

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**DEEP LEARNING APPROACHES FOR MULTIPLE SCLEROSIS LESION
SEGMENTATION USING MULTI-SEQUENCE 3D MR IMAGES**



Ph.D. THESIS

Beytullah SARICA

Department of Applied Informatics

Geographical Information Technologies Programme

MARCH 2023

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**DEEP LEARNING APPROACHES FOR MULTIPLE SCLEROSIS LESION
SEGMENTATION USING MULTI-SEQUENCE 3D MR IMAGES**

Ph.D. THESIS

**Beytullah SARICA
(706142007)**

Department of Applied Informatics

Geographical Information Technologies Programme

Thesis Advisor: Prof. Dr. Dursun Zafer ŞEKER

MARCH 2023

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**ÇOK SEKANSLI 3B MR GÖRÜNTÜLERİ KULLANILARAK MULTİPLE
SKLEROZ LEZYON BÖLÜTLEMESİ İÇİN DERİN ÖĞRENME
YAKLAŞIMLARI**

DOKTORA TEZİ

**Beytullah SARICA
(706142007)**

Bilişim Uygulamaları Anabilim Dalı

Coğrafi Bilgi Teknolojileri Programı

Tez Danışmanı: Prof. Dr. Dursun Zafer ŞEKER

MART 2023

Beytullah SARICA, a Ph.D. student of ITU Graduate School student ID 706142007, successfully defended the thesis entitled “DEEP LEARNING APPROACHES FOR MULTIPLE SCLEROSIS LESION SEGMENTATION USING MULTI-SEQUENCE 3D MR IMAGES”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Dursun Zafer ŞEKER**
Istanbul Technical University

Jury Members : **Prof. Dr. Hande DEMİREL**
Istanbul Technical University

Prof. Dr. Bülent BAYRAM
Yildiz Technical University

Prof. Dr. Zaide DURAN
Istanbul Technical University

Prof. Dr. Ash GARAGON
Boğaziçi University

Date of Submission : 17 February 2023

Date of Defense : 30 March 2023





To my spouse and my family,



FOREWORD

I would like to express my deepest gratitude to my supervisor, Prof. Dr. Dursun Zafer ŞEKER, for his encouragement, invaluable guidance, and insightful comments that have enabled me to complete this thesis. His guidance and encouragement have been instrumental in my growth as a researcher. I will always be grateful for his support and dedication to my work. I would also like to express my sincere appreciation to my thesis committee members, Prof. Dr. Hande DEMİREL and Prof. Dr. Bülent BAYRAM, for their helpful guidance and feedback.

Last but not least, I am thankful to my spouse, parents, siblings, niece, and friends for their love and support during the completion of this thesis. Their encouragement has been essential to my success.

March 2023

Beytullah SARICA
(Software Engineer)

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD.....	ix
TABLE OF CONTENTS.....	xi
ABBREVIATIONS	xiii
LIST OF TABLES	xv
LIST OF FIGURES	xvii
SUMMARY	xxi
ÖZET	xxiii
1. INTRODUCTION	1
1.1 Purpose of Thesis	5
1.2 Thesis Organization.....	6
2. LITERATURE REVIEW.....	9
2.1 Multiple Sclerosis (MS)	9
2.2 Magnetic Resonance Imaging (MRI)	10
2.3 MRI Analysis for MS	12
2.3.1 Preprocessing of MR images	13
2.3.2 MS lesion segmentation.....	15
2.3.3 MRI in MS lesion segmentation	16
2.3.4 MS lesion activity segmentation	16
2.3.5 CNN-based MS lesion segmentation	18
3. DATA, METHODOLOGY, AND IMPLEMENTATION.....	23
3.1 Data Used	23
3.1.1 ISBI2015 longitudinal MS lesion segmentation dataset.....	23
3.1.2 MSSEG2016 dataset	25
3.1.3 MSSEG-2 dataset.....	27
3.1.4 Data preprocessing and preparation.....	28
3.2 Methodology.....	30
3.2.1 U-Net.....	31
3.2.2 Feature pyramid network (FPN)	32
3.2.3 Residual learning (RL).....	33
3.2.4 Attention gate (AG)	34
3.2.5 Efficient channel attention (ECA).....	35
3.2.6 Atrous spatial pyramid pooling (ASPP)	36
3.2.7 Loss functions	37
3.2.7.1 Binary cross-entropy loss	38
3.2.7.2 Categorical cross-entropy loss	38
3.2.7.3 Dice loss.....	38
3.2.7.4 Binary focal loss	39
3.2.7.5 Categorical focal loss.....	39

3.2.8 Evaluation metrics	39
3.3 Proposed a Dense Residual U-Net for MS Lesion Segmentation and Detection.....	42
3.4 Proposed Deep Residual Attention Gate U-Net for MS Lesion Activity	46
3.5 Implementation	49
4. RESULTS AND DISCUSSIONS	55
4.1 ISBI2015 Dataset Results	55
4.1.1 Ablation study	55
4.2 MSSEG2016 Dataset Results	63
4.3 Cross-dataset Validation Results	67
4.4 MSSEG-2 Dataset Results.....	68
4.5 Discussions	74
5. CONCLUSIONS AND RECOMMENDATIONS.....	85
REFERENCES.....	89
CURRICULUM VITAE.....	103



ABBREVIATIONS

AG	: Attention Gate
ASPP	: Atrous Spatial Pyramid Pooling
BN	: Batch Normalization
BCE	: Binary Cross-Entropy
BET	: Brain Extraction Tool
BSE	: Brain Surface Extractor
CRF	: Conditional Random Field
CNN	: Convolutional Neural Network
convGRUs	: Convolutional Gated Recurrent Units
CNS	: Central Nervous System
DSC	: Dice Similarity Coefficient
DL	: Deep Learning
DLB	: DeepLesionBrain
DMD	: Disease Modifying Drug
ECA	: Efficient Channel Attention
FCN	: Fully Convolutional Networks
FLAIR	: Fluid-Attenuated Inversion Recovery
FN	: False Negative
FP	: False Positive
FPN	: Feature Pyramid Networks
GAP	: Global Average Pooling
GPU	: Graphic Processing Unit
KDE	: Kernel Density Estimation
LFPR	: Lesion-Wise False Positive Rate
LTPR	: Lesion-Wise True Positive Rate
MRI	: Magnetic Resonance Imaging
MS	: Multiple Sclerosis
PPV	: Positive Predictive Value
PD-w	: Proton Density-Weighted
ReLU	: Rectified Linear Unit
RL	: Residual Learning
S_e	: Sensitivity
S_p	: Specificity
T	: Tesla
T1-w	: T1-weighted
T2-w	: T2-weighted
TE	: Echo Time
TN	: True Negative
TP	: True Positive
TR	: Repetition Time
VIP	: Virtual Imaging Platform

WM : White Matter



LIST OF TABLES

	<u>Page</u>
Table 4.1 : Ablation study results with different variants of our model trained on the ISBI2015 training set and tested on the ISBI2015 testing set. To be ranked first, the ISBI score, Dice score, PPV, and LTPR are expected to have high numerical values, while the LFPR and VD are expected to have low numerical values.	57
Table 4.2 : Comparison of the result of this study with other state-of-the-art methods published results for the ISBI2015 dataset. To be ranked first and second, the ISBI score, Dice score, PPV, and LTPR are expected to have high numerical values, while the LFPR and VD are expected to have low numerical values.	62
Table 4.3 : Results comparison of the proposed model for each center on the MSSEG2016 test set. To be ranked first, all metrics are expected to have high numerical values.	63
Table 4.4 : Comparison of the results of this study with the manual delineations of experts for the MSSEG2016 testing set. All metrics are expected to have high numerical values to be considered the best and second-best scores.	64
Table 4.5 : Results comparison of cross-dataset, namely, training on the ISBI2015 and testing on the MSSEG2016 or vice versa. Dice score, PPV, and sensitivity are expected to have high numerical values.	68
Table 4.6 : Results comparison with other methods when trained on the MSSEG2016 dataset and tested on the ISBI2015 testing set. To be ranked first and second, the ISBI score, Dice score, PPV, LTPR, and Cor are expected to have high numerical values, while the LFPR is expected to have low numerical values.	68
Table 4.7 : Prediction results of evaluating the challenge test data set published on the challenge website. Bold and italic values are the highest and the second-best scores among other proposed methods excluding the experts, respectively. Dice and F1 Score are expected to be a high numerical value while the Number of and Volume of Lesions are expected to be a low numerical value. The Number of and Volume of Lesions metrics are calculated for no new lesion cases.	70
Table 4.8 : Prediction results of evaluating the challenge test data set published on the challenge website for other useful metrics. Bold and italic values are the highest and the second-best scores among some of the proposed methods and the experts, respectively. Sensitivity, Specificity, and PPV are expected to be a high numerical value while Surface Distance is expected to be a low numerical value.	73

Table 4.9 : The evaluation results of the proposed method with different components using a subset of the MSSEG-2 test dataset.	74
---	-----------



LIST OF FIGURES

	<u>Page</u>
Figure 1.1 : The workflow of the thesis.....	4
Figure 2.1 : The nerve affected by MS. Image extracted from https://www.news-medical.net/health/Types-of-Multiple-Sclerosis-%28MS%29.aspx	10
Figure 2.2 : A 3D brain MRI visualization with its three plane orientations. The red arrow shows an axial slice, the green arrow shows a coronal slice, and the blue arrow shows a sagittal slice.....	11
Figure 2.3 : An axial view of different MR images of the brain from the ISBI2015 dataset: a) FLAIR b) T1-w, c) T2-w, and d) PD-w.	12
Figure 3.1 : Illustration of the preprocessed ISBI2015 dataset training samples with axial, sagittal, and coronal views of FLAIR, T1-w, and T2-w modalities with their corresponding manual delineations by two raters, Rater #1 and Rater #2.	25
Figure 3.2 : Illustration of the preprocessed MSSEG2016 dataset training samples with axial, sagittal, and coronal views of FLAIR, T1-w, and T2-w modalities with their consensus manual delineation.	26
Figure 3.3 : The raw, preprocessed, and delineated mask slices including two-time points for the new MS lesions segmentation task.....	27
Figure 3.4 : A framework for MS lesion segmentation using the proposed model with 2D slices extracted from 3D MR images of multiple sequences including the output of the final 3D binary mask.	28
Figure 3.5 : The illustration of U-Net architecture with a contracting path, a bottleneck, and an expansive path.....	32
Figure 3.6 : The illustration of FPN architecture and a building block showing the lateral connection and the top-down pathway.....	33
Figure 3.7 : Original residual unit with identity mapping. x_l and x_{l+1} are the input and output of the l -th unit, respectively.	34
Figure 3.8 : Schematic of the attention gate.....	35
Figure 3.9 : Atrous Spatial Pyramid Pooling (ASPP). ASPP uses multiple parallel filters with varying rates to take advantage of multi-scale features in order to classify the central pixel (shown in red). The different Field-Of-Views are depicted in different colors.....	37
Figure 3.10 : The dense residual U-Net with AG, ECA, and ASPP using multi-planar 2D slices of FLAIR, T1-w, and T2-w sequences. The number of filters starts from 32 and increases two times up to 512. In the residual block, x_l and x_{l+1} denote the input and output of the l -th unit. In the attention gate, x_l , g , and α denote the input feature map of the layer l , the gating signal providing contextual information, and the attention coefficient used to scale the x_l , respectively.....	45

Figure 3.11	: a) Diagram of the ECA module. GAP is used to obtain the aggregated features from the given input. In ECA, a fast 1D convolution of size k is performed to obtain channel weights, where a mapping of the channel dimension C is used to adaptively calculate k . b) Diagram of the ASPP module. From given the features, ASPP performs multi-rate convolution operations to obtain multi-scale information.	46
Figure 3.12	: A residual unit with identity mapping. x_l and x_{l+1} are the input and output of the l -th unit, respectively.....	46
Figure 3.13	: The architecture of the proposed model combines U-Net, residual learning, attention gate, and a slice-based approach. In AG adapted from Oktay et al., x^l is the input features, α is the attention coefficients used to scale the x^l , and g collected from a coarser scale is the gating signal which provides contextual information.....	48
Figure 3.14	: The proposed pipeline of new MS lesions segmentation using a slice-based approach including the majority voting for the final 3D segmentation output using the predicted 2D axial, sagittal, and coronal slices.....	52
Figure 4.1	: Evaluation metric results on the ISBI2015 testing set for each variant of the proposed model.....	56
Figure 4.2	: The boxplot of Dice and PPV scores for each patient across raters on the ISBI2015 testing set for the proposed model and its variants. Asterisks indicate statistical significance (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$) when using a Wilcoxon test compared to the proposed model (denoted as A).....	59
Figure 4.3	: The segmentation results of each variant of the proposed model on the ISBI2015 testing set. A slice of axial, sagittal, and coronal views of FLAIR from timepoint 1 of test subject 9.	60
Figure 4.4	: The segmentation results of the proposed model on the MSSEG2016 testing set compared to the manual delineation of experts and their consensus ground truth. A slice of axial, sagittal, and coronal views of FLAIR from each center is demonstrated.	65
Figure 4.5	: Evaluation metric results on the MSSEG2016 testing dataset for our model and experts. The figure is sorted in descending order of the Dice score.	66
Figure 4.6	: The boxplot of Dice score, F1 score, PPV, and sensitivity evaluation metrics of each patient for experts and the proposed model on the MSSEG2016 testing set. Asterisks and ns indicate statistical significance (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$, and ns (non-significance)) when using a Wilcoxon test compared to the proposed model.	67
Figure 4.7	: The best and worst performances of the proposed model compared to the consensus and each expert segmentation for F1 and dice scores. A slice of axial view from patients 6 and 2 for the F1 score and patients 60 and 53 for the dice score is presented. ..	72
Figure 4.8	: Analysis of differences in detection and segmentation by using F1 and dice scores for each expert and each team that participated in the challenge, respectively.	78

Figure 4.9 : Analysis of the number and volume of lesions detection for each expert and each team that participated in the challenge (The data of volume of tested lesions was scaled by $\log_{10}$).....	79
--	-----------





DEEP LEARNING APPROACHES FOR MULTIPLE SCLEROSIS LESION SEGMENTATION USING MULTI-SEQUENCE 3D MR IMAGES

SUMMARY

Multiple Sclerosis (MS) is a chronic inflammatory, immune-mediated, neurodegenerative, and demyelinating disease that impacts the Central Nervous System (CNS). The disease can cause permanent damage or deterioration (demyelination) to the nerves in the CNS. This damage results in the formation of lesions or plaques in the nervous system, leading to a wide range of symptoms such as problems with vision, loss of coordination, muscle weakness, and cognitive impairment. Early diagnosis and monitoring of MS are crucial since diagnosing the disease in its advanced stages can be more challenging. Therefore, effective methods for diagnosing and monitoring MS in its early stages are needed to improve patient quality of life and treatment outcomes. Magnetic Resonance Imaging (MRI) is widely used for monitoring, measuring, detecting, and characterizing MS lesions. T1-weighted (T1-w), T2-weighted (T2-w), and Fluid-Attenuated Inversion Recovery (FLAIR) sequences are commonly exploited in MS diagnosis as they provide different information about the brain tissues and the presence of lesions. Thereby, MRI is a useful tool for diagnosing and monitoring MS. Recently, Deep Learning (DL) methods have achieved remarkable results in the automated segmentation of MS lesions from MRI data, potentially improving the accuracy and efficiency of MS diagnosis and monitoring. Although automated methods for MS lesion segmentation have usually been performed on individual MRI scans, tracking lesion activity for quantifying and monitoring MS disease progression, especially detecting new lesions, has become an important biomarker in recent years. This Ph.D. thesis aims to develop novel and fully automated DL approaches for detecting and segmenting MS lesions from a single time-point brain MRI of a patient and also new MS lesions between two time points brain MRI of a patient. DL techniques simplify the feature extraction process from the given input data. Therefore, in this thesis, DL approaches were investigated and examined, then exploited to improve the segmentation and detection of MS lesions for both challenging tasks. Accordingly, a novel dense residual U-Net model that combines Attention Gate (AG), Efficient Channel Attention (ECA), and Atrous Spatial Pyramid Pooling (ASPP) is proposed to enhance the performance of the automatic MS lesion segmentation using 3D MRI sequences. Similarly, a unique pipeline with a deep neural network that combines U-Net, attention gate, and residual learning is proposed to perform better MS new lesion segmentation using baseline and follow-up 3D FLAIR MR images for lesion activity determination.

In the proposed novel dense residual U-Net model, convolution layers in each block of the U-Net architecture are replaced by residual blocks and connected densely. Then,

AGs are exploited to capture salient features passed through the skip connections. The ECA module is appended at the end of each residual block and each downsampling block of U-Net. Later, the bottleneck of U-Net is replaced with the ASSP module to extract multi-scale contextual information. Furthermore, 3D MR images of FLAIR, T1-w, and T2-w are exploited jointly to perform better MS lesion segmentation. The proposed model is validated on the publicly available ISBI2015 and MSSEG2016 challenge datasets. This model produced an ISBI score of 92.75, a mean Dice score of 66.88%, a mean Positive Predictive Value (PPV) of 86.50%, and a mean Lesion-Wise True Positive Rate (LTPR) of 60.64% on the ISBI2015 testing set. Also, it achieved a mean Dice score of 67.27%, a mean PPV of 65.19%, and a mean sensitivity of 74.40% on the MSSEG2016 testing set. The results show that the proposed model performs better than the results of some experts and some of the other state-of-the-art methods realized related to this particular subject. Specifically, the best Dice score and the best LTPR are obtained on the ISBI2015 testing set by using the proposed model to segment MS lesions.

On the other hand, the generated model for the lesion activity determination within the proposed pipeline has a similar architecture to U-Net and is formed from residual units which facilitate the training of deep networks. Networks with fewer parameters are designed with better performance through the skip connections of U-Net and residual units, which facilitate information propagation without degradation. AGs also learn to focus on salient features of the target structures of various sizes and shapes. The MSSEG-2 challenge dataset was used for training and testing the proposed pipeline, and the results were compared with those of other proposed pipelines of the challenge and experts who participated in the same challenge. According to the results obtained from the testing set, the lesion-wise F1 and Dice scores were obtained as a mean of 48 and 44.30%. For the no-lesion cases, the number of tested and volume of tested lesions were obtained as a mean of 0.148 and 1.488, respectively. The proposed pipeline outperformed 22 proposed pipelines and ranked 8th in the challenge for the Dice and F1 scores. It was also ranked 4th and 5th for the number of tested and volume of tested lesions, respectively.

ÇOK SEKANSLI 3B MR GÖRÜNTÜLERİ KULLANILARAK MULTIPLE SKLEROZ LEZYON BÖLÜTLEMESİ İÇİN DERİN ÖĞRENME YAKLAŞIMLARI

ÖZET

Multiple Skleroz (MS), Merkezi Sinir Sistemini (MMS) etkileyen, otoimmün, kronik ve demiyelinizan bir hastalıktır. Bu hastalıkta nöronlar etrafındaki koruyucu kılıf (miyelin kılıf), bağışıklık sistemindeki (immün sistem) antikörler tarafından yabancı bir madde olarak algılanır, yok edilmeye çalışılır ve sonuç olarak koruyucu kılıf hasar görür ve sertleşmiş MS plakları oluşur. Hastalık sinirlerde kalıcı hasara veya bozulmaya neden olabilir. Miyelin tabaka üzerinde oluşan fiziksel tahribat nedeniyle beyinden yollanan mesajların iletilmesi engellenmektedir. MS hastalarında fonksiyonel yeti kayıpları görülebilir ve bu hastalığın ileri safhalarında geri döndürülemez beyin hasarları da oluşabilir. MS kesin tedavisi olmayan bir hastalıktır ve belirtileri hastadan hastaya değişiklik gösterebilmektedir. Bazı hastalarda yaşam konforu önemli ölçüde düşebilmekte, hasta bakıma muhtaç hale gelebilmektedir. Yapılan çalışmalar, MS hastalığının genellikle 20-45 yaş aralığında ortaya çıktığını göstermiştir. 25-35 yaş aralığında ise MS hastalığının görülme sıklığının yüksek olduğu ifade edilmiştir. Hastalığın kronik olması ve kesin tedavisinin olmaması, hastaların uzun seneler boyunca bu hastalıkla mücadele etmesi anlamına gelir. İleri evrelerde tanı daha zor hale gelebileceğinden, erken evrede MS tespiti ve sürecinin izlenmesi için etkili yöntemlere ihtiyaç vardır.

Manyetik Rezonans Görüntüleme (MRG), MS teşhisinde en önemli ve duyarlı görüntüleme yöntemidir. MRG, MS hastalığının tanı koyma sürecinde etkin olduğu gibi klinik olarak diğer hastalıklardan ayırıcı tanının yapılmasında da önemlidir. Ayrıca tedavi sürecinde kontrol görüntüleme ile tedaviye verilen cevabın değerlendirmesinde önemli bir yer tutmaktadır. MS tanısında genellikle T1 ağırlıklı (T1-w), T2 ağırlıklı (T2-w), Sıvı Zayıflatılmış İnversiyon Kurtarma (FLAIR) ve Proton Dansite ağırlıklı (PD-w) MRG sekansları kullanılmaktadır. MS plakları tipik olarak T1-w sekansta beyaz cevhere göre benzer veya düşük sinyalde, T2-w ve FLAIR sekansta ise yüksek sinyalde izlenmektedir.

Son yıllarda, derin öğrenme yöntemlerinin kullanılmasıyla, MS lezyonlarının MR görüntülerinden otomatik olarak segmentasyonunda dikkate değer sonuçlar üretilmiştir. MS lezyon segmentasyonu için otomatik yöntemler genellikle bireysel MRI taramalarında gerçekleştirilmiş olsa da son zamanlarda MS hastalığının ilerlemesini ölçmek ve izlemek için lezyon aktivitesinin izlenmesi, özellikle yeni lezyonların saptanması, önemli bir biyobelirteç haline gelmiştir.

Doktora tez çalışmasının amacı, MS lezyonlarının tek bir zaman noktalı beyin MR görüntülerinden ve ayrıca bir hastanın iki farklı zamanda elde edilen beyin MR

görüntülerinden yeni MS lezyonlarının saptanması ve segmentasyonu için yeni ve tam otomatik derin öğrenme yöntemleri geliştirmektedir. Derin öğrenme teknikleri, girdi verilerinden özellik çıkarma sürecini basitleştirir. Bu nedenle, bu tezde, derin öğrenme yaklaşımları araştırılmış ve ardından her iki görev için MS lezyonlarının segmentasyonunu ve tespitini iyileştirmek için kullanılmıştır. Buna göre, 3B MRG dizileri kullanılarak otomatik MS lezyon segmentasyon performansını artırmak için dikkat geçidini (AG), verimli kanal dikkatini (ECA) ve Atrous uzamsal piramit havuzlamayı (ASPP) birleştiren yeni bir yoğun rezidüel U-Net modeli önerilmiştir. Benzer şekilde, lezyon aktivitesinin belirlenmesi için temel ve takip eden 3B FLAIR MR görüntülerini kullanarak, yeni MS lezyon segmentasyonunu daha iyi gerçekleştirmek için U-Net, dikkat kapısı ve artık öğrenmeyi birleştiren derin bir sinir ağına sahip yeni bir işlem akışı önerilmiştir.

Tez kapsamında, 3B görüntülerinin her bir düzleminden elde edilen 2B kesitler, modellerin eğitilmesinde girdi verisi olarak kullanılmıştır. Her bir düzlemde elde edilen 2B kesitlerin tahmin edilmesiyle, her bir düzlem için 3B segmentasyon çıktısı üretilmiştir. Her bir düzlem için elde edilen 3B çıktılar arasında çoğunluk oylaması yapılarak nihai 3B segmentasyon çıktısı üretilmiştir. 3B görüntüyü bütün olarak işlemek yüksek bilgisayar kaynakları gerektirmektedir. Derin öğrenme modellerinde az veri kullanmak aşırı öğrenmeye sebep olabilir. Medikal alanda da veri azlığı olduğundan 3B medikal görüntüyü bütün olarak kullanmak aşırı öğrenmeye yol açabilir. Diğer yandan, yama tabanlı elde edilen küçük boyutlu 3B girdiler veri sayısını arttırsa bile görüntünün genel yapısal bilgisini ihmal etmektedir. Bu çalışmada hem aşırı öğrenmeden kaçınmak, hem de genel yapısal bilgi kaybını azaltmak için koronel, aksiyel ve sagittal düzlemlerden elde edilen 2B kesitler bütün olarak birlikte kullanılmıştır. Ayrıca bu yöntem daha az bilgisayar kaynağı gerektirdiği için pratikte oldukça etkili bir yöntemdir. Bu tezde kullanılan yöntemin, literatürdeki 3B yama tabanlı ya da 3B tüm görüntünün işlenmesiyle elde edilen sonuçlara yakın hatta daha iyi sonuçlar aldığı gözlemlenmiştir.

Önerilen yeni yoğun rezidüel U-Net modelinde, U-Net mimarisinin her bloğundaki evrişim katmanları, rezidüel bloklarla değiştirilmiş ve yoğun bir şekilde bağlanmıştır. Daha sonra, atlama bağlantılarından geçen önemli özellikleri yakalamak için AG'lerden yararlanılmıştır. ECA modülü, U-Net'in her rezidüel bloğunun ve her aşağı örnekleme bloğunun sonuna eklenmiştir. Daha sonra, çok ölçekli bağlamsal bilgileri çıkarmak için U-Net'in darboğazı ASSP modülüyle değiştirilmiştir. Ayrıca, FLAIR, T1-w ve T2-w 3B MR görüntüleri, daha iyi MS lezyon segmentasyonu gerçekleştirmek için birlikte kullanılmıştır. 3B MR görüntülerinden elde edilen çok düzlemli (koral, aksiyel ve sagittal) 2B kesitler, tüm düzlem yönlerinde bağlamsal bilgileri elde etmek için birlikte kullanılmış ve daha sonra tahmin edilen 2B kesitler, çoğunluk oylamasıyla nihai bir 3B segmentasyon çıktısı üretmek için toplanmıştır. Önerilen model, halka açık ISBI2015 ve MSSEG2016 veri setlerinde doğrulanmıştır. ISBI2015 veri seti 19 hastaya ait farklı zamanlarda elde edilmiş 3B FLAIR, T1-w ve T2-w MR görüntülerinden oluşmaktadır. Eğitim veri seti için 5 hastaya ait toplam 21 tane 3B MR görüntüleri bulunmaktadır. Bu görüntülerin iki uzman tarafından manuel olarak işaretlenmiş etiketli verisi bulunmaktadır. Geri kalan 14 hastaya ait 61 tane 3B MR görüntüsü ise önerilen model ve yöntemleri değerlendirmek için kullanılmaktadır. ISBI2015 test veri setinin değerlendirilmesi için ISBI organizasyonu tarafından sağlanan web sitesinin kullanılması gerekmektedir. Test seti verilerine ait etiketlenmiş veri araştırmacıların kullanımına sunulmamaktadır. MSSEG2016 veri

seti farklı merkezlerden ve farklı MR cihazlarından elde edilen 53 hastaya ait 3B MR görüntülerinden oluşmaktadır. Her hasta için tek bir zamanda elde edilen 3B FLAIR, T1-w ve T2-w MR görüntüleri etiketli verileri ile araştırmacıların kullanımına sunulmuştur. 38 hastaya ait veriler önerilen otomatik yöntemlerin performanslarını doğrulamak ve test etmek için kullanılmaktadır. Bu model ISBI2015 test veri setinde 92,75 ISBI skoru, %66,88'lik ortalama Dice skoru, %86,50'lik ortalama Pozitif Tahmin Değeri (PPV) ve %60,64'lük ortalama Lezyon Bazında Gerçek Pozitif Oran (LTPR) üretmiştir. MSSEG2016 test veri setinde ise %67,27'lik ortalama Dice skoru, %65,19'lük ortalama PPV ve %74,40'luk ortalama hassasiyet elde edilmiştir. Önerilen modelden elde edilen sonuçlar, bazı uzmanlardan elde edilen sonuçlara göre ve diğer bazı son teknoloji yöntemlerden elde edilen sonuçlara göre daha iyidir. Spesifik olarak, MS lezyonlarını segmentlere ayırmak için önerilen bu model ile, ISBI2015 test veri setinde en iyi Dice skoru ve en iyi LTPR elde edilmiştir.

Lezyon aktivitesinin belirlenmesi için önerilen iş akışı içinde yeni geliştirilen model, U-Net'e benzer bir mimariye sahiptir ve derin ağların eğitimini kolaylaştıran rezidüel birimlerden oluşur. Daha az parametrelili ağlar, U-Net ve rezidüel birimlerinin atlama bağlantıları sayesinde daha iyi performansla tasarlanır ve bu da bozulma olmadan bilgi yayılımını kolaylaştırır. Ayrıca, AG'ler çeşitli boyut ve şekillerdeki hedef yapıların önemli özelliklerine odaklanmayı öğrenir ve modelin daha keskin segmentasyon yapmasına yardımcı olur. MSSEG-2 veri seti, önerilen iş akışını eğitmek ve test etmek için kullanılmıştır. MSSEG-2 veri seti toplam 100 hastaya ait, her bir hasta için iki farklı zaman diliminden elde edilmiş 3B FLAIR MR görüntülerinden oluşmaktadır. Bu hastalardan 40'ına ait etiketli veri eğitim seti olarak paylaşılmaktadır. Geriye kalan 60 hastaya ait görüntüler ise önerilen algoritmaların segmentasyon performanslarını ölçmek için kullanılmaktadır. Bu test setinde, 32 hastanın 2. zaman dilimindeki görüntülerinde yeni lezyon varken 28 hastaya ait görüntülerde ise yeni lezyon bulunmamaktadır. Böylece, lezyon olmayan hasta verileri için de önerilen algoritmalar değerlendirilmektedir. Eğitim veri setindeki 3B MR görüntüleri ham olarak verildiği için, bu veri setine bazı ön veri işlemleri gerçekleştirilmiştir. Örneğin, MS lezyonlarının görülmediği beyin tası gibi bölgeler görüntülerden çıkartılmıştır. Elde edilen sonuçlar yarışma kapsamında önerilen diğer işlem akışı sonuçları ve bu yarışmadaki uzmanların etiketli verisinden elde edilen sonuçlar ile karşılaştırılmıştır. Test veri setinden elde edilen sonuçlara göre, lezyon bazında F1 ve Dice skorları sırasıyla ortalama %48 ve %44,30 olarak elde edilmiştir. İkinci zaman noktasından elde edilen görüntülerde yeni lezyon görülmeyen vakalar için, test edilen lezyonların sayısı ve test edilen lezyonların hacmi sırasıyla ortalama 0,148 ve 1,488 olarak bulunmuştur. Yeni geliştirilen model ile önerilen iş akışı diğer önerilen 22 iş akışından daha iyi performans göstermiştir. Önerilen iş akışı elde edilen Dice ve F1 skorlarına göre yarışmada 8. sırada, lezyon olmayan görüntüler için, test edilen lezyonların sayısına ve test edilen lezyonların hacmine göre sırasıyla 4. ve 5. sırada yer almıştır.



