuzzy C-Means, 3D CNN, U-Net, DenseUNet dhe një arkitekturë hibride e personalizuar u vlerësuan duke përdorur Koeficientin e Ngjashmërisë Dice (DSC) si metrikë vlerësimi. Modeli me performancën më të mirë ishte një variant i U-Net (D5), i cili arriti një vlerë mesatare DSC prej 0.7508, falë trajnimit me blloqe të vogla (patch-based), përdorimit të lidhjeve të drejtpërdrejta (skip connections) dhe strukturës së balancuar encoder-decoder. Pavarësisht këtij suksesi, kufizime si madhësia relativisht e vogël e dataset-it, pabarazia mes klasave dhe mungesa e post-proçesimit ndikuan në segmentimin e strukturave të nën-përfaqësuar. Punët e ardhshme duhet të fokusohen në përdorimin e dataset-eve më të mëdha dhe multimodale, në aplikimin e modeleve me arkitekturë të tipit transformator dhe në integrimin e metodave të post-proçesimit për të përmirësuar saktësinë e segmentimit dhe përgjithësueshmërinë.

Fjalë kyçe: Segmentimi i Indeve të Trurit, Imazheria Mjekësore, Mësimi i Thellë, U-Net, MRI.

*Dedicated to my family, friends and mentors who have supported me throughout
this journey.*

TABLE OF CONTENTS

ABSTRACT.....	3
ABSTRAKT.....	4
LIST OF TABLES.....	9
LIST OF FIGURES.....	11
SECTION 1.....	13
INTRODUCTION.....	13
1.1. Overview of Medical Imaging and Brain Tissue Segmentation.....	13
1.2. From Manual to AI: Brain Tissue Segmentation.....	14
1.3. AI-Driven Advances in Brain Tissue Segmentation.....	15
1.4. Problem Statement and Research Objectives.....	16
SECTION 2.....	18
LITERATURE REVIEW.....	18
2.1. Machine Learning.....	18
2.2. Machine Learning and Deep Learning.....	20
2.3. Deep Learning.....	21
SECTION 3.....	32
MATERIALS AND METHODS.....	32
3.1. Dataset.....	32
3.2. Preprocessing.....	34
3.3. K-Means Clustering.....	35
3.3.1. Purpose.....	35
3.3.2. Methodology.....	35
3.3.3. Model Variants.....	36
3.4. Fuzzy C-Means Clustering.....	37

3.4.1. Purpose.....	37
3.4.2. Methodology.....	37
3.4.3. Model Variants.....	38
3.5. Convolutional Neural-Network.....	39
3.5.1. Purpose.....	39
3.5.2. Methodology.....	39
3.5.3. Model Variants.....	40
3.6. U-Net.....	41
3.6.1. Purpose.....	41
3.6.2. Methodology.....	42
3.6.3. Model Variants.....	43
3.7. DenseUNet.....	44
3.7.1. Purpose.....	44
3.7.2. Methodology.....	45
3.7.3. Model Variants.....	46
3.8. Hybrid DenseUNet with U-Net Decoder.....	47
3.8.1. Purpose.....	47
3.8.2. Methodology.....	48
3.8.3. Model Configuration.....	49
SECTION 4.....	51
RESULTS AND DISCUSSION.....	51
4.1. Performance Metrics.....	51
4.2. K-Means.....	51
4.3. Fuzzy C-Means.....	53
4.4. Convolutional Neural-Network.....	54
4.5. U-Net.....	56
4.6. DenseUNet.....	57

4.7. Hybrid DenseUNet with U-Net Decoder.....	58
4.8. Model Comparison and Segmentation Visualization.....	60
4.9. Discussion.....	63
SECTION 5.....	66
CONCLUSION.....	66
REFERENCES.....	68

LIST OF TABLES

Table 1. State-of-the-Art Table of Brain Tissue Segmentation Studies. This table summarizes the studies discussed in the literature review.....	23
Table 2. Distribution of IBSR18 MRI Scans by Set and Voxel Dimensions. This table outlines the allocation of IBSR18 MRI scans by voxel dimensions across training, validation, and testing sets, detailing their specific image IDs.....	31
Table 3. K-Means clustering configurations with varying cluster count, initialization, and background masking.....	34
Table 4. FCM clustering configurations with varying numbers of clusters, fuzziness coefficients, and background masking options.....	36
Table 5. Configuration summary of the two baseline 3D CNN models.....	38
Table 6. Key architectural and training differences across U-Net variants D1–D6.....	41
Table 7. A summary of key architectural differences across DenseUNet variants E1–E3.....	44
Table 8. A summary of the architectural and training configuration of the hybrid DenseUNet–U-Net model.....	46
Table 9. Mean DSC scores for all K-Means configurations across validation volumes.....	49
Table 10. Per-volume DSC scores for K-Means configuration A2.....	49
Table 11. Mean DSC scores for all FCM configurations across validation volumes.....	50
Table 12. Per-volume DSC scores for FCM configuration B3.....	51
Table 13. Mean DSC scores for CNN configurations across validation volumes.....	51
Table 14. Per-volume DSC scores for CNN configuration C1.....	52
Table 15. Mean DSC scores for U-Net configurations (validation set).....	53
Table 16. Per-volume DSC scores for U-Net configuration D5.....	53
Table 17. Mean DSC scores for DenseUNet configurations (validation set).....	54
Table 18. Per-volume DSC scores for DenseUNet configuration E3.....	55
Table 19. Mean DSC scores for Hybrid model F1 (validation set).....	55
Table 20. Per-volume DSC scores for Hybrid model F1.....	56

Table 21. Summary of best-performing models and their top-scoring volumes.....	57
Table 22. Average DSC across validation set for best configuration of each model.....	58

LIST OF FIGURES

Figure 1. Visualization of IBSR18 Dataset. Graphical description of IBSR 01 volume and corresponding ground truth labels in axial, sagittal, and coronal view. Images generated using ITK-SNAP.....	33
Figure 2. MRI Data Preprocessing for IBSR18. This diagram outlines the key preprocessing steps applied to the IBSR18 dataset, showcasing the transformation from raw to preprocessed MRI data.....	35
Figure 3. Schematic representation of the baseline 3D CNN architecture (C1). The diagram illustrates the overall model structure, showing the flow of operations and corresponding tensor dimensions at each stage. This visual complements the architectural summary provided in Table 5.....	41
Figure 4. Schematic of the U-Net architecture used in model D5, showing encoder–decoder structure, skip connections, and corresponding tensor dimensions.....	44
Figure 5. Architecture diagram of DenseUNet with attention gates (E3), shown with 3-level depth for representational clarity, while the actual model uses four levels.....	47
Figure 6. Block diagram of the hybrid segmentation model architecture with labeled volume sizes, channel dimensions, and layer operations.....	50
Figure 7. Predicted segmentations for best-performing configurations on their top validation volumes, alongside ground truth and input MRI.....	61
Figure 8. Predicted segmentations on three test volumes (IBSR02, IBSR10, IBSR15) using the U-Net D5 model.....	62

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
MRI	Magnetic Resonance Imaging
CSF	Cerebrospinal Fluid
GM	Gray Matter
WM	White Matter
DSC	Dice Similarity Coefficient
CNN	Convolutional Neural Network
FCM	Fuzzy C-Means
IBSR18	Internet Brain Segmentation Repository 18
CLAHE	Contrast Limited Adaptive Histogram Equalization

SECTION 1

INTRODUCTION

1.1. Overview of Medical Imaging and Brain Tissue Segmentation

Medical imaging is an integral part of modern healthcare, enabling accurate diagnosis and effective treatment of various health conditions through non-invasive techniques [1]. Among the numerous imaging methods, such as X-rays, computed tomography (CT), and positron emission tomography (PET), magnetic resonance imaging (MRI) stands out due to its superior ability to capture high-resolution soft tissue contrasts while avoiding ionizing radiation [2]. This makes MRI particularly valuable in neuroimaging, where detailed anatomical visualization is critical for diagnosing and monitoring neurodegenerative diseases such as Alzheimer’s disease and multiple sclerosis, [5].

A fundamental application of MRI in neuroimaging is brain tissue segmentation, which classifies the brain into distinct tissue types: gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Accurate segmentation plays a crucial role in both clinical and research settings, aiding in disease detection, progression tracking, and treatment planning. Traditionally, segmentation relied on manual annotation and early computational techniques, which, despite being effective, proved time-consuming, prone to observer variability, and limited in scalability [3].

Over time, computational advances have significantly improved brain tissue segmentation, paving the way for artificial intelligence (AI). By automating key steps with advanced algorithms, AI delivers more precise, consistent results while minimizing human intervention. This evolution aligns with broader innovations in medical imaging, solidifying AI’s role in modern neuroimaging and driving ongoing improvements in reliability and future research directions [4].

1.2. From Manual to AI: Brain Tissue Segmentation

Medical imaging began with Wilhelm Conrad Roentgen's discovery of X-rays in 1895, revolutionizing medicine by enabling visualization of internal human anatomy for the first time. Over the following decades, further advancements emerged, including ultrasound and computed tomography (CT), each significantly enhancing diagnostic capabilities [1]. Yet, it was the development of magnetic resonance imaging (MRI) in the late 20th century that marked a major breakthrough, offering unparalleled detail in soft tissue imaging [2]. Its superior imaging quality laid the essential groundwork for more advanced analytical methods, particularly in brain imaging.

As imaging technologies became more sophisticated and widely utilized, the complexity and sheer volume of imaging data highlighted the need for systematic interpretation and analysis [1]. This necessity initiated the development of image segmentation techniques, which involve identifying and categorizing distinct anatomical structures within medical images. Given its complex anatomy and critical clinical significance, the brain became the principal focus for these segmentation advancements [2].

Initially, brain tissue segmentation was performed through manual delineation by trained radiologists. Although accurate, this manual approach was labor-intensive, subjective, and prone to variability among observers, making it increasingly impractical with growing imaging demands [2]. These inherent limitations prompted researchers to explore automated segmentation solutions.

Early automated approaches, primarily intensity-based methods like thresholding and region-growing algorithms, sought to simplify and speed up segmentation by identifying tissue boundaries based on image intensity values. These methods improved efficiency, but their accuracy still frequently suffered by variations in image quality, noise, and artifacts [3].

While imaging data continued to grow in complexity and size, the limitations of purely intensity-based methods became increasingly apparent. To address these challenges, researchers began turning to artificial intelligence (AI), exploring machine learning techniques capable of automatically recognizing complex patterns within imaging data [4], [6]. Traditional machine learning methods thus emerged as the next evolutionary step, enhancing the consistency and

reproducibility of segmentation results. However, these early AI-driven approaches often required manual feature selection and considerable parameter tuning, limiting their widespread applicability and scalability.

The introduction of deep learning techniques marked a major turning point in brain tissue segmentation. Capable of automatically learning complex features directly from imaging data, deep learning methods dramatically increased segmentation accuracy and speed, reducing reliance on manual input and expertise [5]. Following the success of deep learning, hybrid methods soon emerged, combining deep learning with traditional machine learning approaches to leverage the complementary strengths of both techniques and further optimize segmentation performance.

Despite these advancements, the field continues to face challenges, pushing further innovations in brain tissue segmentation. The adoption of sophisticated AI techniques is crucial to these efforts, enhancing both the accuracy and efficiency of diagnostic processes.

1.3. AI-Driven Advances in Brain Tissue Segmentation

The integration of artificial intelligence (AI), specifically machine learning (ML) and deep learning (DL), has greatly advanced brain tissue segmentation, improving accuracy and efficiency in neuroimaging analyses.

Initially, ML approaches automated basic pattern recognition tasks in brain imaging, laying the foundation for more advanced techniques. Deep learning then advanced these capabilities significantly, with neural networks identifying complex features that traditional methods could not, providing insights previously unattainable through manual segmentation. Building on these deep learning foundations, several key architectures have further elevated segmentation capabilities. Convolutional neural networks (CNNs) excel at processing hierarchical image features, ideal for the nuanced demands of medical imaging [6]. U-Net architectures enhance CNN performance by refining feature localization and analyzing context at multiple scales, while transformer models have recently extended these capabilities by providing comprehensive analysis of spatial relationships, crucial for accurate segmentation [4]. Many of these deep

learning methods also support multidimensional data, including volumetric (3D) scans, allowing robust and detailed analyses of complex anatomical structures [3].

Overall, AI-driven segmentation significantly enhances diagnostic accuracy and efficiency, fundamentally improving neurology treatment planning and patient outcomes. Such progress not only elevates current medical practices but also creates opportunities for continued innovation.

1.4. Problem Statement and Research Objectives

Regardless of remarkable advances in AI-driven brain tissue segmentation, several core challenges persist. First, data scarcity and variability remain critical constraints; high-quality, well-annotated MRI datasets are limited, and differing imaging protocols across institutions block model generalizability [2]. Second, the anatomical complexity of the brain, shaped by factors like age, pathology, and genetics, further complicates consistent segmentation. Third, noise and artifacts introduced by patient movement or varying scanner settings can significantly degrade segmentation accuracy. Fourth, interpretability and robustness of AI models remain evolving concerns, as systems that perform well in controlled settings may falter with atypical cases or unseen data [6]. Lastly, extensive computational resources are often required to train and deploy these advanced algorithms, including powerful hardware (e.g., GPUs or high-performance clusters) and large-scale data storage, making adoption and experimentation more challenging.

Collectively, these issues underscore the urgent need for reliable, adaptable, and data-efficient segmentation techniques. Ongoing research must address not only algorithmic precision but also reproducibility, clinical applicability, and accessibility of computational infrastructure, ensuring that AI solutions provide genuine improvements in real-world healthcare settings.

This thesis provides a thorough analysis and comparison of brain tissue segmentation techniques, evaluating traditional machine learning models alongside advanced deep learning algorithms. The objective is to determine the most effective methods, with a particular focus on enhancing segmentation performance and exploring the impacts of recent advancements in deep learning.

The study enhances brain tissue segmentation, which is vital for the early detection and effective treatment of neurological disorders. Additionally, it broadens our understanding of artificial intelligence in healthcare, highlighting how different deep learning models can be applied to improve diagnostic and therapeutic outcomes [4]. With this context in mind, the next section presents a structured review of important developments and research done in brain tissue segmentation.

