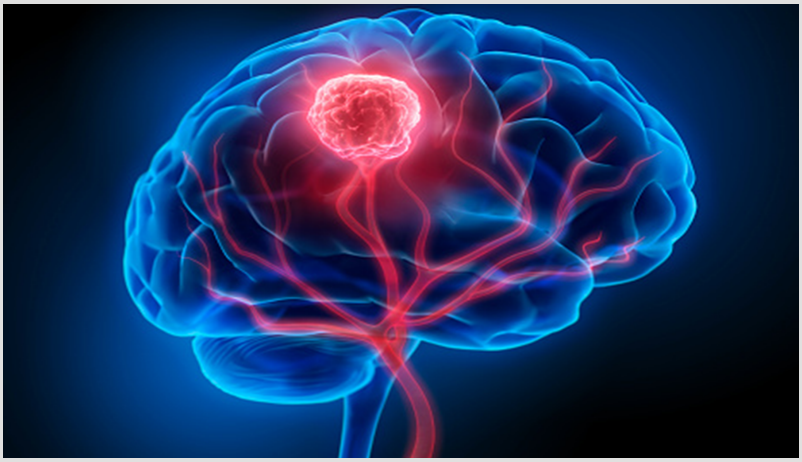
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING

Moona Mazher

UNIVERSITAT ROVIRA i VIRGILI

Self-supervised Advanced Deep Learning for Characterization of Brain Tumor Aggressiveness and Prognosis Analysis Through Multimodality MRI Imaging

Author:
Moona Mazher



DOCTORAL THESIS
December, 2023

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Self-supervised Advanced Deep Learning for Characterization of Brain Tumor Aggressiveness and Prognosis Analysis Through Multimodality MRI Imaging

DOCTORAL THESIS

Moona Mazher

Supervisors:
Domenec Savi Puig Valls
Mohamed Abdelnasser

Department of Computer Engineering and Mathematics

UNIVERSITAT ROVIRA i VIRGILI

December, 2023

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

UNIVERSITAT ROVIRA i VIRGILI

Departament d'Enginyeria Informàtica i Matemàtiques

Av. Països Catalans, 26
43007 Tarragona
Tel. +34 977 *55 9703*
Fax. +34 977 *55 9710*

I STATE that the present study entitled 'Self-supervised Advanced Deep Learning for Characterization of Brain Tumor Aggressiveness and Prognosis Analysis Through Multimodality MRI Imaging" presented by Moona Mazher, for the award of the degree of Doctor, has been carried out under my supervision at the Departament d'Enginyeria Informàtica i Matemàtiques of this university.

Tarragona, 14th August 2023

Doctoral Thesis Supervisor/s

PUIG VALLS DOMÈNEC SAVI - 39869760L
2023.08.13 12:21:30 +02'00'

Dr. Domènec Savi Puig Valls

Mohamed Abdelnasser Mohamed Mahmoud - DNI Y3121427L (TCAT)
Digitally signed by Mohamed Abdelnasser Mohamed Mahmoud - DNI Y3121427L (TCAT)
Date: 2023.08.26 18:00:52 +03'00'

Dr. Mohamed AbdelNasser

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Abstract

Gliomas are the main common primary brain malignancies and are the most common type of brain cancer. It comes from glial brain cells that support nerve cells for a broad category of brain and spinal cord tumors. Identification of the sub-regions of gliomas before surgery is critical, which may affect the survival of patients. Different medical imaging techniques have been used to distinguish brain tumor lesions such as X-ray, Ultrasound, Positron Emission Tomography (PET), Electroencephalography (EEG), Magnetic Brain Wave Graph (MEG), Single-Photon Emission Computed Tomography (SPECT), Computed Tomography (CT), and Magnetic Resonance Imaging (MRI). However, MRI is the most comprehensive technique that provides better image quality and facilitates determining the exact volume and size of the malignant brain tumor. MRI modalities include Fluid-Attenuated Inversion Recovery (FLAIR), T1-weighted (T1), T2-weighted (T2), and T1w contrast-enhanced (T1CE). Each MRI modality provides different structural information about brain tissue. Gliomas have complicated pathological changes such as hemorrhage, necrosis, and edema. Due to these complex pathological changes, gliomas show intricate changes in brightness and texture on MRI scans. Different tissues may have the same gray matter values, which makes it difficult to have an accurate, repeatable, and stable segmentation of gliomas.

Early detection, automatic delineation, and volume estimation are vital tasks for survival prediction and treatment planning. However, gliomas are often difficult to localize and delineate with conventional manual segmentation due to their high variation of shape, location, and appearance. Moreover, manual mark delineation is laborious and time-consuming work for a neurosurgeon. In addition, it is difficult to replicate the segmentation results due to certain practical operation factors. The accurate segmentation of brain tumors can facilitate diagnosis and help assess the prognosis and severity of the disease.

In recent years, convolution neural networks (CNNs) are widely used for the automated classification and segmentation of medical images. For 2D segmentation, the input data is passed slice by slice to the CNN model while in 3D segmentation the whole volumetric data is fed into the CNN model. The focus of the present thesis is to develop a system for automating brain tumor analysis (such as brain tumor segmentation, and survival prediction), using deep learning techniques and applying them to the MRI images for segmenting the brain tumor classes (Enhancing Tumor, Non-enhancing Tumor, and Peritumoral Edema) and estimating the survival days of the patients for prognosis analysis.

Firstly, we applied the standard 2DCNNmodels including base UNET, Residual UNET (ResUnet), and DeepLabv3 and then we proposed depth-wise channel attention (BrainSeg-DCANet) with Binary Cross Entropy (BCE) loss function for brain tumor segmentation on the BraTS2020 dataset. The proposed BrainSeg-DCANet model with BCE loss function yields the best result to segment the combined tumor regions (Enhancing Tumor, Tumor Core, Whole Tumor, in MRI images with overall dice scores coefficients (DSCs) around 87%, 91%, and 93%, and Hausdorff distance (HD) around 2.08, 1.32, and 1.94, respectively, outperforming the common semantic segmentation methods (UNET and Residual UNET).

Secondly, a multiview (axial, sagittal, and coronal) deep inception residual network segmentation was applied to the 2D MRI images to segment Brain tumor regions (TC, WT, ET) with the interdisciplinary dataset for fetal brain segmentation (FETA2021) using Combo loss function. This method takes 2D images as input and generates a 3D mask (stacking of 2D masks) as volumetric output. Moreover, survival days prediction for the BraTS dataset and gestational age prediction for the FETA dataset were given using multiple feature extraction techniques including Deep CNN Features, Radiomics Features, Auto Encoder Features, and clinical features by applying variant regressions methods (Linear Regression, Random Forest, Extreme Gradient Boosting).

The proposed multiview deep inception residual network segmentation model outperformed the state-of-the-art methods for brain tumor region segmentation as well as for fetal brain segmentation. The generalization capabilities of the proposed method were also tested on Head and Neck tumor segmentation and patients' survival days prediction. The radiomics features yield promising results for survival days prediction tasks along with the deep CNN features. Thereafter, we proposed a new 3D CNN method for computing the brain tumor segmentation for all tumor regions including TC, WT, and ET.

Two-stage Self-supervised Contrastive Learning using Parallel Multiview Multiscale Attention-based CNN Transformers for 3D brain tumor volumetric segmentation was introduced. This method works efficiently for small labeled and large unlabeled data challenges. The proposed model results were compared with the various state-of-the-art methods for 3D brain tumor segmentation on MRI images. The results showed, our proposed self-supervised using Parallel Multiview Multiscale Attention-based CNN Transformers with contrastive learning approach outperformed the state-of-the-art methods by yielding the best DCS and HD scores of 88%, 92%, 93%, 4.78, 5.33 and 8.99 for ET, TC, and WT tumor regions.

Finally, a 3D MR image-based survival prediction was performed. Multiple feature extraction techniques are used to extract the features from the 3D volumetric MRI image and then different regression techniques are applied to the extracted features. It is observed that radiomics features produced very interesting results like in the previously given 2D MR images case. In this study, a combination of radiomics features with 3D CNN deep features from the proposed 3D CNN model (self-supervised contrastive learning method with semi-supervised approach) and given clinical features produced the highest concordance (C-index) value as compared to the other feature extraction methods. We have validated our proposed solution extensively for real-time medical imaging datasets and won various competitions.

Overall, this thesis presents the automated deep learning methods for brain tumor segmentation and survival prediction for prognosis analysis; This thesis's findings are promising and showed that the proposed techniques can produce a clinically helpful computer-aided tool for brain tumor segmentation and survival prediction by MRI Images.

Resum

Els giliomes són els principals acomiadaments comuns del cervell primari i són el tipus més comú de càncer cerebral. Prové de cèllules cerebrals brillants que recolzen les cèllules nervioses per a una àmplia categoria de tumors de cervell i medulla espinal. La identificació de les subregions de gliomas abans de la cirurgia és crítica, que pot afectar la supervivència dels pacients. S'han utilitzat diferents tècniques de processament d'imatges mèdiques per distingir lesions del tumor cerebral, com ara radiografia, Ultrasound, Tomografia d'emissió Positron (PET), Electroencefalografia (EEG), Magnetic Brain Wave Graph (MEG), Tomografia computada per a l'Emissió Única (ESPECT), Tomografia Computada (CT) i Resonance Magnetic Imaging (MRI) ... No obstant això, la ressonància magnètica és la tècnica més completa que proporciona una millor qualitat d'imatge i facilita determinar el volum i la mida exactes del tumor cerebral maligne. Les modalitats de ressonància magnètica inclouen Recuperació de la inversió (FLAIR), T1 ponderada (T1), T2-ponderat (T2) i T1w, amb contrast millorat (T1CE). Cada modalitat ressonància magnètica ofereix diferents dades estructurals sobre el teixit cerebral. Els glumes tenen complicats canvis patològics com l'hemorràgia, la necrosi i l'edema. A causa d'aquests complexos canvis patològics, els gliomas mostren canvis intricats en la brillantor i la textura en els escàners de ressonància magnètica. Els diferents teixits poden tenir els mateixos valors de matèria grisa, cosa que dificulta tenir una segmentació exacta, repetible i estable de gliomas.