1. INTRODUCTION

Multiple Sclerosis (MS) is a chronic inflammatory, immune-mediated, neurodegenerative, and demyelinating disease that impacts the Central Nervous System (CNS) [1–3]. In MS, the myelin sheath that protects and surrounds nerve fibers is attacked by the immune system, which causes the nerves to either be damaged or deteriorated. This damage causes the formation of lesions (also called plaques) in the brain and spinal cord. The resulting plaques cause communication issues between the brain and the rest of the body by blocking the transmission of messages sent from the brain due to the physical destruction of the myelin layers in the brain and spinal cord [2]. MS is a disease without definitive treatment, and its symptoms can vary from patient to patient. For some patients, the disease may significantly impact their comfort in life and the patient may need care. The incidence of MS is increasing in the age range of 20-45 years, reaching its highest point between the ages of 25 and 35, decreasing gradually, and becoming rare in those over 50 years of age [2, 4]. The lack of chronic and definitive treatment for the disease means that the patient has been fighting it for many years. Therefore, effective methods are needed for diagnosing and monitoring the MS disease process. Early diagnosis of the disease is necessary because diagnosis in its advanced stages can become more difficult. Magnetic Resonance Imaging (MRI) techniques are widely used to detect MS lesions, the quantitative evaluation of inflammatory activity, and lesion load [4, 5]. In clinical practice, MRI data can be used to diagnose and assess MS lesions, which helps physicians better understand the natural history of MS [4, 6]. MRI sequences (modalities) such as T1-weighted (T1-w), T2-weighted (T2-w), Proton Density-weighted (PD-w), and Fluid-Attenuated Inversion Recovery (FLAIR) are used in the diagnosis of MS because they have different signal properties and can provide different information [7]. In particular, these techniques facilitate lesion segmentation to acquire quantitative features such as the number and volume of lesions [8].

Change detection of MS lesions from MRI data is another essential task to track lesion activity, which is defined as the emergence of new lesions or the expansion of existing lesions [9]. The most important biomarker for monitoring inflammatory changes and disease progression in MS is to track lesion activity between two longitudinal MR images of a patient [6, 10]. Recently, the delineation of new MS lesions on T2/FLAIR by comparing two time-points MRI data has gained attraction. Determination of new lesions has become even more important than identifying the total number and volume of lesions as it allows clinicians to determine whether a given anti-inflammatory Disease Modifying Drug (DMD) is effective for the patient [11]. However, the detection and delineation of new lesions appearing at the second time point are particularly challenging and intra- and inter-rater variability are unavoidable due to small and subtle new lesions [12]. Therefore, automating the detection of these new lesions will also be a significant improvement in assessing the disease activity of a patient.

Manual MRI segmentation is challenging since delineating 3D MRI is tedious, time-consuming, and prone to both inter-observer variability and intra-observer variability [7, 13]. Therefore, accurate automated segmentation methods are required to perform MS lesion detection [14, 15]. In fact, an automatic computer-aided diagnosis system using MRI data could be developed to support physicians in diagnosing, tracking, and treating MS diseases effectively by reducing assessment time and both intra- and inter-rater variability. To design such a system, automatic segmentation approaches have been widely studied in the last two decades for the MS lesion segmentation problem [4], such as [16–19]. Therefore, accurate MS lesion segmentation is an important task to gather useful information about the disease's diagnosis and progress [14]. However, automated lesion segmentation in MS is a challenging task because the size and location of the lesions vary, the lesion boundaries are not very well characterized, and MR images might have low resolution or artifacts.

To yield more effective results, machine learning methods were previously proposed to make faster computations and reduce expert bias [7, 20]. Recently, Deep Learning (DL) methods, especially those relying on Convolutional Neural Networks (CNNs) [21], have shown state-of-the-art results in medical image analysis compared to previous conventional methods. These methods have produced satisfactory results in different

problems of medical imaging, such as brain extraction [22], segmentation of neuronal structures [23], brain tumors [24], MS lesion segmentation [17, 18, 25], and MS lesion activity, especially new lesions appearing on the follow-up images [26–28].

CNN-Based medical image segmentation can be accomplished through either patch-based or image-based approaches [17]. Patch-based segmentation takes a pixel as a center and generates small patches as a classifier input. Then, a CNN is trained with these patches to detect whether each central pixel/voxel is classified as a healthy region. The advantage of this technique is that it exploits better contextual information around pixels and is also preferred in medical image analysis to obtain more training samples, reducing class imbalance problems [29]. However, a longer training time and omitting global structure information are disadvantages of this approach due to repeated computations and small patch sizes. On the other hand, image-based methods exploit the global structure when processing the entire image, and they have higher computational efficiency due to one forward propagation to classify all pixels from the input image [29–31]. Image-based segmentation can be performed with either slice-based or 3D-based segmentation methods. Slice-based segmentation is defined as converting each 3D MRI into 2D slices along the x, y, and z axes, then processing each slice individually as an input to a CNN. Afterward, the segmented slices are fused again to form a 3D binary segmentation. This approach can be further expanded to a whole-brain slice-based segmentation by combining all of the 2D slices extracted from the three planes, resulting in the acquisition of contextual information in all directions. In 3D-based segmentation, extracting meaningful information from original 3D images is obtained via a CNN with 3D kernels. A standard 3D segmentation is prone to a high risk of overfitting due to having many parameters when training small datasets, which is common in the medical field due to limitations in data collection and manual delineation [30]. To handle this issue, Liu et al. [32] and Tetteh et al. [33] proposed 3D cross-hair convolution. They defined three 2D filters for each of the three orthogonal plane orientations around the voxel, and then the sum of the results obtained from each convolution is given to the central voxel. This reduces the number of training parameters, resulting in a shorter training time when compared to a standard 3D segmentation. Nevertheless, this method has three times more parameters for each

layer in the network when compared to slice-based methods, making it more prone to overfitting in the presence of smaller datasets [17].

In this thesis, it is argued that a computer-aided diagnosis system would be a significant improvement in the diagnosis, follow-up, treatment, and monitoring of therapies for MS disease. Automatic lesion segmentation approaches have been widely investigated in the last two decades for the MS lesion segmentation problem, making it a crucial aspect in gathering relevant information for MS diagnosis and progression. Hereby, different CNN-based approaches were investigated and examined to find better solutions for accurate MS lesion segmentation and detection. To yield more effective results, fully automated deep learning-based models were proposed for segmenting and detecting MS lesions from a single time point brain MRI of a patient and also newly formed MS lesions in the follow-up images of two different time points. Moreover, 2D slices obtained from three plane directions of 3D MR images were utilized to alleviate the computational burden. Afterward, the processed 2D slices were exploited to produce the final 3D binary segmentation output using a majority voting method. Figure 1.1 shows the workflow of the study.

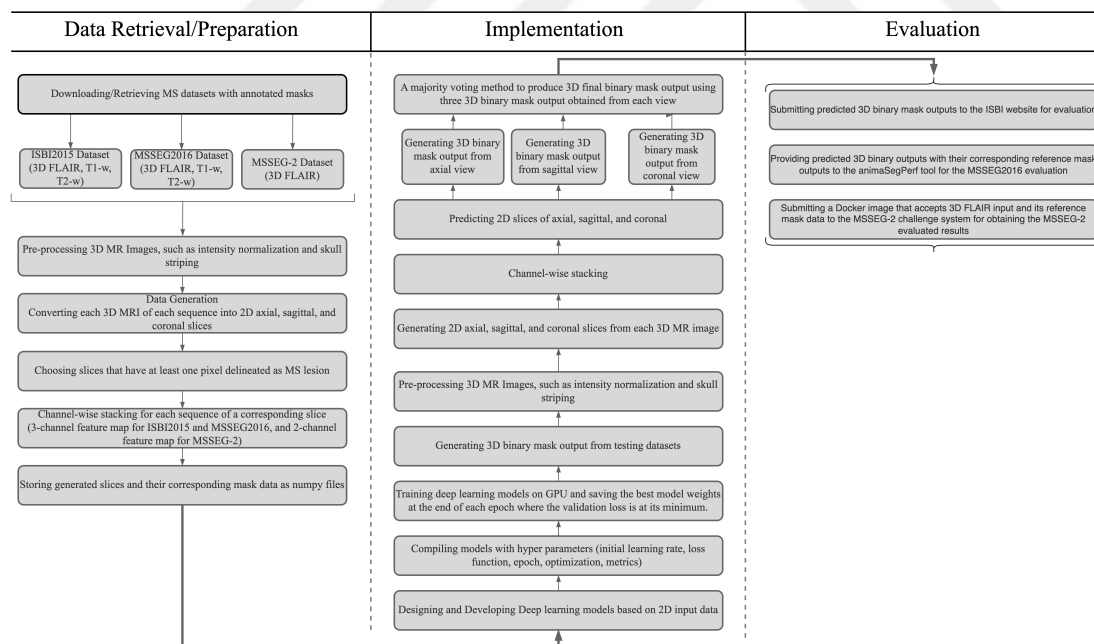


Figure 1.1 : The workflow of the thesis.

1.1 Purpose of Thesis

In this study, the detection and segmentation of MS lesions from 3D MRI data using deep learning and image processing techniques are investigated and assessed to achieve more accurate and robust segmentation, especially with CNN-based deep learning approaches. The aim of the study is to reduce the workload of manual MRI segmentation performed by experts since manual delineation of 3D MRI data by experts is costly, time-consuming, and prone to both inter-observer variability and intra-observer variability [7, 34, 35]. Furthermore, developing novel deep learning methods for the detection and segmentation of MS lesions would allow experts to better understand the natural history of the MS disease. In this study, different 3D MRI sequences are exploited to extract more features related to MS lesions. Instead of using whole 3D MRI or patches extracted from 3D MRI, a whole-brain 2D slice-based segmentation is preferred to avoid both the overfitting that occurs in 3D-based methods and the loss of global structural information that occurs in patch-based methods [8,30,36]. Moreover, it is aimed to compare the performances of existing deep learning architectures for the detection and segmentation of MS lesions, to present an improved deep learning model on a single time-point brain MRI of a patient based on the results obtained, and to present an improved deep learning model to monitor newly formed lesions in the follow-up MR images.

This study resulted in the publication of two SCI-Expanded papers in this field [25,26]. The first paper proposed a unique pipeline with a deep neural network to perform better new MS lesion segmentation using baseline and follow-up 3D FLAIR MR images. The second paper proposed a novel dense residual U-Net model to enhance the automatic MS lesion segmentation performance by exploiting different 3D MRI sequences on a single time point brain MRI of a patient. Eventually, this study has been prepared to include these two published studies.

Additionally, the purpose of this thesis is to seek answers to the following questions.

- What is the performance of DL algorithms in segmenting MS lesions?
- How can DL algorithms be optimized for better MS lesion segmentation?

- What is the impact of imaging sequences on the accuracy of DL algorithms in MS lesion segmentation?
- What is the effect of data augmentation techniques on the performance of DL algorithms in MS lesion segmentation?
- How can transfer learning be applied to enhance DL algorithms in MS lesion segmentation?
- How do process 3D MRI scans with computational efficiency approaches?
- What is the performance of DL algorithms in MS lesion activity segmentation?
- How to handle class imbalance problem for MS lesion segmentation?

1.2 Thesis Organization

This thesis is structured as follows:

- **Chapter 1:** This chapter provides an overview of the definition, diagnosis, and treatment of MS, highlighting the crucial role of MR imaging and the challenges faced by physicians during image analysis. Later, the relevance of deep learning techniques is highlighted and an overview of CNN-based medical image segmentation approaches is presented. Finally, it introduces the purpose of the thesis, presents the intellectual foundations of the thesis study, and outlines the main hypothesis, objectives, research questions, and structure of the thesis.
- **Chapter 2:** This chapter provides a comprehensive overview of MS, including an in-depth explanation of MRI, its four sequences, and the analysis of MR images for MS lesion segmentation. It delves into the preprocessing steps involved in MR image analysis, covers MS lesion segmentation, and examines the use of MRI in this process. Additionally, DL approaches are surveyed, especially those relying on CNNs, for the MS lesion activity segmentation between two time points brain MRI of a patient and for MS lesion segmentation on a single time point brain MRI of a patient.
- **Chapter 3:** This chapter describes the data used in the thesis, specifically, three well-known challenge datasets in MS lesion segmentation, namely, ISBI2015,

MSSEG2016, and MSSEG-2. It also outlines the recent DL approaches applied in the proposed models for both MS lesion segmentation on individual MR images and new MS lesion segmentation on longitudinal MR images. Additionally, the common loss functions used in MS lesion segmentation are explained in detail. The evaluation metrics used to assess the performance of the proposed models for MS lesion segmentation at a single time point MRI of a patient and MS lesion activity segmentation between two different time points MRI of a patient, along with their mathematical formulas, are presented. Finally, a dense residual U-Net for MS lesion segmentation on individual MRI scans, deep residual attention gate U-Net for MS lesion activity between two different MRI scans of a patient, and their implementations are explained in detail.

- **Chapter 4:** This chapter presents the results obtained from the proposed models using the three datasets. It compares the results of the proposed models with other state-of-the-art methods and with the results of experts. Additionally, cross-dataset validation results between ISBI2015 and MSSEG2016 are presented. The results of the two proposed models are also thoroughly discussed based on the three datasets. Finally, the results of the research questions have been evaluated.
- **Chapter 5:** This chapter presents the conclusion of the studies presented with future studies.



2. LITERATURE REVIEW

2.1 Multiple Sclerosis (MS)

MS is a chronic autoimmune disease that affects the CNS, which consists of the brain, spinal cord, and optic nerves. It is characterized by the immune system attacking and damaging the myelin sheaths that protect the nerve fibers in the CNS [1, 14, 37]. In MS, the body's immune system targets and damages the myelin sheath, resulting in difficulties in communication between the brain and the rest of the body. Eventually, the disease can lead to long-term (chronic) damage or deterioration of the axons, known as nerve fibers. This damage (demyelination) results in various neurological symptoms such as muscle weakness, numbness, coordination and balance issues, vision problems, and cognitive impairment. MS has been considered a demyelinating disease as the damage to the myelin surrounding nerve fibers in the CNS causes disruptions in the normal communication between the nerve cells. This can lead to various symptoms, depending on the location and severity of the damage [14, 38, 39].

To make a definitive MS diagnosis, there is a set of guidelines known as the McDonald criteria, which were published by an international panel in 2001 and updated several times since then, with the most recent revision occurring in 2017. These criteria incorporate clinical, laboratory, and imaging findings with the goal of enhancing the accuracy and consistency of MS diagnosis. The McDonald criteria are commonly used in clinical settings and have been demonstrated to improve the accuracy of MS diagnosis, especially in the early stages of the disease [40–43].

The cause of MS is not fully understood; however, it is thought to be a combination of genetic and environmental factors [2, 14]. Although no cure is currently available for MS, various treatments are available to manage the symptoms and slow the progression of the disease. To summarize, MS is a complex and variable disease that affects millions of people worldwide and can have a significant impact on their quality of life. Research into the underlying causes and potential treatments for MS is ongoing,

with the goal of improving the diagnosis and management of the disease [2]. Figure 2.1 shows the nerve affected by MS.

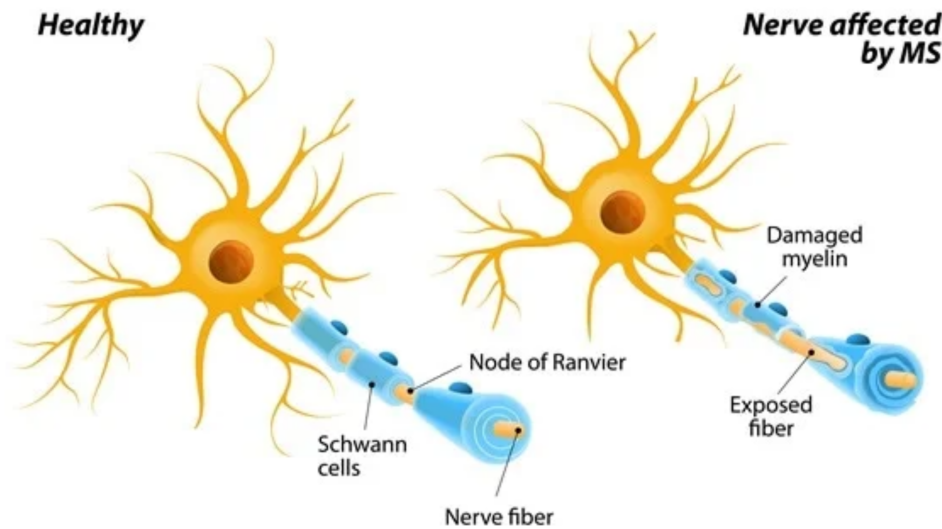


Figure 2.1 : The nerve affected by MS. Image extracted from <https://www.news-medical.net/health/Types-of-Multiple-Sclerosis-%28MS%29.aspx>

2.2 Magnetic Resonance Imaging (MRI)

MRI is a medical imaging technique that employs radio waves and strong magnetic fields to generate three-dimensional, detailed anatomical images of the human body [44, 45]. MRI scans are a very safe and efficient method to diagnose a variety of medical conditions since they are painless, noninvasive, and do not involve ionizing radiation. The images generated by MRI can be utilized to diagnose a variety of conditions, including problems with the spinal cord, the brain, joints, the heart, bones, breasts, and blood vessels. In other words, the results of an MRI scan can be exploited to aid in planning treatments, diagnosing conditions, and evaluating how effective previous treatments have been [42, 46]. There are different types of MRI scanners that can be identified by their magnetic field strength, which is measured in teslas (T). The strength of the magnet can vary from 0.5T to 7.0T. A 1.5T MRI scanner delivers high-quality images, quick scanning times, and the capability to assess the functioning of specific body structures. It is also currently the most frequently used MRI for diagnosing MS. The 3.0T MRI scanner is ideal for observing very small details, including blood vessels in the brain and heart [47]. A 3D MRI volume is

generated by combining multiple 2D slices obtained in three distinct orientations: axial, coronal, and sagittal. Figure 2.2 shows these plane orientations of the brain MRI.

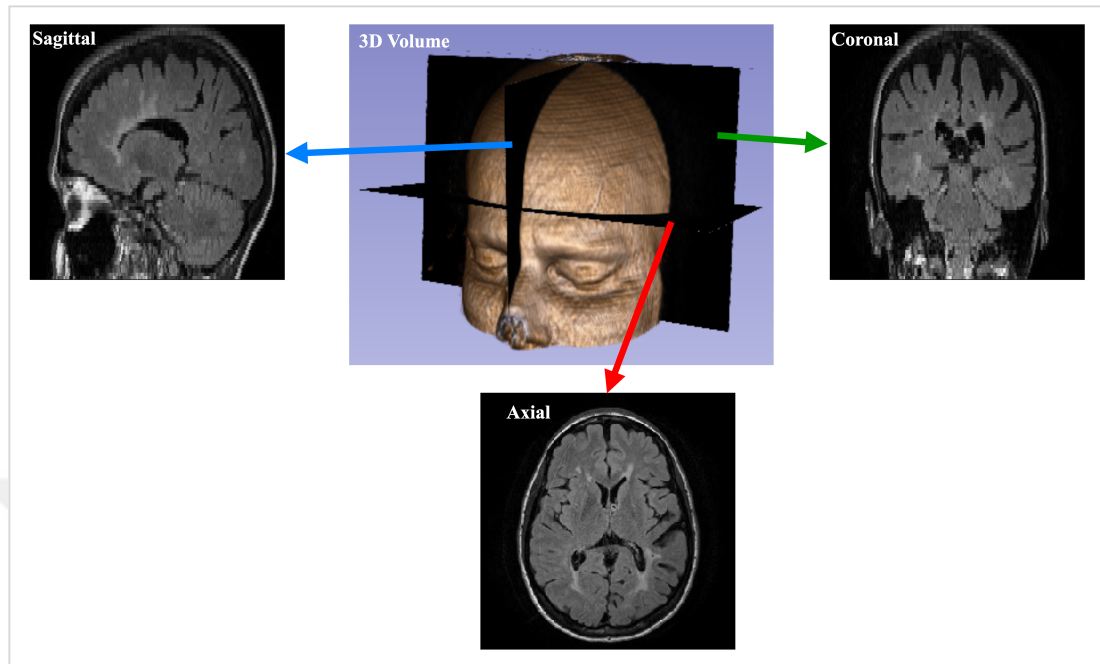


Figure 2.2 : A 3D brain MRI visualization with its three plane orientations. The red arrow shows an axial slice, the green arrow shows a coronal slice, and the blue arrow shows a sagittal slice.

There are several MRI sequences that can be used to identify and characterize lesions associated with MS [7, 48]. Each of them utilizes a different set of settings and techniques to produce images that highlight different aspects of the body's anatomy. Additionally, these MRI sequences can be used individually or in combination to enhance the sensitivity and specificity of MS lesion detection and characterization [4, 48, 49]. These sequences are shown in Figure 2.3 and described in detail as follows:

- **T1-weighted (T1-w):** This technique employs a short Repetition Time (TR) and short Echo Time (TE) to produce images with good contrast between different tissue types. T1-w images are often utilized to identify areas of inflammation and demyelination in the brain and spinal cord.
- **T2-weighted (T2-w):** This technique employs a longer TR and TE to produce images with good contrast between different fluid-filled spaces. T2-w images are often utilized to identify areas of edema or swelling in the brain and spinal cord.

- **Fluid-Attenuated Inversion Recovery (FLAIR):** This technique employs an inversion recovery pulse and a long TR to suppress cerebrospinal fluid signal, which makes it simpler to visualize abnormalities in the brain and spinal cord. FLAIR images are often utilized to identify areas of inflammation and demyelination in the brain's white matter.
- **Proton Density-weighted (PD-w):** This technique employs the tissue's proton density to produce images, giving information about the relative number of protons in a given area of tissue. This information can be beneficial for identifying certain types, such as fluid, muscle, and fat.

Accordingly, each sequence exploited to visualize MS lesions in the brain has its own unique advantages and disadvantages. T1-w imaging, for example, is known for its elevated anatomical resolution, which provides clear images of the brain's structural details, while T2-w imaging is particularly useful for the detection of MS lesions due to its elevated sensitivity to the enhanced water content that is frequently found in the affected tissue. Furthermore, PD-w imaging can reveal both anatomy and disease [47, 50].

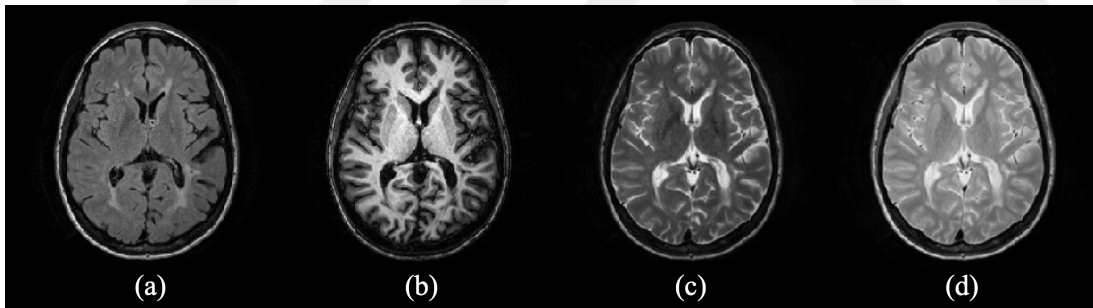


Figure 2.3 : An axial view of different MR images of the brain from the ISBI2015 dataset: a) FLAIR b) T1-w, c) T2-w, and d) PD-w.

2.3 MRI Analysis for MS

MRI is exploited to identify and characterize MS lesions in the CNS. These lesions are often associated with a range of neurological symptoms, and their presence and distribution can be employed to diagnose and monitor the disease. MRI can be exploited to acquire different types of images of the brain and spinal cord, including

T1-w, T2-w, FLAIR, and PD-w images. These images are sensitive to different tissue properties and can be employed to distinguish different types of tissue, such as brain tissue, blood vessels, and MS lesions. Manual analysis of MR images is a challenging and time-consuming task due to the vast number of MRI slices that make up the 3D information for each patient. Furthermore, it is subject to inter-observer variability and intra-observer variability. Since the early 1990s, a variety of methods have been developed for preprocessing MRI and segmentation of lesions to the requirement for manual input and the variations that come with manual delineations [13,51,52].

2.3.1 Preprocessing of MR images

The automatic analysis of MR images is challenging due to a variety of factors, including variations in imaging parameters, blurred edges, normal anatomical variations, noise, overlapping intensities, susceptibility artifacts, and motion artifacts [53]. These images, taken directly from the scanner, often include the entire head and sometimes the neck, and may have issues with intensity non-uniformity. Therefore, preprocessing techniques are essential to address these issues before performing any automated methods for further processing or analysis. These include aligning different MR images to a standard coordinate system for comparison with others or for use in image registration, and performing various preprocessing techniques to enhance the quality of the images for automated analysis [20]. Some common preprocessing steps for brain MR images are described as follows:

- **Brain extraction:** This technique is exploited to remove non-brain tissues, such as the skull, from MRI scans. The segmentation performance of both tissues and lesions is negatively affected by non-brain tissues which have an impact on the distribution of intensity values in the image [54]. Brain Surface Extractor (BSE) [55], Brain Extraction Tool (BET) [56], and most recently BEaST [57] and ROBEX [58] are the most commonly used methods for skull-stripping.
- **Bias field correction:** It is exploited to remove intensity inhomogeneities or "bias fields" from medical images, including MR scans. Various factors, such as irregularities in the scanner or the anatomy of the patient, can lead to these bias fields. The aim of bias field correction is to enhance the quality of the images by eliminating the intensity inhomogeneities, producing images with a more uniform

intensity and accurate representation of the underlying tissue. This can improve the accuracy of image segmentation, registration, and analysis, as well as the reliability of the results. The N3 (nonparametric nonuniformity normalization) correction technique for intensity inhomogeneity was introduced by Sled et al. [59]. It uses an iterative approach to discover the smooth multiplicative field that enhances the high-frequency content of the tissue intensity distribution. Later, Tustison et al. [60] attempt to enhance this method by replacing its B-spline smoothing technique with a more advantageous solution that tackles significant problems highlighted in earlier evaluations of N3, known as N4. The iterative optimization scheme was also changed, leading to improved convergence. These approaches for bias field correction are currently the most widely exploited.

- **Histogram matching:** Due to variations in the scanner and other factors, MR images of a patient taken at different times using the same scanner may appear dissimilar from one another, which makes the absolute intensity values difficult to interpret. Thus, histogram matching aims to adjust the brightness and contrast of an MR image to match the appearance of a reference image by analyzing the pixel intensity values to overcome this issue [61]. One of the most commonly employed tools for intensity normalization in MRI is the method developed by Nyúl et al. [61]. This method works by transforming the image histograms so that they match an average histogram that has been established through training. Apart from histogram matching for intensity normalization, other approaches that normalize the data to a specific range are also utilized, including z-score normalization, min-max normalization, and percentile normalization [49].
- **Image registration:** Once the brain extraction, the bias field correction, and histogram matching are completed, the registration process, which is the process of aligning two images that are in different spaces, can be performed. It plays a vital role in many automated brain MRI applications and is an essential step in both inter-subject and intra-subject analyses. The process of aligning various sequences of images from the same subject, which is also called co-registration, is achieved using intra-subject registration, while inter-subject registration is utilized to align images from different subjects, often from the same sequence, or to register a subject image to a standard template, which is also referred to as

atlas registration [62, 63]. The registration methods employed share a similar approach, which is to optimize the deformation of one image to best match the other image using an energy function and a transformation model. In particular, the registration techniques can be categorized into two different steps: affine and rigid registration and nonrigid registration according to several reviews and surveys [64–66]. Consequently, image registration is an important tool in the field of medical imaging, as it enables the comparison and combination of information from multiple images, thereby improving the diagnostic accuracy and precision of the imaging process.

Ultimately, preprocessing steps are chosen based on the type of image and the task. The choice of preprocessing steps, their order, and the parameters used can notably affect the final outcome of the segmentation.

2.3.2 MS lesion segmentation

MS lesion segmentation is the process of identifying and delimiting the boundaries of lesions that are characteristic of MS within an MR image. These lesions can be difficult to detect due to their small size and the presence of other brain structures, such as blood vessels and tissue. Several methods have been proposed for MS lesion segmentation, including manual, semi-automatic, and fully automatic methods [4, 15]. Manual segmentation, which is performed by a radiologist using a software tool to manually delineate the lesions on an MRI scan, is considered the gold standard for accurately identifying and quantifying MS lesions [67]. However, this approach is time-consuming and subject to observer variability [13]. Therefore, semi-automated and fully automatic methods have been developed to enhance the accuracy and efficiency of the segmentation process.

The McDonald criteria, discussed in section 2.1, aims to use MRI scans as evidence for the presence of lesions in various locations and over a period of time, which are necessary for a definite diagnosis of MS. Therefore, automatic methods would be beneficial in MS lesion segmentation for both diagnosing the condition and monitoring a patient’s progress and response to treatment. DL algorithms, such as CNNs, have been applied to this task and have shown promising results in terms of accuracy and efficiency. In particular, MRI and MS lesion segmentation are important tools in the

diagnosis and management of MS, and have greatly improved the understanding of the disease and its effects on the CNS.