SECTION 2

LITERATURE REVIEW

This section carefully compares and analyzes key studies on brain tissue segmentation to highlight developments in computational techniques within this specialized area of medical imaging. The review is structured into three main subsections: Machine Learning, Machine Learning and Deep Learning, and Deep Learning. Initially, the focus will be on machine learning-based studies, predominantly from the period before the widespread adoption of deep learning technologies. Although the frequency of purely machine learning applications has decreased with the rise of deep learning, their foundational impact is significant. The next subsection examines studies that employed both machine learning and deep learning, offering comparative analyses and documenting the transition and integration of these technologies. The final subsection delves into modern deep learning approaches, showcasing their effectiveness and current applications in brain tissue segmentation. This chronological examination includes a broad spectrum of research, emphasizing significant advancements and their impact on the field. The section concludes with a State-of-Art table that summarizes the methodologies, findings, and impacts of the studies reviewed, providing a comprehensive snapshot of current and evolving trends.

2.1. Machine Learning

Starting off with their 2014 study, Valverde et al. [7] evaluated ten brain tissue segmentation methods using the ISBR18 and ISBR20 datasets, which consist of T1-weighted MRI scans from healthy subjects. The methods assessed included FAST, SPM5, SPM8, GAMIXTURE, ANN, FCM, KNN, SVPASEG, FANTASM, and PVC, utilizing machine learning techniques to improve segmentation accuracy. They utilized the Dice Similarity Coefficient (DSC) to measure performance, with a particular focus on the impact of including or excluding sulcal cerebrospinal fluid (SCSF) voxels, traditionally mislabeled as gray matter in the datasets. The results indicated significant variations in method performance based on the handling of SCSF, suggesting a need

for revising ground-truth annotations in IBSR datasets to avoid bias in comparative studies. Notably, when SCSF was not considered, methods like SVPASEG and SPM8 emerged as more suitable for gray matter segmentation, highlighting their robustness and adaptability to dataset specifics and preprocessing differences. This study sets a precedent for the critical evaluation of segmentation methods in medical imaging, paving the way for further exploration of how different approaches handle complex imaging data in subsequent analyses.

In a subsequent study from 2015, Sucharitha and Geetha [8] used BrainWeb, a simulated MRI database with high-resolution T1-weighted phantom images, to evaluate fuzzy clustering techniques under varying salt-and-pepper and Gaussian noise levels. They introduced Reformulated Fuzzy Local Information C-Means (RFLICM), which replaces the spatial distance factor in FLICM with a local coefficient of variation for enhanced noise robustness. Compared with Fuzzy C-Means (FCM) and Fuzzy Local Information C-Means (FLICM), RFLICM achieved higher Segmentation Accuracy (SA) for GM, WM, and CSF segmentation, peaking at 99.86% under 9% salt-and-pepper noise. This approach requires no precomputation and effectively preserves image details despite noise. However, the reliance on simulated data and the need for additional kernel-based refinements suggest that future work could further validate RFLICM's performance on large-scale clinical datasets

More recently, in their 2022 study, Li et al. [16] investigated clustering-based segmentation of MRI brain images using algorithms like K-means, fuzzy c-means, and mean-shift clustering. The research focused on automated segmentation of complex brain tissues such as white matter, gray matter, and cerebrospinal fluid without specifying a particular dataset. This approach aimed to refine the accuracy and efficiency of diagnostic processes over traditional manual methods. The study demonstrated that fuzzy c-means was particularly effective, offering robust and reproducible results despite challenges related to setting precise algorithm parameters. This advancement in segmentation technology could significantly enhance clinical diagnostics and research, optimizing both time and resource use in medical imaging.

Lastly, in their 2023 study, Van Opbroek et al. [24] developed an automated brain-tissue segmentation method using a multi-feature SVM classifier, leveraging the MRBrainS13 dataset. This method utilized T1-weighted, IR, and FLAIR scans, enriched with Gaussian-scale and derivative features to enhance segmentation smoothness and accuracy. Tested on the

MRBrainS13 challenge images, the approach achieved a notable Dice coefficient of up to 88.3% for white matter. Despite its efficacy, limitations included errors in CSF segmentation due to oversized brain masks and misclassification near large white-matter lesions. This technique marks a significant step forward in automated MRI brain-tissue segmentation using machine learning, effectively handling variations in image quality.

2.2. Machine Learning and Deep Learning

Building on the intersection of machine learning and deep learning techniques, Kumar et al. [9] compared Fuzzy C-Means (FCM) with modern architectures such as SegNet and U-Net using the IBSR-18 dataset. They then proposed U-SegNet, a hybrid that combines SegNet's efficiency and U-Net's skip connection to preserve fine boundary details while mitigating random noise. By employing a patch-based strategy ($40 \times 40 \times 3$), U-SegNet incorporates localized 3D context for segmenting GM, WM, and CSF, ultimately achieving an average Dice ratio of 89.74%, outperforming both SegNet and U-Net. Although FCM demonstrated respectable performance, U-SegNet's success underscores how integrating deep learning innovations can refine brain tissue segmentation accuracy. Nevertheless, the study acknowledges limitations in training on a relatively small dataset (IBSR-18) and calls for broader validation on diverse clinical data. Overall, this work illustrates a transition from traditional machine learning techniques to more advanced neural network deep learning solutions, while emphasizing the advantages of more robust deep learning algorithms.

Continuing on advancements in deep learning segmentation, Zhang et al. [13] in his 2021 study introduced DDSeg, a method that directly segments brain tissues such as white matter, gray matter, and cerebrospinal fluid from diffusion MRI data, sidestepping the need for inter-modality registration with anatomical MRI. Employing a convolutional neural network and contrasting its performance with traditional machine learning approaches such as SVM, the DDSeg approach addresses common challenges like distortions in diffusion MRI. Trained on high-quality Human Connectome Project data and tested across various other acquisitions, DDSeg demonstrated high accuracy, enhancing its utility in both clinical and research settings. However, its effectiveness

depends significantly on the quality of diffusion MRI data, which may affect its applicability across different clinical protocols.

2.3. Deep Learning

Wang et al. [10] introduced a purely deep learning strategy called RP-Net for 3D brain segmentation using T1-weighted MRI scans from the CANDI and IBSR datasets. Their network integrates recursive residual blocks, in which each new residual path connects directly to the first convolutional layer, alongside pyramid pooling to capture multi-scale volumetric context, and deep supervision at intermediate layers to accelerate convergence. Despite training only on T1-weighted images, RP-Net achieved impressive mean Dice scores—90.49% on IBSR18 and 84.96% on IBSR20, surpassing comparable 3D CNNs like 3D U-Net and VoxResNet. While the study’s modest dataset and single-modality approach may limit generalization, the method underscores the potential of volumetric CNNs, particularly for capturing global and local context to enhance segmentation accuracy in challenging neuroimaging scenarios.

Expanding on these deep learning strategies, Yamanakkanavar and Lee [22] proposed a patch-wise M-Net model for brain tissue segmentation using MRI scans from the OASIS and IBSR datasets. Their model improved localization by processing 2D slices resized to 256×256 into 128×128 patches, allowing for finer-grained segmentation. The architecture incorporates multi-scale dilated convolutional kernels and skip connections within an encoder-decoder structure, enhancing detail retention. It outperformed traditional models like SegNet and U-Net, achieving a Dice coefficient of 0.95 on OASIS, although performance was slightly reduced on IBSR due to dataset variability. Overall, the approach struck a balance between segmentation accuracy and computational efficiency, with smaller patches enhancing precision at the cost of increased training time.

Following the momentum, Ghosal et al. [11] introduced the Multi-headed U-Net with Residual Inception (MhURI), evaluated on neonatal MRI from the iSeg-2017 and adult MRI from IBSR18. MhURI combines Residual Inception 2-Residual (RI2R) blocks with a Morphological Gradient channel, boosting segmentation quality across gray matter, white matter, and cerebrospinal fluid. Demonstrating high Dice scores and superior performance over prior

models, MhURI showed particular robustness in handling the structural variability of brain MR images. However, the study also highlighted limitations tied to its reliance on specific modalities and the quality of training data, emphasizing the need for more diverse datasets to improve generalizability and clinical applicability.

Extending this line of advancement, in their 2021 publication, Ding et al. [23] developed the Fuzzy-informed Deep Learning Segmentation (FI-DL-Seg) network, targeting multimodal infant brain MRI segmentation. Leveraging the iSeg-2017 dataset, the model addressed the segmentation of brain tissues. The network architecture integrates Volumetric Fuzzy Pooling (VFP) and Fuzzy-enabled Multiscale Feature Learning (F-MSL) modules, which manage the uncertainties inherent in pediatric MRI data by enhancing feature extraction across multiple scales. Evaluated on metrics such as the Dice Similarity Coefficient, Modified Hausdorff Distance, and Average Surface Distance, FI-DL-Seg demonstrated improvements in segmentation precision, achieving high DSC values. This methodology not only increased accuracy but also reduced computational complexity, presenting a significant advance in infant brain imaging by mitigating challenges related to poor image quality and high noise levels.

In the same year, Long et al. [12] proposed the MSCD-UNet. This model was designed to address the limitations of traditional U-Net architectures by integrating multi-branch pooling (MP) and multi-branch dense prediction (MDP) techniques. Tested on MRBrainS13, IBSR18, and ISeg2017 datasets, MSCD-UNet effectively captures multi-scale spatial and channel information, significantly enhancing the segmentation accuracy of brain tissues. The performance metrics employed, confirmed its superior capability to handle complex anatomical variations compared to previous models. Despite its advancements, the MSCD-UNet also encounters challenges related to computational demands and data diversity.

Furthermore, In 2021, Basnet et al. [14] introduced a 3D deep convolutional neural network architecture. Their model, designed for volumetric analysis of MRI scans, targets efficient computational usage by significantly reducing parameters from 40% to 98% compared to previous models. Evaluated on IBSR18 and iSeg-2017 datasets, their method demonstrates superior performance in segmenting brain tissues, addressing the challenges of computational intensity and data diversity in medical image processing. The use of performance metrics like

Dice Similarity Coefficient and Modified Hausdorff Distance highlighted its accuracy and efficiency, making it a viable tool for both clinical diagnostics and research applications.

In a related effort focused on fetal brain analysis Zhao et al. [15] developed a 3D U-Net model to automate fetal brain segmentation. The study utilized a robust dataset comprising MR images from 106 fetal studies, with fetuses ranging from 23 to 39 weeks of gestation, including those diagnosed with congenital heart disease (CHD). This model directly segmented MR scans of both healthy and CHD-diagnosed fetuses without manual intervention. Despite its results in achieving an average Dice score of 0.897, the study acknowledged potential limitations in training diversity, which might affect its applicability to pathological cases. However, the method significantly reduced segmentation time and improved reproducibility, presenting a substantial advancement for clinical diagnostics and research in fetal brain development.

Similarly, Munir, Frezza, and Rizzi [17] focused on hybrid deep learning techniques derived from the baseline U-Net model. Utilizing the MICCAI BraTS 2019 dataset, they explored various advanced U-Net architectures. These included recurrent and inception layers combined with depth-wise separable convolutions. The study aimed to boost segmentation accuracy, achieving significant improvements in Dice Similarity Coefficient and sensitivity-specificity metrics. While these models showcased enhanced tumor identification capabilities, they also faced challenges with data variability and high computational demands. Ultimately, the hybrid models demonstrated superior performance, indicating the effectiveness of evolving deep learning techniques in refining medical imaging diagnostics.

Likewise, Chattopadhyay and Maitra further [18] explored the application of convolutional neural networks (CNNs) to brain tumor detection. Utilizing the BRATS dataset, their study involved training on diverse MRI images to handle varied tumor characteristics effectively. They developed a CNN model, employing techniques like SVM classifiers and various activation algorithms. Aimed at automating tumor detection, their method achieved a notable accuracy of 99.74%. This model significantly aids in enhancing diagnostic speeds and accuracy, offering many benefits in medical treatment workflows. Despite achieving high accuracy, the study faced challenges related to the high computational demands and managing data variability.

Additionally, Waleed Al Shehri's [19] study advances Alzheimer's disease diagnostics using deep learning. Employing DenseNet-169 and ResNet-50 architectures, Al Shehri's approach utilizes a comprehensive dataset of MRI images to classify stages from Non-Dementia to Moderate Dementia. The study achieved a testing accuracy of 88.7% with DenseNet-169, demonstrating superior performance over ResNet-50. This research highlights the potential of deep learning to enhance diagnostic accuracy, despite facing challenges such as high computational demands and data variability.

Subsequently, Xiaona Huang [20] advanced fetal MRI brain tissue segmentation in January 2023...Utilizing a dataset of high-resolution T2-weighted MRI scans from 80 fetuses, this research employs a novel CNN architecture enhanced with a contextual transformer block for precise segmentation into seven distinct tissue categories. Evaluated through metrics like Dice similarity coefficient, achieving an average of 83.79%, the study highlights significant improvements in accuracy and efficiency over manual methods. Despite challenges such as fetal movements affecting image quality, this approach demonstrates deep learning's transformative potential in prenatal diagnostics, offering faster, error-reduced analysis crucial for early intervention.

In parallel Van de Leemput et al. [21] developed a fully convolutional neural network for multiclass segmentation in 4D CT. Using a dataset of 42 4D CT scans from stroke-suspected patients, the model processed subvolumes with 19 timepoints to segment brain tissues and vessels. The architecture employed multiresolution encoding with shortcut connections and batch normalization, optimized for sparse 2D annotations. Evaluated via Dice coefficient, contour mean distance, and absolute volume difference, the model achieved a mean DSC of 0.87 ± 0.04 . This approach enhances stroke diagnosis by enabling detailed, automated analysis of 4D CT data, though challenges remain in generalization and handling limited annotations.