La detecció precoç, la delineació automàtica i l'estimació de volum són tasques vitals per a la predicció de supervivència i la planificació del tractament. No obstant, els gliomas sovint són difícils de localitzar i delinear amb la segmentació manual convencional per la seva alta variació de forma, ubicació i aspecte. A més, la delineació manual de marques és un treball laboriós i amb un temps de consum per a un neurocirurgià. A més, és difícil replicar els resultats de la segmentació per determinats

factors de funcionament pràctic. La segmentació exacta dels tumors cerebrals pot facilitar el diagnòstic i ajudar a avaluar el pronòstic i la gravetat de la malaltia.

En els últims anys, les xarxes neuronals de convolució (CNN) són molt utilitzades per a la classificació automatitzada i la segmentació de les imatges mèdiques. Per a la segmentació en 2D, les dades d'entrada es passen sector per sectors al model CNN mentre que en la segmentació en 3D s'alimenten totes les dades volumètriques en el model CNN. El focus de la tesi present consisteix a desenvolupar un sistema per automatitzar l'anàlisi dels tumors cerebrals (com la segmentació del tumor cerebral i la predicció de supervivència), usant tècniques d'aprenentatge profundes i aplicant-les a les imatges de ressonància magnètica per segmentar les classes de tumor cerebral (Potenciar el tumor, el tumor no intensificador i l'Edema Peritumoral) i estimar els dies de supervivència dels pacients per a l'anàlisi de pronòstic.

En primer lloc, vam aplicar els models estàndard 2DCNN que inclouen UNET base, Resdual (ResUnet) i DeepLabv3 i després vam proposar atenció al canal en sentit invers (BrainSeg-DCANet) amb funció de pèrdua de la Creu binària Entropy (BEC) per a la segmentació del tumor cerebral al conjunt de dades de BraTS200. El model proposat per a BrainSeg-DCANet amb funció de pèrdua de l'EI dóna el millor resultat per segmentar les regions de tumor combinades (tumor millorat, Tumor Core, Tumor Core, Tumor Sencer Tumor, en imatges de ressonància magnètica amb coeficients globals de daus (DSC) al voltant del 87%, el 91% i el 93%, i la distància Hausdorff (HD) al voltant del 2.08, 1.08, i 1.94, respectivament, superant els mètodes de segmentació semàntica comuna (UNET i UNET Residual).

En segon lloc, es va aplicar a les imatges 2D de ressonància magnètica (axial, sagittal i coronal) una profunda estreta de la xarxa residual de creació (TC) per segmentar regions tumorals cerebrals (TC, WT, ET) amb el conjunt de dades interdisciplinàries per a la segmentació del cervell fetal (FETA2021) usant la funció de pèrdua de Combo. Aquest

mètode agafa les imatges en 2D com a entrada i genera una màscara en 3D (apilada de màscares en 2D) com a resultat volumètric. A més, les prediccions dels dies de supervivència per al conjunt de dades BraTS i la predicció de l'edat gestacional per al conjunt de dades FETA s'han donat mitjançant múltiples tècniques d'extracció de característiques, incloent característiques d'extracció de característiques profundes CNN, trets de Radiomics, característiques de codificador automàtic i característiques clíniques aplicant mètodes de regressions de la variant (Regressió lineal, Bosc Aleatori, buidatge de gradient extremeny).

El model de segmentació de la xarxa residual proposada per a l'inici profund de la xarxa va superar els mètodes d'última generació per a la segmentació de la regió del tumor cerebral així com per a la segmentació del cervell fetal. Les capacitats de generalització del mètode proposat també es van posar a prova en la segmentació del tumor de cap i coll i en la predicció dels dies de supervivència dels pacients. Les característiques de la ràdio donen resultats prometedors per a les tasques de predicció dels dies de supervivència juntament amb les profundes característiques de la CNN. A partir de llavors, vam proposar un nou mètode de la CNN 3D per a la informàtica de la segmentació del tumor cerebral per a totes les regions tumorals, incloses el TC, WT i ET.

S'ha introduït l'aprenentatge contradictori amb l'autosupervisió de dues etapes mitjançant la multivisió parallela dels transformadors de la CNN multiescala d'atenció de la CNN per a la segmentació volumètrica del tumor cerebral en 3D. Aquest mètode funciona eficientment per a petits reptes de dades etiquetats i grans sense etiquetar. Els resultats del model proposat es van comparar amb els diversos mètodes d'última generació per a la segmentació del tumor cerebral 3D en imatges de ressonància magnètica. Els resultats mostrats, la nostra proposta autosupervisada utilitzant els transformadors de la multipantalla múltiple ParalLel per a l'Atenció dels Transformadors de la CNN amb un enfocament d'aprenentatge contrastat superava els mètodes d'última generació al cedir les millors puntuacions de DCS i HD del 88%, 92%, 93%,

4.78, 5.33 i 8.99 per a ET, TC i WT, TC i WT. regions tumorals.

Finalment, es va fer una predicció de supervivència basada en imatges en 3D MR. S'utilitzen múltiples tècniques d'extracció de característiques per extreure les característiques de la imatge de la imatge de la imatge de la imatge de la imatge de la imatge de la imatge de la imatge volumètrica 3D i, a continuació, s'apliquen diferents tècniques de regressió a les característiques extretes. S'observa que les característiques de les radiòmiques produïen resultats molt interessants com en el cas d'imatges 2D MR prèviament donat. En aquest estudi, una combinació de trets radiofònics amb trets profunds de la CNN 3D a partir del model proposat pel 3D CNN (mètode d'aprenentatge contrastat autosupervisat amb enfocament semisupervisat) i donat que els trets clínics van produir el valor més alt de concordança (C-index) en comparació amb els altres mètodes d'extracció de característiques. Hem validat àmpliament la nostra proposta de solució per a conjunts de dades d'imatge mèdica en temps real i hem guanyat diverses competicions.

En conjunt, aquesta tesi presenta els mètodes d'aprenentatge profunds automatitzats per a la segmentació del tumor cerebral i la predicció de supervivència per a l'anàlisi de pronòstic. Les troballes d'aquesta tesi són prometedores i mostren que les tècniques proposades poden produir una eina clínicament ajudada per ordinador per a la segmentació del tumor cerebral i la predicció de supervivència per part d'Imatges de ressonància magnètica.

Resumen

Los gliomas son las principales neoplasias cerebrales primarias comunes y son el tipo más común de cáncer cerebral. Proviene de las células del cerebro glial que soportan las células nerviosas para una amplia categoría de tumores del cerebro y la médula espinal. La identificación de las subregiones de los gliomas antes de la cirugía es crítica, lo que puede afectar la supervivencia de los pacientes. Se han utilizado diferentes técnicas de imagen para distinguir lesiones tumorales cerebrales, como la radiografía, la ecografía, la tomografía por emisión de positrones (PET), la electroencefalografía (EEG), el gráfico de ondas cerebrales magnéticas (MEG), la tomografía computarizada por emisión de fotón único (SPECT), la tomografía computarizada (TC) y la resonancia magnética nuclear (RMN). Sin embargo, la RM es la técnica más completa que proporciona una mejor calidad de imagen y facilita la determinación del volumen y el tamaño exactos del tumor cerebral maligno. Las modalidades de RM incluyen recuperación de inversión atenuada con líquido (FLAIR), T1 (T1), T2 (T2) y T1w con contraste mejorado (T1CE). Cada modalidad de RM proporciona información estructural diferente sobre el tejido cerebral. Los gliomas tienen cambios patológicos complicados, como hemorragia, necrosis y edema. Debido a estos cambios patológicos complejos, los gliomas muestran cambios intrincados en el brillo y la textura en las resonancias magnéticas. Los diferentes tejidos pueden tener los mismos valores de materia gris, lo que dificulta una segmentación precisa, repetible y estable de los gliomas.

La detección temprana, la delineación automática y la estimación del volumen son tareas vitales para la predicción de la supervivencia y la planificación del tratamiento. Sin embargo, los gliomas a menudo son difíciles de localizar y delinear con la segmentación manual convencional debido a su alta variación de forma, ubicación y apariencia. Además, la delineación manual de marcas es un trabajo laborioso y que

consume mucho tiempo para un neurocirujano. Además, es difícil reproducir los resultados de la segmentación debido a ciertos factores de operación prácticos. La segmentación precisa de los tumores cerebrales puede facilitar el diagnóstico y ayudar a evaluar el pronóstico y la gravedad de la enfermedad.

En los últimos años, las redes neuronales de convolución (CNNs) son ampliamente utilizadas para la clasificación automatizada y segmentación de imágenes médicas. Para la segmentación 2D, los datos de entrada se pasan porción a porción al modelo CNN, mientras que en la segmentación 3D todos los datos volumétricos se introducen en el modelo CNN. El objetivo de la presente tesis es desarrollar un sistema para automatizar el análisis de tumores cerebrales (como la segmentación de tumores cerebrales, y la predicción de supervivencia), utilizando técnicas de aprendizaje profundo y aplicándolas a las imágenes de RM para segmentar las clases de tumores cerebrales (tumor potenciador, tumor no potenciador, y edema peritumoral) y estimar los días de supervivencia de los pacientes para el análisis de pronóstico.

En primer lugar, se aplicaron los modelos 2DCNN estándar incluyendo UNET base, UNET residual (ResUnet), y DeepLabv3 y luego se propuso la atención de canal profundo (BrainSeg-DCANet) con la función de pérdida de entropía cruzada binaria (BCE) para la segmentación de tumores cerebrales en el conjunto de datos BraTS2020. El modelo propuesto de BrainSeg-DCANet con función de pérdida de BCE produce el mejor resultado para segmentar las regiones combinadas del tumor (Enhancing Tumor, Tumor Core, Tumor Completo) en imágenes de RM con coeficientes de puntuación de dados (DSCs) globales alrededor del 87%, 91% y 93%, y la distancia de Hausdorff (HD) alrededor de 2,08, 1,32 y 1,94, respectivamente, superando los métodos de segmentación semántica comunes (UNET y UNET residual).