2.3.3 MRI in MS lesion segmentation

In the context of MS, MRI is often utilized to detect and characterize lesions, or areas of abnormal tissue, within the brain and spinal cord [48]. As discussed previously, MS is a chronic autoimmune disorder that affects the CNS and can lead to a wide range of neurological symptoms, including muscle weakness, difficulty with balance and coordination, and problems with vision. Several approaches have been developed for automated or semi-automated lesion segmentation in MRI images of patients with MS. These approaches typically involve the use of algorithms that analyze the intensity, texture, and shape of the lesions in the image. Some methods also incorporate prior knowledge about the typical location and appearance of MS lesions in order to improve the accuracy of the segmentation. To sum up, MRI plays a key role in the diagnosis and management of MS, and lesion segmentation is an important aspect of this process. In what follows the next two subsections, a review of MS lesion segmentation between the two time points, also known as lesion activity, and MS lesion segmentation in a single time point is presented.

2.3.4 MS lesion activity segmentation

The emergence of new lesions or the expansion of existing lesions is referred to as lesion activity [9]. The most important bio-marker for monitoring inflammatory changes and disease progression in MS is to track lesion activity between two longitudinal MR images [6, 10]. Recently, the delineation of new MS lesions on T2/FLAIR by comparing two time-points MRI data has gained attraction. The determination of new lesions has become even more important than identifying the total number and volume of lesions as it allows clinicians to determine whether a given anti-inflammatory disease modifying drug (DMD) is effective for the patient [11]. However, detection and delineation of new lesions appearing at the second-time point are particularly challenging and intra- and inter-rater variability are unavoidable due to small and subtle new lesions [12]. Therefore, automating the detection of these new lesions will be a significant improvement in assessing the disease activity of a patient.

Recently, DL methods, especially relying on CNNs [68], have improved the performance of brain lesion segmentation tasks [69]; such as brain tumor segmentation [24], brain extraction [22], and MS lesion segmentation [8, 17, 18]. Most of these methods rely on encoder-decoder networks, which take MRI data as an input and generate a segmentation output for each pixel [48]. Many CNN-based methods and their variations have also been proposed with different input strategies, such as multi-scale [30], multi-branch [17], and cascaded [36] approaches. However, these together with most of the classical methods perform lesion segmentation on a single MRI data. For determining MS lesion activity, classical image processing approaches have usually been preferred such as image differences, intensity-based approaches, and deformation fields [70–73]. However, some of these approaches have high variability and inconsistency as they use two different segmentation outputs obtained from the baseline and follow-up images to produce the lesion activity [74]. To perform better lesion activity segmentation, DL approaches relying on CNNs are essential which take these two images as input; however, these methods have been so far limited for MS lesion activity segmentation. Salem et al. [75] who used a classical approach in their previous study proposed the first CNN-based longitudinal approach for detecting new T2-w lesions in brain MRI. In their study, intensity- and deformation- based features from two time-points data were incorporated into the proposed network and trained within an end-to-end procedure. The performance of the method was evaluated on the MRI data of 60 MS patients obtained from the Vall d’Hebron Hospital’s center. While 36 patients of the follow-up scans have new T2-w lesions, 24 patients do not have new lesions on their follow-up scans. Gessert et al. [76] have proposed a CNN-based method using two FLAIR images acquired at two different times to detect lesion activity. They used two-path architectures with attention-guided interactions to process two time-points MRI data. Furthermore, they extended their work to full 4D deep learning using a history of MRI volumes and proposed a 3D ResNet-based multi-encoder-decoder network in which temporal aggregation was performed by convolutional gated recurrent units (convGRUs) for lesion activity segmentation [77]. The performance of the proposed method was evaluated on a private dataset obtained from the University Hospital of Zurich, Switzerland. A 3.0T Philips Ingenia Scanner was exploited for image acquisition using similar acquisition parameters for all scans.

A baseline, follow-up, and history scans of FLAIR modality were used for MS lesion assessment in their studies. However, the dataset of these studies consists of MR images from the same scanner, which decreases the generalizability of these methods towards the intensity and texture characteristics variations, which can be inherited if the data is obtained from different scanners. Thus, there is a need for new deep learning approaches to cope with variation problems that may arise through the use of data from multiple scanners as well.

2.3.5 CNN-based MS lesion segmentation

In the literature, many CNN-based methods are proposed for the automatic segmentation of MS lesions with various input data strategies and networks. Vaidya et al. [78] proposed a 3D patch-based CNN consisting of only four layers with sub-sampling methods and sparse convolutions to improve the segmentation performance. A post-preprocessing step was added by applying a white matter mask to the output prediction in order to enhance the performance of the proposed model. The ISBI2015 dataset [79], including FLAIR, T1-w, T2-w, and Pd-w sequences, has been employed to assess the performance of their method. 2D patch-based CNNs from multiple images, multiple views, and multiple time points have been proposed by Birembaum and Greenspan [16] to take advantage of longitudinal data for MS lesion segmentation. Preprocessing, candidate extraction, and CNN prediction are the three phases of their proposed segmentation approach. In the first phase, co-registration, brain extraction, bias field correction, and intensity normalization were performed on the MR images. In the second phase, masks based on FLAIR and white matter (WM) prior were computed and applied to the MR images. In the last phase, the multi-view CNN predicted the probability of a lesion for each voxel in the MR image. The proposed segmentation method was evaluated using four MRI sequences of the ISBI2015 dataset. Brosch et al. [30] proposed a whole-brain segmentation approach using a 3D CNN with a single shortcut connection between the first and last layers of the network. This shortcut connection allows the proposed network to integrate high- and low-level features to obtain information on the structure of MS lesions. However, their proposed network could not benefit from mid-level features, affecting the segmentation performance [23]. They evaluated their proposed method on the MICCAI2008 [80] and ISBI2015 datasets. Additionally, a clinical private dataset was

used to evaluate the proposed method with five methods widely used for MS lesion segmentation. A pipeline for white matter lesion segmentation using a cascade of two 3D patch-based CNNs has been proposed by Valverde et al. [36]. In their study, the first network was designed to detect potential lesion voxels with high sensitivity, while the second network focuses on decreasing the number of false positive results produced by the first network. In fact, the second network obtained the input features from the first network which was designed to select features. Their proposed network consisted of a 7-layer CNN model and used multi-sequence 3D patches from training images. They evaluated their proposed method on two private clinical MS datasets and the MICCAI2008 dataset, using three MRI sequences for each. Roy et al. [8] developed a 2D patch-based fully CNN model for MS lesion segmentation. Their network contains two pathways for FLAIR and T1-w modalities and then concatenates the outputs of each path to generate a member function for MS lesions. Their proposed method was evaluated on the ISBI2015 dataset and a private dataset. A 3D fully convolutional densely connected network (FC-DenseNet) using a 3D patch-based CNN has been proposed by Hashemi et al. [81]. The 3D patches with 50% overlap from different MRI sequences were employed as the training input for the proposed method. In addition, they used an asymmetric similarity loss layer based on the Tversky index to handle unbalanced data issues in medical imaging. Their networks were trained and evaluated on the ISBI2015 and MSSEG2016 datasets.

Although the patch-based methods improve lesion segmentation and perform well, global structural information is not used; namely, the global brain structure and lesion locations are not part of the segmentation. To address this issue, whole-brain slice-based methods for MS segmentation using CNNs based algorithms are also proposed in the literature. For example, Aslani et al. [17, 82] proposed a deep end-to-end 2D encoder-decoder CNN utilizing the slice-based segmentation approach for 3D MRI data of different modalities. In their first study, they exploited a stack of 3-channel slices extracted from each plane of each corresponding modality and a modified ResNet50 architecture. Their other study used single-channel slices obtained from each plane of each modality using multiple networks separately. Their proposed network includes a multi-branch downsampling path and multi-scale feature fusion blocks to merge features from multimodal MRI data. To overcome the overfitting issue

in 3D-based segmentation and the deficiency of global structure in the patch-based method, they preferred a whole-brain slice-based approach in their research. Their proposed models were evaluated on the ISBI2015 dataset and a private dataset. Zhang et al. [18] proposed a method using a fully convolutional densely connected network and stacked adjacent slices from the orthogonal planes (axial, sagittal, coronal) of different modalities along the x, y, and z axes as 2.5D input data. They took advantage of the global and local context from slices, and these slices increased training samples to make accurate segmentation as well. Their proposed model was evaluated on the ISBI2015 and the in-house datasets. Kang et al. [83] presented a 3D attention context U-Net (ACU-Net) which is a novel end-to-end segmentation framework to cope with the challenge of MS lesion segmentation. To expand the perception field and guide contextual information, a 3D context-guided module was used in the encoding and decoding stages of 3D U-Net. They used a 3D spatial attention block to enhance feature representation in the skip connection phase as well. The proposed model was evaluated on the ISBI2015 dataset with the provided four MRI sequences. During the training, the size of the input image cropped from these MRIs was $160 \times 192 \times 160$. To enhance the data, they performed the rotation along the axial view and flipped the original and the rotated volumes horizontally.

Recently, Zhang et al. [84] proposed a deep CNN model based on 3D U-Net to perform fast and accurate MS lesion segmentation that uses FLAIR, T1-w, and T2-w sequences. Anatomical information obtained using distance transformation mapping and lesion-wise loss function were integrated to obtain anatomical structure information and improve small lesion detection, respectively. The proposed method was assessed on a larger in-house Cornell MS and the ISBI2015 datasets. For the ISBI2015 dataset, each 3D volume of each modality was resized to the size of $217 \times 217 \times 217$ voxels for all subjects. Data augmentation was performed by applying elastic deformation, random changes in intensity, and random scaling. As for the larger Cornell MS dataset, the training samples were randomly extracted from the original images with the size of $128 \times 128 \times 48$ voxels. Kamraoui et al. [19] proposed a DeepLesionBrain (DLB) method which is a novel and robust method for domain shift using the 3D CNN model. A spatially distributed strategy of multiple compact 3D CNNs with large overlapping receptive fields was employed to

generate consensus-based segmentation robust to domain shift. To effectively combine both generic and specialized features, they trained the model using hierarchical specialization learning. To increase training data variability, they proposed a novel image quality data augmentation as well. The generalization ability of DLB was tested in cross-dataset experiments with two to four MRI modalities using the MSSEG2016 dataset, the ISBI2015 dataset, and an internal dataset.

Additionally, Weeda et al. [85] aimed to test the CNN-based nicMSLesions software and compare it with manual and other automatic segmentation to investigate its performance using an independent dataset. They focused on five segmentation methods, which are LessionTOADS, LST-LPA default, LST-LPA adjusted-threshold, BIANCA, and nicMSLesions single-subject, respectively. Their results show that the nicMSLesions method can be easily trained with only one manual delineation and achieves better results than others. Most of these studies have focused on accurate MS lesion segmentation using MRI data from the same domain. However, quantifying MS lesions through the use of MRI data from different centers and scanners has become essential for evaluating segmentation performance. Therefore, new deep learning-based segmentation approaches are still required for accurate MS lesion segmentation and generalization capabilities toward different centers and scanners.



3. DATA, METHODOLOGY, AND IMPLEMENTATION

3.1 Data Used

In medical imaging analysis, obtaining labeled datasets is challenging as manual delineation of MRI data by several experts is a costly and time-consuming process. To address this, challenges have been organized for MS lesion segmentation and detection to provide publicly available labeled datasets associated with the MS disease including various sequences such as FLAIR, T1-w, T2-w, and PD-w. Most of these datasets also provide both raw and preprocessed images for all sequences. Additionally, these challenges serve a fair comparison platform for automated methods. Therefore, in this study, three challenge datasets were exploited to evaluate the segmentation performance, generalization, robustness, and lesion activity. The evaluation of proposed deep learning methods was also performed using the provided platforms. To obtain these datasets, each challenge has its own website to download them by registering and approving their data usage policies. In this thesis, three publicly available datasets, the ISBI2015 MS lesion segmentation challenge dataset which will be denoted as ISBI2015, the MSSEG2016 challenge dataset which will be denoted as MSSEG2016, and the MSSEG-2 challenge dataset which will be denoted as MSSEG-2, were exploited to evaluate proposed models. Furthermore, some of these datasets enable assessing the generalization ability and robustness of the proposed models on unseen datasets since they present high variability in terms of acquisition sites, resolution, preprocessing, and clinical cases.

3.1.1 ISBI2015 longitudinal MS lesion segmentation dataset

The ISBI2015 dataset, which is publicly available and downloadable from the Challenge Evaluation website¹, was used to evaluate the MS lesion segmentation task in the proposed models. It has five training and fourteen testing subjects with 4 to 6

¹Challenge website: <https://smart-stats-tools.org/lesion-challenge-2015>

follow-up 3D scans per subject and a mean of 4.4 time-points. T1-w, T2-w, PD-w, and FLAIR sequences are provided for each time-point with data acquired on a 3.0T MRI scanner [79]. 181 (axial and sagittal) and 217 (coronal) slices make up the volumes with 1mm cubic voxel resolution. The training set comprises 21 3D scans that belong to five patients with white matter lesions associated with MS and has been annotated by the two expert raters, while 14 patients of the testing dataset comprising 61 3D scans do not have their delineated masks available for the public. Rater #1 has four years of experience delineating lesions, while rater #2 has ten years of experience in manual segmentation and 17 years in structural MRI analysis [79]. Moreover, in the training set, subject 2 is with high lesion load while subject 3 is with low lesion load. The performance of the proposed methods can be evaluated via the ISBI2015 challenge website² by submitting the 3D binary masks obtained from the ISBI2015 testing set. This dataset has also been widely used as a benchmark in many research studies for the automatic MS lesion segmentation [49].

The preprocessed and raw of images for all modalities are given by the challenge organizers. The provided preprocessed dataset, in which the brain is stripped from the skull using the BET [56], has been employed in this study. First, intensity normalization was performed on each 3D MRI using Kernel Density Estimation (KDE) with the Gaussian kernel. Then, slices along three orthogonal directions were extracted from all modalities for both raters. The size of each slice – axial, sagittal, and coronal – is 181×217 , 217×181 , and 181×181 , respectively. To acquire the same size for each plane view (224×224), a zero-padding technique was applied by centering the brain without considering its orientation. Furthermore, the slices that have MS lesions, including at least one pixel, were preferred to remove non-informative samples and excessively unbalanced data when feeding the models. Figure 3.1 shows the training sample of subject 2, which has a high lesion load, for FLAIR, T1-w, and T2-w modalities of each plane orientation.

²Evaluation: <https://smart-stats-tools.org/lesion-challenge-upload-results>

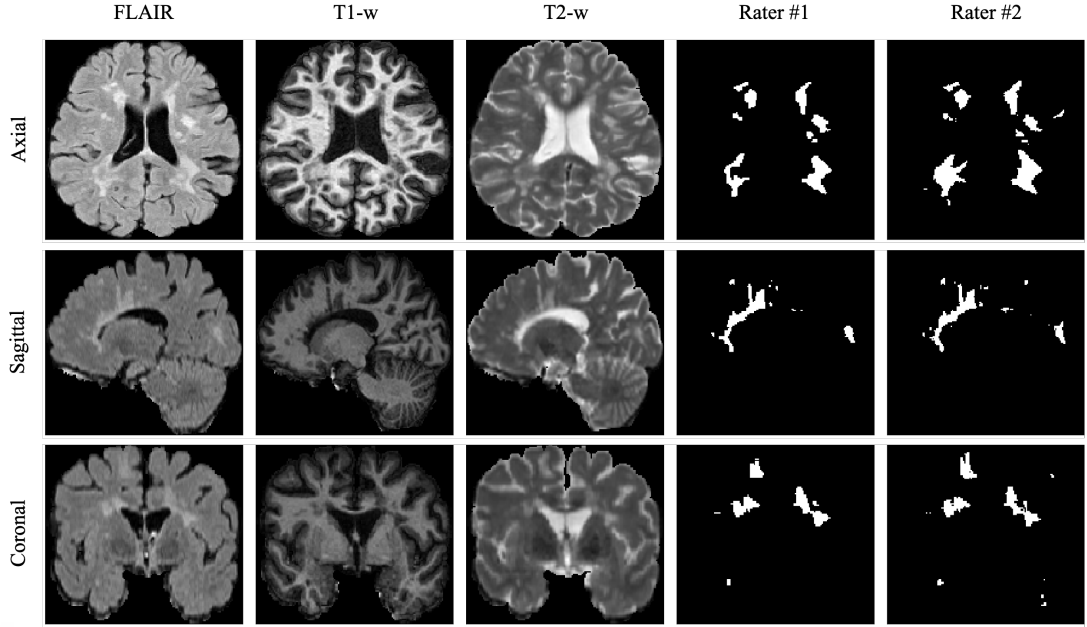


Figure 3.1 : Illustration of the preprocessed ISBI2015 dataset training samples with axial, sagittal, and coronal views of FLAIR, T1-w, and T2-w modalities with their corresponding manual delineations by two raters, Rater #1 and Rater #2.

3.1.2 MSSEG2016 dataset

The MSSEG2016 dataset is composed of 3D MR images of 53 MS patients gathered from four different clinical centers (Center01, Center03, Center07, and Center08) and four MRI scanners (1.5T or 3T). 3D FLAIR, 3D T1-w, 3D T1-w GADO, 2D DP, and 2D T2 sequences are provided for each patient [86]. These images have been divided into two subsets, 15 patients for training and 38 patients for testing. The MRI images from Center03 are only included in the testing set to evaluate the generalization and robustness of models. The MS lesions have been manually delineated by seven experts to evaluate not only the performance of automatic methods but also the inter-expert variability of manual segmentation, and consensus mask data has been computed from their outputs for each patient. The voxel size of each MRI scan ranges in size from $1 \times 0.5 \times 0.5$ to $1.25 \times 1.04 \times 1.0 \text{ mm}^3$ in this dataset. Raw and preprocessed MR images for each patient have been provided by the challenge organizers as well.

In this study, the provided preprocessed dataset was exploited to evaluate the generalization and robustness of the proposed model on unseen centers and scanners.

For the provided preprocessed data, the non-local means algorithm for denoising, the block-matching registration approach for a rigid registration, brain extraction using the volBrain platform, and the N4 algorithm for bias field correction were performed by the challenge organizers [86]. In addition to these steps, intensity normalization was performed on each 3D MRI as in the ISBI2015 dataset. Afterward, each extracted 2D slice from each 3D MRI, which has different spatial dimensions, is resized to a 224×224 shape using the nearest interpolation with keeping the original range of values [87] since the proposed models depend on the fixed input size. Similar to the ISBI2015 dataset, slices only having at least one lesion were chosen to remove excessively unbalanced data and non-informative samples. Figure 3.2 shows the training sample for this dataset.

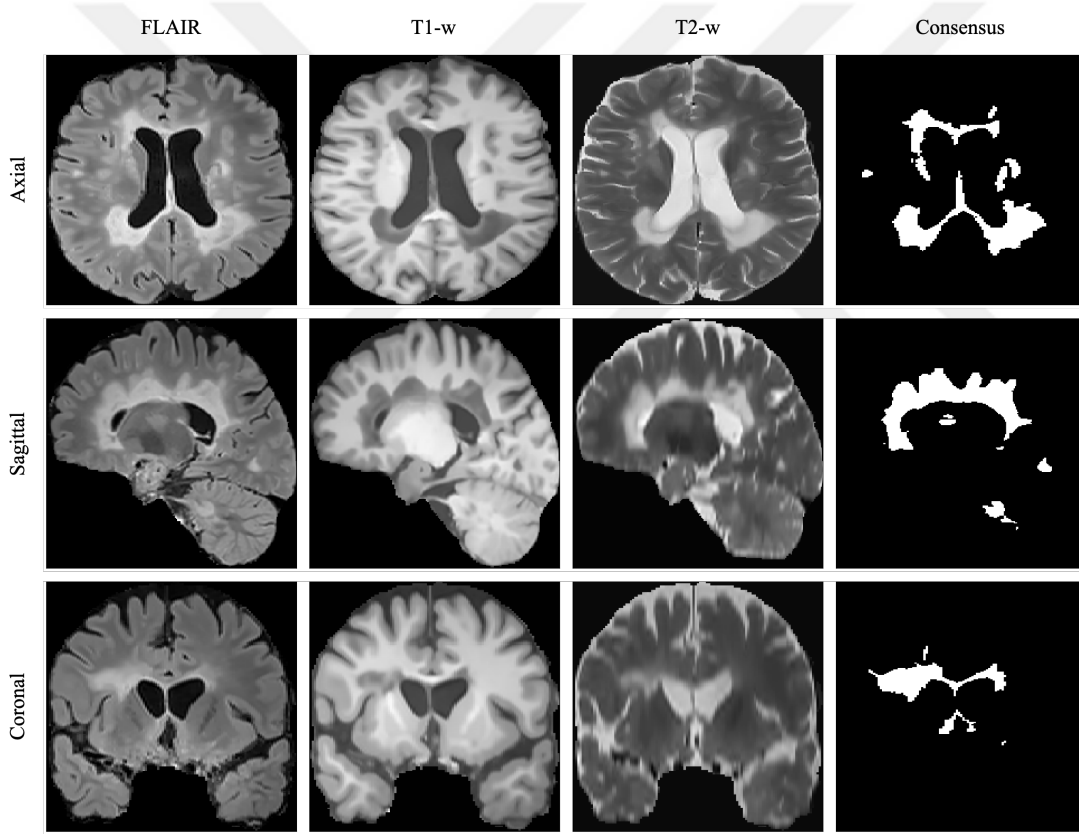


Figure 3.2 : Illustration of the preprocessed MSSEG2016 dataset training samples with axial, sagittal, and coronal views of FLAIR, T1-w, and T2-w modalities with their consensus manual delineation.

3.1.3 MSSEG-2 dataset

In this study, a total of 100 patients' MRI data that was associated with MS disease provided by the MSSEG-2 challenge³ was utilized for the lesion activity detection and segmentation. The voxel size of each MRI data in this dataset varies from $0.5 \times 0.5 \times 0.5$ to $1.2 \times 1.2 \times 1.2 \text{ mm}^3$. The dataset was divided into two groups for training and testing. 40 image pairs were used for the training and the remaining were used for testing. For each patient, raw 3D T2/FLAIR MRI pairs were obtained from 15 different MRI scanners at 1.5T and 3T. A rigid registration was applied to these images to bring them into a middle point in which the ground truth data was calculated by the challenge organizers. Thereafter, a consensus delineated ground truth data for the follow-up images were formed by a majority voting among the four experts and validated by a senior expert neuroradiologist. Figure 3.3 shows the raw and preprocessed input data for the two time points dataset with the delineated ground truth data.

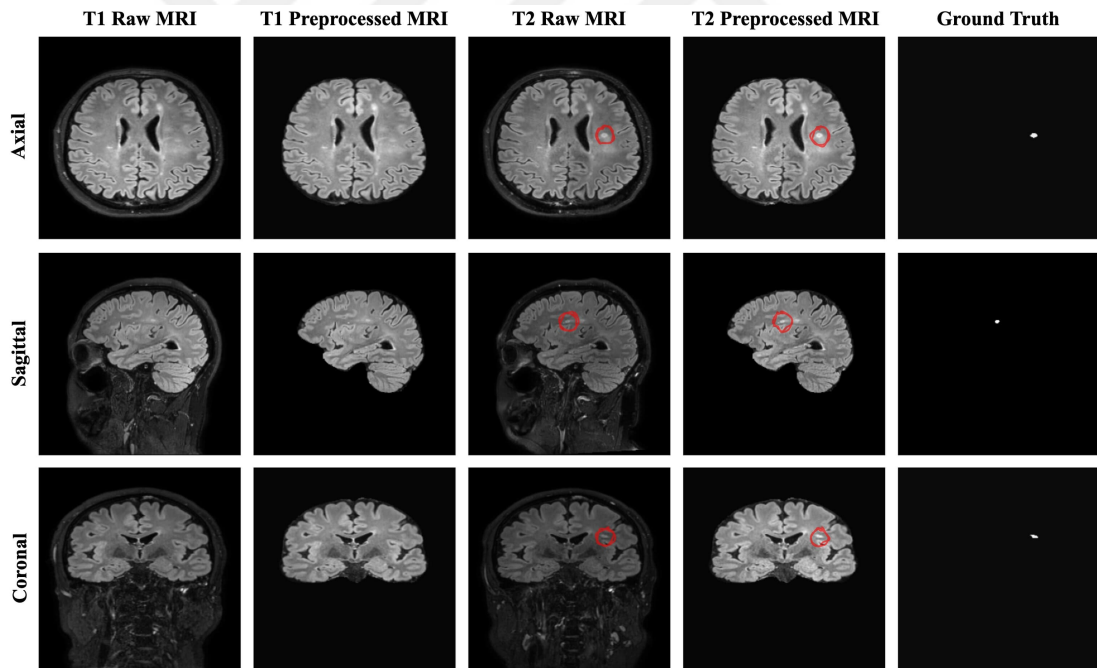


Figure 3.3 : The raw, preprocessed, and delineated mask slices including two-time points for the new MS lesions segmentation task [26].

³Challenge Data: <https://portal.fli-iam.irisa.fr/msseg-2/data/>

3.1.4 Data preprocessing and preparation

Figure 3.4 shows the whole framework of how to process 3D MRI data associated with MS disease for the ISBI2015 and MSSEG2016 datasets. Although the FLAIR sequence has a higher contrast between lesions and white matter than others, other modalities also provide useful features like location and shape. Thus, a stack of 2D slices was generated using 3D volumetric scans of FLAIR, T1-w, and T2-w sequences per channel with their corresponding plane orientations along the x, y, and z axes. Then, stacked 2D slices of all plane orientations were concatenated to form the training input data. For the ISBI2015 dataset, two manual delineations are available, so these two delineation masks were concatenated to create a single training set. The total number of 2D slices extracted from 3D MRI data of each rater was 5197 and 5716, resulting in a total of 10913 2D slices. Later, the total input data was divided into 90% for the training and 10% for the validation data which derived a total of 9821 slices for the training set and a total of 1092 slices for the validation set before feeding the proposed model. For the MSSEG2016 dataset, a total of 5414 2D slices from 15 3D MR images in the provided training set were obtained. Afterward, the data division for training and validation sets was performed similarly to the ISBI2015 dataset, resulting in a total of 4872 slices for the training set and a total of 542 slices for the validation set.

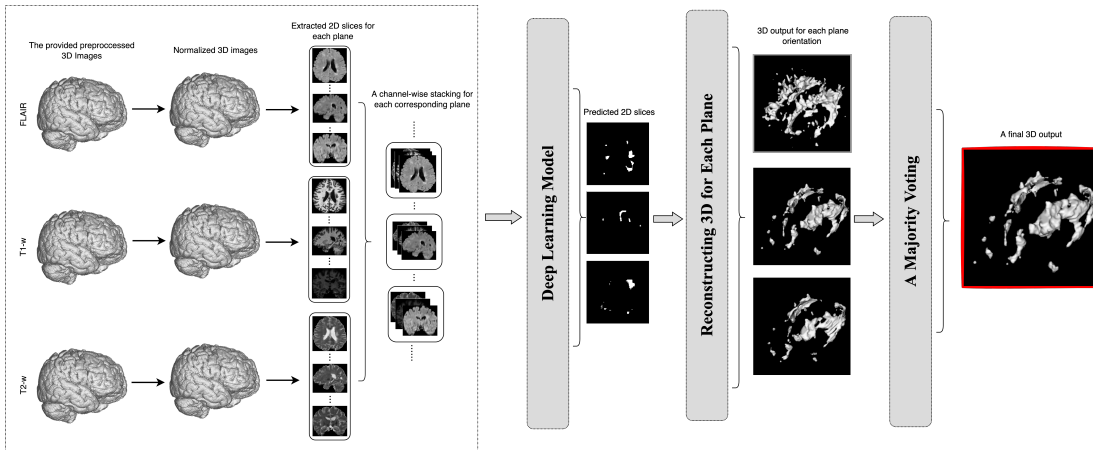


Figure 3.4 : A framework for MS lesion segmentation using the proposed model with 2D slices extracted from 3D MR images of multiple sequences including the output of the final 3D binary mask [25].

Data augmentation is a process of artificially increasing the amount of data by performing random realistic transformations to ensure making the training step is robust, properly improving the prediction accuracy, and reducing overfitting. In this study, data augmentation was performed for the ISBI2015 and MSSEG2016 datasets during the training using Albumentations, an open-source library for image augmentations [88]. In particular, random 90 degrees rotation ($p=0.5$), vertical flip, horizontal flip, random crop, transpose ($p=0.5$), shift scale rotate (shift_limit=0.01, scale_limit=0.04, rotate_limit=0, $p=0.25$), random brightness contrast ($p=0.5$), random gamma ($p=0.25$), emboss ($p=0.25$), blur ($p=0.01$, blur_limit=3), and one of these on elastic transform ($p=0.5$, alpha=120, sigma= 120×0.05 , alpha_affine= 120×0.03), grid distortion ($p=0.5$), optical distortion ($p=1$, distort_limit=2, shift_limit=0.5) were used.

CNN transfer learning from another domain (not medical) to medical domains is usually exploited to initialize training weights in the segmentation tasks due to the data scarcity in this domain. However, in this study, training weights obtained from the ISBI2015 dataset (medical domain) were exploited to initialize training weights on the MSSEG2016 as a transfer learning strategy to improve the segmentation performance, generalization, and robustness of the models.

For the MSSEG-2 challenge, the MSSEG-2 dataset has been gathered to detect and segment new lesions appearing in the follow-up images. This dataset contains 60 3D FLAIR raw images for the training. Data preprocessing is a crucial step for the segmentation task in medical image processing since the raw MRIs may contain irrelevant information like non-brain tissues and skulls. Thus, brain extraction followed by N4 bias field correction [60] was performed on these raw 3D images using the Anima MS longitudinal preprocessing script⁴. Intensity normalization was performed on each 3D MRI scan using the 99th percentile and KDE with the Gaussian kernel similar to one described by Reinhold et al. [89] and Zhang and Oguz [49]. Then, early fusion was performed on the baseline and follow-up images to produce 2-channel input data allowing the proposed model to obtain temporal features from MRI sequences.