Lastly, in their 2024 study, Sorour et al. [25] developed a range of deep learning models for the early detection of Alzheimer's Disease using MRI data. Utilizing a diverse set of models, including CNNs both with and without data augmentation, CNN-LSTM, CNN-SVM, and transfer learning with VGG16-SVM, the team targeted improved diagnostic accuracies. Their models were tested on MRI datasets, focusing on early pathological features of Alzheimer's. The CNN-LSTM model notably excelled, achieving an accuracy of 99.92%, with perfect precision.

and specificity, illustrating the power of integrating convolutional and recurrent neural networks to manage complex patterns in medical imaging data. This approach significantly enhances the capability for early disease detection, vital for conditions like Alzheimer's where early intervention is crucial. However, again the complexity of deep learning models and the substantial data required pose ongoing challenges.

An overview of the studies discussed above is provided in Table 1, summarizing the models, datasets, and performance metrics evaluated in the literature.

Table 1. State-of-the-Art Table of Brain Tissue Segmentation Studies. *This table summarizes the studies discussed in the literature review.*

Source	Model	Dataset	Input	Performance Metric	Performance Score
[7]	ANN FCN KNN	IBSR18 IBSR20	MRI	Dice Score Coefficient (DSC)	IBSR18 ANN: GM: 0.706 +/- 0.07 WM: 0.876 +/- 0.03 CSF: 0.116 +/- 0.06 FCN GM: 0.706 +/- 0.06 WM: 0.886 +/- 0.03 CSF: 0.116 +/- 0.06 KNN GM: 0.796 +/- 0.03 WM: 0.866 +/- 0.03 CSF: 0.166 +/- 0.07
[8]	FCM FLICM RFLICM	BrainWeb	MRI	Segmentation Accuracy (SA)	FCM: 97.64% FLICM: 98.24% RFLICM: 99.86%
[9]	FCM SegNet U-Net U-SegNet	IBSR18	MRI	Dice Score Coefficient (DSC)	FCM: 85.13% SegNet: 85.92% U-Net: 85.40% U-SegNet: 89.64%
[10]	3D-UNet 3D-like-FCN VoxResNet	CANDI IBSR18 IBSR20	MRI	Dice Score Coefficient (DSC)	IBSR18 3D-UNet: 88.64% 3D-like-FCN: 88.80%

	RP-Net				VoxResNet: 85.95% RP-Net: 90.49%
[11]	U-Net Residual U-Net Inception U-Net RI-U-Net SegNet MhURI	iSeg-2017 IBSR18	MRI	Dice Score Coefficient (DSC)	IBSR18 U-Net: 0.8183 Residual U-Net: 0.8226 Inception U-Net: 0.8254 RI-U-Net: 0.8273 Segnet: 0.8061 Proposed: 0.846
[12]	U-Net MSCD-UNet	MRBrainS13 IBSR18 ISeg2017	MRI	Dice Score Coefficient (DSC), Hausdorff Distance (HD), Absolute Volume Difference (AVD)	IBSR18-DSC U-Net - GM: 85.39% - WM: 89.08% - CSF: 88.14% MSCD-UNet - GM: 88.42% - WM: 90.31% - CSF: 90.57
[13]	SVM CNN DDSeg	Human Connectome Project (HCP)	MRI dMRI	Accuracy (ACC)	SVM: 76.8% CNN: 89.35% DDSeg: 89.85%
[14]	FastSurfer VoxResNet 3D-like-FCN 3D-UNet U-SegNet RP-Net DenseUNet 3D-DCNN	iSeg-2017 IBSR18	MRI	Dice Similarity Coefficient (DSC). Modified Hausdorff Distance (MHD), Average Surface	IBSR18 - DSC FastSurfer: 0.7906 VoxResNet: 0.8595 3D-like-FCN: 0.888 3D-UNet: 0.8864 U-SegNet: 0.8205 RP-Net: 0.9049 DenseUNet: 0.9281 3D-DCNN: 0.9315

				Distance (ASD)	
[15]	3D-UNet	106 fetal imaging studies	MRI	Dice Similarity Coefficient (DSC)	DSC 3D-UNet: 0.8965
[16]	K-Means FCM MEC GMM HC Mean-Shift	Random data - No specific source	MRI	Segmentation Accuracy (SA)	K-Means: 0.7283 FCM: 0.7312 MEC: 0.7282 GMM: 0.7203 HC: 0.7024 Mean-Shift: 0.7276
[17]	U-Net Recurrent- inception U-net Depth-wise Separable MI-Unet Hybrid Model Depth-wise separable Hybrid model	BraTS	MRI	Dice Similarity Coefficient (DSC) Sensitivity Specificity	DSC U-Net: 0.723 Recurrent-inception U-net: 0.798 Depth-wise Separable MI-Unet: 0.8313 Hybrid Model: 0.8201 Depth-wise separable Hybrid model: 0.8775
[18]	CNN - Sigmoid - RMSProp Optimizer CNN - SoftMax - AdaMax Optimizer CNN - SoftMax - RMSProp Optimizer	BraTS	MRI	Accuracy (ACC)	CNN - Sigmoid - RMSProp Optimizer: 97.63% CNN - SoftMax - AdaMax Optimizer: 98.1% CNN - SoftMax - RMSProp Optimizer: 99.74%

[19]	DenseNet-169 ResNet-50	Combination of different sources	MRI	Accuracy (ACC)	Training DenseNet-169: 0.977% ResNet-50: 0.838% Testing DenseNet-169: 0.887% ResNet-50: 0.819%
[20]	V-Net DMFNet 3D U-Net CNN-Transformer	Fetus Scans	MRI	Dice Similarity Coefficient (DSC) Sensitivity Specificity Precision Volume Similarity (VS) Hausdorff95 Distance (HD95)	DSC V-Net CSF: 64.32 +/- 2.45 GM: 32.36 +/- 4.63 WM: 80.89 +/- 2.3 DMFNet CSF: 73.59 +/- 2.3 GM: 59.73 +/- 7.05 WM: 85.21 +/- 3.47 3D U-Net CSF: 80.19 +/- 2.83 GM: 69.04 +/- 4.96 WM: 89.94 +/- 2.48 CNN-Transformer CSF: 85.49 +/- 3.16 GM: 71.2 +/- 4.87 WM: 89.73 +/- 1.73
[21]	Multiresonational - CNN 3D U-Net	Not specified	4D CT	Dice Similarity Coefficient (DSC) Contour Mean Distances (CMD) Absolute Volume Differences (AVD)	DSC Multiresonational - CNN: 0.82 +/- 0.04 3D U-Net: 0.68 +/- 0.15

[22]	SegNet	OASIS IBSR18	MRI	Dice Similarity Coefficient (DSC) Jaccard Index (JI) Precision Recall Specificity Mean Squared Error (MSE)	MSE - IBSR18
	U-Net				SegNet: 0.013
	M-Net				U-Net: 0.009
	Patch-wise M-Net				M-Net: 0.007
					Patch-wise M-Net: 0.005
					MSE - OASIS
					SegNet: 0.021
					U-Net: 0.008
					M-Net: 0.006
					Patch-wise M-Net: 0.003
[23]	FC-Semi-DenseNet1	iSeg-2017	MRI	Dice Similarity Coefficient (DSC) Modified Hausdorff Distance (MHD) Average Surface Distance (ASD)	DSC
	FC-Semi-DenseNet2				FC-Semi-DenseNet1: 0.927
	D-SkipDenseSeg				FC-Semi-DenseNet2: 0.927
	V-3D-UNet				D-SkipDenseSeg: 0.927
	HyberDense				V-3D-UNet: 0.930
	FC-DenseNet				HyberDense: 0.926
	H-DenseNet				FC-DenseNet: 0.931
	FI-DL-Seg (GMF)				H-DenseNet: 0.927
	FI-DL-Seg (SMF)				FI-DL-Seg (GMF): 0.943
					FI-DL-Seg (SMF): 0.951

[24]	SVM	MRBrainS13	MRI	Dice Similarity Coefficient (DSC)	DSC SVM: 95.05 +/- 0.53
				Modified Hausdorff Distance (MHD)	MHD SVM: 2.79 +/- 0.82
				Absolute Volume Differences (AVD)	AVD SVM: 3.51 +/- 1.61
[25]	CNN-LSTM-with-Aug	Not Specified	MRI	Accuracy (ACC)	Accuracy CNN-LSTM-with-Aug: 99.92%
	CNN-with-Aug			Precision	CNN-with-Aug: 99.61%
	CNN-without-Aug			Recall	CNN-without-Aug: 99.22%
	CNN-SVM-with-Aug			F1-Score	CNN-SVM-with-Aug: 99.14%
	Vgg16-SVM-With-Aug			Specificity	Vgg16-SVM-With-Aug : 98.67%

This literature review has examined the progression of both machine learning and deep learning methods in brain tissue segmentation, highlighting key model architectures, performance across various datasets, and clinical relevance. The reviewed studies illustrate the growing impact of artificial intelligence in medical imaging, while also pointing to persistent challenges such as dataset variability, computational cost, and generalizability.

With this foundation in place, the remainder of the thesis is structured as follows. Section 3 describes and analyzes the dataset used in this study, detailing the necessary preprocessing steps for preparing the dataset and discussing the models and techniques chosen for brain tissue segmentation. Section 4 presents and compares the performance of these models using standard evaluation metrics, with interpretation of their strengths and limitations. Finally, Section 5 summarizes the key findings, discusses current limitations, and outlines potential directions for future research in automated brain tissue segmentation.

SECTION 3

MATERIALS AND METHODS

3.1. Dataset

This study uses the Internet Brain Segmentation Repository 18 (IBSR18), developed by the Center for Morphometric Analysis at Massachusetts General Hospital, USA [10]. IBSR18 is a publicly available neuroimaging dataset specifically designed to facilitate the evaluation and improvement of automated brain segmentation methods.

IBSR18 comprises 18 volumetric T1-weighted magnetic resonance imaging (MRI) scans with dimensions of $256 \times 128 \times 256$ voxels. The MRI scans feature voxel dimensions ranging from $0.84 \times 0.84 \times 1.5 \text{ mm}^3$ to $1.00 \times 1.00 \times 1.5 \text{ mm}^3$, providing variations in image resolution, as detailed in Table 2. The scans are divided into training (10), validation (5), and testing (3) subsets; however, the specific criteria for this division and demographic details of the subjects have not been disclosed.

Manual expert annotations are provided in the training and validation sets, distinguishing cerebrospinal fluid (CSF), gray matter (GM), white matter (WM), and background, thus providing reliable ground-truth labels essential for supervised algorithm training. The test set initially lacks these annotations to enable unbiased performance evaluations during segmentation challenges.

The dataset is provided in the widely accepted NIfTI format, allowing convenient handling and seamless integration with image-analysis software. Importantly, IBSR18 is already preprocessed, saving significant preprocessing efforts and enabling greater focus on analysis and algorithm development for brain tissue segmentation. The following subsection will provide a detailed description of the preprocessing steps utilized.

IBSR18 was selected due to its standardized format, detailed and high-quality annotations, and established credibility in both academic research and educational applications. Nevertheless, the dataset presents limitations, including a relatively small sample size and limited diversity, which

may influence algorithm generalization. Despite these challenges, IBSR18’s easy access, high-quality annotations, and wide use in neuroimaging research justify its selection for this study.

Table 2 presents the distribution of MRI scans within the IBSR18 dataset by voxel resolution and data split, while Figure 1 illustrates a representative MRI volume and its corresponding ground truth labels across three anatomical planes.

Table 2. Distribution of IBSR18 MRI Scans by Set and Voxel Dimensions. This table outlines the allocation of IBSR18 MRI scans by voxel dimensions across training, validation, and testing sets, detailing their specific image IDs.

Set	Voxel Dimensions		
	0.94 x 1.5 x 0.94	1 x 1.5 x 1	0.84 x 1.5 x 0.84
Train	IBSR: 01, 03, 04, 05, 06	IBSR: 07, 08, 09	IBSR: 16, 18
Validation	IBSR: 13, 14	IBSR: 11, 12	IBSR: 17
Test	IBSR: 02	IBSR: 10	IBSR: 15

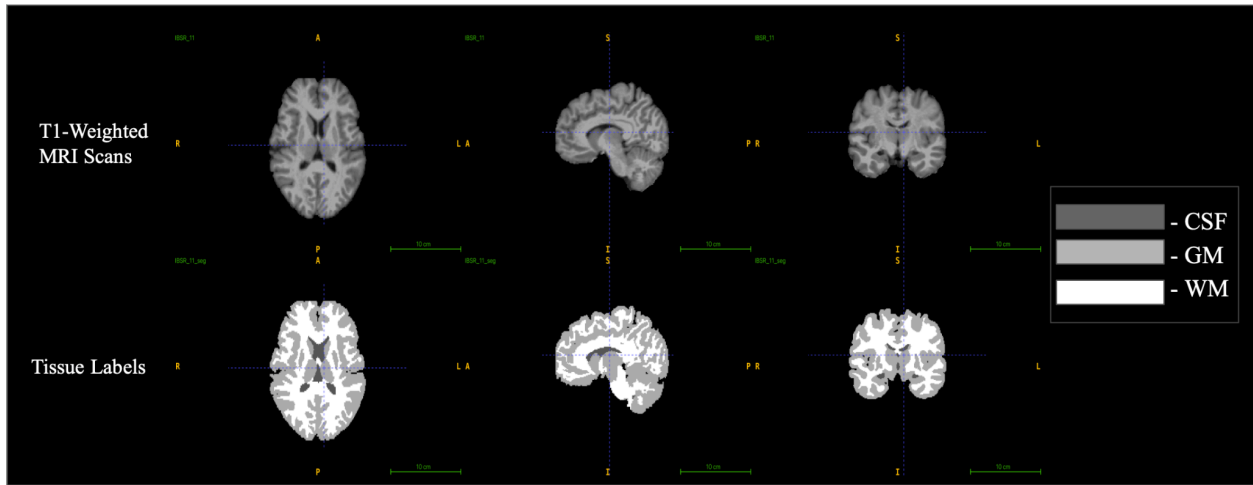


Figure 1. Visualization of IBSR18 Dataset. Graphical description of IBSR 01 volume and corresponding ground truth labels in axial, sagittal, and coronal view. Images generated using ITK-SNAP.