En segundo lugar, se aplicó una segmentación de red residual de inicio profundo multiview (axial, sagital y coronal) a las imágenes de resonancia magnética 2D para segmentar regiones de tumores cerebrales

(CT, WT, ET) con el conjunto de datos interdisciplinarios para la segmentación cerebral fetal (FETA2021) utilizando la función de pérdida Combo. Este método toma imágenes 2D como entrada y genera una máscara 3D (apilamiento de máscaras 2D) como salida volumétrica. Además, la predicción de los días de supervivencia para el conjunto de datos BraTS y la predicción de la edad gestacional para el conjunto de datos FETA se dieron utilizando múltiples técnicas de extracción de características, incluyendo características de CNN profundo, características de radiomics, características de codificador automático, y características clínicas mediante la aplicación de métodos de regresión de variantes (regresión lineal, bosque aleatorio, potenciación de gradiente extremo).

El modelo multiview propuesto para la segmentación de redes residuales de inicio profundo superó a los métodos más avanzados para la segmentación de la región del tumor cerebral así como para la segmentación del cerebro fetal. Las capacidades de generalización del método propuesto también se probaron en la segmentación de tumores de cabeza y cuello y en la predicción de los días de supervivencia de los pacientes. Las características radiómicas arrojan resultados prometedores para las tareas de predicción de días de supervivencia junto con las características de CNN profundo. Posteriormente, se propuso un nuevo método de CNN 3D para calcular la segmentación del tumor cerebral en todas las regiones tumorales, incluyendo CT, WT y ET.

Se introdujo el aprendizaje contrastivo autosupervisado en dos etapas utilizando transformadores CNN basados en la atención multivista paralela para segmentación volumétrica de tumores cerebrales 3D. Este método funciona de manera eficaz para los desafíos de datos sin etiqueta pequeños y grandes. Los resultados del modelo propuesto se compararon con los métodos más avanzados para la segmentación 3D de tumores cerebrales en imágenes de RM. Los resultados mostraron que los transformadores de CNN basados en la atención en múltiples escalas paralelas con enfoque de aprendizaje contrastivo, los cuales fueron autosupervisados, superaron a los métodos de vanguardia al obtener las

mejores puntuaciones de DCS y HD de 88%, 92%, 93%, 4,78, 5,33 y 8,99 para las regiones tumorales ET, TC y WT.

Finalmente, se realizó una predicción de supervivencia basada en imágenes 3D MR. Se utilizan múltiples técnicas de extracción de características para extraer las características de la imagen 3D de resonancia magnética volumétrica y, a continuación, se aplican diferentes técnicas de regresión a las características extraídas. Se observa que las características radiológicas produjeron resultados muy interesantes como en el caso de las imágenes 2D MR previamente dado. En este estudio, una combinación de características radiómicas con características profundas de CNN 3D del modelo propuesto de CNN 3D (método de aprendizaje contrastivo autosupervisado con enfoque semisupervisado) y características clínicas dadas produjeron el valor más alto de concordancia (índice C) en comparación con los otros métodos de extracción de características. Hemos validado ampliamente nuestra solución propuesta para conjuntos de datos de imágenes médicas en tiempo real y hemos ganado varios concursos.

En general, esta tesis presenta los métodos automatizados de aprendizaje profundo para la segmentación de tumores cerebrales y la predicción de la supervivencia para el análisis del pronóstico; Los hallazgos de esta tesis son prometedores y mostraron que las técnicas propuestas pueden producir una herramienta asistida por ordenador clínicamente útil para la segmentación de tumores cerebrales y la predicción de la supervivencia mediante imágenes de RM.

Acknowledgements

In the Name of Allah, the Most Merciful, the Most Compassionate all praise be to Allah, the Lord of the worlds; and prayers and peace be upon Mohamed His servant and messenger. First and foremost, I must acknowledge my limitless thanks to Allah, the Ever-Magnificent; the Ever-Thankful, for His help and blessing. I am totally sure that this work would never have become true without His guidance.

Completion of this thesis would not have been possible without the invaluable help and support of a number of people. I would like to express my deepest gratitude to my supervisors Prof. Dr. Domenec Savi Puig Valls and Dr. Mohamed Abdelnasser, for their useful guidance, insightful comments, and considerable encouragement to complete this thesis. They have guided me to pursue important problems that will have a practical impact and were always available to guide me whenever I approached them. They consistently allowed this thesis to be my work but steered me in the right direction whenever they thought I needed it. This work would not have been completed without their encouragement and patience.

I would like to take this opportunity to say warm thanks to all my beloved friends and family, who have been so supportive along the way of doing my thesis. I also would like to express my wholehearted thanks to my family for the generous support they provided me throughout my entire life, particularly through the process of pursuing the Ph.D. degree. Because of their unconditional love and prayers, I have the chance to complete this thesis.

I owe profound gratitude to my Husband, Abdul Qayyum, and my siblings Saima, Saira, and Asad whose constant encouragement, limitless giving, and great sacrifice, helped me accomplish my degree.

My parents receive my deepest gratitude and love for their dedication and the many years of support during my studies that provided the foundation for this work.

Contents

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

List of Figures

List of Tables

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

List of Abbreviations

ET	**E**nhancing **T**umor
TC	**T**umor **C**ore
WT	**W**hole **T**umor
BraTS	**B**rain **T**umor **S**egmentation
AI	**A**rtificial **I**ntelligence
CNNs	**C**onvolutional **N**eural **N**etworks
ANN	**A**rtificial **N**eural **N**etwork
BCE	**B**inary **C**ross **E**ntropy
BN	**B**batch **N**normalization
MRI	**M**agnetic **R**esonance **I**maging
SE	**S**Squeeze and **E**xcitation
DCAM	Depth-wise **C**hannel **A**ttention **M**odule
DWT	**D**iscrete **W**avelet **T**ransform
GT	**G**round **T**ruth
ED	**E**dema
NCR	**N**ecrotic **T**umor
DSAM	**D**ense **S**patial **A**ttention **M**odule
CL	**C**ontrastive **L**earning
SSL	**S**elf **S**upervised **L**earning
MMPATrans	**M**ultiview **M**ultiscale **P**arallel **A**ttention **T**ransformer
DSC	**D**ice **C**oefficient
HD	**H**ausdorff **D**istance
DL	**D**eep **L**earning
CV	**C**omputer **V**ision

DCNNs	**D**eep **C**onvolutional **N**neural **N**etworks
MSE	**M**ean **S**quared **E**rror
ReLU	**R**ectified **L**inear **U**nit
RMSE	**R**oot **M**ean **S**quare **E**rror
MSAFE	**M**ulti **S**cale-aware **F**eature **E**nhancement
HAM	**H**ierarchica **C**hannel **A**ttention
RF	**R**andom **F**orest
DT	**D**ecision **T**ree
SVM	**S**upport **V**ector **M**achine
LASSO	**L**east **A**bsolute **S**hrinkage and **S**election **O**perator
MLP	**M**ultilayer **P**erceptron

This thesis is dedicated to The sake of Allah, my Creator, and my Master, My great teacher and His messenger, Mohammed (P.B.U.H)
My dearest Husband, who leads me through the valley of darkness with the light of hope and support
My beloved brothers and sisters who stand by me when things look bleak
This work is exceptionally dedicated to my parents, who have always loved me unconditionally and whose good guidance, affection and moral support help me to work hard for the things that I aspire to achieve in my life.

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Part I

Introduction

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Chapter 1

Introduction

The objective of this chapter is to situate the thesis in its clinical and technical context. First, we discuss aspects that are important to understand the clinical motivations of the proposed methods. We start by defining brain tumors, which are the focus of this thesis. We then present the commonly used types of medical images, and we discuss the role of segmentation tasks in neuro-oncology. Second, we discuss important aspects related to deep learning, which form the basis of most of the current state-of-the-art methods for image segmentation. In particular, we discuss its main advantages and inconveniences in the context of the segmentation of brain tumors and organs at risk. Finally, we introduce our main contributions, and we present the organization of the thesis.

1.1 Motivation

1.1.1 Brain Tumor

Brain tumor/cancer is a life-threatening disease involving abnormal proliferation of cells. It originates from one cell which developed several characteristics, often called hallmarks of cancer (Hanahan and Weinberg, 2011). In particular, cancer cells are able to replicate infinitely and autonomously, they invade other tissues and ignore natural regulatory

mechanisms such as programmed cell death. The capacity of invading neighboring tissues and spreading to distant sites of the body (metastasis) distinguishes malignant tumors (cancers) from benign ones. Every year, approximately 9 million people in the world die from different forms of cancer.

Brain tumors can be malignant or benign. A cancerous brain tumor is malignant, meaning it can grow and spread to other parts of the body. The most common types of malignant brain tumors (Mehta, 2011) are gliomas (Schwartzbaum et al., 2006). Gliomas originate from glial cells, which are one of the two main components of the nervous tissue (with neurons) and have several functions related to the support and protection of neurons. Gliomas represent approximately 80 % of primary malignant brain tumors and their malignant forms, such as glioblastoma, are among the most aggressive cancers. More than 11,000 people are diagnosed with a primary brain tumor in the UK each year, of which about half are cancerous. Many others are diagnosed with a secondary brain tumor.

Early detection, automatic delineation, and volume estimation are vital tasks for survival prediction and treatment planning. The work presented in this dissertation focused on automatic solutions for the early detection of malignant brain tumor characterization and survival days prediction for the prognosis analysis of brain tumor patients.

However, gliomas are often difficult to localize and delineate with conventional manual segmentation due to their high variation of shape, location, and appearance. Moreover, manual mark delineation is laborious and time-consuming work for a neurosurgeon.

1.1.2 Medical imaging in neuro-oncology

Medical images are extensively used in oncology for diagnosis, therapy planning, and monitoring of tumors. Oncologists analyze images to locate tumors and assess their different characteristics.

Different types of medical images are used, depending on the task (search of metastases, radiotherapy planning) and the region of interest (brain, lungs, digestive system). The commonly used types of imaging include computed tomography (CT), magnetic resonance imaging (MRI), and positron-emission tomography (PET).

Positron-emission tomographyGambhir, 2002 is based on the injection of a radioactive tracer in the blood of the patient in order to observe the metabolism of different tissues. A commonly used tracer is fludeoxyglucose which is a structural analog of glucose. As cancer cells need an important glucose supply due to their divisions, the tumoral tissues may be detected by their abundant absorption of the radioactive tracer. A PET scan is particularly useful for the diagnosis and staging of tumors, for detecting cancer metastases, and for monitoring the effects of therapy. However, due to physical limitations, PET scans have usually a considerably lower spatial resolution than MRI and CT scans.