⁴Anima scripts: RRID SCR_017072 <https://anima.irisa.fr/>

The resulting 3D MRI data consists of orthogonal plane orientations which yield three views. From this data, the axial, sagittal, and coronal views along the x, y, and z axes were obtained as 2D slices. Since each generated 2D slice has a different size that depends on the orientation, zero padding was applied to obtain a 512 x 512 slice size for all orientations by centering the brain without affecting the original voxel size. As discussed in detail by Hashemi [90], zero padding does not deform the patterns in the image and does not affect the network weights during the backpropagation. To restrict excessively unbalanced data and ignore non-informative samples, the slices which have at least one pixel delineated as a new lesion on the follow-up MR images were chosen to create a training subset. As a result, a total of 2637 2D slices for each time point were derived to be used for training and validation sets. Afterward, the baseline and follow-up images were stacked to generate a 2-channel feature map for each plane orientation. Finally, all 2D stacked slices extracted from all three planes were aggregated to generate a single training input, which allowed to increase training samples and use the contextual information in all directions.

3.2 Methodology

DL has achieved state-of-the-art performance in many computer vision tasks, including medical image analysis, and has become a fundamental tool in artificial intelligence. CNNs have developed into a beneficial approach for the automated segmentation of various tasks in medical image analysis with the advent of deep learning. CNNs are commonly used for these tasks due to their ability to automatically learn and extract features from the given images. Long et al. [91] proposed fully convolutional networks (FCNs) to enhance the performance of CNNs. The fully connected layers in traditional CNNs were replaced with convolutional layers to allow the network to output feature maps at the same resolution as the input image. Subsequently, deconvolutional layers for upsampling are employed to increase the resolution of the feature maps and produce a dense pixel-wise classification. Although FCNs have been successful in many computer vision tasks, such as semantic segmentation, instance segmentation, and object detection, they have only a deconvolutional operation for upsampling, causing significant information loss in the decoder and affecting the segmentation performance. Later, Ronneberger et al. [23] modified and extended this network to generate the

U-Net architecture, which works with fewer training images and combines feature maps from multiple levels to enhance segmentation accuracy. Recent advancements in deep learning techniques have led to many potential improvements in architectural designs, particularly with regard to U-Net which is a commonly utilized architecture for image segmentation. Furthermore, several CNN architectures have been developed specifically for the task of image segmentation, such as FPN and ResNet. These architectures have achieved good performance on benchmark datasets and have the potential to improve the accuracy and efficiency of the segmentation tasks. In this section, deep learning approaches are explained in detail.

3.2.1 U-Net

U-Net, an encoder-decoder convolutional neural network architecture for semantic segmentation with skip connections, was developed for medical image segmentation by Ronneberger et al. [23]. It has been widely used in medical image segmentation tasks and has shown competitive results in the medical field. This network concatenates features from different levels to enhance segmentation performance. It consists of a contracting path (encoder), a bottleneck (bridge), and an expansive path (decoder). In the contracting path, the input image is encoded into the feature representations at multiple different levels with convolution blocks followed by a max pooling operation, and its spatial resolution was reduced by halving its size. In the expansive path, upsampling operations are performed to expand the feature dimensions to concatenate with the corresponding features from the contracting path through the skip connections to better learn representations. The bottleneck typically consists of two 3x3 convolutions that enable propagating features from the contracting path to the expansive path while also compressing the high-level feature representations learned by the encoder into a smaller representation that can be easily propagated through the decoder. This compression yields a more effective computational process while still enabling it to gather the necessary information for accurate semantic segmentation. The skip connections between the contracting and expansive paths allow the network to fuse both low-level and high-level features for the segmentation task. They also help to alleviate the problem of vanishing gradients, which can occur when training deep CNNs. The standard U-Net architecture consists of four downsampling and four upsampling blocks with skip connections added between each corresponding block.

Each block of U-Net consists of two 3x3 convolutions each followed by a Rectified Linear Unit (ReLU) activation. Besides, a 2x2 max pooling operation is applied for each downsampling block while a 2x2 convolution is applied for each upsampling block to change the spatial dimensions of the input feature map. The final layer of the U-Net utilizes a 1x1 convolution with a sigmoid activation function to predict each pixel value ranging from 0 to 1 [23]. The U-Net is illustrated in Figure 3.5.

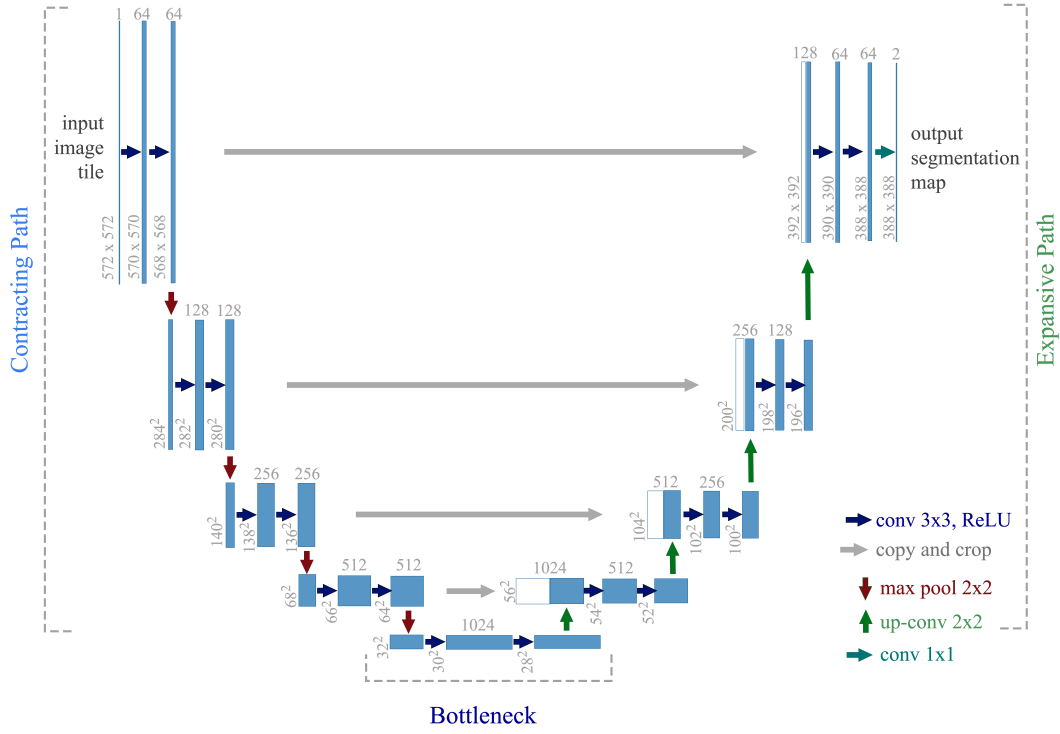


Figure 3.5 : The illustration of U-Net architecture with a contracting path, a bottleneck, and an expansive path.

3.2.2 Feature pyramid network (FPN)

FPN is designed to extract features from a single-scale image of any size, by producing proportionally sized feature maps at multiple levels through a fully convolutional process [92]. FPN has shown competitive performance in instance segmentation and has been employed in many object detection tasks. It consists of bottom-up and top-down pathways as shown in Figure 3.6. The bottom-up pathway of the network is a standard CNN used for image classification, while the top-down pathway generates pyramid-shaped features that can be utilized for object detection at different scales. This pathway takes input from the final layer of the bottom-up path and then increases the resolution of the output from the previous layer as it moves up through the pyramid,

adding in the output from the bottom-up path at each level via lateral connections. As a result, each layer in the top-down path generates features that are specifically designed to detect objects of different sizes.

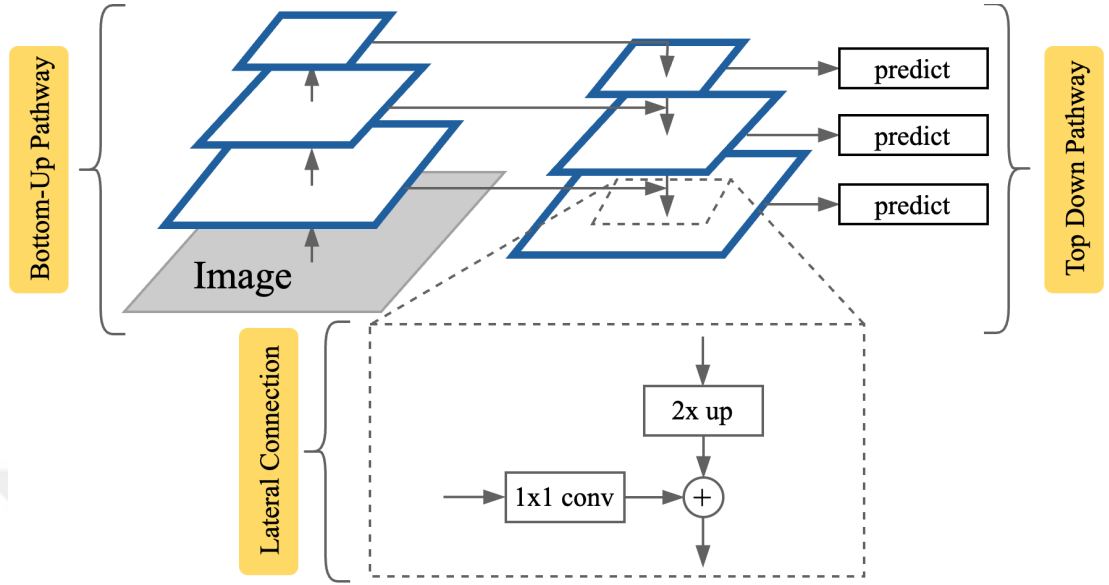


Figure 3.6 : The illustration of FPN architecture and a building block showing the lateral connection and the top-down pathway.

3.2.3 Residual learning (RL)

Adding more layers to build a deeper neural network could enhance the performance of networks; however, increasing the depth of the network may slow down the training process, perhaps resulting in a degradation problem [93]. Deep residual learning uses several residual blocks together in which an identity mapping is created to handle the performance problem, and also address the degradation problem [93]. Indeed, the key idea behind RL is that it allows for the training of much deeper neural networks without encountering the problem of vanishing gradients. RL has been successful in addressing the problem of vanishing gradients and has enabled the training of very deep neural networks with good performance. It has also been applied to a wide range of tasks, including image classification, object detection, and image segmentation. The residual unit consists of two 3x3 convolutional blocks, each with a convolutional layer, BN, and a ReLU activation, as well as an identity mapping that combines the input and output of the residual unit. Figure 3.7 shows the original residual unit in [93].

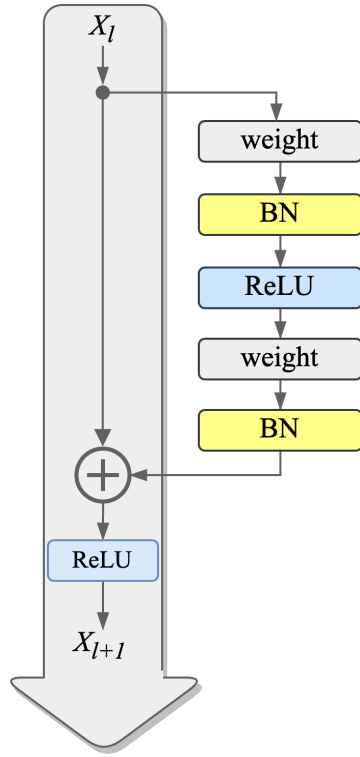


Figure 3.7 : Original residual unit with identity mapping. x_l and x_{l+1} are the input and output of the l -th unit, respectively.

Each residual unit is formulated according to He et al. [94] as follows:

$$y_l = h(x_l) + F(x_l, W_l) \quad (3.1)$$

$$x_{l+1} = f(y_l) \quad (3.2)$$

where x_l and x_{l+1} are the input and output of the l -th unit while F , f , and h indicate the residual function, activation function, and identity mapping, respectively.

3.2.4 Attention gate (AG)

In the context of image segmentation, attention is a technique to suppress feature activations in irrelevant regions of the image during training, reducing the computational resources due to focusing on the relevant feature activations of the given input image. In CNNs, an attention mechanism is a useful tool for feature selection since it simultaneously captures crucial local information and suppresses noisy features [95]. The attention mechanism is categorized into two groups: hard attention and soft attention. Hard attention, such as iterative region proposal and cropping, is typically non-differentiable and depends on reinforcement learning for parameter updates. Due to its non-differentiable, the network can either pay attention to or not to a given region

in an input image. Since standard back-propagation cannot be performed for hard attention, Monte Carlo sampling is required to compute the accuracy over different stages of back-propagation. Contrary to hard attention, soft attention is probabilistic and does not require Monte Carlo sampling; instead, it uses standard back-propagation. In soft attention, different parts of a given image are weighted based on their relevance. High-relevance regions are multiplied with larger weights, while low-relevance regions are multiplied with smaller weights. During training, the model focuses more on the higher-weighted regions [96]. In the standard U-Net, low-level and high-level semantic features are combined via skip connections; however, the high-level features represent poor representation. Therefore, to improve the performance of the U-Net architecture, AGs are used at skip connections before concatenating features that are passed through skip connections and upsampling operations. This allows the network to focus more on features related to different sizes and shapes in the target structure. Indeed, AGs help the CNN models focus on learning the salient features beneficial for specific tasks while avoiding unnecessary regions in an input image [96]. Contextual information provided by the gating signal (g) which is obtained from a coarser scale is used to achieve feature selectivity in AGs [96]. Figure 3.8 shows the additive attention gate approach proposed by Oktay et al. [96].

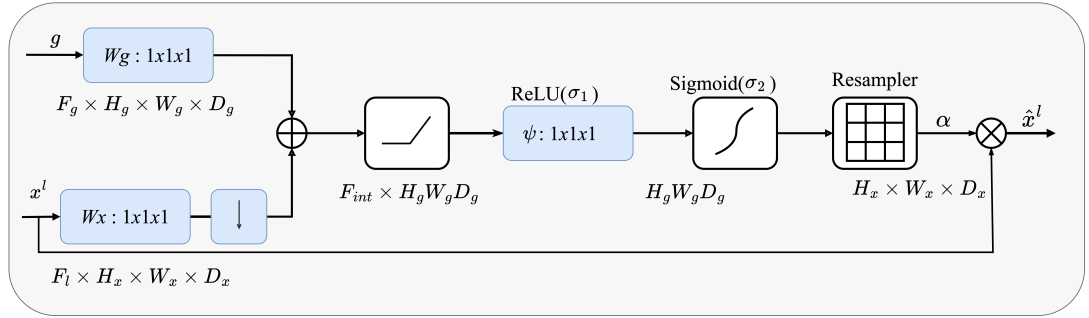


Figure 3.8 : Schematic of the attention gate.

3.2.5 Efficient channel attention (ECA)

The performance of deep CNNs is continuously being improved by many research studies [93, 97, 98]. Squeeze-and-excitation networks (SENet) [99] are one of the notable approaches that improve the performance of CNNs by selectively focusing on important features in the input data. It does this by learning to pay attention to specific channels in each convolutional block, resulting in a clear improvement in

performance across different CNN architectures. Specifically, the SE block takes in input features and first uses global average pooling (GAP) for each channel separately. It then uses two fully-connected layers with non-linearity and a sigmoid function to generate channel weights. These two layers are used to capture non-linear interactions between channels, while also reducing dimensionality to control the complexity of the model. Some research studies have improved the SE block by capturing more complex dependencies among channels or by incorporating additional spatial attention [100–103]. While these methods have produced better accuracy, they also tend to increase the complexity of the model and result in an increased computational burden. Furthermore, dimensionality reduction has a negative impact on the ability to predict channel attention and it is not necessary or efficient to capture dependencies among all channels [104]. Therefore, Wang et al. [104] introduce a new module called ECA for deep CNNs which captures cross-channel interaction efficiently without reducing the dimensionality. The ECA captures local cross-channel interaction by looking at every channel and its k closest neighboring channels. This method has been shown to be both efficient and effective. The ECA can be implemented using fast 1D convolution of size k , where k represents the number of neighboring channels involved in attention prediction for one channel. To avoid manually tuning k via cross-validation, they suggest a method to adaptively determine k , where the coverage of interaction (i.e., kernel size k) is proportional to the channel dimension. Consequently, the ECA module is able to learn effective channel attention without dimensionality reduction and by efficiently capturing cross-channel interactions.

3.2.6 Atrous spatial pyramid pooling (ASPP)

ASPP is a deep learning-based approach for semantic image segmentation proposed by Chen et al. [105]. It was built on the idea of "spatial pyramid pooling" [106–108], which has been proven to be effective in classifying regions of various sizes by resampling features at different scales. ASPP involves using dilated convolutions, also known as atrous convolutions, to expand the network's receptive field without adding more parameters. Also, a fully connected conditional random field (CRF) is employed to enhance the segmentation results. Specifically, ASPP uses multiple parallel atrous convolutional layers with different dilation rates to capture multi-scale context information. The outputs of these atrous convolutional layers are then

concatenated and passed through a final 1×1 convolutional layer to reduce the dimension before upsampling to the original image size. This allows the network to effectively capture both fine and coarse context information, resulting in improved performance on semantic segmentation tasks. ASPP can be used in the context of any CNN architecture and has been shown to improve the performance of CNNs on a variety of tasks, including image classification, object detection, and semantic segmentation. Indeed, ASPP is a useful approach for improving the ability of CNNs to capture contextual information in images and has been widely used in the field of computer vision. Figure 3.9 shows the proposed ASPP module in [105].

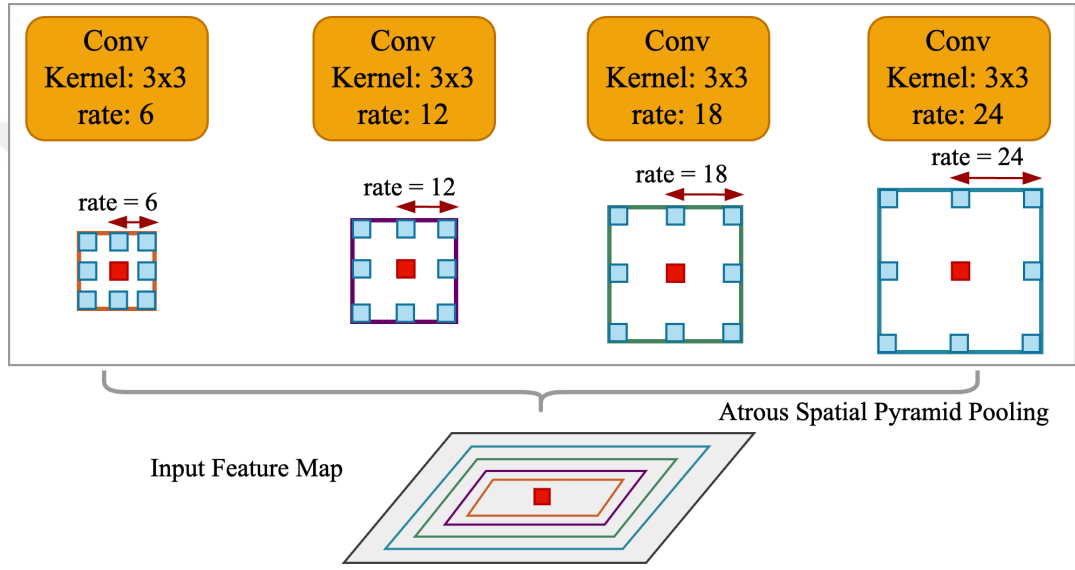


Figure 3.9 : Atrous Spatial Pyramid Pooling (ASPP). ASPP uses multiple parallel filters with varying rates to take advantage of multi-scale features in order to classify the central pixel (shown in red). The different Field-Of-VIEWS are depicted in different colors.

3.2.7 Loss functions

In DL, a loss function is a measure of how well a neural network is able to predict the correct output for a given input. When training a neural network, the goal is to minimize the loss function by adjusting the weights and biases of the network in order to improve its performance. There are many different types of loss functions that can be used in medical image analysis, depending on the specific task and the desired properties of the model [109]. The most commonly used loss functions are as follows:

3.2.7.1 Binary cross-entropy loss

The binary cross-entropy (BCE) loss function is commonly utilized for binary classification tasks in medical image analysis. It measures the difference between the predicted probability of the positive class and the true label (0 or 1). The BCE loss function can be calculated for each pixel, and the total loss for the model can then be calculated using the average loss over all the pixels. It is formulated as,

$$L(gt, pr) = -(gt \log(pr) + (1 - gt) \log(1 - pr)) \quad (3.3)$$

where gt and pr indicate ground truth and predicted segmentation, respectively.

3.2.7.2 Categorical cross-entropy loss

This loss function is similar to BCE loss; however, it is used for multi-class classification tasks, where the output can belong to one of multiple classes. It is formulated as,

$$L(gt, pr) = -gt \log(pr) \quad (3.4)$$

where gt and pr indicate ground truth and predicted segmentation, respectively.

3.2.7.3 Dice loss

The Dice Similarity Coefficient (DSC) is used to measure the similarity between the evaluated segmentation and the ground truth. Later, it was also adapted as a Dice loss function [110]. This loss function is often exploited for image segmentation tasks to measure the overlap between the predicted segmentation and ground truth, and is particularly useful for imbalanced datasets. Indeed, it is a variant of the standard cross-entropy loss function and is designed to improve the performance of deep learning models on tasks in which the classes are imbalanced. Milletari et al. [110] indicate that the dice loss function can improve the performance of a deep learning model on a medical image segmentation task, compared to using the standard cross-entropy loss function, and can be expressed as,

$$L(gt, pr) = 1 - \frac{2gtpr + 1}{gt + pr + 1} \quad (3.5)$$

where gt and pr indicate ground truth and predicted segmentation, respectively. 1 is added in the numerator and denominator to prevent the function from being not undefined in edge-case scenarios such as when $gt = pr = 0$.

3.2.7.4 Binary focal loss

The binary loss function was introduced by Lin et al. [111] and designed to address the problem of class imbalance in binary classification tasks when the number of examples belonging to one class is noticeably greater than the number of examples belonging to the other class. The binary focal loss function modifies the standard cross-entropy loss function by introducing a weighting term that gives easy samples less weight and hard samples more weight. This weighting term is based on the prediction probability of the model, and is designed to focus the model's attention on the hard examples that are more likely to be misclassified. It can be defined as follows:

$$L(gt, pr) = -(gt\alpha(1 - pr)^\gamma \log(pr) + (1 - gt)\alpha pr^\gamma \log(1 - pr)) \quad (3.6)$$

where gt and pr indicate ground truth and predicted segmentation, respectively. The default values of α and γ are 0.25 and 0.2, respectively.

3.2.7.5 Categorical focal loss

The categorical focal loss function is designed to address the class imbalance problem in multi-class classification problems. It is a modification of the standard cross-entropy loss function and can be defined as follows:

$$L(gt, pr) = -gt\alpha(1 - pr)^\gamma \log(pr) \quad (3.7)$$

where gt and pr indicate ground truth and predicted segmentation, respectively. The default values of α and γ are 0.25 and 0.2, respectively.

3.2.8 Evaluation metrics

The ISBI2015, MSSEG2016, and MSSEG-2 challenges use several evaluation metrics to assess segmentation quality. Accordingly, common metrics in these challenges are Dice and positive predictive value (PPV) scores. In addition to these metrics, Lesion-Wise True Positive Rate (LTPR), Lesion-Wise False Positive Rate (LFPR), and absolute volume difference (VD) metrics are provided for the ISBI2015

challenge while the F1 score, sensitivity, and specificity metrics are provided for the MSSEG2016 and MSSEG-2 challenges. The results of the ISBI2015 can be obtained via the online website by submitting predicted 3D binary masks. However, to evaluate the model performance on the MSSEG2016 testing set, the segmentation performance analyzer tool available in Anima (animaSegPerfAnalyzer)⁵ is used. To obtain evaluation metric results, each final 3D binary mask obtained from a majority voting and its corresponding ground truth data provided by the MSSEG2016 challenge organizers are given as input to the animaSegPerfAnalyzer tool [35]. To evaluate the model performance on the MSSEG-2 testing set for those who attended the challenge, a Docker or Singularity image was supposed to publish on the Virtual Imaging Platform (VIP)⁶, a web-based portal specialized in medical simulations and image data analysis. Additionally, the segmentation performance analyzer tool can be utilized if the MSSEG-2 testing set is available.

DSC is the most commonly used metric in validating MS lesion segmentation. It is generally exploited in image segmentation tasks across different domains, including medical image analysis [110, 112]. DSC is employed to measure the overall segmentation accuracy between the predicted segmentation and the ground truth, and is formulated as follows:

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (3.8)$$

where TP , FP , and FN indicate the true positive, false positive, and false negative voxels, respectively.

To compare the proposed model with other models used in the literature for the ISBI2015 testing dataset, LTPR and LFPR are also exploited for all experiments. LTPR is expected to be a higher percentage while LFPR is expected to be a lower percentage, and these metrics are calculated as follows:

$$LTPR = \frac{LTP}{RL} \quad LFPR = \frac{LFP}{PL} \quad (3.9)$$

where LTP is the number of lesions in the ground truth that overlap with a lesion in the predicted segmentation. RL denotes the total number of lesions in the ground truth.

⁵Anima scripts: RRID SCR_017072 <https://anima.irisa.fr/>

⁶VIP:<https://vip.creatis.insa-lyon.fr/>

LFP is the number of lesions in the predicted segmentation that do not overlap with a lesion in the ground truth reference mask. PL denotes the total number of lesions in the predicted segmentation.

The ISBI Challenge website also provides additional metrics for performance evaluation, such as PPV and VD [79]. The PPV and VD metrics can be formulated as follows:

$$PPV = \frac{TP}{TP + FP} \quad VD = \frac{|TP_s - TP_{gt}|}{TP_{gt}} \quad (3.10)$$

where TP_s and TP_{gt} indicate the number of segmented voxels in the automatic segmentation output and the ground truth, respectively.

The ISBI challenge exploits some of these metrics in the ISBI score calculation. The total score used on the official ISBI website is calculated as follows:

$$Score = \frac{1}{|R| \cdot |S|} \cdot \sum_{R,S} \left(\frac{DSC}{8} + \frac{PPV}{8} + \frac{1 - LFPR}{4} + \frac{LTPR}{4} + \frac{Cor}{4} \right) \quad (3.11)$$

where S is all test subjects (14 test subjects consisting of 61 3D scans), R is all raters (rater 1 and rater 2), and Cor is Pearson's correlation coefficient of the volumes. The inter-rater score was computed as a score of 90, which means a score of 90 or higher can be considered a comparable score to a human rater [79].

In addition to these metrics, the F1 score, sensitivity, and specificity were employed to evaluate the segmentation performance of the proposed models on the MSSEG2016 and MSSEG-2 datasets. The F1 score is a detection and lesion-wise metric that focuses on the number of lesions correctly recognized without considering the precision of their contours. Lesion sensitivity (S) and lesion positive predictive (P) are exploited to calculate the F1 score defined in equation 3.12 according to [35]:

$$S = \frac{TP_G}{M} \quad P = \frac{TP_A}{N} \quad F_1 = \frac{2SP}{S + P} \quad (3.12)$$

where M , N , TP_G , and TP_A indicate the number of lesions in the ground truth, in the automatic segmentation, overlapped in the ground truth, and overlapped in the automatic segmentation, respectively.

Sensitivity (S_e) and Specificity (S_p) are defined as overlap metrics that consider the voxel-based overlap of the automatic segmentation (A) and the ground truth (G). These

two metrics are computed as follows:

$$S_e = \frac{A \cap G}{G} \quad S_p = \frac{B - A \cap G}{B - G} \quad (3.13)$$

where B denotes the entire image.

Mean Surface Distance (S) is also an overlap metric and is formulated as follows:

$$S = \frac{\sum_{i \in A_S} d(x_i, G_S) + \sum_{j \in G_S} d(x_j, A_S)}{N_A + N_G} \quad (3.14)$$

where d indicates the minimal Euclidean distance of a point of one surface to the other surface. N_A and N_G reveal the number of points of each surface, respectively.

In addition to these evaluation metrics used for the MSSEG-2 challenge, two simple metrics are developed based on the number and volume (mm^3) of lesions detected for cases that do not have any new lesions on their follow-up images. This is usual in clinical cases, and this challenge also has similar cases in both training and test data sets. For example, the testing set consists of 28 patients with no new lesions and 32 patients with at least one or more new lesions. The number and volume of new lesions were used as evaluation metrics as well. The volume of new lesions was calculated by multiplying the number of voxels in the segmentation with the voxel volume. A value of zero is the optimal value for these metrics.

3.3 Proposed a Dense Residual U-Net for MS Lesion Segmentation and Detection

The proposed model integrates different components into the modified U-Net architecture. U-Net was chosen since it performs well over different domains in image segmentation, especially in the medical domain [23]. It is an encoder-decoder architecture for semantic segmentation with skip connections and consists of a contracting path, a bottleneck, and an expansive path. In the contracting path, the input image is encoded into the feature representations at multiple different levels with convolution blocks followed by a max pooling operation. In the expansive path, upsampling operations are performed to expand the feature dimensions to concatenate with the corresponding features from the contracting path through the skip connections to better learn representations. The bottleneck, consisting of two 3×3 convolutions, enables propagating features from the contracting path to the expansive path. The standard U-Net architecture consists of four downsampling and four upsampling

blocks with skip connections added between each corresponding block. Each block of U-Net consists of two 3x3 convolutions each followed by a ReLU activation function. Besides, a 2x2 max pooling operation is applied for each downsampling block while a 2x2 convolution is applied for each upsampling block to change the spatial dimensions of the input feature map. At the final layer of U-Net, each pixel value is predicted in the range of 0 to 1 by employing a 1x1 convolution followed by a sigmoid activation function [23]. To improve the performance of MS lesion segmentation, the U-Net architecture was modified by batch normalization (BN), spatial dropout (SD), an exponential linear unit (ELU) for the activation function, strided convolutions for the pooling operations, and transposed convolutions for the upsampling operations. BN standardizes the input to a layer for each mini-batch, enhances convergence, and reduces overfitting. The ELU activation used in this study is given in equation 3.15 as follows [113]:

$$ELU = f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(\exp(x) - 1), & \text{if } x \leq 0 \end{cases} \quad (3.15)$$

where α is a positive scale factor to control negative x values.