3.2. Preprocessing

Although the IBSR18 dataset was provided in a preprocessed format, briefly outlining the preprocessing techniques is essential, as these directly influence segmentation accuracy and overall quality of the analysis.

Initially, skull stripping was performed to remove non-brain structures such as the skull and scalp [7]. This procedure ensures that segmentation methods are applied exclusively to relevant brain tissues, simplifying computations and reducing potential segmentation errors. Subsequently, image registration was conducted, aligning all MRI scans into a common coordinate system. This alignment enables reliable comparisons across different images and enhances the consistency of analyses.

Next, several techniques were employed to correct intensity inconsistencies and reduce image noise. Bias Field Correction (BFC) using the M4 method was applied to address the uneven intensity distributions often observed in MRI scans, thereby normalizing the intensity across images and clearly differentiating between tissue types. Additionally, median filtering was implemented to effectively reduce noise while preserving important anatomical details and tissue boundaries.

Finally, intensity normalization and contrast enhancement techniques were applied. Min-Max normalization scaled voxel intensity values to a standardized range, ensuring uniform input conditions for segmentation algorithms. Contrast Limited Adaptive Histogram Equalization (CLAHE) was also used to enhance local contrast within the images, significantly improving the visibility of brain structures, particularly in regions initially exhibiting low contrast. Collectively, these preprocessing steps provided a consistent, high-quality dataset, thereby establishing a solid foundation for accurate and effective brain tissue segmentation, as illustrated in Figure 2 [10].

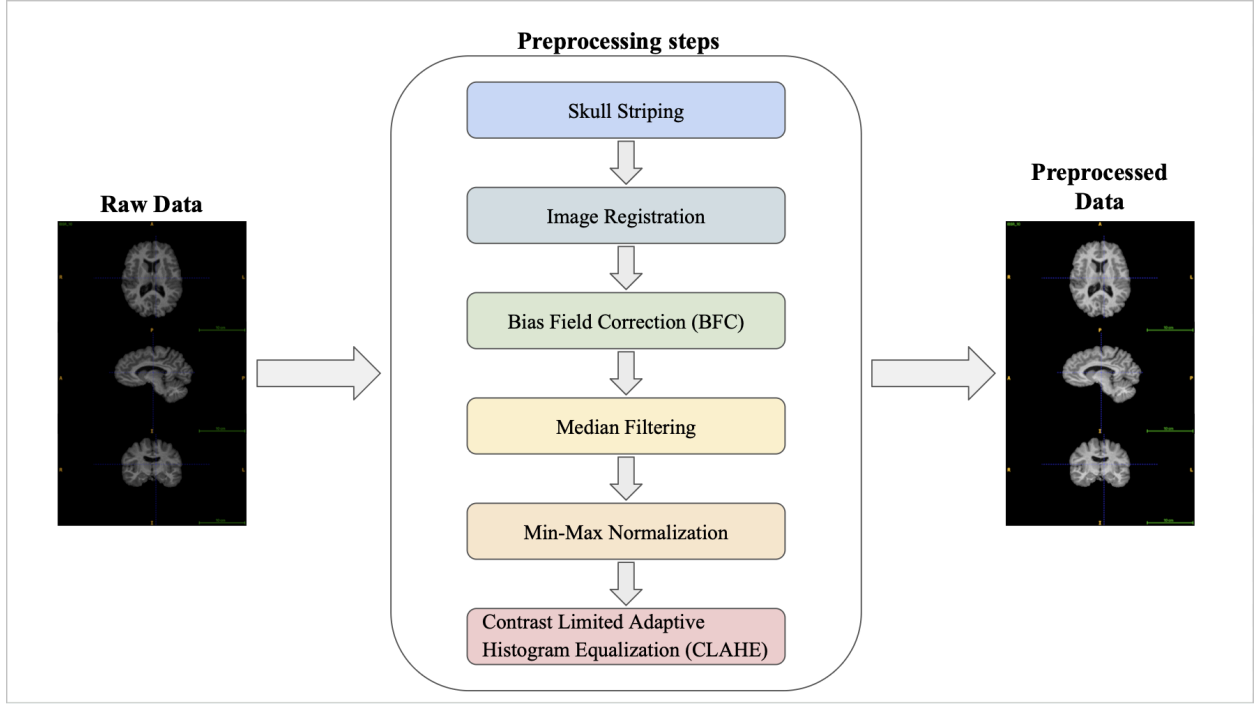


Figure 2. MRI Data Preprocessing for IBSR18. This diagram outlines the key preprocessing steps applied to the IBSR18 dataset, showcasing the transformation from raw to preprocessed MRI data.

3.3. K-Means Clustering

3.3.1. Purpose

K-Means clustering was chosen as the initial segmentation method due to its foundational role as an unsupervised, hard clustering algorithm widely used in early brain tissue segmentation studies. Its simplicity and non-parametric nature make it an ideal baseline for evaluating the performance of more advanced methods. In this study, K-Means provides a classic unsupervised benchmark for segmentation performance in comparison to both soft clustering techniques like FCM and deep learning-based approaches

3.3.2. Methodology

K-Means clustering was applied as an unsupervised, non-parametric method for segmenting brain tissue based solely on voxel intensity values. Each 3D MRI volume was flattened into a

one-dimensional array of intensities, with no spatial priors or anatomical constraints applied. Clustering was performed independently for each subject volume, and the number of clusters K was predefined per experiment.

The algorithm was implemented using Elkan’s optimized variant of K-Means combined with K-Means++ initialization, as provided by scikit-learn. A fixed random seed was used for reproducibility. No spatial regularization or post-processing was applied; voxels were clustered purely based on intensity, independent of neighboring context. When background masking was enabled, it was manually applied based on the configuration. As an unsupervised method, K-Means did not involve any data augmentation, loss function, or optimizer. All MRI volumes were processed in memory, as computational resources were sufficient to handle the full resolution ($256 \times 128 \times 256$) without requiring tiling or out-of-core computation.

3.3.3. *Model Variants*

Seven configurations (A1–A7) were designed to explore the impact of cluster count, centroid initialization, and background masking. These configurations are summarized in Table 3. The experiments assess the sensitivity of K-Means to initialization stability and evaluate how masking non-brain voxels influences segmentation quality.

Table 3. *K-Means clustering configurations with varying cluster count, initialization, and background masking.*

Model	n_clusters	n_init	mask_background
A1	4	1	False
A2	4	10	False
A3	4	20	False
A4	3	1	True
A5	3	10	True
A6	3	20	True
A7	5	10	False

Cluster counts of $K = 3, 4$, and 5 were selected based on prior anatomical expectations: $K = 3$ targets the core tissue classes (gray matter, white matter, cerebrospinal fluid), while $K = 4$ includes background as a separate cluster. $K = 5$ explores whether a finer subdivision of tissue intensity could improve segmentation. Background masking was additionally tested to assess its effect on CSF detectability and overall segmentation behavior.

3.4. Fuzzy C-Means Clustering

3.4.1. Purpose

Fuzzy C-Means (FCM) clustering was included in the segmentation pipeline as an unsupervised soft-clustering baseline, offering a direct comparison to hard clustering methods like K-Means. Unlike K-Means, FCM assigns partial memberships to each voxel across all clusters, allowing it to model uncertainty and tissue overlap, particularly valuable near anatomical boundaries where partial volume effects are common. This makes FCM a meaningful and interpretable baseline, well-suited for assessing how soft assignment impacts segmentation quality compared to both hard clustering and deep learning models.

3.4.2. Methodology

FCM performs soft clustering by assigning each voxel a degree of membership to every cluster, with all memberships summing to 1. The fuzziness coefficient $m > 1$ controls the degree of overlap: higher values of m increase membership smoothness, while values closer to 1 behave more like hard clustering. The algorithm minimizes a distance-weighted objective function and iteratively updates both the voxel-cluster memberships and the cluster centroids until convergence. Each 3D MRI scan was flattened into a one-dimensional intensity array and processed as a full volume (no patching). Background masking was optionally applied to exclude non-brain voxels from the clustering input.

All FCM experiments were implemented using the standard algorithm without acceleration techniques (e.g., Elkan optimization or approximation), as CPU memory was sufficient to handle full-resolution volumes without tiling or sampling. No spatial regularization or post-processing

was applied; voxels were clustered independently based solely on intensity. When background masking was enabled, it was applied manually using predefined masks. Each subject was clustered independently, without inter-volume dependencies, and a fixed random seed ensured reproducibility across runs.

3.4.3. Model Variants

Six configurations (B1–B6) were designed to explore the impact of cluster count, fuzziness level, and background masking. These configurations are summarized in Table 4 below. The experiments assess how changes in the fuzziness coefficient m , the number of clusters K , and the inclusion or exclusion of background voxels influence segmentation behavior.

Table 4. FCM clustering configurations with varying numbers of clusters, fuzziness coefficients, and background masking options.

Model	n_clusters	m	mask_background
B1	4	2	False
B2	4	2.5	False
B3	4	1.8	False
B4	3	2	True
B5	3	2.5	True
B6	5	2	False

Cluster counts of $K=3,4$ and 5 were chosen based on anatomical expectations: $K=3$ targets the core brain tissues (GM, WM, and CSF), while $K=4$ includes background as a distinct cluster. The configuration with $K=5$ was included to investigate whether an additional cluster could capture finer-grained intensity variations. The fuzziness parameter m was set to 1.8, 2.0, and 2.5, reflecting commonly used values in FCM literature. These values were selected to examine how different levels of cluster softness affect segmentation, with 2.0 serving as the standard default, 1.8 providing a sharper, more K-Means-like behavior, and 2.5 increasing uncertainty to test robustness in highly ambiguous regions. Background masking was tested to assess its impact on excluding irrelevant voxels and improving segmentation quality at tissue boundaries

3.5. Convolutional Neural-Network

3.5.1. Purpose

A three-dimensional convolutional neural network (CNN) was introduced as a minimal deep learning baseline for brain MRI tissue segmentation. This simple CNN approach follows the earlier unsupervised methods (K-Means and FCM) and provides a reference point for deep learning performance. The goal of including this baseline CNN was to establish a point of comparison before progressing to more sophisticated segmentation architectures, such as encoder–decoder networks like U-Net. By starting with a simplified CNN, the study allows later improvements to be attributed directly to architectural advancements.

3.5.2. Methodology

The baseline CNN was implemented using TensorFlow and trained from scratch with randomly initialized weights. Each variant consists of a sequence of three convolutional blocks, where each block includes a 3D convolution (kernel size $3 \times 3 \times 3$), followed by batch normalization, ReLU activation, and max-pooling to progressively reduce spatial resolution. A bottleneck layer increases feature depth, and a 1×1 convolution projects the features into voxel-wise class logits.

Two architectural variants were developed. C1 is a minimal 3D CNN with three convolutional blocks and no dropout; it performs a single $8 \times$ upsampling via nearest-neighbor interpolation to restore the original volume shape after three downsampling stages. C2 introduces dropout regularization (rate = 0.2) and replaces the fixed upsampling with a learned three-stage transposed convolutional pipeline, incorporating residual bottleneck structures for enhanced feature refinement.

Each model was trained independently on full-resolution 3D MRI volumes ($256 \times 128 \times 256$ voxels), without patching or input downsampling, as memory resources were sufficient. Models were trained from scratch without any form of pretraining or transfer learning. No data augmentation techniques (e.g., flipping, rotation, intensity shifts) were applied during training to

ensure that performance differences reflect purely architectural factors. Both variants were trained using the AdamW optimizer (learning rate = 1×10^{-4} , weight decay = 1×10^{-4}), minimizing Sparse Categorical Cross Entropy loss. Training was capped at 100 epochs, with early stopping applied (patience = 25) based on validation loss.

3.5.3. Model Variants

Two CNN configurations (C1 and C2) were evaluated to assess the effect of architectural complexity and upsampling strategy on segmentation performance. The key characteristics of each configuration are summarized in Table 5, while Figure 3 illustrates the overall structure and layer flow of the baseline C1 architecture.

Table 5. Configuration summary of the two baseline 3D CNN models.

Model	Encoder Structure	Filters	Dropout	Upsampling	Training
C1	3× [Conv3D → BN → ReLU → MaxPool]	32 → 64 → 128	None	8× Nearest Neighbor (fixed)	Sparse CE Loss, AdamW, 100 epochs, BS = 1
C2	3× [Conv3D → BN → ReLU → MaxPool]	32 → 64 → 128	0.2	Conv3DTranspose ×2×2×2 + residual bottlenecks	Sparse CE Loss, AdamW, 100 epochs, BS = 1

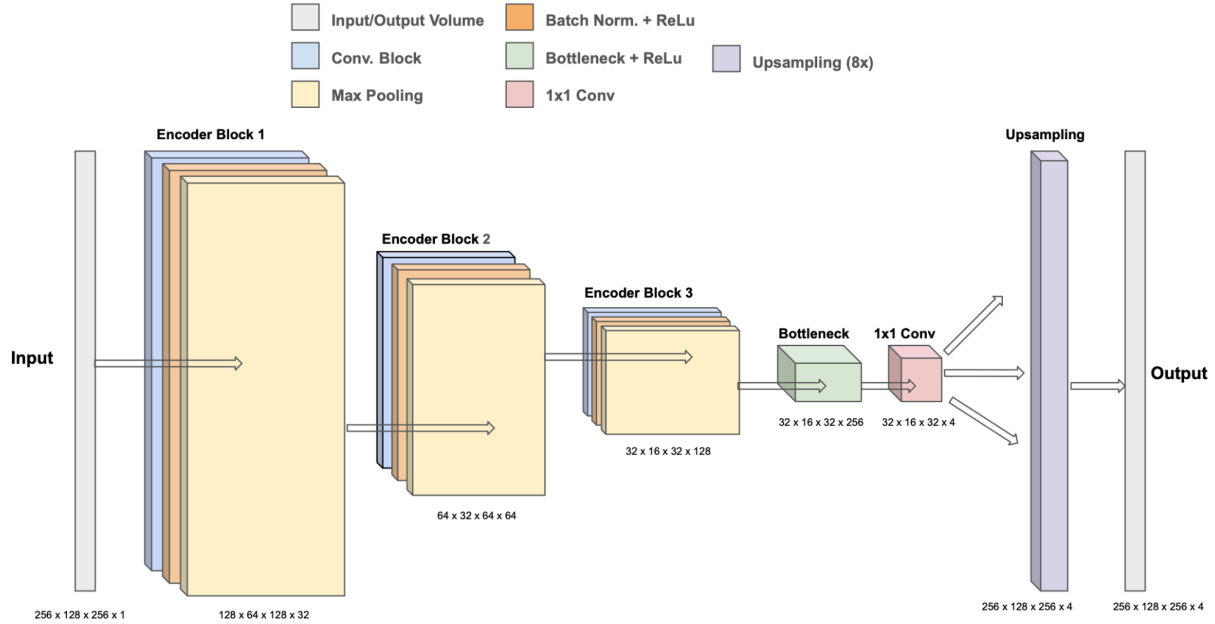


Figure 3. Schematic representation of the baseline 3D CNN architecture (C1). The diagram illustrates the overall model structure, showing the flow of operations and corresponding tensor dimensions at each stage. This visual complements the architectural summary provided in Table 5.