Computed tomography measures the absorption of X-rays of different tissues in the body. The radiation is emitted from different angles in order to acquire a series of 2D radiographic images from which a 3D scan is then reconstructed. Even if CT scans have generally a better spatial resolution than MRI, they offer a significantly weaker contrast between soft tissues such as the ones present in the brain. Moreover, exposure to X-rays may induce cancers by damaging the DNA of body cells.

Acquisition of MRI Atlas, 2009 is based on the detection of signals emitted by the nuclear magnetic resonance of atoms in the body. The detected signal is usually produced by protons of hydrogen, present in abundance in the human body (water, fat). The atoms are set in a strong magnetic field and are then perturbed by a radio wave, called pulse sequence. By modifying the parameters of the pulse sequence and pulsed field gradients, different contrasts are obtained, corresponding to specific MRI sequences. The MRI sequences commonly used for brain tumor imaging are T1, T2, and FLAIR (T2 with suppression of fluids). T1

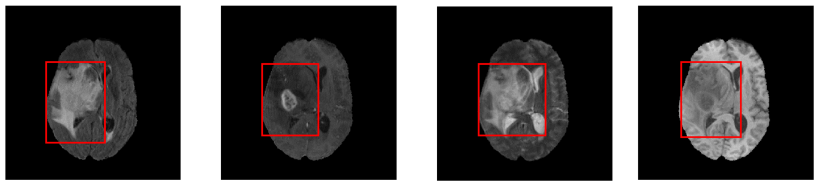


Figure 1.1: Four images from T1, T2, T1CE, and Flair MRI modalities with brain tumors having different shapes, sizes, and textures (Brain tumor is highlighted in red boxes).

is often acquired after injection of a gadolinium-based contrast agent Zhou et al., 2013 in the blood of the patient, in particular, to highlight the tumor angiogenesis, i.e. creation of new vascular networks by the tumor.

MRI modalities include Fluid-Attenuated Inversion Recovery (FLAIR), T1-weighted (T1), T2-weighted (T2), and T1w contrast-enhanced (T1CE). Each MRI modality provides different structural information about brain tissue. Some brain tumor types such as gliomas and glioblastoma cells spread rapidly and are difficult to segment due to low contrast region as compared to meningiomas that can segment easily. Gliomas show the progression of infiltrative growth of tumor boundaries and do not have a consistent growth pattern. These tumors have complicated pathological changes such as hemorrhage, necrosis, and edema. Due to these complex pathological changes, gliomas show intricate changes in brightness and texture on MRI scans. Depending on its severity and aggressiveness, it is divided into four grades ranging from grade I to grade IV(Grade I, II are Low-Grade Glioma(LGG) and grades III and IV are High-Grade Glioma(HGG)). Different tissues may have the same gray matter values, which makes it difficult to have an accurate, repeatable, and stable segmentation of gliomas. Figure1.1 shows examples of brain tumors in different MRI modalities.

Magnetic resonance images are particularly suitable for imaging of

brain tumors and organs. In particular, they offer a high contrast between soft tissues in the brain (compared to other types of imaging) and the use of different MRI sequences offers the possibility to highlight different tumoral compartments (edema, tumor vascularisation, necrosis).

1.1.3 Deep learning in medical imaging & Survival Prediction

The methods presented in this thesis are mainly based on deep learning, which is a branch of machine learning. In this section, we briefly present the general principles of deep learning, we motivate its use for segmentation tasks in neuro-oncology and we discuss its limitations, some of which are addressed in this thesis.

Given an input space X and a label space Y, the objective of supervised machine learning is to find a predictive function $f : X(Y)$, using a database of training examples (x_i, y_i), where $x_i \in X$ and $y_i \in Y$. To achieve this goal, three main elements have to be defined:

• Family of candidate functions fθ, parametrized by a vector of parameters $\theta \in \Theta$.

• Loss function L: $\Theta \rightarrow R$, which quantifies the mismatch between the outputs predicted by a candidate function $f\theta$ and the ground truth.

• Training algorithm, which minimizes the loss function (with respect to the parameters θ) over the training data.

The main particularity of deep learning is the nature of the considered candidate functions. The term deep is related to multiple compositions of functions. The composed functions are differentiable and organized in layers, with the idea to progressively transform the input vector, extracting more and more complex information. The term neural network is related to the considered family of functions, represented typically by a graph. Training of the model (minimization of the loss function) is typically based on iterative optimization with variants of the stochastic gradient descent.

Convolutional Neural Networks (CNN) (LeCun, Bengio, et al., 1995) are a commonly used type of neural network for image processing and analysis (classification, segmentation). They exploit spatial relations between pixels (or voxels, in 3D) and are based on the application of local operations such as convolution, pooling (maximum, average), and upsampling. The objectives of such a design are to limit the number of parameters of the network and to limit computational costs, as images correspond generally to very large inputs. In fact, an important property of the operations used in CNNs is that they can be parallelized and efficiently computed on a Graphical Processing Unit (GPU).

CNNs for image segmentation are usually trained in an end-to-end manner, i.e. their input is the image and the output is the segmentation. With end-to-end training, the model automatically learns to extract relevant information from images, using the training database. This property is particularly important for very challenging tasks in medical imaging, such as tumor segmentation. Most of the current state-of-the-art methods for image segmentation are based on CNNs, in particular the methods for brain tumor segmentation (Kamnitsas et al., 2017a).

However, even if CNNs have recently obtained state-of-the-art results in many recognition tasks, they still have important limitations in the context of segmentation in medical imaging. The objective of the methodological contributions of this thesis is to address these limitations. Despite the progress of GPU capacities, computational costs still severely limit the potential of CNNs for segmentation tasks in medical imaging. A typical segmentation network, such as U-net Ronneberger, Fischer, and Brox, 2015a performs thousands of convolutions, max-poolings, and upsampling operations. Outputs of these operations have to be stored in the memory of the GPU during each iteration of the training, in order to compute gradients of the loss function by the Backpropagation algorithm. A typical MRI is composed of several millions of voxels. Training of neural networks for an end-to-end segmentation on entire MRIs requires therefore a huge amount of GPU memory and is

often may impossible using the currently available GPUs.

For this reason, current segmentation models are usually trained on subvolumes of limited size and have limited receptive fields. Another important problem is the cost of the ground truth annotations necessary to train neural networks, and machine learning models in general. Manual segmentation of tumors is particularly costly as it is not only time-consuming but also requires medical expertise and therefore has to be performed by experienced clinicians. Other difficulties are related to the use of multimodal data. Usually, different types of images (e.g. different MRI sequences) are used in oncology. Most of the current CNN-based models consider MRIs as 4D tensors and assume the presence of all modalities for all patients in the training database, which is rarely the case in practice. Finally, commonly used segmentation CNNs may produce spatially inconsistent results, as they are based on individual classification of voxels given their receptive fields. It means that, in general, the model does not explicitly analyze aspects related, for instance, to the connectivity of the output segmentation and the spatial relations between the different segmentation classes.

The accurate segmentation of brain tumors can facilitate diagnosis and help assess the prognosis and severity of the disease. Automatic segmentation and survival rate prediction models will help the diagnosis and treatment to be much more accurate and faster.

1.1.4 Survival Prediction

Survival for brain tumors depends on different factors likewise it is different for adults and children. Different types of brain tumors respond differently to treatment. Some respond better to radiotherapy than others. Some types are likely to spread into the surrounding brain tissue. This may make them difficult to remove with surgery.

Glioblastoma (GBM) is the most common malignant primary brain tumor in adults. Prognosis is generally very poor, with a median overall survival (OS) of less than 15 months, and a 5-year OS rate of only 10%, even when aggressively treated. Magnetic resonance (MR) images of glioblastoma patients contain vast amounts of information about the disease, some of which may carry prognostic value. The literature on imaging biomarkers for glioblastoma survival prediction is currently dominated by radiomics, an approach in which hundreds or even thousands of features are extracted from delineated tumor regions of MR images, each quantifying some shape, texture, wavelet, or histogram property. This approach has shown good performance in predicting survival in many studies likely stemming from the correlation between the tumor's texture in MR images and its intratumoral heterogeneity and aggressiveness. However, despite good prediction performance, radiomics suffers from three issues impeding wide-scale practical adoption:

• Lack of interpretability: Radiomic features, instead of aiming to be interpretable, are designed to be many, to maximize the chance of some having correlation to the target variable. Consequently, many radiomic features are seemingly arbitrary and hard to connect in a meaningful way to the nature of the disease. However, the interpretability of features is important: If a model cannot give interpretable explanations of its predictions, clinicians may not trust the model enough to factor its predictions into their decisions, even if the model is accurate. Interpretable models may also uncover patterns in the data that give valuable new insight into the disease, and inspire future research.

• Difficulties generalizing: The reproducibility of studies using radiomics has been shown to be less than ideal, with results failing to generalize well across scanners and software implementations. Since many radiomic features depend directly on raw image intensities, they are sensitive to subtle changes in scanning equipment and image acquisition parameters. Furthermore, both textural and shape features depend on

the segmentation mask that is used, underlining the importance of using image segmentation methods that are robust with respect to such sources of variation.

• Focus on pre-operative data: Compared to pre-operative images of glioblastoma, relatively little attention has been given to radiomics and other biomarkers in post-operative images. The reason may be that post-operatively, tumor shape and textural features are less easily detectable, as a large part of the tumor is usually removed. Nevertheless, post-operative images are collected closer to the time of disease progression and contain information about the success of the operation, making them important to consider in a survival model.

In this dissertation, we proposed deep learning and machine learning-based approaches that aim to start addressing these shortcomings.

1.2 Thesis overview & Organization

The objective of this thesis is to propose efficient methods for the segmentation of brain tumors and survival prediction for prognosis analysis. The four main chapters correspond to journal articles that have been published or submitted during the preparation of the thesis while one chapter is related to the book chapter published on the challenge from Medical Image Computing and Computer Assisted Intervention (MICCAI 2021). The manuscript is organized as follows.