In addition to these modifications, the 3x3 convolution in each block was replaced by a residual block that handles performance and degradation problems [93]. A full pre-activation residual unit was implemented according to He et al. [94], and each residual block in each block was densely connected as presented in Figure 3.10. AGs, which allow the model to focus on relevant features while avoiding irrelevant regions in an input image [96], were modified by adding ELU and BN. To suppress feature activations in irrelevant regions of the image, AGs are implemented at skip connections before concatenating features passed through skip connections and upsampling operations as shown in Figure 3.10. Furthermore, the ECA module proposed by Wang et al. [104] was incorporated into the proposed model as given in Figure 3.11(a) in detail. This module was appended to the end of each downsampling block and each residual block as shown in Figure 3.10. This module produces channel weights based on the features aggregated from GAP by applying a fast 1D convolution operation of size k , which is adaptively determined by channel dimension. Later, BN and a sigmoid activation function are performed, respectively. Another important component is the modified ASPP semantic segmentation module. The bottleneck of U-Net was replaced by this module which resamples a given feature layer at multiple

rates before the convolution operation [105]. The ASPP has demonstrated promising results on numerous segmentation tasks by providing multi-scale information. Hence, ASPP was exploited to acquire useful multi-scale information for the MS lesion segmentation task. Figure 3.11(b) shows the implemented ASSP component in this study. This component consists of four 3x3 convolutional layers with dilation rates of 1, 6, 12, and 18 each followed by BN. Later, the extracted feature maps from each convolutional layer are added, and a 1x1 convolution, BN having L2 regularization of gamma and beta, and ELU are performed, respectively.



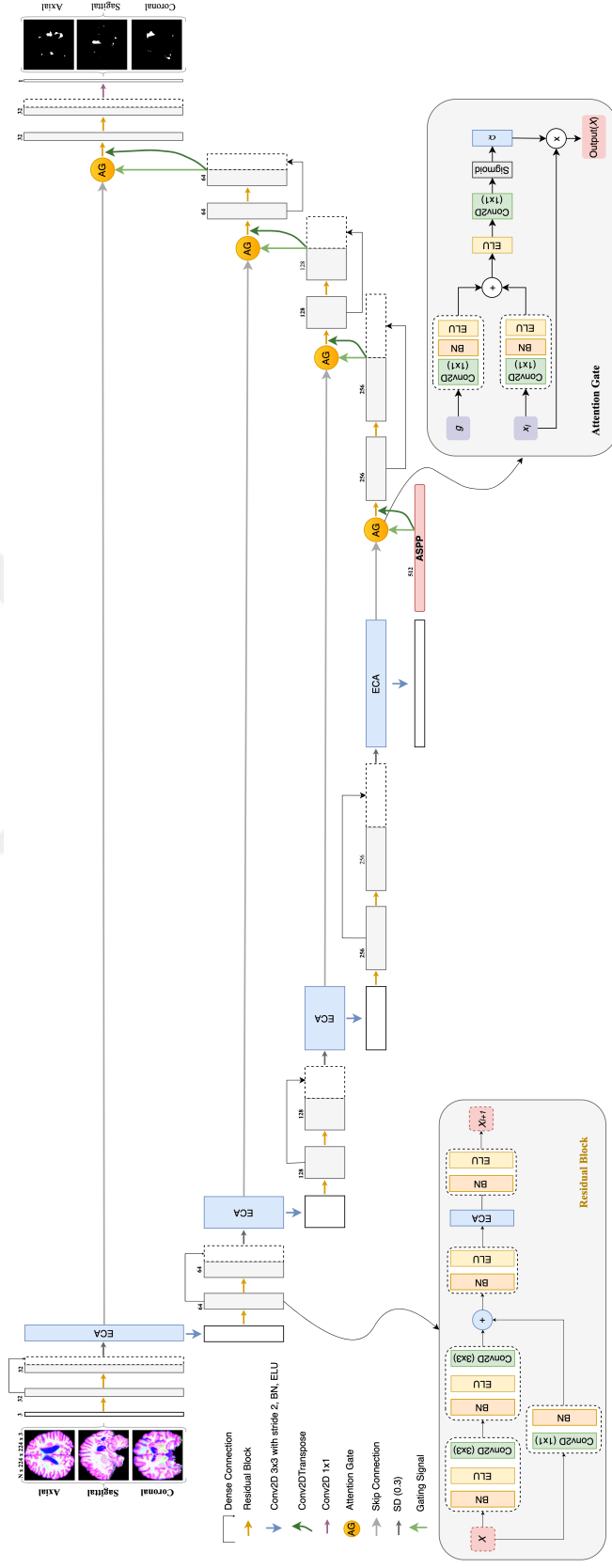


Figure 3.10 : The dense residual U-Net with AG, ECA, and ASPP using multi-planar 2D slices of FLAIR, T1-w, and T2-w sequences. The number of filters starts from 32 and increases two times up to 512. In the residual block, x_l and x_{l+1} denote the input and output of the $l - th$ unit. In the attention gate, x_l , g , and α denote the input feature map of the layer l , the gating signal providing contextual information, and the attention coefficient used to scale the x_l , respectively [25].

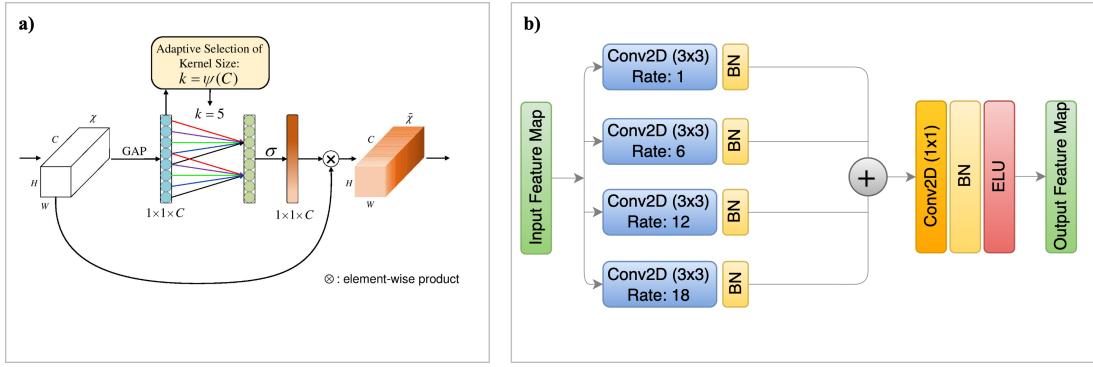


Figure 3.11 : a) Diagram of the ECA module. GAP is used to obtain the aggregated features from the given input. In ECA, a fast 1D convolution of size k is performed to obtain channel weights, where a mapping of the channel dimension C is used to adaptively calculate k [104]. b) Diagram of the ASPP module. From given the features, ASPP performs multi-rate convolution operations to obtain multi-scale information [25].

3.4 Proposed Deep Residual Attention Gate U-Net for MS Lesion Activity

In this study, the combination of U-Net, deep residual learning, and AG was proposed for the new MS lesion segmentation task. In this combination, the residual unit will facilitate the network training. Information will be able to propagate without degradation thanks to the skip connections within a residual unit and between low and high levels of the network. Thus, deep neural networks are built with fewer parameters while still achieving a competitive segmentation performance. As such, the standard blocks were replaced with residual blocks in the proposed model. Figure 3.12 shows the residual unit including identity mapping within the proposed model. He et al. [94] also recommended a full pre-activation as demonstrated in Figure 3.12. In this study, a full pre-activation residual unit was used to construct and design the deep residual attention gate U-Net.

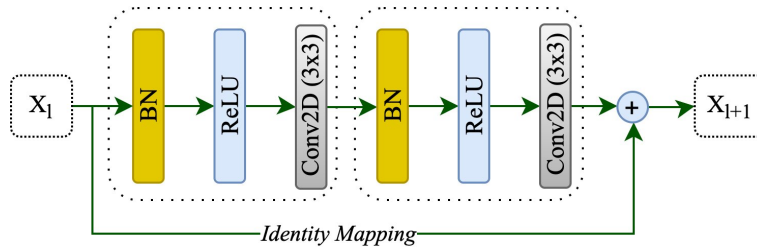


Figure 3.12 : A residual unit with identity mapping. x_l and x_{l+1} are the input and output of the l -th unit, respectively [26].

AGs, modified by adding BN and a ReLU activation for both input features before convolutional operations, were added between the corresponding encoding part and the upsampling of feature maps produced from the lower level. Thus, allowing the model to learn to focus on salient features of various shapes and sizes. Figure 3.13 demonstrates the details of the designed network with the input data formed by the axial, sagittal, and coronal views extracted from the baseline and follow-up 3D MRI for the new MS lesion segmentation. Additionally, Figure 3.13 shows the overview of the attention gate mechanism used within the proposed model.



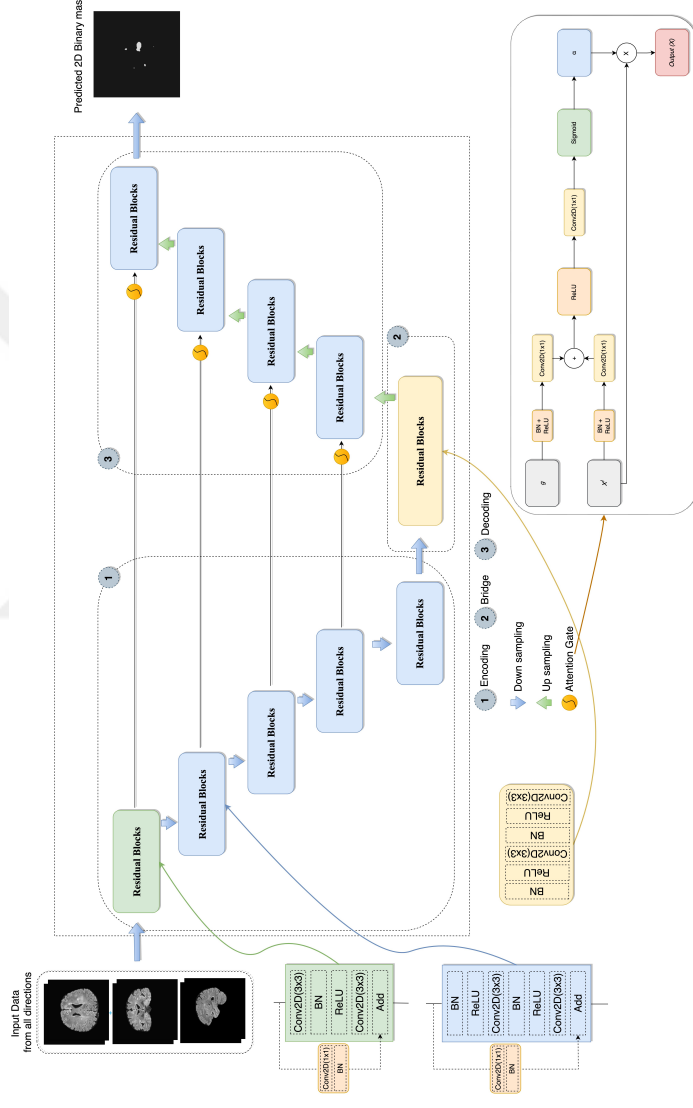


Figure 3.13 : The architecture of the proposed model combines U-Net, residual learning, attention gate, and a slice-based approach. In AG adapted from Oktay et al. [96], α is the input features, α' is the attention coefficients used to scale the α' , and g collected from a coarser scale is the gating signal which provides contextual information [26].

3.5 Implementation

The ISBI2015 and MSSEG2016 training sets are composed of 21 and 15 3D MR images, respectively. FLAIR, T1-w, and T2-w sequences are also provided for both datasets. To prepare training input data, 2D slices were extracted from these 3D MR images for each sequence according to the three orthogonal directions. Then, a three-channel input feature map was generated by leveraging each corresponding 2D slice obtained from plane orientations of the three sequences as discussed previously. Afterward, the obtained total 2D slices were divided into the training and validation sets with a ratio of 90 to 10 for each dataset, respectively. The proposed model was implemented in the Python programming language⁷ using Keras⁸ running on top of TensorFlow⁹ [114, 115]. All experiments were conducted on Google Colab Pro, which provides an NVIDIA Tesla P100 GPU with 16GB memory [116]. The models were trained by using the Adam optimizer [117] with an initial learning rate of 1e-4 (adjusting with patience = 10, factor = 0.1, cooldown = 10, and min_lr = 1e-5 during the training), and a batch size of 8 over 300 epochs. BN was implemented by gamma and beta regularizers with L2 (1e-4). Class imbalance is a common problem in MS lesion segmentation, as lesions only constitute a minority of the MRI volume [31]. Therefore, a hybrid loss function, the addition of Dice and focal losses [111, 118], was exploited as the training loss function to handle the class imbalance problem. The final loss function is calculated in equation 3.16.

$$L(gt, pr) = \left(1 - \frac{2gt\,pr + 1}{gt + pr + 1}\right) + 1 \times (-gt\alpha(1-pr)\gamma\log(pr) - (1-gt)pr\gamma\log(1-pr)) \quad (3.16)$$

where gt and pr indicate the ground truth and the prediction segmentation, respectively. The values of α and γ are 0.25 and 2.0, respectively.

The training dataset was divided into training and validation sets to adjust the network weights and early stopping criterion. The validation loss was monitored to choose the best model, and model weights were saved at the end of the epoch where the validation

⁷<https://www.python.org/>

⁸<https://keras.io/>

⁹<https://www.tensorflow.org/>

loss was at its minimum during the training. Moreover, early stopping, used to avoid overfitting, was exploited to enable the training to be automatically stopped when the validation loss stopped improving for 50 epochs. To yield the final 3D segmentation output, the predictions of 3D binary output generated from 2D slices of each plane orientation were fused using a majority voting method as shown in Figure 3.4.

For the MSSEG-2 challenge, the training set comprised 3D FLAIR images of 40 patients and only 29 had new lesions in their follow-up images. These 29 MR images were divided into the training and validation sets (24 patients for training and 5 patients for validation). To prepare input data, each 3D image was divided into its axial, sagittal, and coronal views. Two-channel input feature data was created using each corresponding 2D slice from both time points as discussed previously. Keras and TensorFlow libraries were used for the model development in the Python language [114, 115]. The Google Colaboratory, having a Tesla K80 GPU, was used for the training procedure [116]. The proposed model was trained by using the Adam optimizer [117], an initial learning rate of 1e-4 (adjusting with patience=10 and factor=0.1 during the training), and batch size of 8 over 200 epochs, respectively. The validation dice score was also monitored to choose the best model, and model weights were saved based on the best validation dice score during the training. Early stopping (patience=50) was exploited to prevent overfitting as well. Hashemi et al. [119] used the sum of dice loss with a 1.5 coefficient and binary cross entropy loss as a custom loss function for MS lesion segmentation. Similarly, in this study, a hybrid loss function consisting of binary focal loss and dice loss (dice loss + (1 × binary focal loss)) was employed in order to handle unbalanced labeled data between lesion and background since lesion pixels constitute a minor portion of the whole image. The total loss function is defined as follows:

$$L_t = \left(1 - \frac{2gtpr + 1}{gt + pr + 1}\right) + (1 \times (-gt\alpha(1 - pr)^\gamma \log(pr) - (1 - gt)\alpha pr^\gamma \log(1 - pr))) \quad (3.17)$$

where gt denotes the ground truth, and pr indicates prediction. 0.25 and 2.0 default values were used for the parameters of α and γ , respectively.

Additionally, the Keras data generator was used for performing real-time data augmentation such as vertical flipping, horizontal flipping, random rotation, and shift range to increase the number of training samples. Figure 3.14 shows the proposed

pipeline for new lesion segmentation of MS activity. First, 3D MRIs were converted into their plane orientations along the x, y, and z axes. Then, 2D slices of two-time points were fused together to create a single input training data for the proposed model. Predicted 2D slices based on the axial, sagittal, and coronal views were converted into the 3D binary segmentation output, and then the final output segmentation mask was generated by using the majority voting among the 3D binary outputs obtained from each view.



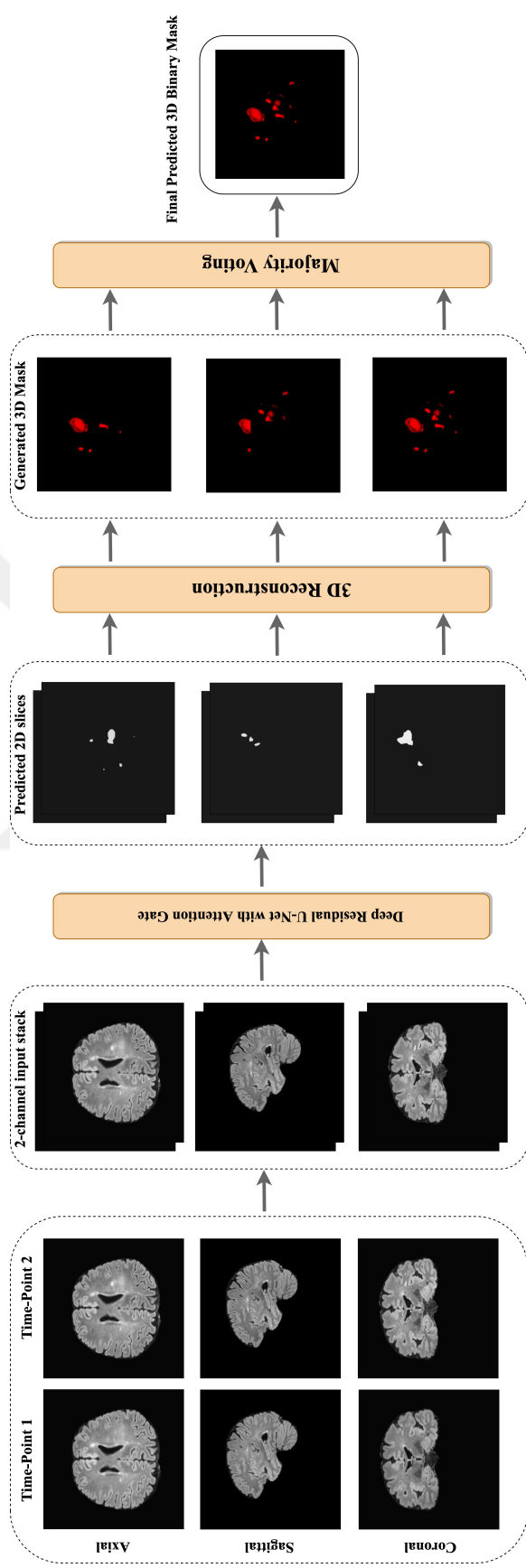


Figure 3.14 : The proposed pipeline of new MS lesions segmentation using a slice-based approach including the majority voting for the final 3D segmentation output using the predicted 2D axial, sagittal, and coronal slices [26].

To compare components of the designed network, a testing subset was created from the MSSEG-2 test dataset provided by the challenge organizers. This subset comprised MRI data of 7 patients by considering the different scanners and new lesion loads. Satisfactory results with the MSSEG-2 dataset could not be obtained by the implementation of the original U-Net. Therefore, this implementation was modified with transpose upsampling instead of a simple upsampling operation, and BN to make the neural network more stable. A hybrid loss function, the summation of binary focal and Dice losses, was used for all models.

To generate a 3D binary prediction output, a 3D image reconstruction approach was utilized. The slices from each view were used to reconstruct the final 3D binary segmentation output. The 3D binary segmentation was produced using the 2D predicted slices from each plane orientation. Then, a majority voting was applied to these 3D segmentation outputs to generate the final 3D binary segmentation as shown in Figure 3.14.



4. RESULTS AND DISCUSSIONS

To evaluate the performance of the proposed model [25] both quantitatively and qualitatively, several experiments were conducted on the ISBI2015 and MSSEG2016 datasets. First, an ablation study was performed to demonstrate the effectiveness of each component used to build the proposed model. Second, a comparison with other methods in the literature was performed based on the results of both datasets. Finally, comparisons between datasets were performed to evaluate the generalization ability and robustness of the model, including transferring weights from one dataset to another.

Regarding the MSSEG-2 challenge, evaluations were conducted based on the results published on the VIP platform for the proposed pipelines. First, the detection and segmentation metrics F1 and Dice scores were assessed. Second, the number and volume of lesions detected for cases with no new lesions were evaluated. Third, a comprehensive evaluation was performed using the additional metrics sensitivity, specificity, PPV, and surface distance, discussed in Section 3.2.8. Lastly, an ablation study was conducted to demonstrate the impact of each component in the proposed model [26].

4.1 ISBI2015 Dataset Results

4.1.1 Ablation study

The ISBI2015 dataset was employed to conduct the ablation study due to performing a fair comparison of the results obtained from the challenge website. Figure 4.1 and Table 4.1 summarize the performance of different components of the proposed model [25] according to Dice score, PPV, LTRP, LFPR, and VD evaluation metrics. As seen in Figure 4.1 and Table 4.1, the proposed model, which contains all components, outperformed variants of other components used to build the proposed model. Dice score is one of the most important evaluation metrics for MS lesion segmentation, as

such this proposed model produced a mean Dice score of 66.88% and also an ISBI score of 92.75. Eventually, this model achieved the highest Dice and ISBI scores compared to its variants. Additionally, this model achieved the best PPV of 0.8650 and the best LFPR of 0.2617 among its variants.

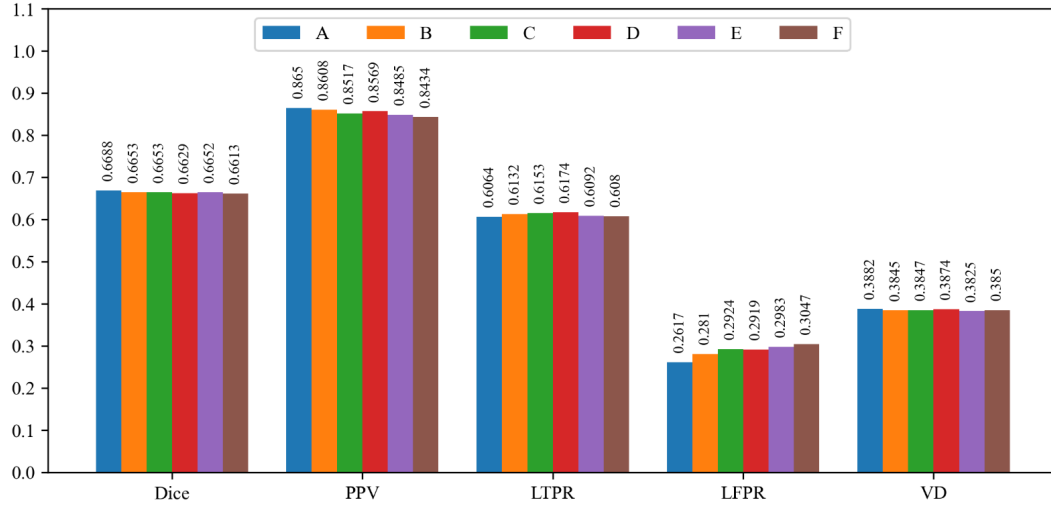


Figure 4.1 : Evaluation metric results on the ISBI2015 testing set for each variant of the proposed model [25].

Table 4.1 : Ablation study results with different variants of our model trained on the ISBI2015 training set and tested on the ISBI2015 testing set. To be ranked first, the ISBI score, Dice score, PPV, and LTPR are expected to have high numerical values, while the LFPR and VD are expected to have low numerical values.

Model	ISBI Score	Dice Score	PPV	LTPR	LFPR	VD
(A) Dense Residual U-Net with AG, ECA, and ASPP	92.75	0.6688	0.8650	0.6064	0.2617	0.3882
(B) Dense Residual U-Net with AG and ECA	92.63	0.6653	0.8608	0.6132	0.2810	0.3845
(C) Dense Residual U-Net with AG	92.52	0.6653	0.8517	0.6153	0.2924	0.3847
(D) Dense Residual U-Net	92.54	0.6629	0.8569	0.6174	0.2919	0.3874
(E) Dense U-Net	92.41	0.6652	0.8485	0.6092	0.2983	0.3825
(F) U-Net	92.31	0.6613	0.8434	0.6080	0.3047	0.3850

¹ This table is sorted in descending order of the ISBI score.

² The bold value indicates the best score among the generated models.

³ Each model will be denoted by letters between A-F in future analyses.

Figure 4.2 shows the boxplot of the Dice score and PPV for each patient data of two raters to make comparisons visually, as well. The proposed model surpassed its variants with statistical significance in Dice score with p-values $p \leq 0.0001$ and in PPV with p-values ranging from $p \leq 0.05$ and $p \leq 0.0001$ according to a Wilcoxon test. Figure 4.3 demonstrates the predicted samples for the FLAIR modality of three views for each variant of the proposed model. MS lesions on the output segmentation can be easily distinguished and give a hint of where possible MS lesions could be located. The prediction outputs would assist physicians in monitoring MS disease progression or measuring lesion volume in a short time.



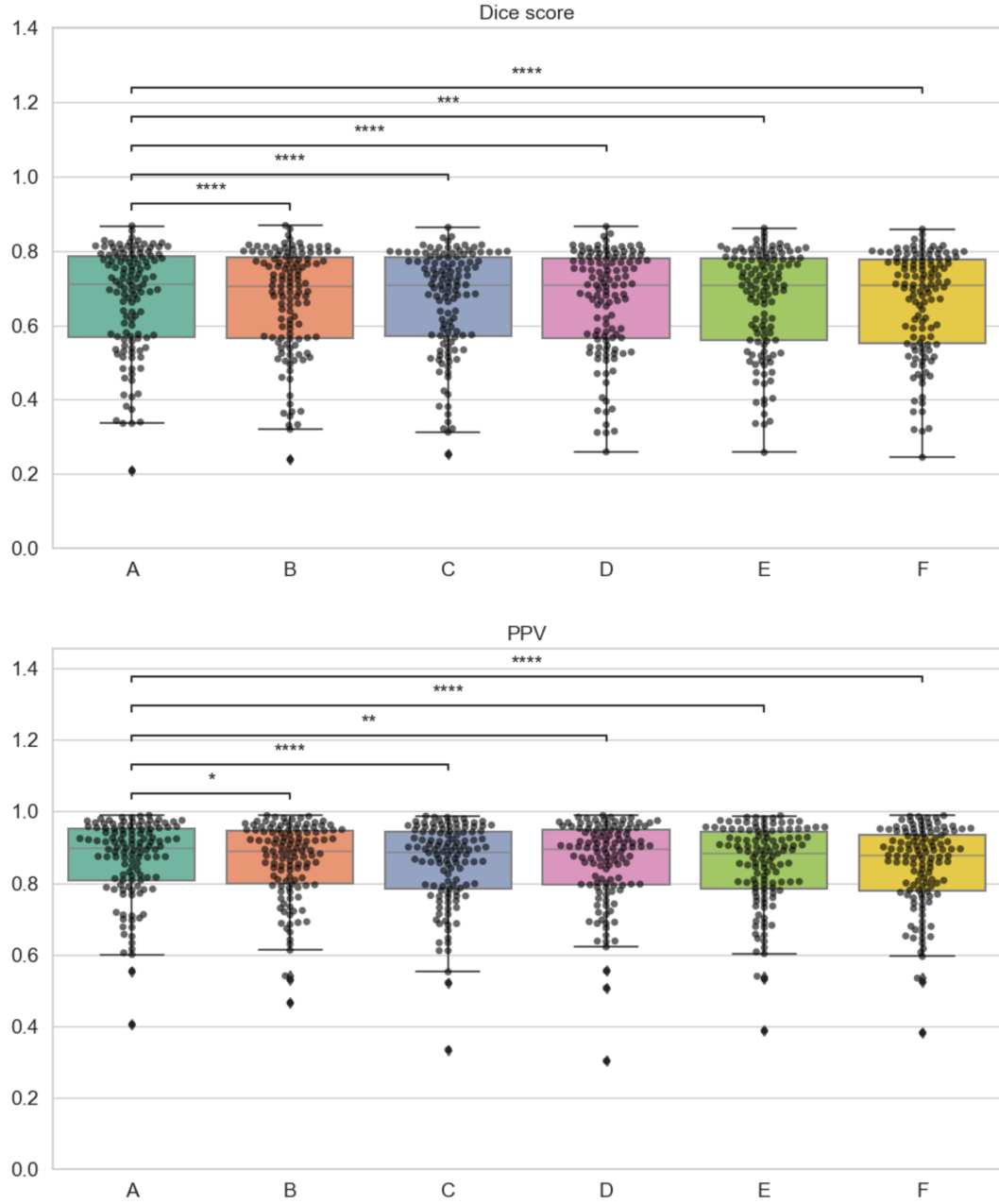


Figure 4.2 : The boxplot of Dice and PPV scores for each patient across raters on the ISBI2015 testing set for the proposed model and its variants. Asterisks indicate statistical significance (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$) when using a Wilcoxon test compared to the proposed model (denoted as A) [25].