3.6. U-Net

3.6.1. Purpose

U-Net was adopted as a more advanced deep learning architecture tailored for biomedical image segmentation tasks, particularly effective in volumetric medical data such as brain MRI. Unlike the baseline CNN, U-Net features a symmetric encoder–decoder structure with lateral skip connections that directly propagate high-resolution features from the encoder to the decoder. This design helps preserve fine-grained anatomical detail that is critical for segmenting brain tissues. Its proven success in similar medical segmentation benchmarks, combined with its ability to learn from limited data through strong augmentation, made U-Net a natural next step beyond the baseline models. The goal was to evaluate whether U-Net’s architectural

enhancements, including hierarchical feature extraction and multi-scale feature fusion, could lead to higher segmentation accuracy, especially for thin structures like CSF that tend to be underrepresented and poorly captured by simpler models.

3.6.2. Methodology

The U-Net models were implemented in TensorFlow and trained from scratch with randomly initialized weights. Each model follows a symmetric encoder–decoder architecture adapted for volumetric brain MRI segmentation. The encoder consists of repeated blocks of two 3D convolutions ($3\times3\times3$, same padding), followed by batch normalization, a nonlinear activation (ReLU or LeakyReLU), and $2\times2\times2$ max pooling to reduce spatial resolution and increase feature depth. The decoder mirrors this structure, using transposed convolutions for upsampling and concatenating encoder features via skip connections. Dropout (rate = 0.2) was applied in the bottleneck for most models, and in all encoder blocks for the baseline.

A final $1\times1\times1$ convolution maps the decoder output to four logits (background, CSF, GM, WM), followed by softmax activation. All models were trained using the AdamW optimizer and Sparse Categorical Cross Entropy loss (SCCE), with class weighting applied in select variants to address tissue imbalance. Mixed-precision training and early stopping (patience = 25) were used throughout. Architectures varied in encoder–decoder depth (3–5 levels), filter progression, and input dimensions. Training was conducted either on full-resolution volumes or 3D patches. Patch-based models used random augmentation, including flipping, rotation, and intensity scaling, to improve generalization.

Patch-based training was adopted under the assumption that 3D patches retain sufficient anatomical context for accurate segmentation, while also helping to manage memory constraints. Input dimensions were selected to balance spatial coverage and computational efficiency. All models were trained solely on the provided dataset, without pre-training or post-processing, based on the assumption that the dataset was sufficiently representative. Augmentation was limited to the patch-based models, with the expectation that simple geometric and intensity perturbations simulate realistic anatomical variability.

3.6.3. Model Variants

To investigate how model depth, input size, loss weighting, and activation functions influence segmentation outcomes, six U-Net variants (D1–D6) were developed. All variants followed the general U-Net structure described above. The baseline model (D1) used a 4-level encoder–decoder with full-volume input, while D2–D5 adopted 3-level designs optimized for patch-based training. A deeper 5-level model (D6) was included to assess the impact of increased capacity. The key architectural and training differences are summarized in Table 6, with Figure 4 depicting the encoder–decoder structure and skip connections of the D5 variant.

Table 6. Key architectural and training differences across U-Net variants D1–D6.

Model	Input Shape	Depth	Filters/Block	Activation	Loss Function	Training Type
D1	256x128x256	4/4	32 \rightarrow 512	ReLU	SCCE	Full volume
D2	96×96×96	3/3	16 \rightarrow 128	LeakyReLU ($\alpha=0.1$)	SCCE	Patch-based + Augment.
D3	64×64×64	3/3	16 \rightarrow 128	LeakyReLU ($\alpha=0.1$)	Weighted SCCE	Patch-based + Augment.
D4	96×96×96	3/3	16 \rightarrow 128	LeakyReLU ($\alpha=0.1$)	Weighted SCCE	Patch-based + Augment.
D5	128×128×128	3/3	16 \rightarrow 128	LeakyReLU ($\alpha=0.1$)	SCCE	Patch-based + Augment.
D6	128×128×128	5/5	16 \rightarrow 512	LeakyReLU ($\alpha=0.1$)	Weighted SCCE	Patch-based + Augment.

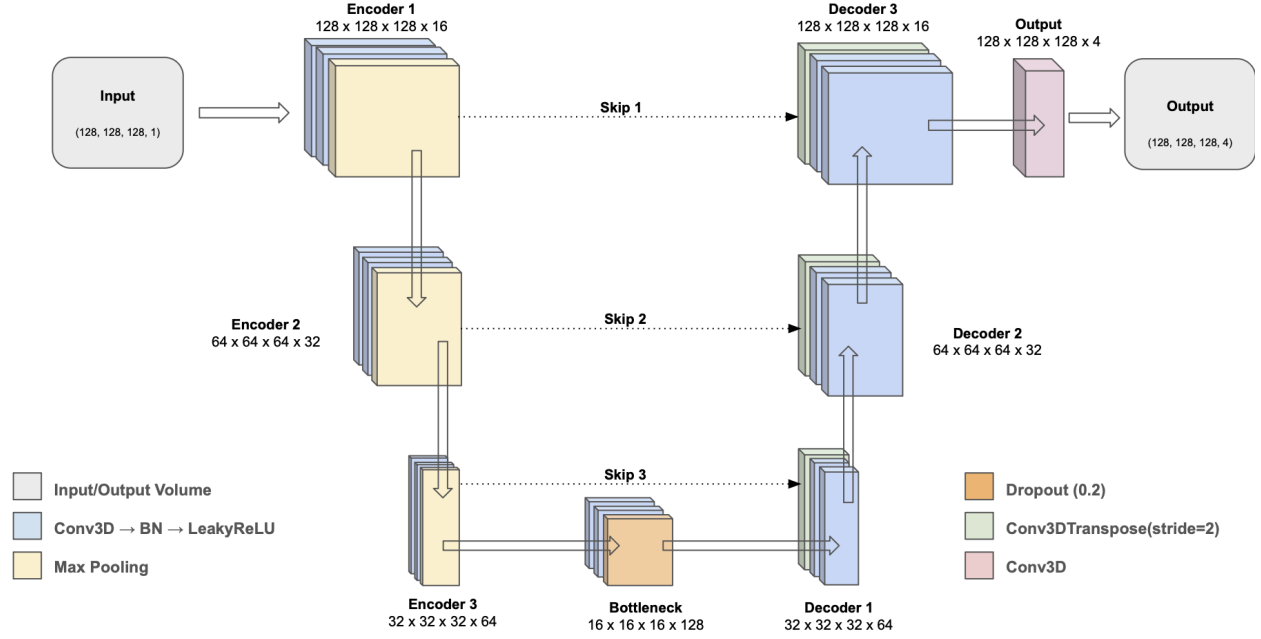


Figure 4. Schematic of the U-Net architecture used in model D5, showing encoder–decoder structure, skip connections, and corresponding tensor dimensions.

3.7. DenseUNet

3.7.1. Purpose

DenseUNet was explored as a more advanced deep learning architecture that extends the encoder–decoder foundation of U-Net by incorporating densely connected convolutional blocks. The key motivation was to improve feature reuse, training efficiency, and gradient flow, particularly in settings with limited training data such as the IBSR18 dataset. Dense connections allow each layer to receive inputs from all preceding layers within the same block, promoting implicit deep supervision and mitigating vanishing gradients. This structure enables the model to learn richer representations with fewer parameters compared to plain stacked layers. In the context of brain tissue segmentation, DenseUNet was expected to improve segmentation quality, especially for underrepresented structures like CSF, which often exhibit low contrast in MRI scans, though empirical outcomes are discussed in the subsequent Results section.

3.7.2. Methodology

Several DenseUNet models were implemented in TensorFlow/Keras by modifying the standard U-Net architecture to incorporate densely connected convolutional blocks. Each dense block consisted of multiple 3D convolutional layers (kernel size $3 \times 3 \times 3$), each preceded by batch normalization and Leaky ReLU activation ($\alpha = 0.1$), and followed by dropout for regularization. Within each block, layer outputs were concatenated with all preceding feature maps, promoting multi-scale feature reuse and enhancing gradient propagation. Growth rates and dropout strategies were adjusted across models based on architectural depth and complexity.

Downsampling in the encoder was performed via transition down modules ($1 \times 1 \times 1$ convolution followed by $2 \times 2 \times 2$ max pooling), while the decoder used transposed convolutions (stride 2) for upsampling. Skip connections were used at all levels to preserve spatial detail. In one model (E3), attention gates were integrated at each skip connection to selectively filter encoder features based on decoder context.

All models were trained from scratch on randomly extracted $128 \times 128 \times 128$ voxel patches from the IBSR18 dataset, using a batch size of 2. Patch-based training was augmented with random flips, rotations, and intensity scaling to improve generalization. Mixed-precision training (float16) was employed for computational efficiency, and all models were optimized using AdamW with the Sparse Categorical Cross Entropy (SCCE) loss function. Training was scheduled for 200 epochs, but all models converged earlier due to early stopping with a patience threshold of 25. No pre-training or post-processing techniques were applied.

These models were trained under the assumption that 128^3 voxel patches provide sufficient anatomical and contextual information for accurate brain tissue boundary learning while remaining memory efficient. The IBSR18 dataset was considered representative despite its small size, with dense connectivity and regularization expected to prevent overfitting. No external datasets or pretrained weights were used. Data augmentation techniques were intended to simulate plausible anatomical variability and scanner conditions. For the attention-enhanced model (E3), it was assumed that attention gates would improve segmentation by refining the spatial focus of decoder inputs.

3.7.3. Model Variants

Three DenseUNet variants (E1–E3) were developed to investigate how architectural depth, growth rate, and attention mechanisms affect segmentation performance. All models follow the same densely connected encoder–decoder structure described earlier, with E3 uniquely incorporating attention gates in the skip connections. The architectural differences between these three configurations are detailed in Table 7.

Table 7. A summary of key architectural differences across DenseUNet variants E1–E3.

Model	Input Shape	Depth (Enc/Dec)	Dense Block Layers	Growth Rate	Dropout	Attention Gates?
E1	128x128x128	3/3	2–2–2 + 2 (bottleneck)	16	0.2 (bottleneck)	No
E2	128x128x128	4/4	4–6–8–10 + 4 (bottleneck)	32	0.2 (all)	No
E3	128x128x128	4/4	3–3–3 + 4 (bottleneck)	16	0.2 (all)	Yes

These configurations reflect distinct design trade-offs. E1 serves as a mild baseline with limited capacity, E2 pushes depth and feature growth to the limit, and E3 maintains moderate complexity while integrating attention mechanisms to refine spatial focus at skip connections. This architectural setup is illustrated in Figure 5, which depicts the E3 model using a 3-level structure for visual clarity.

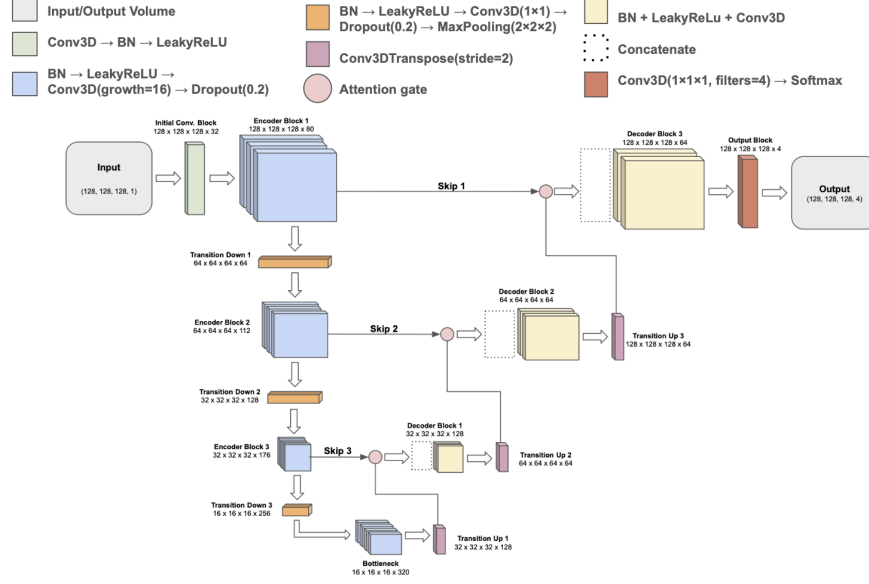


Figure 5. Architecture diagram of DenseUNet with attention gates (E3), shown with 3-level depth for representational clarity, while the actual model uses four levels.

3.8. Hybrid DenseUNet with U-Net Decoder

3.8.1. Purpose

The hybrid model was designed to integrate the dense feature extraction capabilities of DenseNet-style encoders with the spatial reconstruction strengths of a U-Net-style decoder. The motivation comes from the need to propagate low-level features through the network efficiently while simultaneously preserving high-resolution anatomical detail throughout the segmentation process.

In the encoder, densely connected convolutional blocks enable feature reuse by allowing each layer to receive inputs from all preceding layers. This enhances gradient flow, promotes deep supervision, and reduces redundancy. Meanwhile, the decoder leverages U-Net-style skip connections to restore fine spatial detail by directly reintroducing early-layer features during upsampling. Together, these architectural principles aim to achieve accurate tissue boundary delineation by maintaining both abstract semantic understanding and spatial precision across the network.