In Chapter 2, published as a journal article 'BrainSeg-DCANet: Robust Multimodality Brain Tumor MR Image Segmentation Using Deep Depth-Wise Channel Attention' (Accepted: Intelligent Systems 2023), we introduce a CNN-based system for brain tumor segmentation. Most existing deep learning-based methods cannot precisely identify small-scale tumors due to the insufficient feature extraction capability of utilized models. To handle this issue, we propose BrainSeg-DCANet,

which is a fully automated yet robust deep learning-based multimodality brain tumor MR image segmentation method. In particular, an efficient depth-wise attention module is proposed to capture small brain tumor-relevant features and to localize the small tumor regions accurately. This module can efficiently capture the local cross-channel interaction to produce brain tumor contextual feature information using a dynamic 1D convolution layer. Besides, a discrete wavelet transformed (DWT) is employed to enlarge the receptive fields. The performance of BrainSeg-DCANet is assessed on the BraTS2020 multimodal brain tumor segmentation dataset and obtained one of the best performances over available methods.

In Chapter 3, corresponding to the second journal article (Mazher et al., 2022), This chapter proposed an end-to-end automatic yet effective method for a multi-tissues fetal brain segmentation model called IRMMNET. It included the Inception-Residual encoder block (EB) and the dense-spatial attention (DSAM) block that facilitates extracting multi-scale fetal brain tissue-relevant information from the multi-view MRI images and enhances feature reuse and substantially reduces the number of parameters of the segmentation model. Later, this proposed technique is also applied on the BraTS2020 dataset for examining its outcome on brain tumor segmentation. Additionally, we proposed three methods for predicting gestational age (GA) prediction using the 3D autoencoder, using radiomics features, and using the IRMMNET segmentation model deep features. Our experiments were performed on a dataset of 80 manually segmented pathological and non-pathological fetal magnetic resonance brain volume reconstructions across a range of gestational ages (20 to 33 weeks) into 7 different tissue categories. Similar to the segmentation task, these gestational age prediction methods were also deployed on the BraTS2020 dataset for patients' survival days prediction.

In Chapter 4, corresponding to the third journal article (Two-stage

Self-supervised Contrastive Learning using Parallel Multiview Multiscale Attention-based CNN Transformers for 3D Brain Tumour Volumetric Segmentation), To reduce the requirement for annotated training data, self-supervised pre-training strategies on non-annotated data were designed. Especially contrastive learning schemes operating on dense pixel-wise representations have been introduced as an effective tool. In this work, we expand on this strategy and leverage inherent anatomical similarities in medical imaging data. We apply our approach to the task of semantic segmentation in a self-supervised setting with limited amounts of annotated volumes. Trained alongside a segmentation loss in one single training stage, a contrastive loss aids to differentiate between salient anatomical regions that conform to the available annotations. We proposed an efficient parallel transformer module using Multiview multiscale features and depth-wise features. The multiEncoder-based proposed transformer is trained on a self-supervised approach using contrastive loss. First, we prepared the proposed transformer with an unlabelled dataset and fine-tuned one encoder coming from the first stage and a second encoder trained with a few annotated segmentation masks. Further, we concatenated the features of both encoders for final brain tumor segmentation. MultiEncoder-based transformers achieved excellent results in various medical image segmentation tasks.

In Chapter 5, corresponding to the fourth journal article (Accepted: Survival Prediction of Brain Tumor Patients Using Attention-Guided 3D Deep Learning with Radiomics Features from Multimodality MRI Images), a 3D MR image-based survival prediction task was performed. Multiple feature extraction techniques including 3D radiomics features, 3D regressor features, and 3D deep CNN features were used to extract the features from the volumetric MRI images and then different regression techniques were applied to the extracted features. This method was used on BraTS 2020 dataset. Also, it was applied to the Head & Neck Tumor dataset (HECTOR2021) to test its generalization capability.

Chapter 6 frames the validation of the proposed solutions on real-time medical imaging challenges published as book various chapters to authenticate the generalization capability of the proposed solutions which can be applied as automated disease analysis and progression systems.

Finally, in Chapter 7, we summarize the contributions of the thesis and we propose directions for future research works.

1.3 Problem statements and challenges

1. Early detection, automatic delineation, and volume estimation are vital tasks for survival prediction and treatment planning. However, gliomas are often difficult to localize and delineate with conventional manual segmentation due to their high variation of shape, location, and appearance.

2. Manual mark delineation is laborious and time-consuming work for a neurosurgeon.

3. The accurate segmentation of brain tumors is crucial in facilitating diagnosis and helping assess the prognosis and severity of the disease.

4. Survival prediction for brain tumor patients relies on multiple factors that vary from person to person and from child to adult.

5. Different types of brain tumors respond differently to treatment. Some respond better to radiotherapy than others. Some types are likely to spread into the surrounding brain tissue. This may make them difficult to remove with surgery. Therefore, a manual assessment for survival days is time-consuming, costly, and often not an accurate solution

6. Automatic segmentation and survival rate prediction models will help the diagnosis and treatment to be much more accurate and faster.

1.4 Objectives and scope of the research

The main objective of this thesis is to create an advanced ADA (Automatic Disease Analysis) system for the segmentation, characterization, and survival days prediction for the diagnosis and prognosis of brain tumor patients from multimodality MRI imaging.

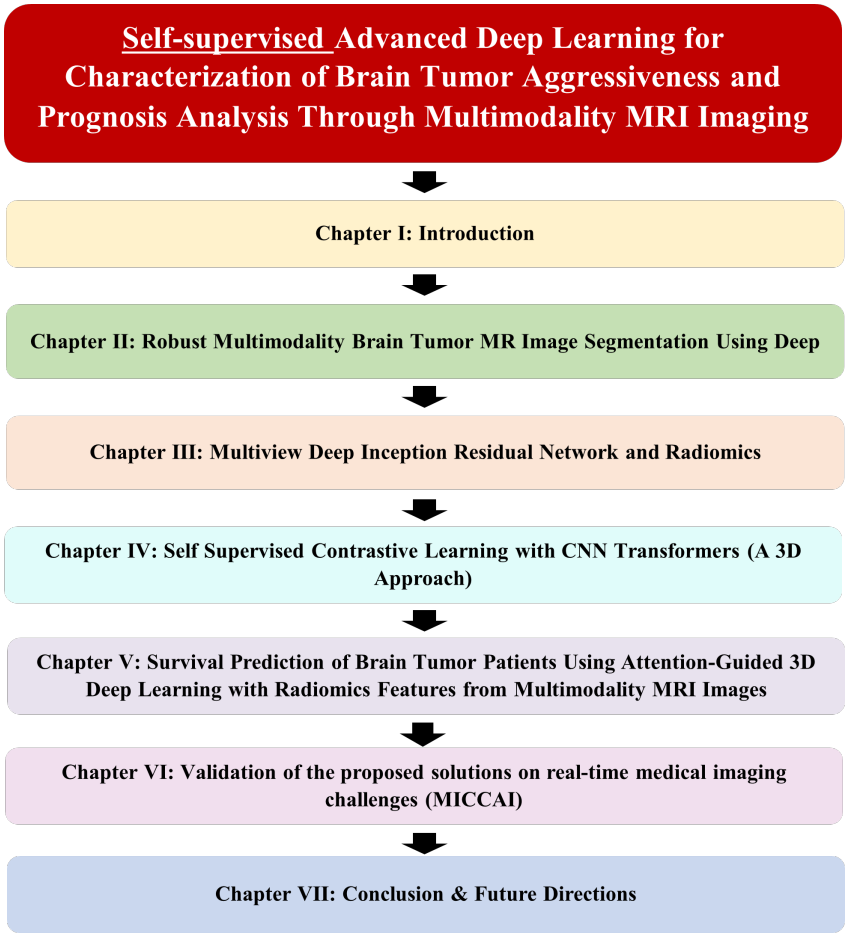


Figure 1.2: Organization of the thesis.

The significant objective of this thesis detailed is as follows:

1. Developed deep learning models for volumetric multimodalities MRI dataset for end-to-end automatic brain tumor segmentation. The manual labeling for supervised segmentation models is time-consuming and a costly procedure for training large deep-learning models. To overcome this issue, we have deployed a method that used a large unlabelled and a small labeled MRI dataset for brain tumor segmentation.

Medical image segmentation has seen significant progress using supervised deep learning. Hereby, large annotated datasets were employed to segment anatomical structures reliably. We proposed a Two-stage Self-supervised Contrastive Learning using Parallel Multiview Multiscale Attention-based CNN Transformers for 3D Brain Tumour Volumetric Segmentation for handling large unlabelled datasets to efficiently reduce the procedure cost and time.

We have also proposed 2D and 3D segmentation models with attention-guided and depth-wise attention approaches for efficient segmentation of brain tumors.

2. We have covered all possible features from the input MRI dataset of brain tumor patients for survival days prediction that can contribute to generating vital results including clinical features, radiomics features, and CNN-based features.

Developed survival time predictive models using radiomics features, deep features extracted from 3D regressor models, latent features from 3D autoencoder models basic imaging, and clinical features using machine learning models. We have customized parameters in the classical machine learning models like the random forest, gradient boosting, and support vector machine.

3. To analyze the more contributed features for diagnosis and prognosis of the brain tumor from risky patients using multimodalities brain tumor dataset. To measure and analyze the expandability of proposed models in feature space and used the most prominent features for decision-making.

4. To validate and measure the performance of proposed segmentation models, participated in several real-time MICCAI (Medical Image Computing and Computer Assisted Intervention) challenges like brain, heart, and abdominal CT and MRI datasets and ranked top position in those challenges.

1.5 Scientific Dissemination

1.5.1 Journal

1. BrainSeg-DCANet: Robust Multimodality Brain Tumor MR Image Segmentation Using Deep Depth-Wise Channel Attention (IEEE Intelligent Systems) (2023).

2. Fetal Brain Tissue Annotation and Segmentation Challenge Results. (Medical Image Analysis)(2023).

3. Deep Learning Segmentation of the Right Ventricle in Cardiac MRI: The M&ms Challenge. (Journal of Biomedical and Health Informatics)(2023).

4. Effective Approaches to Fetal Brain Segmentation in MRI and Gestational Age Estimation by Utilizing a Multiview Deep Inception Residual Network and Radiomics. (Entropy)(2022).

5. Semi-supervised 3D-InceptionNet for segmentation and survival prediction of head and neck primary cancers. (Engineering Applications of Artificial Intelligence)(2023).