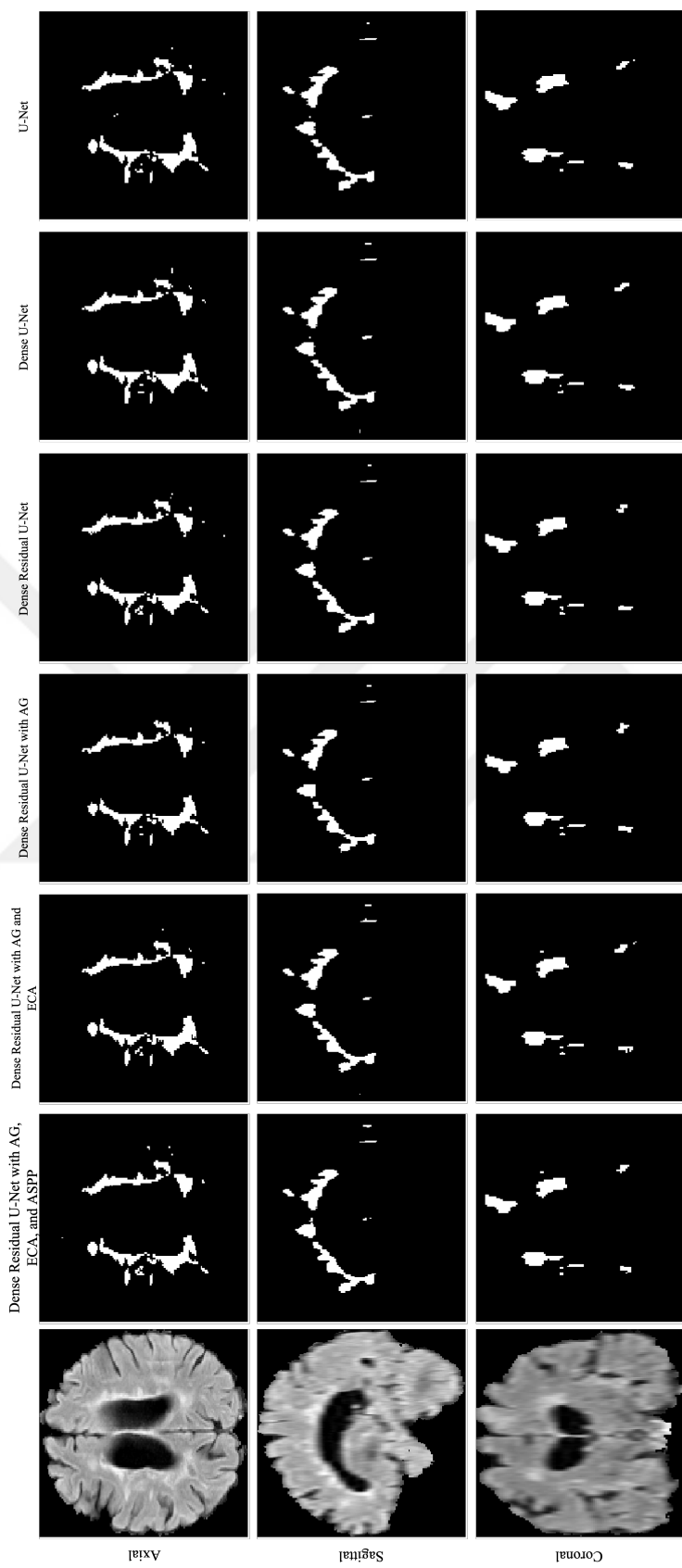


Figure 4.3 : The segmentation results of each variant of the proposed model on the ISBI2015 testing set. A slice of axial, sagittal, and coronal views of FLAIR from timepoint 1 of test subject 9 [25].

In addition to the ablation study, Table 4.2 summarizes the comparison of the proposed and other state-of-the-art approaches for MS lesion segmentation on the ISBI2015 testing dataset. The results of the recently proposed methods can also be accessed through the ISBI official website¹. The ISBI score, Dice score, PPV, LTPR, LFPR, and VD metric results were exploited for the comparison of models. Accordingly, our method achieved a competitive performance among all other methods, especially the Dice score and LTPR. The proposed method achieved the best Dice score by improving the second-best score by 3.43% (from 0.6451 to 0.6688). In terms of LTPR, the proposed method achieved the best score, while the second-best score was achieved by Zhang et al. [84]. As such, the LTPR score was improved by an average of 13.77% from 0.5330 to 0.6064 compared to the second-best score. The best PPV and LFPR of 0.9207 and 0.0866 were achieved by Hashemi et al. [81] while the second-best PPV and LFPR of 0.9140 and 0.1220 were achieved by Zhang et al. [84]. The best VD of 0.3653 was obtained by Maier and Handels [120] while the proposed method achieved the second-best VD of 0.3882. In terms of the ISBI score, the best and second-best performances in the overall scores which are 93.32 and 93.21 were achieved by Zhang et al. [84] and Zhang et al. [18], respectively.

¹Results: <https://smart-stats-tools.org/top25>

Table 4.2 : Comparison of the result of this study with other state-of-the-art methods published results for the ISBI2015 dataset. To be ranked first and second, the ISBI score, Dice score, PPV, and LTPR are expected to have high numerical values, while the LFPR and VD are expected to have low numerical values.

Model	ISBI Score	Dice Score	PPV	LTPR	LFPR	VD
Zhang et al. [84]	93.32	0.6390	0.9140	0.5330	0.1220	-
Zhang et al. [18]	<u>93.21</u>	0.6430	0.9080	0.5200	0.1240	0.4200
Kamraoui et al. [19]	92.85	0.6451	0.8907	0.5022	0.1470	0.4225
Proposed Model	92.75	0.6688	0.8650	0.6064	0.2617	<u>0.3882</u>
Kang et al. [83]	92.60	0.6345	0.8682	0.4787	0.1299	-
Hashemi et al. [81]	92.48	0.5841	0.9207	0.4135	0.0866	0.4972
Aslani et al. [17]	92.12	0.6114	0.8992	0.4103	0.1393	0.4537
Maier and Handels [120]	90.28	0.6050	0.7746	0.3672	0.2657	0.3653
Aslani et al. [82]	89.85	0.4864	0.7402	0.3034	0.1708	0.4768

¹ This table is sorted in descending order of the ISBI score.

² All metric scores are the averages of the two raters.

³ Bold and underlined values indicate the highest and second-highest scores among models, respectively.

4.2 MSSEG2016 Dataset Results

Table 4.3 summarizes the performance of the proposed model [26] on the MSSEG2016 testing dataset for each center. This dataset was gathered from four different centers to evaluate the generalization and robustness of the model for unseen centers. Hence, the training set contains data from three centers, excluding the center03. Although the testing data consists of 38 patients, one of the cases is an outlier due to no visible lesions delineated by five experts out of seven. Therefore, this outlier case (Center07-Patient08) was excluded while measuring all metrics used in this study for experts and our model. The proposed model produced the highest Dice score on the testing data of center01. In terms of F1 score and sensitivity, the proposed model achieved the best scores of 0.6556 and 0.7793 on the Center07 data, respectively. The best PPV was obtained with a mean of 0.7315 on the Center01 data. In terms of specificity, the proposed model achieved a similar score for all centers around 0.9998.

Table 4.3 : Results comparison of the proposed model for each center on the MSSEG2016 test set. To be ranked first, all metrics are expected to have high numerical values.

Centers	Dice Score	F1 Score	PPV	Sensitivity	Specificity
Center01	0.7069	0.6003	0.7315	0.6952	0.9998
Center03 (Unseen center)	0.6360	0.4913	0.5674	0.7683	0.9997
Center07	0.6916	0.6556	0.6663	0.7793	0.9997
Center08	0.6564	0.6059	0.6421	0.7334	0.9998
Average	0.6727	0.5883	0.6519	0.7440	0.9997

¹ This table is ordered according to the center name.

² The bold value indicates the highest score among centers.

Table 4.4 summarizes the comparison of the proposed model and the experts on the MSSEG2016 testing dataset. The proposed model performed similarly to the segmentation output of experts and even better than the results of some experts. As such, the average Dice score of the proposed model was obtained as 0.6727 where expert3, expert6, and expert7 produced a mean Dice score of 0.6724, 0.6717, and 0.6690, respectively. In terms of PPV, sensitivity, and specificity, our model produced a mean score of 0.6519, 0.7440, and 0.9997, respectively. According to the results of expert3, expert6, and expert7, their sensitivity scores were obtained as a mean of 0.7206, 0.6136, and 0.6867, respectively. In terms of specificity, the proposed model

and experts achieved a similar score. Figure 4.4 presents the qualitative results of our proposed method for the MSSEG2016 testing set. Accordingly, the proposed method achieves good performance in detecting MS lesions. Additionally, it is observed that the proposed model can detect more MS lesions in the data of all centers compared to the outputs of some experts.

Table 4.4 : Comparison of the results of this study with the manual delineations of experts for the MSSEG2016 testing set. All metrics are expected to have high numerical values to be considered the best and second-best scores.

Model	Dice Score	F1 Score	PPV	Sensitivity	Specificity
Expert5	0.7819	0.8928	<u>0.7359</u>	0.8518	<u>0.9997</u>
Expert4	<u>0.7590</u>	<u>0.8619</u>	0.6837	<u>0.8677</u>	0.9996
Expert1	0.7428	0.8509	0.7046	0.8090	<u>0.9997</u>
Expert2	0.6961	0.8141	0.5912	0.8736	0.9994
Proposed Model	0.6727	0.5883	0.6519	0.7440	<u>0.9997</u>
Expert3	0.6724	0.6782	0.6575	0.7206	<u>0.9997</u>
Expert6	0.6717	0.7544	0.8051	0.6136	0.9998
Expert7	0.6690	0.6561	0.6980	0.6867	<u>0.9997</u>

¹ The results of the experts can be accessed at <http://doi.org/10.5281/zenodo.1307653>.

² This table is sorted in descending order of the Dice score.

³ Bold and underlined values indicate the best and second-best scores among the proposed method and the experts, respectively.

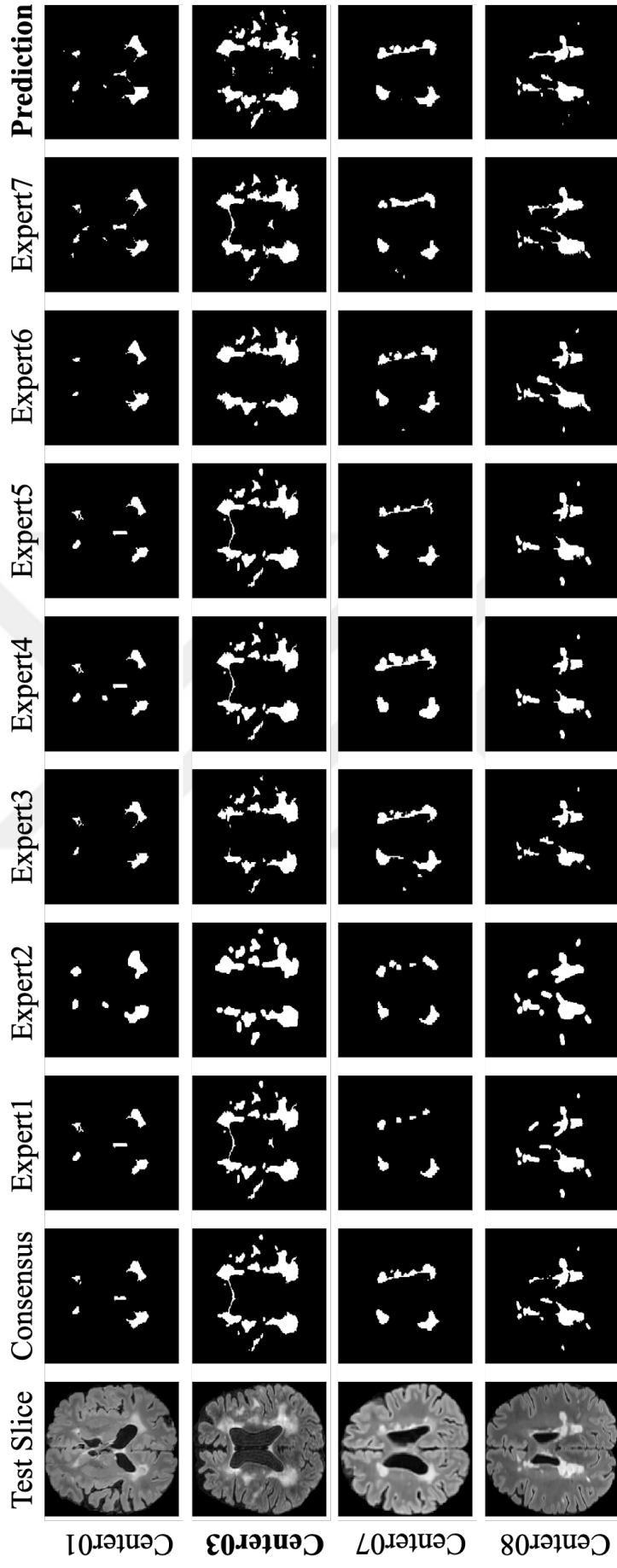


Figure 4.4 : The segmentation results of the proposed model on the MSSEG2016 testing set compared to the manual delineation of experts and their consensus ground truth. A slice of axial, sagittal, and coronal views of FLAIR from each center is demonstrated [25].

Figure 4.5 demonstrates the barplot of Dice score, PPV, sensitivity, and specificity for our model and the experts who delineated the structures, comparing visually as well. In addition, Figure 4.6 shows the boxplot of the Dice score, F1, PPV, and sensitivity of each patient on the testing set for experts and our model, and also shows the statistical significance test.

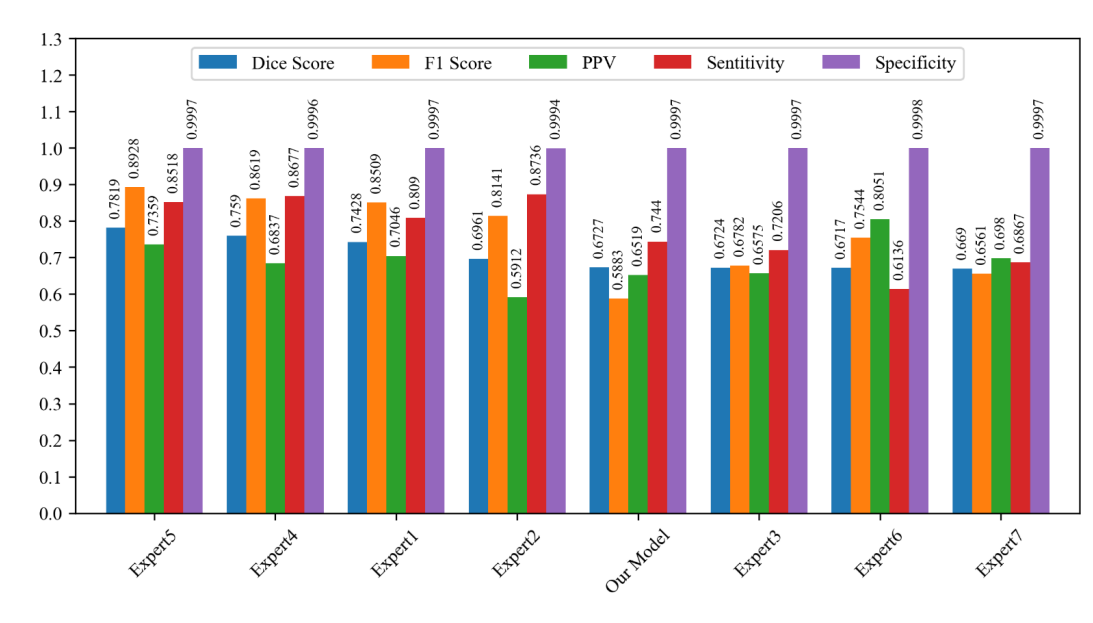


Figure 4.5 : Evaluation metric results on the MSSEG2016 testing dataset for our model and experts. The figure is sorted in descending order of the Dice score [25].

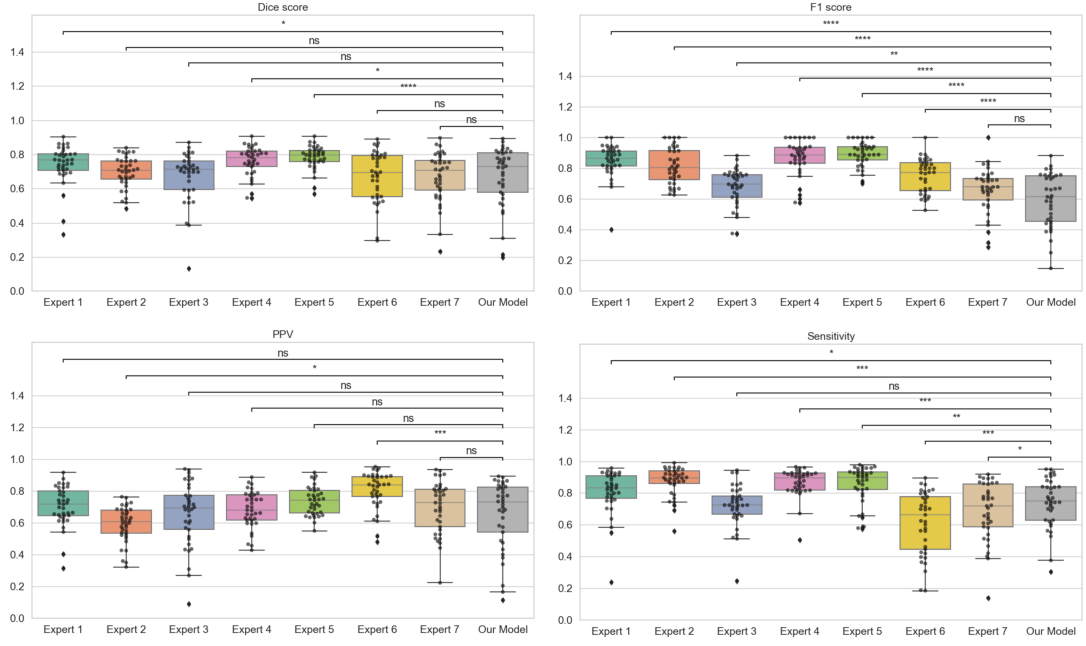


Figure 4.6 : The boxplot of Dice score, F1 score, PPV, and sensitivity evaluation metrics of each patient for experts and the proposed model on the MSSEG2016 testing set. Asterisks and ns indicate statistical significance (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$, and ns (non-significance)) when using a Wilcoxon test compared to the proposed model [25].

4.3 Cross-dataset Validation Results

The cross-dataset robustness and generalization ability of the proposed approach were assessed using the segmentation outputs produced by the proposed model [25] on different datasets. First, the ISBI2015 dataset was trained and tested on the MSSEG2016. Then, the MSSEG2016 dataset was trained and tested on the ISBI2015 via the online challenge website by submitting 3D binary output masks. Table 4.5 summarizes the performance of each dataset in terms of Dice score, PPV, and sensitivity, which are common in both dataset evaluation metrics. Accordingly, the prediction results of the MSSEG2016 on the ISBI2015 testing set achieved a mean Dice score of 0.4819, a mean PPV of 0.9450, and a mean sensitivity of 0.3540, respectively. On the other hand, the prediction results of the ISBI2015 on the MSSEG2016 testing set obtained a mean Dice score of 0.6031, a mean PPV of 0.7011, and a mean sensitivity of 0.5797, respectively.

Table 4.5 : Results comparison of cross-dataset, namely, training on the ISBI2015 and testing on the MSSEG2016 or vice versa. Dice score, PPV, and sensitivity are expected to have high numerical values.

Trained on	Tested on	Dice Score	PPV	Sensitivity
ISBI2015	MSSEG2016	0.6031	0.7011	0.5797
MSSEG2016	ISBI2015	0.4819	0.9450	0.3540

Additionally, the segmentation results of the MSSEG2016 on the ISBI2015 testing set were compared with other methods in the literature. The results were obtained from the challenge website to make a fair comparison with the others. Table 4.6 shows the numerical details of the results compared with the previously proposed two methods. The results of these two methods were obtained from the study of Kamraoui et al. [19]. For the proposed model, the evaluation metric results of ISBI score, PPV, LTPR, LFPR, and Cor were obtained 91.84, 0.4819, 0.9450, 0.1604, and 0.8398, respectively. As a result, the proposed model obtained the highest scores in terms of the ISBI score, PPV, LTPR, and Cor. As for Dice score and LFPR, the highest Dice score of 0.5350 was achieved by Kamraoui et al. [19] while the best LFPR of 0.0750 was obtained by Zhang et al. [18].

Table 4.6 : Results comparison with other methods when trained on the MSSEG2016 dataset and tested on the ISBI2015 testing set. To be ranked first and second, the ISBI score, Dice score, PPV, LTPR, and Cor are expected to have high numerical values, while the LFPR is expected to have low numerical values.

Model	ISBI Score	Dice Score	PPV	LTPR	LFPR	Cor
Proposed Model	91.84	<u>0.4819</u>	0.9450	0.4493	0.1604	0.8398
Kamraoui et al. [19]	<u>89.04</u>	0.5350	0.6970	<u>0.3730</u>	0.3530	<u>0.8350</u>
Zhang et al. [18]	86.69	0.1650	<u>0.9370</u>	0.1600	0.0750	0.2120

¹ This table is sorted in descending order of the ISBI score.

² The results of compared methods were taken from the study of Kamraoui et al. [19].

³ Bold and underlined values indicate the best and second-best scores, respectively.

4.4 MSSEG-2 Dataset Results

The MSSEG-2 challenge aims to segment and detect new MS lesions by comparing the baseline and the follow-up 3D FLAIR images of a patient. 24 teams with a total number of 30 pipelines participated in this challenge. Deep learning approaches, most of them relying on the U-Net architecture, were proposed by most of the participants,

while only one of the teams used a conventional statistical method and the subtraction between two MR images [121]. Table 4.7 shows the average quantitative metric results of some of the methods presented in the challenge, including the results of the experts².



²Quantitative metrics for all proposed pipelines are available at <https://doi.org/10.5281/zenodo.5775523>

Table 4.7 : Prediction results of evaluating the challenge test data set published on the challenge website. Bold and italic values are the highest and the second-best scores among other proposed methods excluding the experts, respectively. Dice and F1 Score are expected to be a high numerical value while the Number of and Volume of Lesions are expected to be a low numerical value. The Number of and Volume of Lesions metrics are calculated for no new lesion cases.

Methods	F1 Score	Dice Score	Number of Tested Lesions	Volume of Tested Lesions(mm ³)
Expert 1	0.712	0.631	0.036	1.453
Expert 3	0.636	0.598	0.000	0.000
Expert 2	0.607	0.536	0.107	3.981
Mediaire-B*	0.541	0.437	0.536	29.235
Empenn	0.532	0.424	0.286	4.258
Mediaire-A	0.525	0.432	0.429	15.908
Expert 4	0.524	0.461	0.036	0.623
LaBRI-IQDA	0.517	0.500	1.143	38.486
SNAC	0.514	0.485	0.321	5.726
MedICL	0.500	0.507	0.536	12.713
LaBRI-D&E	0.498	0.472	1.964	177.131
ITU (Proposed)	0.480	0.443	0.148	1.488
New Brain	0.477	0.451	0.786	12.371
LYLE	0.441	0.409	0.036	0.470
SCAN	0.433	0.403	0.071	5.373
Neuropoly-2	0.410	0.409	0.107	0.498
SCA-withPriors	0.216	0.224	2.464	302.121
IBBM ⁺	0.143	0.155	3.786	123.309

¹ The source data can be accessed at <https://doi.org/10.5281/zenodo.5775523>

² This table is ordered highest to lowest based on the F1 score.

³ * and + indicates the first and last ranks among the participants, respectively.

Four metrics were used to evaluate the proposed pipelines for new MS lesion segmentation and detection. The test data set consists of MR images of 60 patients and 32 of them were used for the calculation of the F1 and dice scores due to possessing new lesions at their follow-up images. The remaining patients' data were used for the calculation of the number of tested lesions and volume of tested lesions. According to the challenge results, the proposed pipeline in this study was ranked as the 8th for F1 and Dice scores among the proposed methods. The proposed pipeline produced a mean score of 48.00% for the F1 score and a mean score of 44.30% for the dice score. For the no-lesion cases, this pipeline was ranked in 5th and 4th places with a mean score of 0.148 and 1.488, respectively for the number of tested and volume of tested lesions. Also, the highest F1 and Dice scores including the expert raters were a mean score of 71.20% and 63.10% respectively, which belonged to expert 1. As for the number of tested and volume of tested lesions, the highest score was 0 which belonged to expert 3. On the other hand, the highest F1 and dice scores for the automated methods belonged to team Mediaire-B and MedICL with a mean score of 54.10% and 50.70%, respectively. The highest score for the number of tested lesions and volume of tested lesions belonged to team LYLE with a mean score of 0.036 and 0.498, respectively. The lowest F1 and dice scores, belonging to the team IBBM, had a mean score of 14.30% and 15.50%, respectively. Figure 4.7 shows the segmentation performance of the proposed model, consensus, and experts on a slice of an axial view of four patients. As seen in the figure, the proposed model had competitive performance compared to the segmentation output of experts.

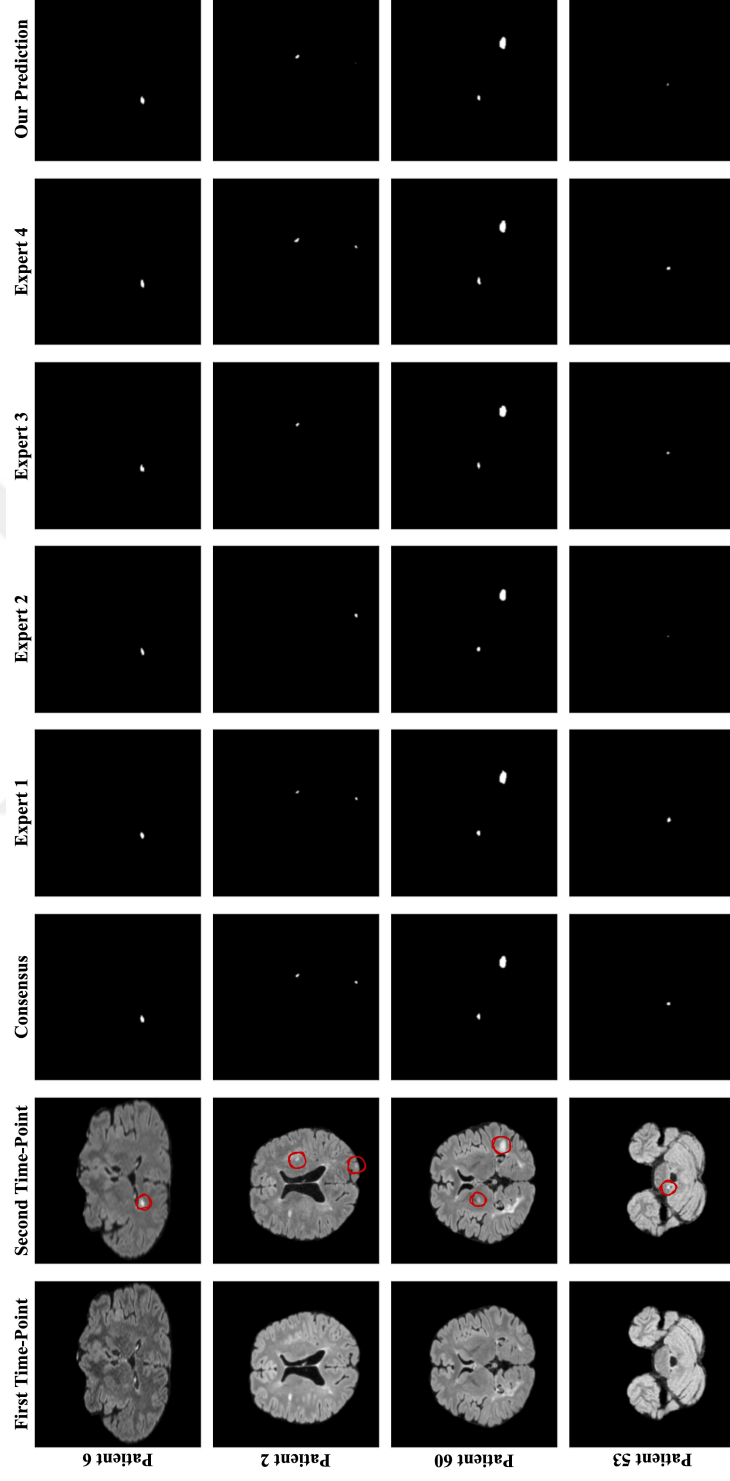


Figure 4.7 : The best and worst performances of the proposed model compared to the consensus and each expert segmentation for F1 and dice scores. A slice of axial view from patients 6 and 2 for the F1 score and patients 60 and 53 for the dice score is presented [26].

The challenge also provides additional metrics discussed in subsection 3.2.8 for a complete evaluation although these metrics were not considered for the ranking. The results obtained from some of the proposed methods and experts for additional metrics are given in Table 4.8. Accordingly, the results of the proposed pipeline with respect to sensitivity, specificity, PPV, and surface distance were a mean score of 0.364, 1.000, 0.675, and 8.548, respectively. This pipeline had competitive performance compared to experts and other proposed pipelines in some of these metrics. For example, the highest PPV score among experts and proposed methods were a mean of 0.813 and 0.703 for expert 1 and the team LYLE, respectively. Also, the highest score for surface distance belonged to expert 2 and the team LYLE with a mean score of 4.543 and 7.210.

Table 4.8 : Prediction results of evaluating the challenge test data set published on the challenge website for other useful metrics. Bold and italic values are the highest and the second-best scores among some of the proposed methods and the experts, respectively. Sensitivity, Specificity, and PPV are expected to be a high numerical value while Surface Distance is expected to be a low numerical value.

Methods	Sensitivity	Specificity	PPV	Surface Distance
Expert 1	0.650	1.000	0.707	<i>5.907</i>
Mediaire-B	<i>0.616</i>	1.000	0.394	8.803
Expert 3	0.589	1.000	0.760	5.990
Expert 2	0.526	1.000	0.813	4.543
MedICL	0.514	1.000	0.556	9.194
Expert 4	0.407	1.000	<i>0.801</i>	7.885
Proposed Model	0.364	1.000	0.675	8.548
LYLE	0.344	1.000	0.703	7.210
SCAN	0.340	1.000	0.678	8.521
IBBM	0.170	1.000	0.242	24.102

¹ The source data can be accessed at <https://doi.org/10.5281/zenodo.5775523>

² This table is ordered according to the highest to the lowest based on the sensitivity score.