3.8.2. Methodology

The network operates on 3D MRI patches of size $128 \times 128 \times 128$. It adopts a symmetric encoder–decoder structure with three downsampling and three upsampling stages. The encoder is composed of three densely connected blocks, each with two convolutional layers. The growth rates of these blocks progressively increase (12, 16, and 24), enabling deeper stages to capture increasingly complex representations. Each 3D convolutional layer employs a $3 \times 3 \times 3$ kernel, followed by batch normalization and a LeakyReLU activation function. Spatial downsampling is performed using $2 \times 2 \times 2$ max pooling after each block, with no compression applied to the feature maps.

At the bottleneck level, the model includes a deeper dense block comprising three convolutional layers with a growth rate of 32. A dropout rate of 0.2 is applied within this block to regularize the network and mitigate overfitting, particularly given the small size of the training dataset.

The decoder mirrors the encoder in structure but replaces dense connectivity with conventional convolutional processing. Each decoder stage begins with a transposed 3D convolution (stride 2) to restore spatial resolution, followed by concatenation with the corresponding encoder block's feature maps via skip connections. These merged feature maps are then refined through two successive $3 \times 3 \times 3$ convolutions with batch normalization and LeakyReLU activations. This configuration is designed to integrate both high-level context and low-level detail, facilitating accurate tissue segmentation without introducing additional complexity such as attention gates.

The final layer of the network applies a $1 \times 1 \times 1$ convolution to reduce the feature maps to four output channels, representing background, CSF, GM, and WM. A softmax activation function converts these into voxel-wise class probabilities.

The model was developed under the assumption that dense connectivity in the encoder improves feature propagation and reuse, allowing early anatomical features to influence deeper layers. The use of skip connections was expected to enhance spatial accuracy during decoding by restoring high-resolution detail from the encoder stages. Dropout regularization in the bottleneck was included to prevent overfitting, and no compression was applied to maintain full feature richness across blocks. Lastly, attention mechanisms were excluded based on the assumption that standard

skip connections and learned convolutional filters would be sufficient for accurate reconstruction without added complexity. The key architectural and training specifications of this hybrid model are summarized in Table 8, while Figure 6 provides a block-level overview of the network's structure, including volume dimensions and layer operations, as described in the following subsection.

3.8.3. Model Configuration

Table 8. *A summary of the architectural and training configuration of the hybrid DenseUNet–U-Net model.*

Component	Details
Input Volume	$128 \times 128 \times 128$ voxels
Encoder Depth	3 levels (dense blocks with 2 layers)
Growth Rates	$12 \rightarrow 16 \rightarrow 24$ (encoder), 32 (bottleneck)
Bottleneck	3 layers + dropout (rate = 0.2)
Decoder Structure	3 transposed conv stages + double Conv3D
Skip Connections	U-Net-style concatenation (no attention)
Output Layer	Conv3D ($1 \times 1 \times 1$), softmax (4 classes)
Optimizer	AdamW
Loss Function	Sparse categorical cross-entropy
Batch Size	2
Precision	Mixed precision (float16)
Early Stopping	25 - Patience-based on validation loss

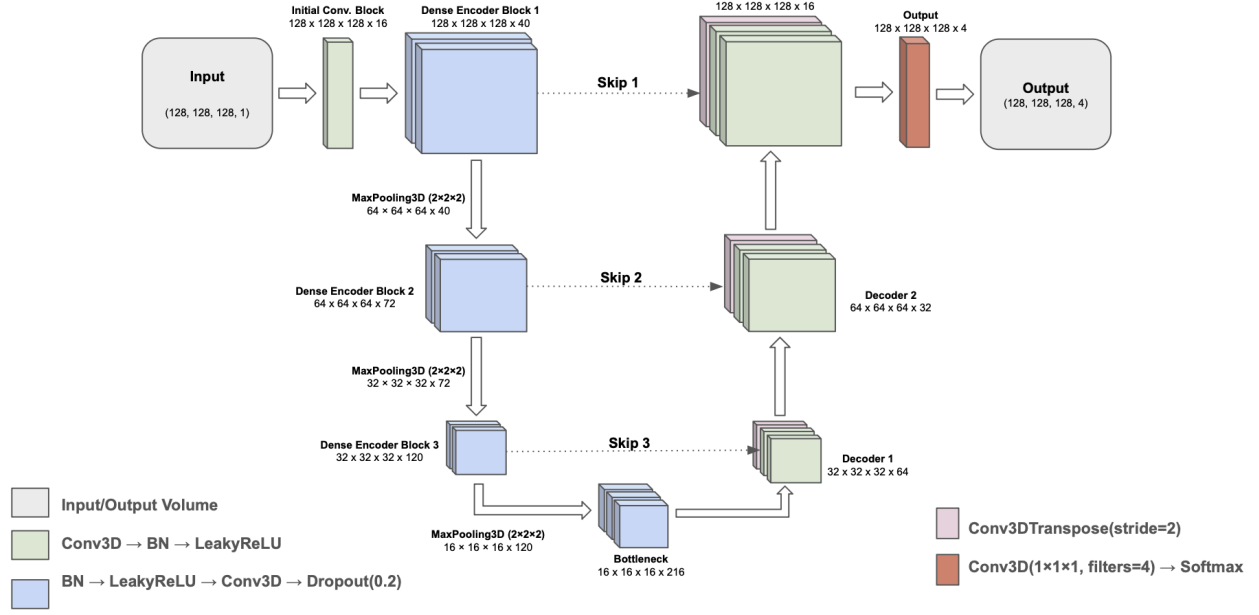


Figure 6. Block diagram of the hybrid segmentation model architecture with labeled volume sizes, channel dimensions, and layer operations.

SECTION 4

RESULTS AND DISCUSSION

4.1. Performance Metrics

Segmentation performance was evaluated using the Dice Similarity Coefficient (DSC), a standard measure of overlap between predicted and ground truth segmentations in medical imaging. It is defined as:

$$DSC = \frac{2 \cdot |P \cap G|}{|P| + |G|} \quad (1)$$

where P and G denote the predicted and ground truth voxel sets for a given tissue class. A score of 1 indicates perfect overlap; 0 indicates no overlap.

DSC was computed for three tissue classes, cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM), with the background class excluded from all evaluations. Class-wise scores were averaged across validation volumes to obtain a mean DSC per configuration. This metric was used to compare model performance, identify the best variant in each family, and assess generalization on the test set.

4.2. K-Means

K-Means performed well for gray matter (GM) and white matter (WM), but consistently failed to segment cerebrospinal fluid (CSF). Table 9 reports the average Dice scores for all tested configurations. The best-performing variant, A2, achieved the highest mean DSC (0.588), with strong results for GM and WM. All configurations produced near-zero scores for CSF, highlighting the algorithm’s limitations in capturing fine-grained, low-intensity tissue boundaries.

Table 9. Mean DSC scores for all K-Means configurations across validation volumes.

Model ID	GM	WM	CSF	Average DSC
A1	0.882 ± 0.021	0.880 ± 0.017	0.000 ± 0.001	0.587 ± 0.007
A2	0.882 ± 0.023	0.883 ± 0.015	0.000 ± 0.001	0.588 ± 0.009
A3	0.880 ± 0.025	0.881 ± 0.014	0.000 ± 0.001	0.587 ± 0.010
A4	0.381 ± 0.331	0.718 ± 0.320	0.086 ± 0.045	0.395 ± 0.086
A5	0.394 ± 0.347	0.882 ± 0.018	0.069 ± 0.055	0.448 ± 0.100
A6	0.536 ± 0.344	0.716 ± 0.319	0.077 ± 0.059	0.443 ± 0.090
A7	0.870 ± 0.020	0.826 ± 0.034	0.000 ± 0.000	0.565 ± 0.013

Configurations A1–A3, which did not apply background masking, consistently outperformed masked configurations (A4–A6), particularly in GM and WM. Increasing the number of initializations from 1 to 20 had little impact, suggesting convergence stability. Background masking introduced variability and degraded performance, especially in GM segmentation, as seen in configuration A4, possibly due to the exclusion of informative intensity ranges. Configurations with five clusters, such as: A7, were also tested to explore finer tissue subdivision but did not yield meaningful improvements, as CSF segmentation remained negligible.

Table 10 presents per-volume DSC scores for configuration A2. As the best-performing configuration, it maintained consistent performance across validation volumes, especially for GM and WM. CSF segmentation remained negligible, reinforcing the model’s limited capacity to capture underrepresented, low-intensity tissue structures.

Table 10. Per-volume DSC scores for K-Means configuration A2.

IBSR Volume	GM	WM	CSF	Average DSC
IBSR-11	0.8735	0.9071	0.0005	0.5937

IBSR-12	0.8407	0.8696	0.0005	0.5703
IBSR-13	0.8967	0.8653	0.0001	0.5874
IBSR-14	0.8970	0.8905	0.0001	0.5959
IBSR-17	0.9004	0.8820	0.0006	0.5944

4.3. Fuzzy C-Means

Fuzzy C-Means (FCM) achieved slightly higher average Dice scores than K-Means, particularly for gray and white matter. Table 11 reports the mean validation performance across all tested configurations. The best-performing variant, B3, obtained an average DSC of 0.591, with the highest WM score and strong performance on GM. As with K-Means, CSF segmentation remained negligible across all configurations.

Table 11. Mean DSC scores for all FCM configurations across validation volumes.

Model ID	GM	WM	CSF	Average DSC
B1	0.892 ± 0.011	0.879 ± 0.016	0.001 ± 0.000	0.591 ± 0.005
B2	0.892 ± 0.012	0.872 ± 0.014	0.001 ± 0.000	0.588 ± 0.007
B3	0.891 ± 0.011	0.882 ± 0.015	0.001 ± 0.000	0.591 ± 0.004
B4	0.357 ± 0.256	0.560 ± 0.409	0.001 ± 0.001	0.306 ± 0.187
B5	0.365 ± 0.247	0.714 ± 0.340	0.001 ± 0.001	0.360 ± 0.174
B6	0.868 ± 0.021	0.818 ± 0.034	0.000 ± 0.000	0.562 ± 0.014

Configurations B1–B3, which used four clusters without background masking, consistently outperformed the masked variants (B4–B5), especially in GM and WM segmentation. Changes in the fuzziness coefficient (m) had minimal impact on final scores, suggesting that the algorithm was relatively robust to this hyperparameter within the tested range. Configurations B4 and B5, which applied background masking, suffered substantial performance drops, most notably in

GM. B6, which increased the number of clusters to five, failed to improve CSF detection and slightly underperformed in overall segmentation accuracy.

Table 12 presents the per-volume Dice scores for configuration B3. Performance was consistent across all subjects for GM and WM, while CSF remained underrepresented, similar to the trend observed in K-Means.

Table 12. Per-volume DSC scores for FCM configuration B3.

IBSR Volume	GM	WM	CSF	Average DSC
IBSR-11	0.8754	0.9020	0.0006	0.5927
IBSR-12	0.8813	0.8941	0.0012	0.5922
IBSR-13	0.8977	0.8635	0.0001	0.5871
IBSR-14	0.8922	0.8658	0.0003	0.5861
IBSR-17	0.9065	0.8866	0.0009	0.5980

Although FCM slightly improved GM and WM segmentation over K-Means, it remained similarly ineffective in detecting CSF, reaffirming the limitations of intensity-only clustering methods.

4.4. Convolutional Neural-Network

Convolutional neural networks (CNNs) were evaluated as baseline deep learning models for brain tissue segmentation. Table 13 summarizes the average Dice scores for the two tested configurations. The best-performing variant, C1, achieved a mean DSC of 0.372, with moderate segmentation quality for gray matter (GM) and white matter (WM). CSF segmentation remained entirely absent in both configurations.

Table 13. Mean DSC scores for CNN configurations across validation volumes.

Model ID	GM	WM	CSF	Average DSC
C1	0.572 ± 0.012	0.542 ± 0.026	0.000 ± 0.000	0.372 ± 0.011

C2	0.638 ± 0.025	0.061 ± 0.047	0.000 ± 0.000	0.233 ± 0.021
----	-------------------	-------------------	-------------------	-------------------

Configuration C2, which incorporated dropout regularization and transposed convolutional upsampling, performed worse overall, particularly in WM, despite slightly higher GM scores. The sharp performance drop in C2 suggests that the increased architectural complexity and use of learned upsampling may have introduced instability when applied without data augmentation. In contrast, C1 maintained more consistent results across volumes using fixed nearest-neighbor upsampling and a simpler convolutional pipeline.

Table 14 presents the per-volume DSC scores for configuration C1. GM and WM segmentation was relatively consistent across all validation cases, though segmentation quality remained low overall. Neither model succeeded in segmenting CSF, mirroring the trend observed in clustering-based methods.

Table 14. Per-volume DSC scores for CNN configuration C1.

IBSR Volume	GM	WM	CSF	Average DSC
IBSR-11	0.574	0.565	0.000	0.380
IBSR-12	0.563	0.570	0.000	0.378
IBSR-13	0.587	0.525	0.000	0.371
IBSR-14	0.583	0.553	0.000	0.379
IBSR-17	0.555	0.495	0.000	0.350

Although the CNN architecture showed some potential, it underperformed across all tissue classes compared to both K-Means (A2, 0.588) and FCM (B3, 0.591). Segmentation of gray matter (GM) and white matter (WM) was limited, while cerebrospinal fluid (CSF) was not detected at all. This highlights the limitations of shallow CNN architectures trained on full-resolution volumes without augmentation or spatial priors. The results indicate that such models lack the representational capacity and spatial precision required for accurate brain tissue segmentation. Consequently, the CNN serves as a minimal deep learning baseline and underscores the need for more advanced encoder–decoder designs, as explored in the subsequent U-Net models.

4.5. U-Net

U-Net models were evaluated to establish a stronger deep learning baseline for brain tissue segmentation. Configurations D1–D6 explored the impact of architectural depth, loss functions, and training strategies such as patch-based learning and data augmentation, which can better be examined in Table 15

Table 15. Mean DSC scores for U-Net configurations (validation set).