6. Two-stage Self-supervised Contrastive Learning using Parallel Multiview Multiscale Attention-based CNN Transformers for 3D Brain Tumour Volumetric Segmentation (2023) (Under-review).

7. Survival Prediction of Brain Tumor Patients Using Attention-Guided 3D Deep Learning with Radiomics Features from Multimodality MRI Images. (Neural Computing) (2023) (Accepted).

8. 3D-IncNet: Head and Neck (H&N) Primary Tumors Segmentation and Survival Prediction. (IEEE Journal of Biomedical and Health Informatics)(2022).

9. Spontaneous Facial Behavior Analysis using Deep Transformer Based Framework for Child–Computer Interaction. (ACM Transactions on Multimedia Computing, Communications, and Applications) (2022).

10. Progressive ShallowNet for large-scale dynamic and spontaneous facial behavior analysis in children. (Image and Vision Computing)(2022).

11. FetReg2021: A Challenge on Placental Vessel Segmentation and Registration in Fetoscopy, " Medical Image Analysis " Journal paper. (Submitted Jun 2022).

1.5.2 Book Chapters

1. Mazher, M., Qayyum, A., Abdel-Nasser, M., & Puig, D. (2023). Automatic Semi-supervised Left Atrial Segmentation Using Deep-Supervision 3DResUnet with Pseudo Labeling Approach for LAScarQS 2022 Challenge. In Left Atrial and Scar Quantification and Segmentation: First Challenge, LAScarQS 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings. (pp. 153-161). Cham: Springer Nature Switzerland.

2. Mazher, M., Qayyum, A., Benzinou, A., Abdel-Nasser, M., & Puig, D. (2022, January). Multi-disease, Multi-view and Multi-center Right Ventricular Segmentation in Cardiac MRI Using Efficient Late-Ensemble Deep Learning Approach. In Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge: 12th International Workshop, STACOM 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Revised Selected Papers (pp. 335-343). Cham: Springer International Publishing.

3. Qayyum, A., Mazher, M., Niederer, S., & Razzak, I. (2023). Segmentation of Intra-operative Ultrasound Using Self-supervised Learning Based 3D-ResUnet Model with Deep Supervision. In Lesion Segmentation in Surgical and Diagnostic Applications: MICCAI 2022 Challenges, CuRIOUS 2022, KiPA 2022 and MELA 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18–22, 2022, Proceedings (pp. 55-62). Cham: Springer Nature Switzerland.

4. Qayyum, A., Mazher, M., Niederer, S., Meriaudeau, F., & Razzak, I. (2023, January). Automatic Cardiac Magnetic Resonance Respiratory Motions Assessment and Segmentation. In Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers: 13th International Workshop, STACOM 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Revised Selected Papers (pp. 485-493). Cham: Springer Nature Switzerland.

5. Qayyum, A., Benzinou, A., Mazher, M., Abdel-Nasser, M., & Puig, D. (2022). Automatic segmentation of head and neck (H&N) primary tumors in PET and CT images using 3D-Inception-ResNet model. In Head and Neck Tumor Segmentation and Outcome Prediction: Second Challenge, HECKTOR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings (pp. 58-67). Cham: Springer International Publishing.

6. Qayyum, A., Ahamed Khan, M. K. A., Benzinou, A., Mazher, M., Ramasamy, M., Aramugam, K., & Suresh, M. (2022). An Efficient 1DCNN–LSTM Deep Learning Model for Assessment and Classification of fMRI-Based Autism Spectrum Disorder. In Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021 (pp. 1039-1048). Singapore: Springer Nature Singapore.

7. Qayyum, A., Benzinou, A., Mazher, M., & Meriaudeau, F. (2022). Efficient multi-model vision transformer based on feature fusion for

classification of dfuc2021 challenge. In Diabetic Foot Ulcers Grand Challenge: Second Challenge, DFUC 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings (pp. 62-75). Cham: Springer International Publishing.

Part II

2D Deep Learning Approach for Brain Tumor Segmentation & Survival Prediction

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Chapter 2

Multimodality Brain Tumor MR Image Segmentation Using Deep Depth-Wise Channel Attention (A 2D Deep Learning Approach)

Manual brain tumor segmentation is a critical and time taking process. Accurate brain tumor segmentation is key to better prognosis and treatment planning. Most existing deep learning-based methods cannot precisely identify small-scale tumors due to insufficient feature extraction capabilities. To handle this, we propose BrainSeg-DCANet, a fully automated yet robust deep learning-based brain tumor MR image segmentation method. In particular, an efficient depth-wise attention module is proposed to capture small brain tumor-relevant features and to localize the small tumor regions accurately. Besides, a discrete wavelet transform (DWT) is employed to enlarge the receptive fields. The performance of BrainSeg-DCANet is assessed on the BraTS2020 multimodal

brain tumor segmentation dataset. BrainSeg-DCANet outperforms the state-of-the-art methods in segmenting the whole tumor, and tumor core, and enhancing tumor (WT, TC, and ET) structures. Interestingly, BrainSeg-DCANet showed promising results when tested with the SegTHOR2019 dataset, and this proves the generalization capability of BrainSeg-DCANet.

2.1 Introduction

Gliomas are the main common primary brain malignancies and are the most common type of brain cancer Saut et al., 2014. It comes from glial brain cells that support nerve cells for a broad category of brain and spinal cord tumors. Identification of the sub-regions of gliomas before surgery is critical, which may affect the survival of patients. Different medical imaging techniques have been used to distinguish brain tumor lesions such as X-ray, Ultrasound, Positron Emission Tomography (PET), Electroencephalography (EEG), Magnetic Brain Wave Graph (MEG), Single-Photon Emission Computed Tomography (SPECT), Computed Tomography (CT), and Magnetic Resonance Imaging (MRI). However, MRI is the most comprehensive technique that provides better image quality and facilitates determining the exact volume and size of the malignant brain tumor. MRI modalities include Fluid-Attenuated Inversion Recovery (FLAIR), T1-weighted (T1), T2-weighted (T2), and T1w contrast-enhanced (T1CE). Each MRI modality provides different structural information about brain tissue.

Some brain tumors such as gliomas and glioblastoma cells spread rapidly and are difficult to segment due to low contrast region as compared to meningiomas that can segment easily Menze et al., 2014. Gliomas show the progression of infiltrative growth of tumor boundaries and do not have a consistent growth pattern. These tumors have complicated pathological changes such as hemorrhage, necrosis, and edema. Due to these complex pathological changes, gliomas show intricate changes in brightness and texture on MRI scans. Different tissues may have the

same gray matter values, which makes it difficult to have an accurate, repeatable, and stable segmentation of gliomas.

The importance for a neurosurgeon is to delineate precisely the tumor region. Unfortunately, manual mark delineation is laborious and time-consuming work for a neurosurgeon. In addition, it is difficult to replicate the segmentation results due to certain practical operation factors. The accurate segmentation of brain tumors can facilitate diagnosis and help assess the prognosis and severity of the disease.

In the last years, 2D and 3D convolutional neural networks (CNNs) are widely used for the classification and segmentation of medical images. For 2D segmentation, the input data is passed slice by slice to the CNN model while in 3D segmentation the whole volumetric data is fed into the CNN model. Çiçek et al., 2016Isensee et al., 2018 have proved the better performance of 3D architectures in comparison to 2D architectures. Although, 3D models are better than 2D architectures but due to having a large number of parameters they are computationally expensive Noori, Bahri, and Mohammadi, 2019.

Deep learning-based brain tumor segmentation techniques can be categorized as follows: 1) CNNs use a small patch-based classification method for brain tumor segmentation and 2) CNNs based on the fully convolutional network (FCN) that comprise encoder and decoder networks. UNet Ronneberger, Fischer, and Brox, 2015b has been widely used to segment biomedical images. The UNet architecture is based on the FCN concept. Havaei et al., 2017 proposed a CNN-based architecture to extract contextual features using convolutional kernels of various sizes for brain tumor segmentation regions on BraTS 2013. They used a patch-based pixel-wise classification technique to segmentation brain tumors and cascaded multiple models using majority voting methods. Their method produced a dice score of 0.84, 0.71, 0.57 for WT, TC, and ET, respectively.

Pereira et al., 2016 used patch-based VGG-Net for automated brain tumor segmentation on BraTS 2013 dataset. They used different patches

based on CNN's for pixel-wise segmentation of brain tumors. They achieved dice scores of 0.88, 0.83, and 0.77 for the whole tumor (WT), Tumor core (TC), and Enhancing tumor (ET), respectively.

The patch-based methods are limited with the need for large storage space and a lack of spatial continuity. Moreover, there are some limitations of patch-wise segmentation methods like they require high computational cost, and their performance is highly associated with the patch size of the image. As compared to patch-based CNN, FCN based methods could produce better performance and handle the spatial continuity problem Long, Shelhamer, and Darrell, 2014.

Dong et al., 2017 presented a 2D UNet segmentation model with real-time data augmentation techniques to improve the performance of brain tumor segmentation on the BraTS 2015 dataset. They obtained dice scores of 0 .86, 0 .86, and 0 .65 for WT, TC, and ET, respectively. Though, with the interpretation of the low memory capability, the U-Net model has limited effective feature learning capacity for complicated images task like brain tumor segmentation. This limitation could be overcome by optimizing the network that has widened parameter space to learn more demonstrative features.

Kong et al., 2018 proposed a UNet-based model with a feature pyramid module by incorporating the detailed location and extracting multi-scale semantic information to improve the precision of brain tumor segmentation on the BraTS2017 dataset. They achieved dice scores of 0.92, 0.80, 0.76 for WT, TC, and ET, respectively. The feature pyramid approach consumes a long time and has a computational load.

Wang et al., 2019a proposed the Test-Time Augmentation (TTA) technique and integrated it with UNet. They used three deep learning models for axial, sagittal, and coronal with TTA. They achieved dice scores of 0 .90, 0 .85, and 0 .79 for WT, TC, and ET, respectively. However, the three models with TTA have higher computational costs than UNet. In Li, Li, and Wang, 2019, dilated convolution, skip connection, dense block, and embedding MultiRes block have been used to enhance the

precision of brain tumor segmentation. The authors of Li, Li, and Wang, 2019 used an inception module that caters to more computational costs in terms of network parameters. They obtained dice scores of 0 .89, 0 .73, and 0.72 for WT, TC, and ET on the BraTS2015 dataset. Moreover, various Attention Gates (AGs) have been integrated with the UNet model to improve the segmentation performance Schlemper et al., 2019. The Squeeze-and-Excitation (SE) blocks have been used in UNet for prostate zonal segmentation on MR images Hu et al., 2017.