Finally, comparisons between U-Net, U-Net with AGs, U-Net with RUs, U-Net with RUs and AGs (two types) were realized for the new MS lesion segmentation. The results of U-Net, U-Net + AGs, U-Net + RUs, and U-Net + RUs + AGs are presented in Table 4.9. As seen in this table, the proposed model achieved the highest dice and F1 scores, a mean score of 58.70% and 61.10%, respectively. U-Net + RUs achieved the highest PPV score, a mean score of 62.40%. Furthermore, this network had fewer training parameters and performed better compared to the U-Net architecture.

Table 4.9 : The evaluation results of the proposed method with different components using a subset of the MSSEG-2 test dataset.

Methods	Dice	F1 Score	PPV	Total Parameters
U-Net + RUs + AGs	0.587	0.611	0.567	4,934,613
U-Net + RUs	0.551	0.441	0.624	4,722,897
U-Net + AGs	0.505	0.592	0.609	7,947,109
U-Net	0.558	0.490	0.467	7,771,585

4.5 Discussions

In this thesis, an automated pipeline for MS lesion segmentation was designed using 3D MRI data of FLAIR, T1-2, and T2-w sequences. The proposed model [25] within this pipeline was developed by modifying the U-Net architecture. First, the network was modified by adding BN, SD, ELU, strided convolutions for the pooling operations, and transposed convolutions for the upsampling operations. Then, dense connections, residual blocks, AG, ECA, and ASSP were incorporated to improve overall MS lesion segmentation performance measured by several common metrics, such as the Dice score and LTPR. This model achieved advantages through the modification of the encoder-decoder network together with modified components, such as replacing the bottleneck of the network with atrous convolutions with different dilated rates to extract multi-scale contextual information from the given feature map. In the ablation study (see Table 4.1 and Figure 4.1), the proposed model, which contains all components, outperformed its variants. Additionally, this model outperformed other proposed methods in the literature in terms of Dice score and LTPR according to the results obtained from ISBI2015 as given in Table 4.2. Moreover, this model outperformed the results of some experts for the MSSEG2016 testing data, especially for the Dice score, PPV, sensitivity, and specificity as presented in Table 4.4.

Although a lower number of modalities might be sufficient to evaluate MS lesions, automated MS lesion segmentation should consider using different modalities such as T1-w and T2-w to improve the segmentation performance. These modalities will provide additional information, such as the location and shape of MS lesions. All MS benchmark datasets provide different modalities for better segmentation. The proposed model achieved a competitive performance using three modalities compared to the state-of-the-art methods published recently, as given in Table 4.2, even on

unseen datasets. According to the results of the ISBI2015 testing set, the proposed model improved the Dice score by an average of 3.43% and the LTPR by an average of 13.77% compared to the second-best scores. Additionally, data augmentation strategies are realized to make a robust model, improve prediction accuracy, and reduce overfitting during the training. Therefore, data augmentation methods as discussed previously rather than using simple strategies (rotating and flipping) were applied to improve the performance, generalization, and robustness of the proposed model.

In this study, the MS lesion segmentation performance was improved due to the use of different components in the encoder-decoder network. Modifications to the U-Net and other components, which were used to build the proposed model, obtained competitive performance in most of the evaluation metrics as presented in Table 4.1, Table 4.2, and Table 4.4. Besides, a whole-brain slice-based approach was exploited as patch-based CNNs suffer from spatial information about MS lesions due to patch size limitations [17,84]. The results indicated that the proposed approach using whole brain slices had a competitive performance for most measures, such as Dice score and LTPR, compared to other methods and experts, as presented in Table 4.2 and Table 4.4. Especially, the results obtained from the MSSEG2016 testing set have shown that the proposed model outperformed the manual delineation of some experts (see Figure 4.5). According to the results of the MSSEG2016 testing set, the proposed model improved the Dice score by 0.5%, the PPV score by 10.26%, and the sensitivity score by 21.25% compared to the results of experts who achieved lower scores among the evaluation metrics. Furthermore, Figure 4.4 exhibits visually detected lesions, as such the proposed model could be able to detect better than the results of some experts, even on an unseen center. We observed that transferring weights obtained from the ISBI2015 training set is an effective approach for training the proposed model on the MSSEG2016 training set. Indeed, this approach improved MS lesion segmentation performance on the MSSEG2016 testing set, as given in Table 4.4. Another observation was that MS lesion segmentation performance improved significantly when corresponding axial, sagittal, and coronal slices, which were obtained from three modalities, were stacked into the input channel, respectively. Although Aslani et al. [17] emphasized that stacking corresponding modalities together into the three-channel input is not an optimal solution, the proposed model and stacking corresponding modalities into the

input channel dimension enhanced the segmentation performance compared with their results presented in Table 4.2. This network exploited the contextual information in all plane directions and obtained useful features of the location and shape of the lesions with channel-wise stacking.

The Dice score between two raters who delineated the same structures is relatively low compared to automated methods evaluated on the ISBI2015 testing set. The average Dice score across raters is approximately 73%, and an ISBI score of 90 or higher can be considered a comparable score to the human rater mentioned in [79]. According to the results obtained from the official ISBI test set, the proposed model and its variants produced a score over 90, as presented in Table 4.1. Moreover, this model and its variants achieved a competitive segmentation performance compared to the results across raters for other metrics, such as a mean Dice score of 0.6688, a mean LTPR of 0.6064, and a mean LFPR of 0.2617, as presented in Table 4.1. Indeed, rater#1 produced a mean LTPR of 0.6450 and a mean LFPR of 0.1740 on the ground truth of rater#2, while rater#2 produced a mean LTPR of 0.8260 and a mean LFPR of 0.3550 on the ground truth of rater#1 [79].

MS appearance can vary significantly based on the manufacturer and imaging protocol. Aggregating additional MRI data would generalize the DL models for better segmentation outputs in clinical setups since DL models would be better trained with a large number of patients' data [15]. Although the segmentation of a 3D scan requires several steps, such as extracting slices as 2D, processing these slices individually, and reconstructing a 3D binary output mask, 2D CNNs using 3D MRI data are still achieving state-of-the-art results compared to 3D-based CNN methods. According to the results of this study, it is observed that processing 3D MRI data by converting it into 2D slices still outperformed others, especially in terms of Dice score and LTPR [19, 83, 84]. According to Altay et al. [122], when clinical raters with different levels of experience assessed MS lesions on an MS dataset for no more than 10 minutes per a study, notable variability among these raters in counting lesions manually was observed. Hence, training the proposed model with a larger dataset would reduce the variability of experts, and even decrease the lesion detection time and effort since processing a 3D scan takes around 35 seconds on a mid-range GPU in our study. This is faster than the approach of Aslani et al. [17], which segments the

input image in approximately 90 seconds. Overall, this study has provided clues about recent techniques regarding accuracy that can be used to guide future research about MS lesion segmentation using multiple datasets.

A deep learning model [26] was developed to handle the problem of identifying new MS lesions using the baseline and the follow-up 3D FLAIR MR images. Activity segmentation particularly for new lesions is a more challenging task compared to lesion segmentation in a single-time MR scan due to small lesion loads. MS lesion segmentation using traditional and deep learning approaches has usually been studied in a single MRI scan in recent years. However, deep learning approaches for MS lesion activity using the baseline and follow-up MR images still remain limited. In most of these studies, the researchers have been using their own datasets making it difficult to compare and reproduce their results with the proposed pipeline. Thus, in this study, comparisons were performed on the automated methods proposed in the challenge. Moreover, comparisons were performed among components used for building the designed network as well. The proposed network, which combines the strengths of U-Net, residual units, and attention gates, has outperformed other methods comprising different combinations of components in terms of dice and F1 scores.

A whole-brain slice-based approach was used as patch-based CNNs lack spatial information about MS lesions due to the patch size limitation [17]. The results indicated that the proposed pipeline with this approach had competitive performance for most measures compared to the other pipelines, as given in Table 4.7. The segmentation performance of new MS lesions improved significantly when baseline and follow-up MRI scans were stacked in the input channel dimension. Thus, baseline and follow-up scans for each patient were stacked as a two-channel input for the proposed pipeline. Furthermore, AGs modified with BN and ReLU allowed the model to focus on small and subtle new lesions.

Figure 4.8 presents the analysis of differences in detection and segmentation for F1 and dice scores for each expert and each team that participated in the MSSEG-2 challenge, respectively. The red box highlights the team performance of this study for these two metrics. According to F1 and dice scores, proposed methods could not reach the expert level; however, some methods were able to outperform experts who revealed varying

scores in different patients³. Based on this observation, it was concluded that the detection and segmentation of MS new lesions in longitudinal studies is a difficult task, even for experts. Therefore, an external reviewer may be needed while analyzing the new lesions with automated methods for lesion activity.

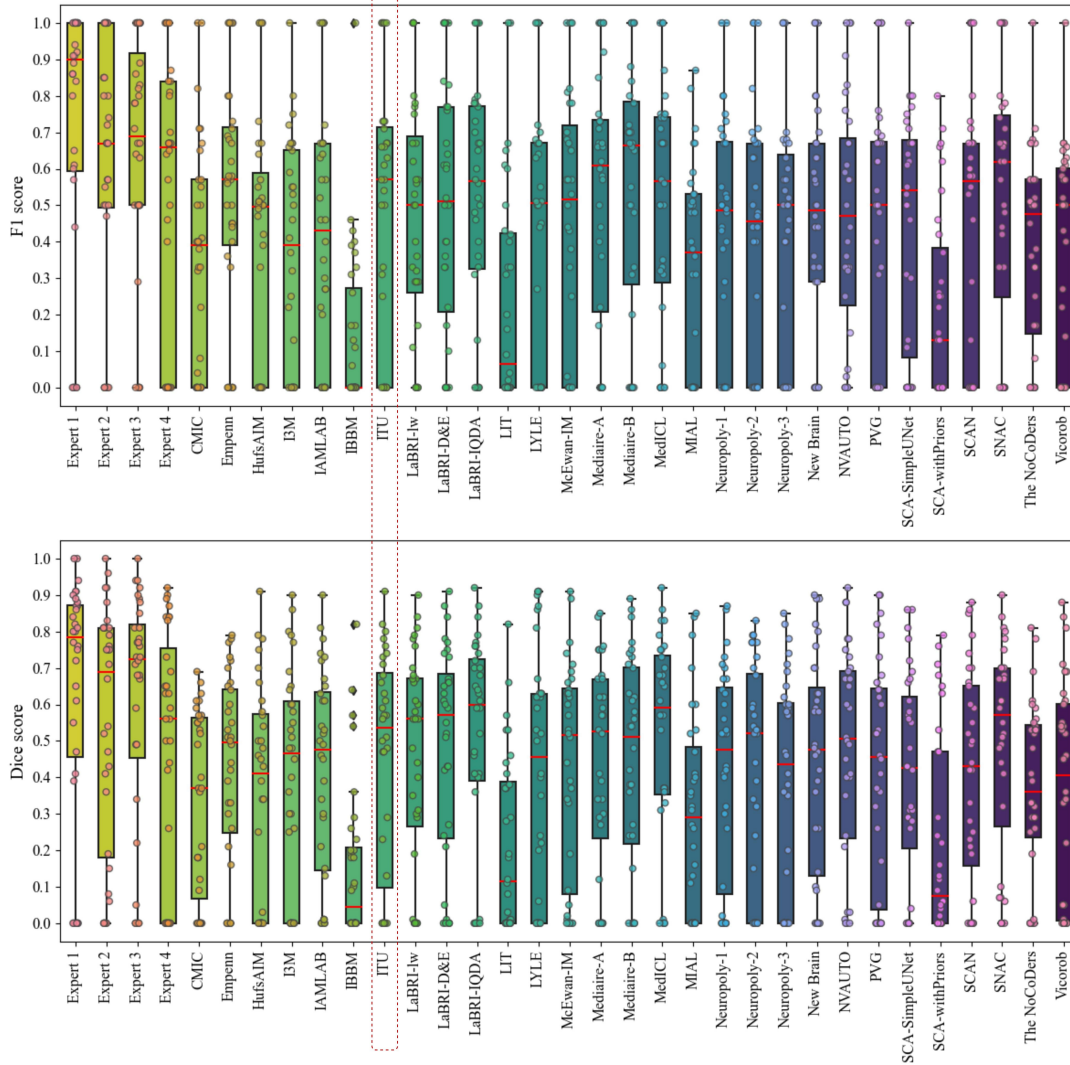


Figure 4.8 : Analysis of differences in detection and segmentation by using F1 and dice scores for each expert and each team that participated in the challenge, respectively [26].

The evaluation metrics for no new lesions are indicated in Figure 4.9. The number of connected components in automatic segmentation was used to find the number of lesions detected. Also, the volume of lesions detected (mm^3) was used to evaluate

³Evaluation results and analysis slides at https://files.inria.fr/empenn/msseg-2/Challenge_Day_MSSEG2_Results_2021.pdf

the segmentation performance of both automated and expert delineation outputs. As seen in Figure 4.9 and Table 4.7, the proposed pipeline outperformed compared to some of the other proposed methods. The dotted red rectangle highlights the proposed pipeline within this study. Accordingly, some of the proposed methods, including ours, outperformed those of some experts.

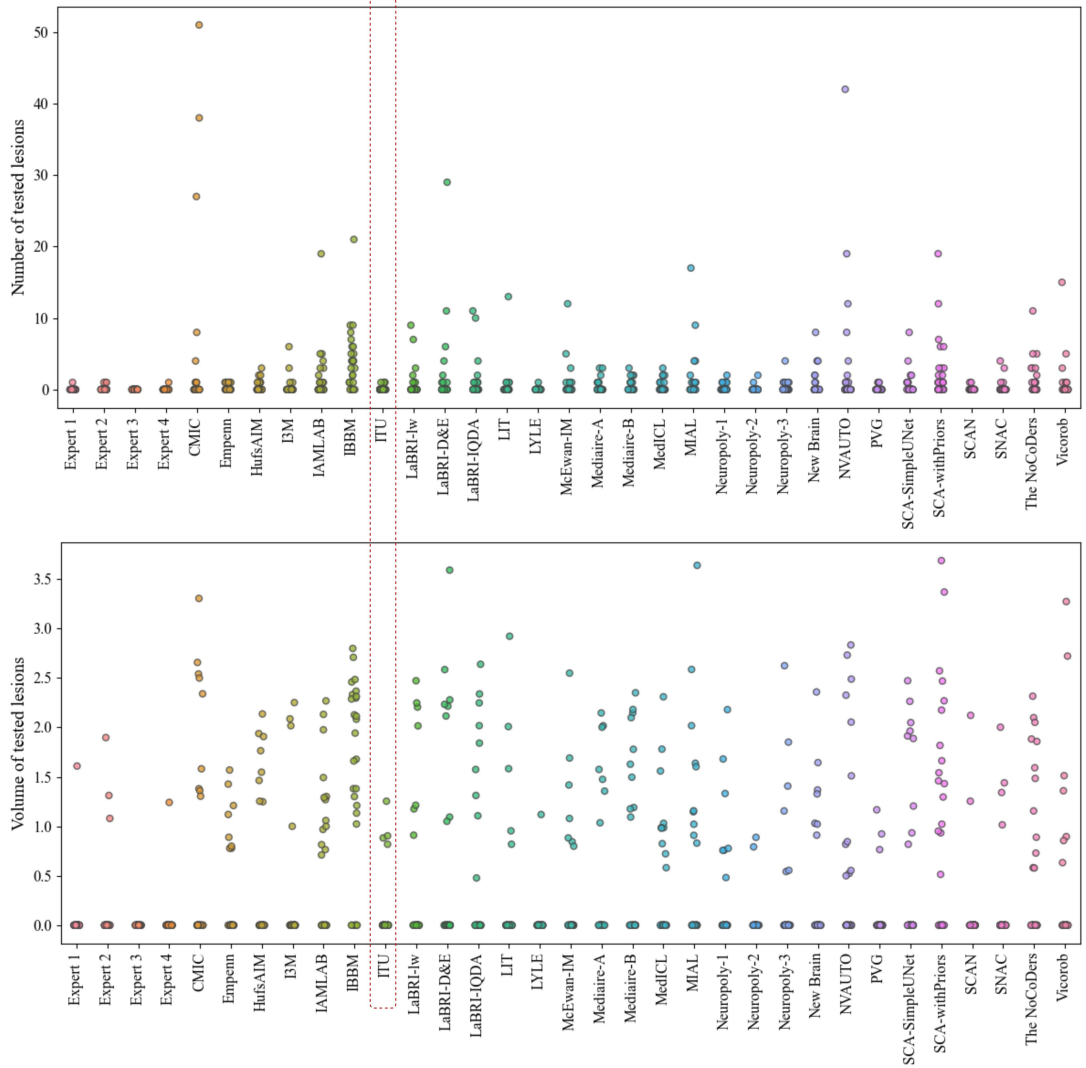


Figure 4.9 : Analysis of the number and volume of lesions detection for each expert and each team that participated in the challenge (The data of volume of tested lesions was scaled by $\log_{10}$) [26].

Instead of using a 3D segmentation approach requiring more computational power and learning parameters, the proposed method and the slice-based approach were used together for detecting and segmenting new lesions on the follow-up images. While the appearance of new lesions is of primary interest for the challenge, enlarged or disappearance of MS lesions could also be studied. Different MRI modalities, such as T1-w and T2-w, can also be incorporated into the given task to extract more features

related to the size or location of new MS lesions, even though the FLAIR images reveal lesions as more intense. To achieve a robust automated model for the given task, large datasets from different scanners are needed; however, it is difficult to obtain such datasets.

Finally, the results of the research questions have been evaluated as follows:

Q1. What is the performance of DL algorithms in segmenting MS lesions?

A1. In this study, a deep learning model based on the U-Net architecture was proposed to enhance the performance of MS lesion segmentation. Accordingly, several experiments were conducted on how DL approaches could perform better for this task. According to the results obtained from the ISBI2015 and MSSEG2016 datasets, the proposed model outperformed the results of experts who performed manual delineation for MS lesions. Table 4.1 shows the results of the proposed model and its variants. An ISBI score of 90 or higher can be considered a comparable score to the human rater mentioned in [79]. As such, DL-based models produced an ISBI score higher than 90.

Q2. How can deep learning algorithms be optimized for better MS lesion segmentation?

A2. Several strategies can be used to optimize DL algorithms for better MS lesion segmentation, such as network architecture, hyperparameter tuning, data augmentation, loss function, and using additional MRI sequences. The choice of DL architecture can significantly impact performance. CNNs have been used for this segmentation task in recent years. For example, U-Net, which is a fully CNN, has been widely used and achieved remarkable results in various medical imaging tasks, such as semantic segmentation of organs, tissues, or lesions. Therefore, the proposed models are based on the U-Net architecture to improve the segmentation performance of MS lesions. This architecture is detailed in section 3.2.1 and its modifications for MS lesion segmentation on both a single time point MRI of a patient and two different time points MRI of a patient are presented in sections 3.3 and 3.4. Data augmentation techniques were heavily used to prevent overfitting and realize a robust model. Additionally, a hybrid loss function, which is the addition

of focal and Dice losses, was employed to overcome the class imbalance problem in MS lesion segmentation. Multiple MRI sequences, such as FLAIR, T1-w, and T2-w, have been used jointly to obtain more features related to MS lesions, such as shape and location. According to the results of these proposed models given in chapter 4, MS lesion segmentation can be improved based on these strategies.

Q3. What is the impact of imaging sequences on the accuracy of DL algorithms in MS lesion segmentation?

A3. Imaging sequences can have a significant impact on the accuracy of DL algorithms in MS lesion segmentation. The use of multiple imaging sequences can also be incorporated to improve the accuracy of DL algorithms in MS lesion segmentation. For this purpose, multiple MRI sequences, such as FLAIR, T1-w, and T2-w, have been used jointly to obtain more features related to MS lesions since they have different signal properties and can provide different information. According to the results presented in the section of results and discussions, using these sequences jointly performed better MS lesion segmentation. For example, the study of Kamraoui et al. [19], in which the results are given in Table 4.2, used T1-w and FLAIR sequences; however, this study performed better in terms of several evaluation metrics. For the MSSEG-2 dataset, it is concluded that different MRI sequences, such as T1-w and T2-w, can be incorporated to perform better segmentation due to extracting more features related to the size or location of new MS lesions given in section 4.5.

Q4. What is the effect of data augmentation techniques on the performance of DL algorithms in MS lesion segmentation?

A4. Data augmentation is used to improve the generalization of a model and reduce overfitting by applying various transformations to the images, such as random rotations, translations, flipping, scaling, and adding random noise. In this study, the data augmentation techniques used are explained in section 3.1.4.

Q5. How can transfer learning be applied to enhance deep learning algorithms in MS lesion segmentation?

A5. Transfer learning can be applied to enhance DL models for MS lesion segmentation as it allows the DL model to utilize the knowledge acquired from a pre-training task, reducing the amount of data needed for training and improving performance on the MS lesion segmentation task. In this study, training weights obtained from the ISBI2015 dataset (same domain) were exploited to initialize training weights on the MSSEG2016 as a transfer learning strategy to improve the segmentation performance, generalization, and robustness of the model. According to the results obtained from the MSSEG2016 testing set given in Table 4.4, it has been observed that transferring weights from one dataset to another in the same domain has outperformed, especially on the MSSEG2016 dataset that contains different acquisition sites, resolutions, preprocessing, and clinical cases.

Q6. How do process 3D MRI scans with computational efficiency approaches?

A6. Processing the entire 3D MRI data can be a challenging task, due to the large volume of data involved. An alternative approach is to process the MRI data based on slices, which can be more computationally efficient because each slice can be processed individually. This method, however, has some drawbacks, such as the chance of information loss during the slicing process. It is also critical to ensure a careful reconstruction of the 3D MRI scan from the processed slices. In this study, the slice-based approach was further expanded to whole-brain segmentation by combining slices extracted from the three planes (axial, sagittal, and coronal), thereby enabling the acquisition of contextual information in all directions. Moreover, a 3D image reconstruction method was introduced using slices from each plane orientation presented in section 3.5, and a majority voting approach was applied to the 3D outputs obtained from each plane to form the final 3D segmentation result. The results showed that his approach achieved competitive performance for most measures, such as Dice score and LTPR, compared to other methods and the results of experts, as presented in Table 4.2 and Table 4.4.

Q7. What is the performance of DL algorithms in MS lesion activity segmentation?

A7. MS lesion activity segmentation is a challenging task due to the complex nature of MS lesions and the variability in their appearance across patients and over time. DL approaches, those relying on CNNs, have demonstrated their efficacy in the task

of MS lesion segmentation. This is due to the ability of CNNs to learn significant features and patterns from a large amount of delineated training data. Therefore, DL approaches can incorporate into the task of MS lesion activity segmentation to achieve better performance. In this study, the DL-based model was proposed to detect and segment new MS lesions appearing at the second time point detailed in section 3.2.1. This study shows that an encoder-decoder-based architecture, namely U-Net, showed acceptable results in detecting and segmenting lesion activity. Also, this proposed model improved the segmentation of lesion activity on MRI data acquired from different scanners. The results of the MSSEG-2 testing set, presented in Table 4.7 and Table 4.8, indicate that the model outperformed the results of some experts based on certain evaluation metrics.

Q8. How to handle class imbalance problem for MS lesion segmentation?

A8. Class imbalance is a common problem in medical image segmentation, particularly in the context of MS lesion segmentation, as lesions are often small, scarce, and only constitute a minority of the MRI volume. Therefore, a hybrid loss function, which is the addition of focal and Dice losses, was used as the training loss function to handle the class imbalance problem during the training procedure. The implementation details are given in section 3.5.



5. CONCLUSIONS AND RECOMMENDATIONS

In this thesis, a novel dense residual U-Net that combines modified AG, ECA, and ASPP modules was proposed to improve the segmentation of MS lesions. Two publicly available datasets, namely, the ISBI2015 and MSSEG2016, were employed to validate the proposed model's segmentation performance, generalization ability, and robustness. 2D axial, sagittal, and coronal slices extracted from 3D volumetric scans of FLAIR, T1-w, and T2-w sequences were exploited jointly to obtain contextual information in all directions and complementary information related to MS lesions. 2D slices extracted from the corresponding orientation of each sequence were stacked to generate a 3-channel input feature map. Then, all the 2D stacked slices were aggregated to form the training input data. Additionally, for the ISBI2015 dataset, the manual delineations of two raters were concatenated to form a single training set, which enables an increase in the training samples.

Data augmentation methods of the Albumentations library were employed to make a robust model, improve prediction accuracy, and reduce overfitting during the training. While manual delineation of MS lesions is time-consuming, costly, and subject to variability across experts, DL methods, which learn features from their input data during the training period, can automatically assist in MS lesion segmentation and detection, reducing cost and variability. Thus, CNN-based deep learning methods were used in this study for accurate automatic MS lesion segmentation. Furthermore, whole-brain slice-based segmentation of MS lesions gave promising results according to most of the metrics in all experiments and comparisons.

The proposed model was also compared with some other proposed methods that employed these two datasets in their research. The results of the MSSEG2016 testing set were used to compare the proposed model's segmentation performance, generalization, and robustness according to the results of experts and each center. Additionally, the results of the ISBI2015 obtained from the official test set via the online website were exploited to make a fair comparison with other methods.

The results showed promising outputs in most of the metrics when compared to the evaluation metric results of the inter-rater and scores from other methods. Also, a quantitative evaluation comparing the consistency of the two datasets was presented. Finally, the ISBI score of 92.75 is considered comparable to a human rater, as discussed previously. As such, the proposed model can be used for different segmentation problems in image analysis.

Additionally, an automated pipeline for new MS lesion segmentation using the baseline and follow-up 3D FLAIR MRI has been designed with a deep learning-based network that fuses the strengths of U-Net, residual learning, and AG. For more accurate segmentation of new MS lesions, this network architecture was designed as a deep encoder-decoder network to enhance the U-Net by replacing plain blocks with residual blocks and adding AGs. These residual blocks replaced with the plain blocks facilitate the training. Skip connections within both residual units and U-Net facilitate the propagation of information in both forward and backward phases during the training procedure. AGs integrated into the proposed model emphasize important features propagated over skip connections.

A hybrid loss function was introduced as the addition of dice loss and $1 \times$ binary focal loss. The input data for the proposed method was prepared by converting 3D scans into their plane orientations of axial, sagittal, and coronal views which yielded 2D slices. Baseline and follow-up slices were stacked to create a two-channel feature mapping for each plane orientation. Then, all slices extracted from all three planes were grouped into a single input to increase training samples and use the contextual information in all directions. The predicted 2D slices for each view were aggregated using a majority voting to generate the final 3D binary output. Although new MS lesion segmentation and detection pose a difficult problem due to small lesion sizes, the proposed method has achieved comparable segmentation performance compared to the experts and top-ranked automated methods in the challenge.

Finding the appropriate data sets and using the existing ones as publicly available will reduce the gap for the data required in these studies and the lack of which is frequently discussed, and will allow different studies to be carried out. This study provides clues about the recent techniques regarding MS lesion activity segmentation that can be used as a guide for future research in this field.

One of the most important outputs of this thesis is to facilitate the diagnosis process performed by physicians who usually need to depict MR scans in 3D in their minds while delineating MS lesions. Since this study provides the exact location of lesions based on the axial, sagittal, and coronal slices, physicians will be able to quickly evaluate their findings using the outputs of this study. Additionally, training the proposed models with a large dataset will aid in the diagnosis of MS disease in its early stages in clinical settings. However, the output of this study is not mature enough to be used in clinical settings due to the lack of a large enough dataset and limited computational resources.

In future studies, different large datasets will be investigated and used to improve the segmentation, generalization ability, and robustness of the deep learning models and proposed models, although it is challenging to obtain such datasets. Also, the outcomes of this study will be integrated into digital twin technology to improve the diagnosis and treatment of disease.



REFERENCES

- [1] **Steinman, L.** (1996). Multiple sclerosis: a coordinated immunological attack against myelin in the central nervous system, *Cell*, 85(3), 299–302, [https://doi.org/10.1016/S0092-8674\(00\)81107-1](https://doi.org/10.1016/S0092-8674(00)81107-1).
- [2] **Goldenberg, M.M.** (2012). Multiple sclerosis review, *Pharmacy and therapeutics*, 37(3), 175.
- [3] **Baecher-Allan, C., Kaskow, B.J. and Weiner, H.L.** (2018). Multiple sclerosis: mechanisms and immunotherapy, *Neuron*, 97(4), 742–768, <https://doi.org/10.1016/j.neuron.2018.01.021>.
- [4] **Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J.C., Quiles, A., Valls, L., Ramió-Torrentà, L. and Rovira, À.** (2012). Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches, *Information Sciences*, 186(1), 164–185, <https://doi.org/10.1016/j.ins.2011.10.011>.
- [5] **Wattjes, M., Steenwijk, M. and Stangel, M.** (2015). MRI in the diagnosis and monitoring of multiple sclerosis: an update, *Clinical neuroradiology*, 25(2), 157–165, <https://doi.org/10.1007/s00062-015-0430-y>.
- [6] **Combès, B., Kerbrat, A., Pasquier, G., Commowick, O., Le Bon, B., Galassi, F., l’Hostis, P., El Graoui, N., Chouteau, R., Cordonnier, E. et al.** (2021). A clinically-compatible workflow for computer-aided assessment of brain disease activity in multiple sclerosis patients, *Frontiers in Medicine*, 8, 740248, <https://doi.org/10.3389/fmed.2021.740248>.
- [7] **Lladó, X., Ganiler, O., Oliver, A., Martí, R., Freixenet, J., Valls, L., Vilanova, J.C., Ramió-Torrentà, L. and Rovira, À.** (2012). Automated detection of multiple sclerosis lesions in serial brain MRI, *Neuroradiology*, 54(8), 787–807, <https://doi.org/10.1007/s00234-011-0992-6>.
- [8] **Roy, S., Butman, J.A., Reich, D.S., Calabresi, P.A. and Pham, D.L.** (2018). Multiple sclerosis lesion segmentation from brain MRI via fully convolutional neural networks, *arXiv preprint arXiv:1803.09172*.
- [9] **McFarland, H.F., Frank, J.A., Albert, P.S., Smith, M.E., Martin, R., Harris, J.O., Patronas, N., Maloni, H. and McFarlin, D.E.** (1992). Using gadolinium-enhanced magnetic resonance imaging lesions to monitor disease activity in multiple sclerosis, *Annals of neurology*, 32(6), 758–766, <https://doi.org/10.1002/ana.410320609>.