Model ID	GM (\pm)	WM (\pm)	CSF (\pm)	Average DSC (\pm)
D1	0.734 ± 0.010	0.592 ± 0.022	0.000 ± 0.001	0.442 ± 0.008
D2	0.813 ± 0.026	0.809 ± 0.049	0.386 ± 0.112	0.669 ± 0.063
D3	0.523 ± 0.038	0.753 ± 0.028	0.361 ± 0.196	0.546 ± 0.072
D4	0.609 ± 0.013	0.742 ± 0.120	0.400 ± 0.185	0.584 ± 0.089
D5	0.881 ± 0.033	0.844 ± 0.082	0.528 ± 0.260	0.751 ± 0.115
D6	0.850 ± 0.026	0.821 ± 0.064	0.385 ± 0.245	0.686 ± 0.111

Model D5 achieved the highest overall Dice score (0.751 ± 0.115), with strong and stable performance on GM and WM and the best CSF segmentation of all models tested. Patch-based training and augmentation techniques were essential to its success. While deeper (D6) or weighted-loss variants (D3, D4) showed promise, they did not surpass D5, likely due to diminishing returns and increased variability on the limited IBSR18 dataset. The full-volume baseline (D1) performed poorly, especially on CSF, underscoring the value of localized patch extraction and augmentation. The per-volume Dice scores for U-Net configuration D5 are presented in Table 16, further illustrating its consistency across most validation subjects and its relative strength in segmenting all three tissue types.

Table 16. Per-volume DSC scores for U-Net configuration D5.

IBSR Volume	GM	WM	CSF	Average DSC
IBSR-11	0.9002	0.9068	0.7104	0.8391

IBSR-12	0.8961	0.8909	0.7018	0.8296
IBSR-13	0.8901	0.8493	0.5285	0.7560
IBSR-14	0.9013	0.8913	0.6334	0.8087
IBSR-17	0.8159	0.6791	0.0661	0.5204

Compared to earlier models: K-Means (0.588), FCM (0.591), and CNN (0.372); the U-Net model family achieved substantial improvements, particularly in detecting low-contrast tissue like CSF. Model D5 sets a strong foundation for the DenseUNet-based architectures evaluated in the next section.

4.6. DenseUNet

Densely connected U-Net variants (DenseUNet) were evaluated to investigate the effect of dense feature propagation and attention-enhanced skip connections on brain tissue segmentation. Configurations E1–E3 varied in network depth, growth rate, and attention usage. Table 17 presents the average Dice scores across validation volumes for each model.

Table 17. Mean DSC scores for DenseUNet configurations (validation set).

Model ID	GM (\pm)	WM (\pm)	CSF (\pm)	Average DSC (\pm)
E1	0.841 \pm 0.012	0.772 \pm 0.138	0.0308 \pm 0.043	0.522 \pm 0.062
E2	0.717 \pm 0.038	0.737 \pm 0.029	0.0062 \pm 0.012	0.486 \pm 0.009
E3	0.829 \pm 0.007	0.823 \pm 0.024	0.2632 \pm 0.220	0.639 \pm 0.087

Among the tested variants, E3 achieved the highest mean DSC (0.639 \pm 0.087), with strong and consistent performance for GM and WM, and a substantial improvement in CSF segmentation. This model employed attention gates in its skip connections, which dynamically filtered encoder outputs based on spatial relevance, helping the decoder recover finer structural details, particularly for underrepresented classes like CSF.

In contrast, E1, a simpler DenseUNet without attention, segmented GM and WM reliably but failed to generalize to CSF. E2, the deepest configuration with aggressive growth and depth, underperformed across all tissues. Its poor CSF results suggest overfitting or insufficient regularization, further affected by the small and imbalanced nature of the IBSR18 dataset. Table 18 reports the per-volume Dice scores for DenseUNet configuration E3, highlighting its improved generalization and relatively robust performance across validation subjects, particularly in CSF segmentation.

Table 18. Per-volume DSC scores for DenseUNet configuration E3.

IBSR Volume	GM	WM	CSF	Average DSC
IBSR-11	0.8364	0.8658	0.5114	0.7379
IBSR-12	0.8205	0.8225	0.5351	0.7261
IBSR-13	0.8340	0.7969	0.1406	0.5905
IBSR-14	0.8320	0.8334	0.1203	0.5952
IBSR-17	0.8225	0.8005	0.0084	0.5438

Compared to the best-performing U-Net model (D5, 0.751 ± 0.115), E3 underperformed slightly (0.639 ± 0.087), particularly in CSF and WM consistency. While DenseUNet introduces richer feature reuse, its benefits may be constrained under data-limited conditions such as IBSR18. These results underscore the importance of combining effective architectural design with dataset-appropriate regularization.

These findings lay the groundwork for the next section, which presents a hybrid architecture that integrates the strengths of both U-Net and DenseUNet models.

4.7. Hybrid DenseUNet with U-Net Decoder

The hybrid model (F1) combines densely connected convolutional encoding blocks with a U-Net-style decoder, aiming to balance feature richness and spatial accuracy without introducing attention mechanisms. This design aimed to improve segmentation, especially for

underrepresented tissues, while maintaining moderate architectural complexity. Table 19 presents the mean validation Dice scores for the hybrid model F1, demonstrating its ability to segment all three tissue types while maintaining a balance between performance and architectural simplicity.

Table 19. Mean DSC scores for Hybrid model F1 (validation set).

Model ID	GM	WM	CSF	Average DSC
F1	0.805 ± 0.084	0.720 ± 0.171	0.266 ± 0.226	0.597 ± 0.141

Model F1 achieved an average Dice score of 0.597 ± 0.141 , placing it among the top-performing models in the study, behind U-Net D5 (0.751 ± 0.115) and DenseUNet E3 (0.639 ± 0.087). Despite its relatively simple architecture, lacking attention and aggressive depth, it produced consistent results for GM and WM and a meaningful gain in CSF segmentation over early models like CNN and K-Means. Its performance highlights the strength of architectural synergy: dense feature reuse in the encoder combined with lightweight spatial reconstruction in the decoder.

The model was not heavily fine-tuned beyond baseline configuration and still managed to perform competitively. This suggests that even a modest integration of dense and U-Net design principles can produce a robust and generalizable model under data-limited conditions like those posed by the IBSR18 dataset. Table 20 provides the per-volume Dice scores for the hybrid model F1, offering further insight into its segmentation consistency across different validation subjects and its variable performance on CSF.

Table 20. Per-volume DSC scores for Hybrid model F1.

IBSR Volume	GM	WM	CSF	Average DSC
IBSR-11	0.8682	0.8706	0.4939	0.7442
IBSR-12	0.8374	0.8289	0.6180	0.7614
IBSR-13	0.8381	0.7383	0.1280	0.5681
IBSR-14	0.8267	0.7704	0.0892	0.5621

IBSR-17	0.6530	0.3898	0.0000	0.3476
---------	--------	--------	--------	--------

Performance varied more across volumes compared to U-Net D5 or DenseUNet E3, especially in CSF segmentation. Still, F1 demonstrated the ability to generalize anatomical features across most cases while remaining computationally efficient. Although it did not outperform the best U-Net model, its results validate the potential of hybrid architectures that blend dense connectivity with spatially-aware decoding.

4.8. Model Comparison and Segmentation Visualization

To enable a direct comparison between approaches, Table 21 lists the best-performing configuration from each model family, evaluated on its top validation volume. Dice scores for GM, WM, and CSF are included along with the average DSC.

Table 21. Summary of best-performing models and their top-scoring volumes

Model Type	Model ID	Best Volume	GM	WM	CSF	Average DSC
K-Means	A2	IBSR-14	0.8970	0.8905	0.0001	0.5959
FCM	B3	IBSR-17	0.9065	0.8866	0.0009	0.5980
CNN	C1	IBSR-11	0.574	0.565	0.0000	0.380
U-Net	D5	IBSR-11	0.9002	0.9068	0.7104	0.8391
DenseUNet	E3	IBSR-11	0.8364	0.8658	0.5114	0.7379
Hybrid	F1	IBSR-12	0.8374	0.8289	0.6180	0.7614

Among all models, U-Net configuration D5 achieved the highest Dice score, followed by the hybrid model F1 and DenseUNet E3. Traditional clustering methods (K-Means, FCM) failed to segment CSF effectively despite strong GM and WM scores. Deep learning models consistently outperformed classical clustering approaches, especially in segmenting complex and low-contrast structures like CSF.

Figure 7 visualizes the segmentation results of each best-performing model on their top validation volumes, highlighting differences in anatomical accuracy and tissue boundary delineation. Complementing this, Table 22 summarizes the average Dice scores across the validation set for the top configuration from each model family.

Figure 7. Predicted segmentations for best-performing configurations on their top validation volumes, alongside ground truth and input MRI.

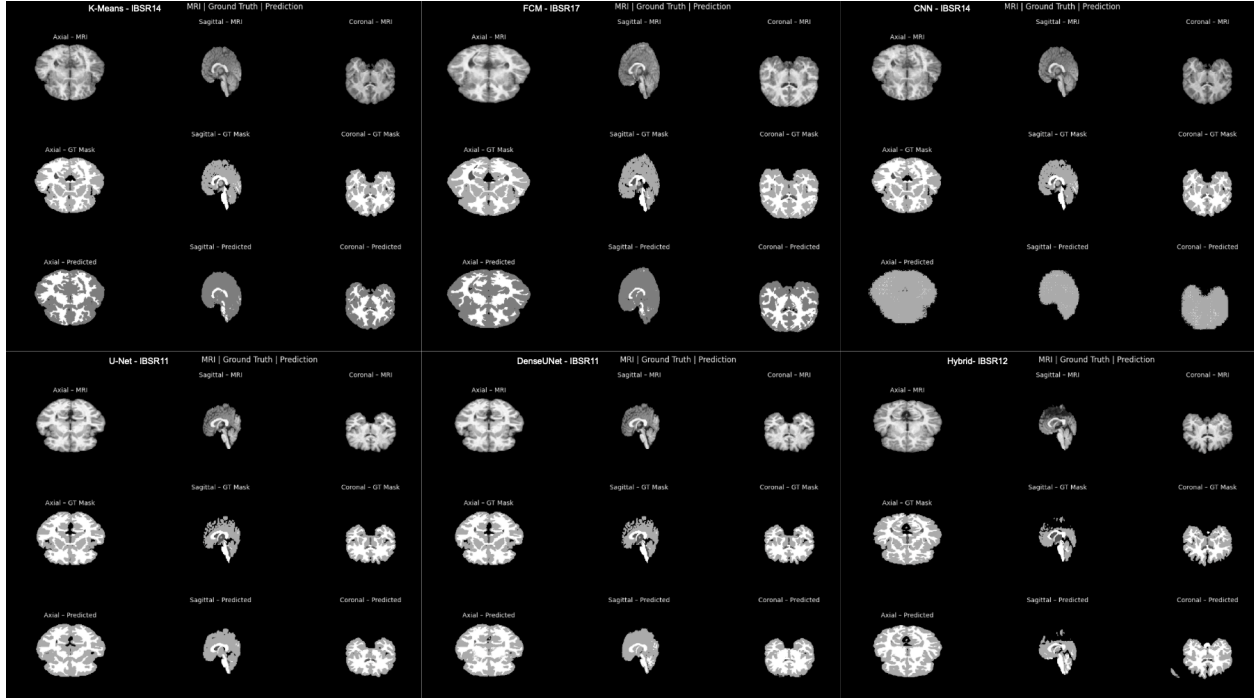


Table 22. Average DSC across validation set for best configuration of each model

Model Type	Model ID	Average DSC	CSF	GM	WM
K-Means	A2	0.588	0.000	0.882	0.883
FCM	B3	0.591	0.001	0.891	0.882
CNN	C1	0.372	0.000	0.572	0.542
U-Net	D5	0.751	0.528	0.881	0.844

DenseUNet	E3	0.639	0.263	0.829	0.824
Hybrid	F1	0.597	0.266	0.805	0.720

While Tables 21 and 22 provide a quantitative overview based on the validation set, qualitative results on the unseen test set further demonstrate the model's generalization ability. The following Figure 8 presents MRI slices and predicted tissue segmentations produced by the best-performing U-Net model (D5) across three test subjects (IBSR02, IBSR10, IBSR15), highlighting consistency and anatomical plausibility in the absence of ground truth.

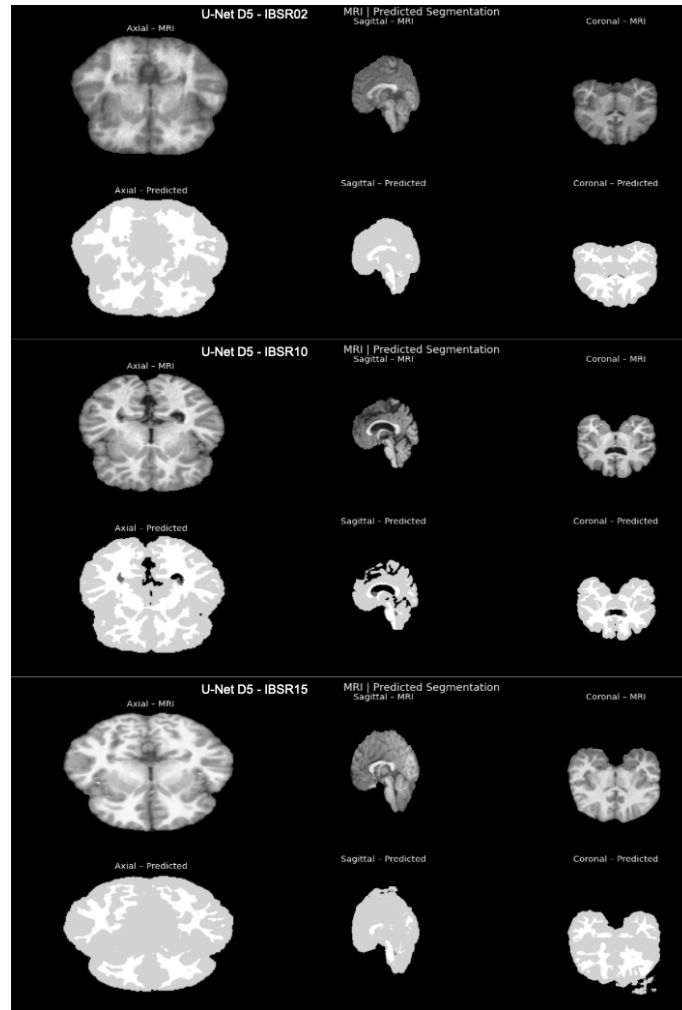


Figure 8. Predicted segmentations on three test volumes (IBSR02, IBSR10, IBSR15) using the U-Net D5 model.