The aforementioned methods somehow produced optimal performances that would not be sufficient to deploy for clinical brain tumor segmentation applications. However, most of the aforementioned methods are not able to identify small brain tumors due to the limited capability of the deep learning feature extractors to locate the small tumors. Moreover, these methods are computationally expensive. In an attempt to handle these two issues, in this chapter, we proposed a depth-wise attention module that helps enhance the extraction of small brain tumor-relevant features and helps efficiently locate the small tumor regions.

Specifically, a fully automated and robust deep learning-based approach named 'Depth-wise Channel Attention' (BrainSeg-DCANet) is presented to accurately segment the Whole Tumor, Tumor Core, and Enhancing Tumor (i.e., WT, TC, and ET) structures. The proposed depth-wise attention module can capture the local cross-channel interaction in an efficient way to produce more powerful contextual feature information using a dynamic 1D convolution layer. Furthermore, a discrete wavelet transformed (DWT) is used to enlarge the receptive fields of the segmentation model. The proposed brain tumor segmentation model can extract the prevailing contextual features and produce a better performance as compared to state-of-the-art brain tumor segmentation methods with low computational cost.

2.2 Proposed tumor segmentation model

This section presents architectural details of the proposed deep learning model BrainSeg–DCANet for brain tumor segmentation.

2.2.1 The proposed BrainSeg-DCANet framework

Figure 2.2 shows the proposed model, which includes encoder and decoder networks. In previous studies, multimodal data are stacked like the multichannel RGB images Pereira et al., 2016Kamnitsas et al., 2017b from four MRI modalities (T1, T1CE, T2, Flair). Similarly, this model receives stacked MRI images as input and produces a mask for each tumor class (enhancing, non-enhancing/necrotic, peritumoral edema) highlighting the tumor regions. As shown in Figure 2.2, there are seven main components in the proposed model: DWT, convolutional block, Max-Pooling blocks, Up-Sampling blocks, 1x1 Conv, depth-wise channel attention module (DCAM) block.

The raw multimodal MRI images are processed using DWT. The output of the DWT is passed to the encoder network. The convolutional block consists of convolutional layers with Batch-Normalization and ReLU activation function to extract the different feature maps from each block in the encoder side. The 2D max-pooling layer has been used to reduce the input image spatial size. The 1x1 convolutional layer with softmax function has been used at the end of the proposed model. In the encoder block, the spatial input size is reduced with increasing the number of layers, while on the decoder side, the input image spatial size is recovered via a 2D upsampling layer using a bilinear upsampling method.

The proposed DCAM block handles the input feature maps extracted from each convolutional block in the encoder side and further passed these feature maps into the proposed residual module (a detailed explanation is given in the next section). The output of each encoder block is

fed into the proposed DCAM (green lines in Figure 2.1). Later, the outputs of DCAMs are passed to the residual blocks before concatenating them with the corresponding decoder blocks (red dotted lines in Figure 2.1). The output of the last encoder block is passed to the DCAM and the resulting feature maps are directly concatenated with the corresponding decoder block. Each max-pooling layer receives the output from DCAM (maroon dotted lines).

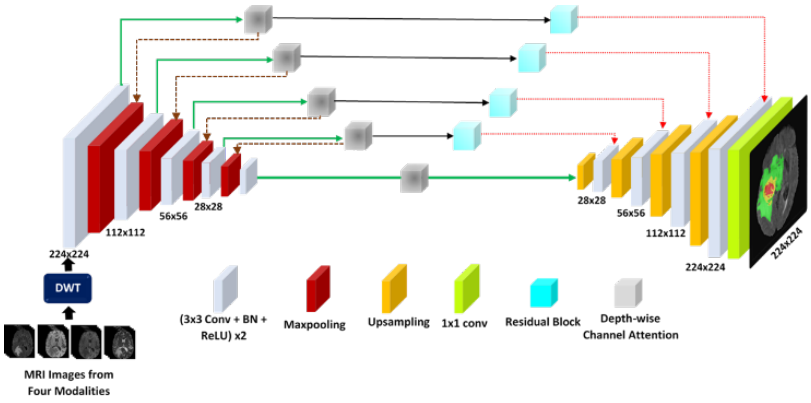


Figure 2.1: Diagram of the proposed BrainSeg-DCANet Model with the proposed depth-wise channel attention module and the residual block.

As shown in Figure 2.1, the spatial size is doubled at each decoder layer of the proposed model. The feature concatenation is done at each encoder and decoder block except the last 1x1 convolutional layer. In the following subsections, the proposed Residual. DWT, DCAM blocks are explained in detail.

2.2.2 Proposed Residual Block

The Residual network (ResNets) He et al., 2015 has been widely used for deep learning-based image classification and segmentation. Each

residual block comprised of two paths, the first path either used identity mapping or used 1x1 convolutional layer with Batch Normalization (BN), and the second path consisted of a series of layers such as ReLU, convolutional, and BN. These two paths are summed together to get the final output from the residual block. The residual block has three advantages as compared to traditional CNN-based models: 1) the gradient can flow continuously in deep networks by updating deeper network parameters, 2) introducing identity operation using a single layer to better optimize parameters of very deep networks. 3), have the robust capability to perform similar operations in neighboring layers in deeper networks.

In this chapter, 1x1 Conv with BN is used as a skip connection with a series of several layers for boundary and structural information preservation in the residual block (see Figure 2.2). The main purpose of such residual blocks is the preservation of the feature maps within convolutional layer blocks that are used before each encoder block which helps bridge the semantic gap between the encoder and decoder while maintaining the same (or little increment) in the computational overhead for providing accurate segmentations. The structural information for feature maps could be restored by the addition of the residual blocks that aimed to preserve the fine-grained structures that would be useful and play an important role in medical image segmentation.

As shown in Figure 2.2, the proposed residual module consisted of 3x3 Conv, 1x1 Conv, BN, and ReLU activation layers. The 1x1 Conv layer is used in skip connection along with the BN layer (the bottom branch). The top branch includes two 3x3 Conv layers with BN and ReLU activation functions. In the end, the resulting outputs of the two branches are summed.

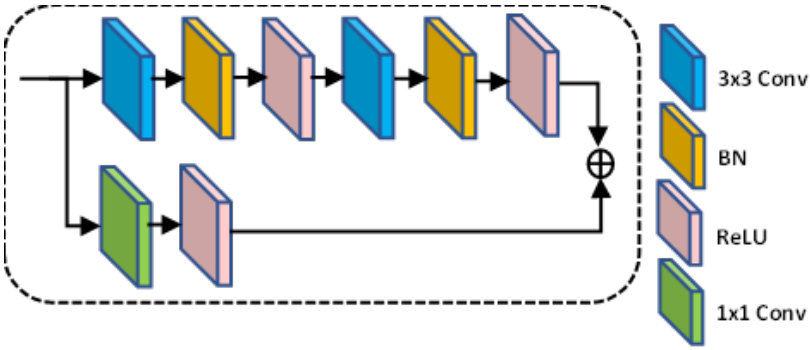


Figure 2.2: Proposed Residual Block.

2.2.3 Proposed Depth-wise Channel Attention Module

Attention mechanisms are used in deep CNNs to enhance the performance of image segmentation and classification models. The SE-Net Hu et al., 2017 was used for the first time as an attention module and achieved better performance using the channel attention mechanism. Furthermore, CBAM Woo et al., 2018a employed both average and max pooling to aggregate features. CBAM also rescales the activations Roy, Navab, and Wachinger, 2018 which are used to recalibrate the input tensors. Double Attention Networks (A2-Nets) Chen et al., 2018 used a function for Non-Local (NL) neural networks for video and image classification tasks. The Dual Attention Network (DAN) Fu et al., 2018 employed spatial and NL channels attention simultaneously for semantic segmentation. However, the NL-based attention modules used few convolutional layer blocks that could provide high computational complexity while implementing semantic segmentation. Noticeably, the aforementioned methods are used to develop a better attention module, but these methods are computationally expensive.

In an attempt to solve this issue, an efficient yet lightweight attention module is proposed in this article for brain tumor segmentation. The proposed DCAM module shown in Figure 2.3 can capture the local

cross-channel interaction in an efficient way and channel-wise convolutions by replacing (fully connected) layers in the channel attention module with a 1D convolution layer after channel-wise global average pooling without dimensionality reduction. The depth-wise convolutional module with BN and Swish activation effectively capture the features as compared to the standard convolutional module. It is inspired by some convolutional-based methods like group convolutions Xie et al., 2016 and depth-wise separable convolutions Ma et al., 2018.

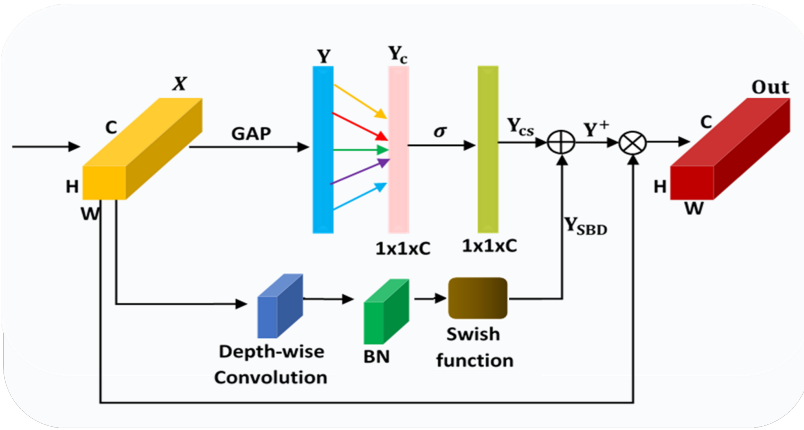


Figure 2.3: Diagram of the proposed depth-wise channel attention module (DCAM).