- [10] **Patti, F., De Stefano, M., Lavorgna, L., Messina, S., Chisari, C.G., Ippolito, D., Lanzillo, R., Vacchiano, V., Realmuto, S., Valentino, P. et al.** (2015). Lesion load may predict long-term cognitive dysfunction in multiple sclerosis patients, *PLoS one*, 10(3), e0120754, <https://doi.org/10.1371/journal.pone.0120754>.
- [11] **Moraal, B., van den Elskamp, I.J., Knol, D.L., Uitdehaag, B.M., Geurts, J.J., Vrenken, H., Pouwels, P.J., van Schijndel, R.A., Meier, D.S., Guttman, C.R. et al.** (2010). Long-interval T2-weighted subtraction magnetic resonance imaging: a powerful new outcome measure in multiple sclerosis trials, *Annals of neurology*, 67(5), 667–675, <https://doi.org/10.1002/ana.21958>.
- [12] **McKinley, R., Wepfer, R., Grunder, L., Aschwanden, F., Fischer, T., Friedli, C., Muri, R., Rummel, C., Verma, R., Weisstanner, C. et al.** (2020). Automatic detection of lesion load change in Multiple Sclerosis using convolutional neural networks with segmentation confidence, *NeuroImage: Clinical*, 25, 102104, <https://doi.org/10.1016/j.nicl.2019.102104>.
- [13] **Sweeney, E.M., Shinohara, R.T., Shiee, N., Mateen, F.J., Chudgar, A.A., Cuzzocreo, J.L., Calabresi, P.A., Pham, D.L., Reich, D.S. and Crainiceanu, C.M.** (2013). OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI, *NeuroImage: clinical*, 2, 402–413, <https://doi.org/10.1016/j.nicl.2013.03.002>.
- [14] **Rolak, L.A.** (2003). Multiple sclerosis: it's not the disease you thought it was, *Clinical Medicine & Research*, 1(1), 57–60, <https://doi.org/10.3121/cmr.1.1.57>.
- [15] **Ma, Y., Zhang, C., Cabezas, M., Song, Y., Tang, Z., Liu, D., Cai, W., Barnett, M. and Wang, C.** (2022). Multiple sclerosis lesion analysis in brain magnetic resonance images: techniques and clinical applications, *IEEE Journal of Biomedical and Health Informatics*, <https://doi.org/10.1109/JBHI.2022.3151741>.
- [16] **Birenbaum, A. and Greenspan, H.** (2016). Longitudinal multiple sclerosis lesion segmentation using multi-view convolutional neural networks, *Deep learning and data labeling for medical applications*, Springer, pp.58–67.
- [17] **Aslani, S., Dayan, M., Storelli, L., Filippi, M., Murino, V., Rocca, M.A. and Sona, D.** (2019). Multi-branch convolutional neural network for multiple sclerosis lesion segmentation, *NeuroImage*, 196, 1–15, <https://doi.org/10.1016/j.neuroimage.2019.03.068>.
- [18] **Zhang, H., Valcarcel, A.M., Bakshi, R., Chu, R., Bagnato, F., Shinohara, R.T., Hett, K. and Oguz, I.** (2019). Multiple sclerosis lesion segmentation with tiramisu and 2.5 d stacked slices, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp.338–346.

- [19] **Kamraoui, R.A., Ta, V.T., Tourdias, T., Mansencal, B., Manjon, J.V. and Coupé, P.** (2022). DeepLesionBrain: Towards a broader deep-learning generalization for multiple sclerosis lesion segmentation, *Medical Image Analysis*, 76, 102312, <https://doi.org/10.1016/j.media.2021.102312>.
- [20] **García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L. and Collins, D.L.** (2013). Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging, *Medical image analysis*, 17(1), 1–18, <https://doi.org/10.1016/j.media.2012.09.004>.
- [21] **LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.** (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86(11), 2278–2324, <https://doi.org/10.1109/5.726791>.
- [22] **Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M. and Biller, A.** (2016). Deep MRI brain extraction: A 3D convolutional neural network for skull stripping, *NeuroImage*, 129, 460–469, <https://doi.org/10.1016/j.neuroimage.2016.01.024>.
- [23] **Ronneberger, O., Fischer, P. and Brox, T.** (2015). U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp.234–241.
- [24] **Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M. and Larochelle, H.** (2017). Brain tumor segmentation with deep neural networks, *Medical image analysis*, 35, 18–31, <https://doi.org/10.1016/j.media.2016.05.004>.
- [25] **Sarica, B., Seker, D.Z. and Bayram, B.** (2023). A dense residual U-net for multiple sclerosis lesions segmentation from multi-sequence 3D MR images, *International Journal of Medical Informatics*, 170, 104965, <https://doi.org/10.1016/j.ijmedinf.2022.104965>.
- [26] **Sarica, B. and Seker, D.Z.** (2022). New MS lesion segmentation with deep residual attention gate U-Net utilizing 2D slices of 3D MR images, *Frontiers in Neuroscience*, 16, <https://doi.org/10.3389/fnins.2022.912000>.
- [27] **Hitziger, S., Ling, W.X., Fritz, T., D’Albis, T., Lemke, A. and Grilo, J.** (2022). Triplanar U-Net with lesion-wise voting for the segmentation of new lesions on longitudinal MRI studies, *Frontiers in Neuroscience*, 16, <https://doi.org/10.3389/fnins.2022.964250>.
- [28] **Basaran, B.D., Matthews, P.M. and Bai, W.** (2022). New lesion segmentation for multiple sclerosis brain images with imaging and lesion-aware augmentation, *Frontiers in Neuroscience*, 16, <https://doi.org/10.3389/fnins.2022.1007453>.

- [29] **Tseng, K.L., Lin, Y.L., Hsu, W. and Huang, C.Y.** (2017). Joint sequence learning and cross-modality convolution for 3D biomedical segmentation, *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp.6393–6400.
- [30] **Brosch, T., Tang, L.Y., Yoo, Y., Li, D.K., Traboulsee, A. and Tam, R.** (2016). Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation, *IEEE transactions on medical imaging*, 35(5), 1229–1239, <https://doi.org/10.1109/TMI.2016.2528821>.
- [31] **Zeng, C., Gu, L., Liu, Z. and Zhao, S.** (2020). Review of deep learning approaches for the segmentation of multiple sclerosis lesions on brain MRI, *Frontiers in Neuroinformatics*, 14, 610967, <https://doi.org/10.3389/fninf.2020.610967>.
- [32] **Liu, S., Zhang, D., Song, Y., Peng, H. and Cai, W.** (2017). Triple-crossing 2.5 d convolutional neural network for detecting neuronal arbours in 3d microscopic images, *International Workshop on Machine Learning in Medical Imaging*, Springer, pp.185–193, https://doi.org/10.1007/978-3-319-67389-9_22.
- [33] **Tetteh, G., Efremov, V., Forkert, N.D., Schneider, M., Kirschke, J., Weber, B., Zimmer, C., Piraud, M. and Menze, B.H.** (2020). Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes, *Frontiers in Neuroscience*, 1285, <https://doi.org/10.3389/fnins.2020.592352>.
- [34] **Egger, C., Opfer, R., Wang, C., Kepp, T., Sormani, M.P., Spies, L., Barnett, M. and Schippling, S.** (2017). MRI FLAIR lesion segmentation in multiple sclerosis: Does automated segmentation hold up with manual annotation?, *NeuroImage: Clinical*, 13, 264–270, <https://doi.org/10.1016/j.nicl.2016.11.020>.
- [35] **Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S.C., Girard, P., Ameli, R., Ferré, J.C. et al.** (2018). Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure, *Scientific reports*, 8(1), 1–17, <https://doi.org/10.1038/s41598-018-31911-7>.
- [36] **Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À., Oliver, A. and Lladó, X.** (2017). Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach, *NeuroImage*, 155, 159–168, <https://doi.org/10.1016/j.neuroimage.2017.04.034>.
- [37] **Calabresi, P.A.** (2004). Diagnosis and management of multiple sclerosis, *American family physician*, 70(10), 1935–1944.

- [38] **Van Waesberghe, J.H.T., Kamphorst, W., De Groot, C.J., Van Walderveen, M.A., Castelijns, J.A., Ravid, R., Lycklama a Nijeholt, G., Van Der Valk, P., Polman, C.H., Thompson, A.J. et al.** (1999). Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability, *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 46(5), 747–754.
- [39] **Mortazavi, D., Kouzani, A.Z. and Soltanian-Zadeh, H.** (2012). Segmentation of multiple sclerosis lesions in MR images: a review, *Neuroradiology*, 54(4), 299–320, <https://doi.org/10.1007/s00234-011-0886-7>.
- [40] **McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C. et al.** (2001). Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis, *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 50(1), 121–127, <https://doi.org/10.1002/ana.1032>.
- [41] **Polman, C.H., Reingold, S.C., Edan, G., Filippi, M., Hartung, H.P., Kappos, L., Lublin, F.D., Metz, L.M., McFarland, H.F., O'Connor, P.W. et al.** (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald Criteria”, *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 58(6), 840–846, <https://doi.org/10.1002/ana.20703>.
- [42] **Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L. et al.** (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria, *Annals of neurology*, 69(2), 292–302, <https://doi.org/10.1002/ana.22366>.
- [43] **Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S. et al.** (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria, *The Lancet Neurology*, 17(2), 162–173, [https://doi.org/10.1016/S1474-4422\(17\)30470-2](https://doi.org/10.1016/S1474-4422(17)30470-2).
- [44] **Khoo, V.S., Dearnaley, D.P., Finnigan, D.J., Padhani, A., Tanner, S.F. and Leach, M.O.** (1997). Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning, *Radiotherapy and Oncology*, 42(1), 1–15, [https://doi.org/10.1016/S0167-8140\(96\)01866-X](https://doi.org/10.1016/S0167-8140(96)01866-X).
- [45] **Fatahi, M., Speck, O. et al.** (2015). Magnetic resonance imaging (MRI): A review of genetic damage investigations, *Mutation Research/Reviews in Mutation Research*, 764, 51–63, <https://doi.org/10.1016/j.mrrev.2015.02.002>.

- [46] **Miller, D., Grossman, R., Reingold, S. and McFarland, H.** (1998). The role of magnetic resonance techniques in understanding and managing multiple sclerosis., *Brain: a journal of neurology*, 121(1), 3–24, <https://doi.org/10.1093/brain/121.1.3>.
- [47] **Hemond, C.C. and Bakshi, R.** (2018). Magnetic resonance imaging in multiple sclerosis, *Cold Spring Harbor perspectives in medicine*, 8(5), a028969, <https://doi.org/10.1101/cshperspect.a028969>.
- [48] **Danelakis, A., Theoharis, T. and Verganelakis, D.A.** (2018). Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging, *Computerized Medical Imaging and Graphics*, 70, 83–100, <https://doi.org/10.1016/j.compmedimag.2018.10.002>.
- [49] **Zhang, H. and Oguz, I.** (2021). Multiple sclerosis lesion segmentation-a survey of supervised cnn-based methods, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6*, Springer, pp.11–29, https://doi.org/10.1007/978-3-030-72084-1_2.
- [50] **Bitar, R., Leung, G., Perng, R., Tadros, S., Moody, A.R., Sarrazin, J., McGregor, C., Christakis, M., Symons, S., Nelson, A. et al.** (2006). MR pulse sequences: what every radiologist wants to know but is afraid to ask, *Radiographics*, 26(2), 513–537, <https://doi.org/10.1148/rg.262055063>.
- [51] **Cline, H.E., Lorensen, W.E., Kikinis, R. and Jolesz, F.** (1990). Three-dimensional segmentation of MR images of the head using probability and connectivity., *Journal of computer assisted tomography*, 14(6), 1037–1045, <https://doi.org/10.1097/00004728-199011000-00041>.
- [52] **Kapur, T., Grimson, W.E.L., Wells III, W.M. and Kikinis, R.** (1996). Segmentation of brain tissue from magnetic resonance images, *Medical image analysis*, 1(2), 109–127, [https://doi.org/10.1016/S1361-8415\(96\)80008-9](https://doi.org/10.1016/S1361-8415(96)80008-9).
- [53] **David, D.S. and Sutton, J.P.** (2001). Towards automated enhancement, segmentation and classification of digital brain images using networks of networks, *Information Sciences*, 138(1-4), 45–77, [https://doi.org/10.1016/S0020-0255\(01\)00130-X](https://doi.org/10.1016/S0020-0255(01)00130-X).
- [54] **Shen, S., Sandham, W. and Granat, M.** (2003). Preprocessing and segmentation of brain magnetic resonance images, *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003.*, IEEE, pp.149–152, <https://doi.org/10.1109/ITAB.2003.1222495>.

- [55] **Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A. and Leahy, R.M.** (2001). Magnetic resonance image tissue classification using a partial volume model, *NeuroImage*, 13(5), 856–876, <https://doi.org/10.1006/nimg.2000.0730>.
- [56] **Smith, S.M.** (2002). Fast robust automated brain extraction, *Human brain mapping*, 17(3), 143–155, <https://doi.org/10.1002/hbm.10062>.
- [57] **Eskildsen, S.F., Coupé, P., Fonov, V., Manjón, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Østergaard, L.R., Collins, D.L., Initiative, A.D.N. et al.** (2012). BEaST: brain extraction based on nonlocal segmentation technique, *NeuroImage*, 59(3), 2362–2373, <https://doi.org/10.1016/j.neuroimage.2011.09.012>.
- [58] **Iglesias, J.E., Liu, C.Y., Thompson, P.M. and Tu, Z.** (2011). Robust brain extraction across datasets and comparison with publicly available methods, *IEEE transactions on medical imaging*, 30(9), 1617–1634, <https://doi.org/10.1109/TMI.2011.2138152>.
- [59] **Sled, J.G., Zijdenbos, A.P. and Evans, A.C.** (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data, *IEEE transactions on medical imaging*, 17(1), 87–97, <https://doi.org/10.1109/42.668698>.
- [60] **Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A. and Gee, J.C.** (2010). N4ITK: improved N3 bias correction, *IEEE transactions on medical imaging*, 29(6), 1310–1320, <https://doi.org/10.1109/TMI.2010.2046908>.
- [61] **Nyúl, L.G., Udupa, J.K. and Zhang, X.** (2000). New variants of a method of MRI scale standardization, *IEEE transactions on medical imaging*, 19(2), 143–150, <https://doi.org/10.1109/42.836373>.
- [62] **Roura, E., Oliver, A., Cabezas, M., Vilanova, J.C., Rovira, A., Ramio-Torrenta, L. and Lladó, X.** (2014). MARGA: multispectral adaptive region growing algorithm for brain extraction on axial MRI, *Computer methods and programs in biomedicine*, 113(2), 655–673, <https://doi.org/10.1016/j.cmpb.2013.11.015>.
- [63] **Roura, E., Schneider, T., Modat, M., Daga, P., Muhlert, N., Chard, D., Ourselin, S., Lladó, X. and Wheeler-Kingshott, C.G.** (2015). Multi-channel registration of fractional anisotropy and T1-weighted images in the presence of atrophy: application to multiple sclerosis, *Functional neurology*, 30(4), 245.
- [64] **Brown, L.G.** (1992). A survey of image registration techniques, *ACM computing surveys (CSUR)*, 24(4), 325–376, <https://doi.org/10.1145/146370.146374>.
- [65] **Pluim, J.P., Maintz, J.A. and Viergever, M.A.** (2003). Mutual-information-based registration of medical images: a survey, *IEEE transactions on medical imaging*, 22(8), 986–1004, <https://doi.org/10.1109/TMI.2003.815867>.

- [66] **Holden, M.** (2007). A review of geometric transformations for nonrigid body registration, *IEEE transactions on medical imaging*, 27(1), 111–128, <https://doi.org/10.1109/TMI.2007.904691>.
- [67] **Simon, J., Li, D., Traboulsee, A., Coyle, P., Arnold, D., Barkhof, F., Frank, J., Grossman, R., Paty, D., Radue, E. et al.** (2006). Standardized MR imaging protocol for multiple sclerosis: Consortium of MS Centers consensus guidelines, *American Journal of Neuroradiology*, 27(2), 455–461.
- [68] **LeCun, Y., Bengio, Y. and Hinton, G.** (2015). Deep learning, *Nature*, 521(7553), 436–444, <https://doi.org/10.1038/nature14539>.
- [69] **Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D.L. and Erickson, B.J.** (2017). Deep learning for brain MRI segmentation: state of the art and future directions, *Journal of digital imaging*, 30(4), 449–459, <https://doi.org/10.1007/s10278-017-9983-4>.
- [70] **Ganiler, O., Oliver, A., Diez, Y., Freixenet, J., Vilanova, J.C., Beltran, B., Ramió-Torrentà, L., Rovira, À. and Lladó, X.** (2014). A subtraction pipeline for automatic detection of new appearing multiple sclerosis lesions in longitudinal studies, *Neuroradiology*, 56(5), 363–374, <https://doi.org/10.1007/s00234-014-1343-1>.
- [71] **Lesjak, Ž., Pernuš, F., Likar, B. and Špiclin, Ž.** (2016). Validation of white-matter lesion change detection methods on a novel publicly available MRI image database, *Neuroinformatics*, 14(4), 403–420, <https://doi.org/10.1007/s12021-016-9301-1>.
- [72] **Salem, M., Cabezas, M., Valverde, S., Pareto, D., Oliver, A., Salvi, J., Rovira, À. and Lladó, X.** (2018). A supervised framework with intensity subtraction and deformation field features for the detection of new T2-w lesions in multiple sclerosis, *NeuroImage: Clinical*, 17, 607–615, <https://doi.org/10.1016/j.nicl.2017.11.015>.
- [73] **Köhler, C., Wahl, H., Ziemssen, T., Linn, J. and Kitzler, H.H.** (2019). Exploring individual multiple sclerosis lesion volume change over time: development of an algorithm for the analyses of longitudinal quantitative MRI measures, *NeuroImage: Clinical*, 21, 101623, <https://doi.org/10.1016/j.nicl.2018.101623>.
- [74] **Krüger, J., Opfer, R., Gessert, N., Ostwaldt, A.C., Manogaran, P., Kitzler, H.H., Schlaefer, A. and Schippling, S.** (2020). Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks, *NeuroImage: Clinical*, 28, 102445, <https://doi.org/10.1016/j.nicl.2020.102445>.
- [75] **Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J., Rovira, À. and Lladó, X.** (2020). A fully convolutional neural network for new T2-w lesion detection in multiple sclerosis, *NeuroImage: Clinical*, 25, 102149, <https://doi.org/10.1016/j.nicl.2019.102149>.

- [76] Gessert, N., Krüger, J., Opfer, R., Ostwaldt, A.C., Manogaran, P., Kitzler, H.H., Schippling, S. and Schlaefer, A. (2020). Multiple sclerosis lesion activity segmentation with attention-guided two-path CNNs, *Computerized Medical Imaging and Graphics*, 84, 101772, <https://doi.org/10.1016/j.compmedimag.2020.101772>.
- [77] Gessert, N., Bengs, M., Krüger, J., Opfer, R., Ostwaldt, A.C., Manogaran, P., Schippling, S. and Schlaefer, A. (2020). 4D deep learning for multiple sclerosis lesion activity segmentation, *arXiv preprint arXiv:2004.09216*.
- [78] Vaidya, S., Chunduru, A., Muthuganapathy, R. and Krishnamurthi, G. (2015). Longitudinal multiple sclerosis lesion segmentation using 3D convolutional neural networks, *Proceedings of the 2015 longitudinal multiple sclerosis lesion segmentation challenge*, 1–2.
- [79] Carass, A., Roy, S., Jog, A., Cuzzocreo, J.L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C.H. et al. (2017). Longitudinal multiple sclerosis lesion segmentation: resource and challenge, *NeuroImage*, 148, 77–102, <https://doi.org/10.1016/j.neuroimage.2016.12.064>.
- [80] Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V. and Warfield, S. (2008). 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation, *MIDAS journal*, 2008, 1–6.
- [81] Hashemi, S.R., Salehi, S.S.M., Erdogmus, D., Prabhu, S.P., Warfield, S.K. and Gholipour, A. (2018). Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection, *IEEE Access*, 7, 1721–1735, <https://doi.org/10.1109/ACCESS.2018.2886371>.
- [82] Aslani, S., Dayan, M., Murino, V. and Sona, D. (2019). Deep 2D encoder-decoder convolutional neural network for multiple sclerosis lesion segmentation in brain MRI, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, Springer, pp.132–141, https://doi.org/10.1007/978-3-030-11723-8_13.
- [83] Kang, G., Hou, B., Ma, Y., Labeau, F., Su, Z. et al. (2020). Acu-net: A 3d attention context u-net for multiple sclerosis lesion segmentation, *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp.1384–1388, <https://doi.org/10.1109/ICASSP40776.2020.9054616>.
- [84] Zhang, H., Zhang, J., Li, C., Sweeney, E.M., Spincemille, P., Nguyen, T.D., Gauthier, S.A., Wang, Y. and Marcille, M. (2021). ALL-Net: Anatomical information lesion-wise loss function integrated into neural network for multiple sclerosis lesion segmentation, *NeuroImage: Clinical*, 32, 102854, <https://doi.org/10.1016/j.nicl.2021.102854>.

- [85] **Weeda, M., Brouwer, I., de Vos, M., de Vries, M., Barkhof, F., Pouwels, P. and Vrenken, H.** (2019). Comparing lesion segmentation methods in multiple sclerosis: Input from one manually delineated subject is sufficient for accurate lesion segmentation, *NeuroImage: Clinical*, 24, 102074, <https://doi.org/10.1016/j.nicl.2019.102074>.
- [86] **Commowick, O., Kain, M., Casey, R., Ameli, R., Ferré, J.C., Kerbrat, A., Tourdias, T., Cervenansky, F., Camarasu-Pop, S., Glatard, T. et al.** (2021). Multiple sclerosis lesions segmentation from multiple experts: The MICCAI 2016 challenge dataset, *NeuroImage*, 244, 118589, <https://doi.org/10.1016/j.neuroimage.2021.118589>.
- [87] **Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E. and Yu, T.** (2014). scikit-image: image processing in Python, *PeerJ*, 2, e453, <https://doi.org/10.7717/peerj.453>.
- [88] **Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M. and Kalinin, A.A.** (2020). Albumentations: fast and flexible image augmentations, *Information*, 11(2), 125, <https://doi.org/10.3390/info11020125>.
- [89] **Reinhold, J.C., Dewey, B.E., Carass, A. and Prince, J.L.** (2019). Evaluating the impact of intensity normalization on MR image synthesis, *Medical Imaging 2019: Image Processing*, volume10949, SPIE, pp.890–898, <https://doi.org/10.1117/12.2513089>.
- [90] **Hashemi, M.** (2019). Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation, *Journal of Big Data*, 6(1), 1–13, <https://doi.org/10.1186/s40537-019-0263-7>.
- [91] **Long, J., Shelhamer, E. and Darrell, T.** (2015). Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.3431–3440.
- [92] **Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S.** (2017). Feature pyramid networks for object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.2117–2125.
- [93] **He, K., Zhang, X., Ren, S. and Sun, J.** (2016). Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.770–778.
- [94] **He, K., Zhang, X., Ren, S. and Sun, J.** (2016). Identity mappings in deep residual networks, *European conference on computer vision*, Springer, pp.630–645.
- [95] **Sun, J., Li, J. and Liu, L.** (2021). Semantic segmentation of brain tumor with nested residual attention networks, *Multimedia Tools and Applications*, 80(26), 34203–34220, <https://doi.org/10.1007/s11042-020-09840-3>.

- [96] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B. *et al.* (2018). Attention u-net: Learning where to look for the pancreas, *arXiv preprint arXiv:1804.03999*.
- [97] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*.
- [98] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015). Going deeper with convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.1–9.
- [99] Hu, J., Shen, L. and Sun, G. (2018). Squeeze-and-excitation networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.7132–7141.
- [100] Woo, S., Park, J., Lee, J.Y. and Kweon, I.S. (2018). Cbam: Convolutional block attention module, *Proceedings of the European conference on computer vision (ECCV)*, pp.3–19.
- [101] Hu, J., Shen, L., Albanie, S., Sun, G. and Vedaldi, A. (2018). Gather-excite: Exploiting feature context in convolutional neural networks, *Advances in neural information processing systems*, 31.
- [102] Chen, Y., Kalantidis, Y., Li, J., Yan, S. and Feng, J. (2018). A²-nets: Double attention networks, *Advances in neural information processing systems*, 31.
- [103] Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z. and Lu, H. (2019). Dual attention network for scene segmentation, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.3146–3154.
- [104] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.11531–11539.
- [105] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE transactions on pattern analysis and machine intelligence*, 40(4), 834–848, <https://doi.org/10.1109/TPAMI.2017.2699184>.
- [106] Grauman, K. and Darrell, T. (2005). The pyramid match kernel: Discriminative classification with sets of image features, *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, IEEE, pp.1458–1465, <https://doi.org/10.1109/ICCV.2005.239>.

- [107] **Lazebnik, S., Schmid, C. and Ponce, J.** (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 2, IEEE, pp.2169–2178, <https://doi.org/10.1109/CVPR.2006.68>.
- [108] **He, K., Zhang, X., Ren, S. and Sun, J.** (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE transactions on pattern analysis and machine intelligence*, 37(9), 1904–1916, <https://doi.org/10.1109/TPAMI.2015.2389824>.
- [109] **Jadon, S.** (2020). A survey of loss functions for semantic segmentation, *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, IEEE, pp.1–7, <https://doi.org/10.1109/CIBCB48159.2020.9277638>.
- [110] **Milletari, F., Navab, N. and Ahmadi, S.A.** (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation, *2016 fourth international conference on 3D vision (3DV)*, IEEE, pp.565–571, <https://doi.org/10.1109/3DV.2016.79>.
- [111] **Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P.** (2017). Focal loss for dense object detection, *Proceedings of the IEEE international conference on computer vision*, pp.2980–2988.
- [112] **Aung, H.L., UzKent, B., Burke, M., Lobell, D. and Ermon, S.** (2020). Farm parcel delineation using spatio-temporal convolutional networks, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp.76–77.
- [113] **Clevert, D.A., Unterthiner, T. and Hochreiter, S.** (2015). Fast and accurate deep network learning by exponential linear units (elus), *arXiv preprint arXiv:1511.07289*.
- [114] **Chollet, F.**, (2015), Keras, <https://github.com/fchollet/keras>.
- [115] **Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. et al.** (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems, *arXiv preprint arXiv:1603.04467*.
- [116] **Bisong, E.** (2019). *Building machine learning and deep learning models on Google cloud platform: A comprehensive guide for beginners*, Apress.
- [117] **Kingma, D.P. and Ba, J.** (2014). Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.
- [118] **Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S. and Jorge Cardoso, M.**, (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, *Deep learning in medical image analysis and multimodal learning for clinical decision support*, Springer, pp.240–248, https://doi.org/10.1007/978-3-319-67558-9_28.

- [119] **Hashemi, M., Akhbari, M. and Jutten, C.** (2022). Delve into multiple sclerosis (MS) lesion exploration: a modified attention U-net for MS lesion segmentation in brain MRI, *Computers in Biology and Medicine*, 145, 105402, <https://doi.org/10.1016/j.compbiomed.2022.105402>.
- [120] **Maier, O. and Handels, H.** (2015). MS lesion segmentation in MRI with random forests, *Proc. 2015 Longitudinal Multiple Sclerosis Lesion Segmentation Challenge*, 1–2.
- [121] **Commowick, O., Cervenansky, F., Cotton, F. and Dojat, M.** (2021). MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure, *MICCAI 2021-24th International Conference on Medical Image Computing and Computer Assisted Intervention*, pp.1–118.
- [122] **Altay, E.E., Fisher, E., Jones, S.E., Hara-Cleaver, C., Lee, J.C. and Rudick, R.A.** (2013). Reliability of classifying multiple sclerosis disease activity using magnetic resonance imaging in a multiple sclerosis clinic, *JAMA neurology*, 70(3), 338–344, <https://doi.org/10.1001/2013.jamaneurol.211>.



CURRICULUM VITAE

Name Surname : Beytullah SARICA

EDUCATION :

- **B.Sc.** : 2009, Middle East Technical University, Department of Computer Education and Instructional Technology
- **M.Sc.** : 2012, Rochester Institute of Technology, Golisano College of Computing and Information Sciences, Department of Information Technology

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2009-2010 Multimedia Developer at Sebit Information & Education Technologies Inc.
- 2010-2012 Fulbright Scholarship by Institute of International Education
- 2010-2012 Graduate Merit Scholarship by Rochester Institute of Technology
- 2011-2012 Teaching Assistance at Rochester Institute of Technology
- 2012-2013 Software Engineer at Energy Curtailment Specialists
- 2014-2015 Java Developer at YAZ Bilgi Sistemleri ve Tic A.Ş.
- 2015-2019 Senior Java Developer at SoftTech A.Ş.
- 2019-2020 Senior Software Engineer at GE Aviation
- 2020-2023 Senior Software Engineer at GE Power
- 2023-Present Senior Software Engineer at GE Vernova

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Sarica, B., & Seker, D. Z.** (2022). New MS lesion segmentation with deep residual attention gate U-Net utilizing 2D slices of 3D MR images. *Frontiers in Neuroscience*, 16. <https://doi.org/10.3389/fnins.2022.912000>
- **Sarica, B., Seker, D.Z. and Bayram, B.** (2023). A dense residual U-net for multiple sclerosis lesions segmentation from multi-sequence 3D MR images, *International Journal of Medical Informatics*, 170, 104965. <https://doi.org/10.1016/j.ijmedinf.2022.104965>

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- **Bicakci, Y. S., Sarica, B., Pakdil, M. E., Yazirli, B., & Demirel, H.** (2017). Spatio-Temporal Analyses to Estimate Speed Information for Transport Forecasting and Scenario Testing. *Fresenius Environmental Bulletin*, 26, 100-106.