Together, these quantitative and qualitative comparisons provide a comprehensive performance profile across all segmentation models. While validation results offer insight into architectural strengths and tissue-specific performance, the qualitative predictions on the IBSR18 test set (Figure 8) further demonstrate the generalization capability of the best-performing U-Net model (D5). The visual outputs across axial, sagittal, and coronal planes highlight the model’s ability to consistently delineate brain tissue structures, particularly gray and white matter, even in the absence of ground truth labels. The following section discusses these findings in greater depth, emphasizing segmentation trends, architectural trade-offs, and areas for future development.

4.9. Discussion

This study systematically evaluated a range of classical and deep learning models for 3D brain tissue segmentation on the IBSR18 dataset. The comparison spanned unsupervised clustering algorithms (K-Means, Fuzzy C-Means), shallow CNNs, and multiple encoder–decoder architectures, including U-Net, DenseUNet, and a custom hybrid model. The results reveal several key trends in segmentation performance, architectural effectiveness, and tissue-specific accuracy.

Classical clustering methods (K-Means, FCM) showed strong performance in segmenting gray matter (GM) and white matter (WM) but consistently failed to identify cerebrospinal fluid (CSF), yielding near-zero Dice scores for this class. Although these models are simple and computationally efficient, their reliance solely on voxel intensity makes them inadequate for detecting low-contrast or spatially diffuse structures like CSF. FCM slightly outperformed K-Means in overall consistency, but the improvement was marginal.

The shallow CNN baseline demonstrated some ability to localize GM and WM but remained limited in both precision and generalizability. Notably, it performed worse than K-Means and FCM in average Dice score and failed to detect CSF altogether. This underscores the limitations of depth-constrained models trained on full volumes without augmentation or spatial priors.

U-Net architectures marked a clear leap in performance, with model D5 achieving the highest overall Dice score (0.751 ± 0.115). Its ability to segment CSF (0.528 ± 0.260) far surpassed that

of all prior models. This success is largely attributed to the use of patch-based training, data augmentation, and skip connections that preserve spatial detail during decoding. Deeper or loss-weighted variants (D3, D4, D6) did not outperform D5, suggesting that overly complex modifications may yield diminishing returns on limited datasets.

DenseUNet models extended U-Net with dense connectivity, enabling deeper feature reuse. While the deepest variant (E2) suffered from overfitting and degraded CSF performance, the balanced model with attention gates (E3) achieved strong results across all tissue types (average DSC: 0.639 ± 0.087). Attention mechanisms helped focus on spatially relevant features, partially compensating for the limited training data. However, the IBSR18 dataset—due to its small size and class imbalance—posed a major constraint for deeper and parameter-heavy architectures like DenseUNet. These models require large and diverse datasets to reach their full potential, and in this study, their capacity often exceeded what the data could support.

The hybrid model (F1), combining DenseNet-style encoders with a U-Net-style decoder, delivered competitive results (average DSC: 0.597 ± 0.141) despite its relatively simple architecture. It outperformed all classical models and the CNN baseline while maintaining computational efficiency. While it did not surpass the best U-Net or DenseUNet variants in this configuration, its performance, particularly in CSF segmentation, demonstrates strong potential. With further tuning and regularization, the hybrid model could feasibly outperform all other models evaluated in this study. Its design effectively blends dense feature reuse with spatial reconstruction, making it well-suited for anatomically complex segmentation tasks.

Across all models, CSF remained the most challenging tissue to segment, due to its low contrast, small volume, and underrepresentation in the training set. Deep learning models consistently benefited from patch-based training and augmentation strategies, which helped capture localized anatomical structures. In contrast, full-volume and parameter-heavy models were prone to overfitting or underutilization of capacity due to data scarcity.

In summary, the findings suggest that balanced encoder–decoder architectures, supported by augmentation and lightweight design, are most effective under data-constrained conditions like those of IBSR18. While U-Net D5 remains the top performer, DenseUNet E3 and hybrid model F1 offer competitive and extensible alternatives. Evaluating models both quantitatively and

qualitatively proved essential in uncovering performance nuances, particularly for tissue-specific segmentation quality and anatomical fidelity.

SECTION 5

CONCLUSION

This thesis explored and evaluated a range of machine learning and deep learning models for brain tissue segmentation on the IBSR18 dataset, focusing on cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM). Starting with classical clustering techniques like K-Means and Fuzzy C-Means (FCM), the study progressed to deep architectures including 3D CNNs, U-Net, DenseUNet, and a custom hybrid model. Each model family was systematically assessed based on the Dice Similarity Coefficient (DSC), a standard metric for evaluating overlap between predicted and ground truth segmentations.

The experimental results confirm that deep learning methods., particularly encoder–decoder architectures such as U-Net and DenseUNet, significantly outperform traditional clustering techniques in capturing complex anatomical structures, especially underrepresented tissues like CSF. Among all models tested, the U-Net configuration D5 achieved the highest average DSC (0.7508), highlighting the effectiveness of patch-based training, skip connections, and architectural balance for volumetric segmentation.

Nonetheless, several limitations emerged. The IBSR18 dataset's small size and limited diversity constrained the ability of deeper models to generalize, particularly for CSF. The absence of post-processing may have affected boundary smoothness, and no transfer learning or cross-dataset validation was conducted, limiting the broader applicability of the results. Additionally, relying solely on DSC may not fully capture all aspects of segmentation performance.

Future work should focus on training with larger, multimodal datasets and exploring transfer learning and domain adaptation techniques to improve robustness in clinical contexts. Attention-enhanced decoding, transformer-based architectures, and semi-supervised learning could further address data scarcity challenges. Incorporating post-processing methods, such as conditional random fields or morphological filtering, may also refine segmentation outputs.

In conclusion, this study contributes a structured, comparative analysis of segmentation models on a benchmark neuroimaging dataset. It reinforces the strength of U-Net-based architectures and emphasizes the importance of data quality, regularization, and model simplicity in developing scalable and clinically meaningful brain segmentation systems.

REFERENCES

- [1] E. Bercovich and M. C. Javitt, “Medical Imaging: From Roentgen to the Digital Revolution, and Beyond,” *Rambam Maimonides Medical Journal*, vol. 9, no. 4, p. e0034, Oct. 2018, doi: <https://doi.org/10.5041/rmmj.10355>.
- [2] I. Despotović, B. Goossens, and W. Philips, “MRI Segmentation of the Human Brain: Challenges, Methods, and Applications,” *Computational and Mathematical Methods in Medicine*, vol. 2015, no. 1, pp. 1–23, 2015, doi: <https://doi.org/10.1155/2015/450341>.
- [3] L. Wu *et al.*, “A survey of MRI-based brain tissue segmentation using deep learning,” *Complex & Intelligent Systems*, vol. 11, no. 1, Dec. 2024, doi: <https://doi.org/10.1007/s40747-024-01639-1>.
- [4] M. Avanzo, J. Stancanello, G. Pirrone, A. Drigo, and A. Retico, “The Evolution of Artificial Intelligence in Medical Imaging: From Computer Science to Machine and Deep Learning,” *Cancers*, vol. 16, no. 21, pp. 3702–3702, Nov. 2024, doi: <https://doi.org/10.3390/cancers16213702>.
- [5] A. Fawzi, A. Achuthan, and B. Belaton, “Brain Image Segmentation in Recent Years: A Narrative Review,” *Brain Sciences*, vol. 11, no. 8, p. 1055, Aug. 2021, doi: <https://doi.org/10.3390/brainsci11081055>.
- [6] D. R. Sarvamangala and R. V. Kulkarni, “Convolutional neural networks in medical image understanding: a survey,” *Evolutionary Intelligence*, vol. 15, no. 1, Jan. 2021, doi: <https://doi.org/10.1007/s12065-020-00540-3>.
- [7] S. Valverde, A. Oliver, M. Cabezas, E. Roura, and X. Lladó, “Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations,” *Journal of Magnetic Resonance Imaging*, vol. 41, no. 1, pp. 93–101, Jan. 2014, doi: <https://doi.org/10.1002/jmri.24517>.
- [8] M. Sucharitha and K. P. Geetha, “Brain tissue segmentation using fuzzy clustering techniques,” *Technology and Health Care*, vol. 23, no. 5, pp. 571–580, Sep. 2015, doi: <https://doi.org/10.3233/thc-151012>.
- [9] P. Kumar, P. Nagar, C. Arora, and A. Gupta, “U-Segnet: Fully Convolutional Neural Network Based Automated Brain Tissue Segmentation Tool,” *International Conference on Image Processing*, Oct. 2018, doi: <https://doi.org/10.1109/icip.2018.8451295>.

- [10] L. Wang, C. Xie, and N. Zeng, “RP-Net: A 3D Convolutional Neural Network for Brain Segmentation From Magnetic Resonance Imaging,” *IEEE Access*, vol. 7, pp. 39670–39679, 2019, doi: <https://doi.org/10.1109/access.2019.2906890>.
- [11] P. Ghosal, T. Chowdhury, A. Kumar, A. K. Bhadra, J. Chakraborty, and D. Nandi, “MhURI: A Supervised Segmentation Approach to Leverage Salient Brain Tissues in Magnetic Resonance Images,” *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105841, Nov. 2020, doi: <https://doi.org/10.1016/j.cmpb.2020.105841>.
- [12] J.-S. Long, G.-Z. Ma, E.-M. Song, and R.-C. Jin, “Learning U-Net Based Multi-Scale Features in Encoding-Decoding for MR Image Brain Tissue Segmentation,” *Sensors*, vol. 21, no. 9, p. 3232, Jan. 2021, doi: <https://doi.org/10.3390/s21093232>.
- [13] F. Zhang *et al.*, “Deep learning based segmentation of brain tissue from diffusion MRI,” *NeuroImage*, vol. 233, p. 117934, Jun. 2021, doi: <https://doi.org/10.1016/j.neuroimage.2021.117934>.
- [14] R. Basnet, M. O. Ahmad, and M. N. S. Swamy, “A deep dense residual network with reduced parameters for volumetric brain tissue segmentation from MR images,” *Biomedical Signal Processing and Control*, vol. 70, p. 103063, Sep. 2021, doi: <https://doi.org/10.1016/j.bspc.2021.103063>.
- [15] L. Zhao *et al.*, “Automated 3D Fetal Brain Segmentation Using an Optimized Deep Learning Approach,” *American Journal of Neuroradiology*, vol. 43, no. 3, pp. 448–454, Mar. 2022, doi: <https://doi.org/10.3174/ajnr.A7419>.
- [16] M. Li, J. Zhou, D. Wang, P. Peng, and Y. Yu, “Application of Clustering-Based Analysis in MRI Brain Tissue Segmentation,” *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–16, Aug. 2022, doi: <https://doi.org/10.1155/2022/7401184>.
- [17] K. Munir, F. Frezza, and A. Rizzi, “Deep Learning Hybrid Techniques for Brain Tumor Segmentation,” *Sensors*, vol. 22, no. 21, p. 8201, Oct. 2022, doi: <https://doi.org/10.3390/s22218201>.
- [18] A. Chattopadhyay and M. Maitra, “MRI-based brain tumour image detection using CNN based deep learning method,” *Neuroscience Informatics*, vol. 2, no. 4, p. 100060, Dec. 2022, doi: <https://doi.org/10.1016/j.neuri.2022.100060>.

- [19] W. Al Shehri, “Alzheimer’s disease diagnosis and classification using deep learning techniques,” *PeerJ Computer Science*, vol. 8, p. e1177, Dec. 2022, doi: <https://doi.org/10.7717/peerj-cs.1177>.
- [20] X. Huang *et al.*, “Deep Learning-Based Multiclass Brain Tissue Segmentation in Fetal MRIs,” *Sensors*, vol. 23, no. 2, pp. 655–655, Jan. 2023, doi: <https://doi.org/10.3390/s23020655>.
- [21] V. De, Midas Meijs, A. Patel, Frederick, Bram Van Ginneken, and Rashindra Manniesing, “Multiclass Brain Tissue Segmentation in 4D CT Using Convolutional Neural Networks,” *IEEE access*, vol. 7, pp. 51557–51569, Jan. 2019, doi: <https://doi.org/10.1109/access.2019.2910348>.
- [22] N. Yamanakkanavar and B. Lee, “Using a Patch-Wise M-Net Convolutional Neural Network for Tissue Segmentation in Brain MRI Images,” *IEEE Access*, vol. 8, pp. 120946–120958, 2020, doi: <https://doi.org/10.1109/access.2020.3006317>.
- [23] W. Ding, M. Abdel-Basset, H. Hawash, and W. Pedrycz, “Multimodal Infant Brain Segmentation by Fuzzy-informed Deep Learning,” *IEEE Transactions on Fuzzy Systems*, pp. 1–1, 2021, doi: <https://doi.org/10.1109/tfuzz.2021.3052461>.
- [24] A. Van opbroek, F. Van der Lijn, and M. De Bruijne, “Automated Brain-Tissue Segmentation by Multi-Feature SVM Classification,” *The MIDAS Journal*, Oct. 2013, doi: <https://doi.org/10.54294/ojfo7q>.
- [25] S. E. Sorour, Amr A. Abd El-Mageed, K. M. Albarrak, A. K. Alnaim, A. A. Wafa, and Engy El-Shafeiy, “Classification of Alzheimer’s disease using MRI data based on Deep Learning Techniques,” *Journal of King Saud University - Computer and Information Sciences*, pp. 101940–101940, Jan. 2024, doi: <https://doi.org/10.1016/j.jksuci.2024.101940>.