The group-wise convolutional layer extracts the high-dimensional channels involved in long-range convolutions to provide the fixed number of groups. Based on the group-based convolutional idea, the kernel size in 1D CNNs is proportional to the channel dimension.

The depth-wise convolutional layer consists of two layers. In the first layer, a single convolutional filter is used for each input channel to achieve lightweight filtering. At the second layer, there is a pointwise convolution is applied by using a 1x1 convolution. This layer builds a

new feature by computing linear combinations of input channels. Depth-wise convolution produced an efficient solution as compared to the standard convolutional layers. The Swish activation function proposed by Ramachandran, Zoph, and Le, 2017a effectively works in automatic search techniques based on reinforcement learning.

The Swish function would provide a better capability of generalization and optimization in training deep learning models Ramachandran, Zoph, and Le, 2017a due to non-monotonic and smoothness properties. As shown in Eq. (2.1), the Swish function is simple and it provides a better improvement in accuracy as it does not suffer from the vanishing gradient problems during training the deep learning models.

$$Swish = x.sigmoid(x) = \frac{x}{1 + e^{-x}} \tag{2.1}$$

The swish activation function is used after the batch normalization in the proposed DCAM module as shown in Figure 2.3.

Assume that $X \in R^{W \times H \times G}$ is the input volume that comes from the residual block in the proposed encoder module, and W, H, and C are the height, width, and the number of channels of input volume X. The channel-wise global average pooling can be expressed as:

$$GAP(X) = \frac{1}{W \times H} \sum_{i=1,j=1}^{W,H} X_{ij} \tag{2.2}$$

$$Y = GAP(X) \tag{2.3}$$

Where Y is the output features achieved from the global average pooling layer.

$$Y_c = Conv1 \times 1(Y) \tag{2.4}$$

Y_c is the output of the 1x1 Conv layer, where there can be used a dynamical number of filters and several features.

$$Y_{cs} = \sigma(Y_c) \tag{2.5}$$

The Y_{cs} is the output of the sigmoid activation function σ.

$$Y_{SBD} = Swish(BN(Convd(X))) \tag{2.6}$$

Where Y_{SBD} is the output of the depth-wise convolutional layer with BN and Swish activation function and *convd* is the depth-wise convolutional layer.

Y^+ is the output of the summation of Y_{cs} and Y_{SBD}.

$$Out = X * Y^+ \tag{2.7}$$

Out is the output of the proposed module.

2.2.4 Enlarging the Receptive Fields Using DWT

DWT uses filter banks for identifying both time and frequency resolutions at the same time. The frequency components information based on DWT is used to process the raw multimodalities of MRI brain images for brain tumor segmentation.

In this chapter, we employed DWT to extract contextual information from the raw multimodalities of MRI brain images. It would preserve the image's spatial and contextual information. Moreover, it will also extend the receptive fields. To decompose the MRI images into four different sub-bands i.eX_{LL}, X_{LH} , X_{HL} , X_{HH} the Haar filters are used. The decomposition process is expressed below:

$$\begin{cases} X_{LL} = X(2i-1,2j-1) + X(2i-1,2j) + X(2i,2j-1) + X(2i,2j) \\ X_{LH} = -X(2i-1,2j-1) - X(2i-1,2j) + X(2i,2j-1) + X(2i,2j) \\ X_{HL} = -X(2i-1,2j-1) + X(2i-1,2j) - X(2i,2j-1) + X(2i,2j) \\ X_{HH} = X(2i-1,2j-1) - X(2i-1,2j) - X(2i,2j-1) + X(2i,2j) \end{cases} \tag{2.8}$$

where X is the stacked image of the images from four MRI modalities (T1, T1CE, T2, Flair).

The DWT method used the low and high pass filters to decompose the MRI input into the time and frequency domain. Hence, the same frequency content is represented by the outputs of the low and high branches as the input MRI image. It is worth mentioning that DWT is associated with dilated filtering and pooling operation Liu et al., 2019. The dilating filtering is related to the decomposition of the input images into their sub-images.

2.2.5 Loss Function

In this chapter, the Combo loss function proposed by Taghanaki et al., 2019a is used for training the proposed models in multiclass class settings for brain tumor segmentation. Combo loss function can be expressed as follows:

$$L = \alpha \left(-\frac{1}{N} \sum_{i=1}^{N} \beta(t_i - \ln p_i) + (1-\beta)[(1-t_i)\ln(1-p_i)] \right) - (1-\alpha) \sum_{i=1}^{K} \left(\frac{2 \sum_{i=1}^{N} p_i t_i + S}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} t_i + S} \right) \tag{2.9}$$

where N is the number of classes and t_i is the ground truth, p_i is the predicted segmentation map, α controls the amount of Dice term contribution in loss function L, and $\beta[0,1]$ controls the level of model that penalized for false positives/negatives. The S term is added to prevent division by zero (S is added in both the denominator and numerator of the dice term). All models are trained using Adam optimizer with a learning rate of 0.000116, $\rho = 0.95$, $\in = 1e-8$, and $decay = 0$.

Experimentally, we found that $\alpha = 0.5$ Dice and cross-entropy terms lead to better segmentation results. We tried different β values with all used brain tumor segmentation datasets, finding that $\beta = 0.5$ is the best value for our proposed model.

The proposed models have been developed in the PyTorch library and trained from scratch. We optimized the hyperparameters of the models and used Adam optimizer with a learning rate of 0.0001. The number of epochs was 500 and the batch sizes were 16. An NVIDIA GTX 1080 GPU having 12GM memory is used for the training and optimization of the proposed model.

2.3 Experimental Results and Discussion

2.3.1 Dataset

Three datasets are used in our experiments to validate the efficacy of the proposed BrainSeg-DCANet model.

1. Brats 2020 dataset: The total number of subjects in the training set of BraTS 2020 with their masks is 369. This set is divided into two sets one for training and one for testing the proposed model. A ratio of 80:20 is used for training and testing set division. The total training set holds 295 subjects and the testing set holds 74 subjects. Each subject has nifti volumes for Flair, T1, T1CE, and T2 MRI modalities. There are three classes of tumor given in the ground truth (GT) masks including enhancing tumor (ET) (labeled as class-4 in GT), peritumoral edema (ED) (labeled as class-2 in GT), and non-enhancing tumor/ narcotic tumor (NCR) (labeled as class-1 in GT). In our experiment, all four MRI modalities are stacked for brain tumor segmentation using the proposed model. Figure 5 shows the Flair, T2, T1, T1CE, and the corresponding segmentation GT of a subject from the BraTS dataset (2020).

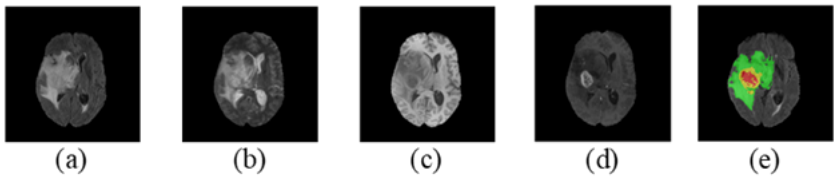


Figure 2.4: The four MRI modalities were used in this study. (a) shows the Flair image, (b) shows the T2 image, (c) shows the T1 image, (d) shows the T1CE image, and (e) shows the ground truth mask.

Here, the green color in the GT segmentation represents the ET class, the yellow color stands for ED, and the red color stands for the NCR tumor class.

2. Brats 2019 dataset: Similarly, the same training and testing sample distribution was used for the training dataset of 335 cases as for the BraTS2020 dataset. BraTS 2019 dataset is a publicly available dataset and can be accessed at *www.med.upenn.edu/cbica/brats*2019/*data.html* Wang et al., 2020.

3. SegTHOR 2019 dataset: The SegTHOR 2019 Ke et al., 2021 dataset contains 40 cases in training and 20 cases in testing. In our experiments, 32 CT (computed tomography) cases (the number of patients) are used for training, and 8 for validation from the training dataset on a ratio of 80:20. There are four classes for each subject in the dataset including the aorta, esophagus, trachea, and heart.

2.3.2 Image pre-processing

To enhance the quality of the MRI scans and to remove the undesired noise, the following preprocessing steps are performed:

1. Center cropping of each slice and the zero slices are removed from all subjects of the BRATS2020 dataset.

2. After center cropping, the dataset is normalized by the standard normalization method with subtraction of the mean and the division

of the standard deviation on non-zero-pixel values in all channels for providing the same scale data. The range of intensities is between 0 and 5000, and after normalization, the range of intensities of each slice becomes [-1,1].

3. BraTS2020 MRI images contain most of zero pixel values in the background. To separate these background zero pixel values from other regions, the histogram normalization techniques have been used to shift the pixel values to another bin range [-1, 1]. Size of bin -9 is used as it has provided better performance in our experiments experiment.

The same steps are repeated for the BraTS 2019 dataset.

2.3.3 Ablation Study

This section will present the proposed BrainSeg-DCANet model and its variants for WT, TC, and ET structures and the different hyperparameters used to measure the robustness of the proposed model.

In this ablation study, different variants of the proposed BrainSeg-DCANet model are studied: 1) RUNet: residual block residual blocks are inserted before decoder blocks, and 2) BrainSeg-DCANet-Last: residual blocks are inserted before decoder block, and DCAMs are inserted after the last encoder block. In the case of the proposed BrainSeg-DCANet model, residual blocks are inserted before each decoder block, and DCAMs are inserted after each encoder block.

The dice scores coefficients (DSCs) and Hausdorff distance (HD) of the variants of the proposed model (RUNet, BrainSeg-DCANet-Last, and BrainSeg-DCANet) as well as UNet and DeepLabv3 models when segmenting the three sub-regions (i.e., ET, WT, and TC) are shown in Table 2.1. The higher the dice coefficient in segmentation means higher performance and the lower the HD distance represents better performance. The BrainSeg-DCANet produced better performance in terms of DSCs for ET, TC, and WT brain tumor classes.

Table 2.1: Investigating the performance of different configurations of the proposed method (mean ± standard deviation). The best results are in bold.

Models	**Dice**			**HD**		
	ET	**TC**	**WT**	**ET**	**TC**	**WT**
BrainSeg-DCANet (proposed)	**0.870**	**0.913**	**0.930**	**2.08 ± 0.571**	**1.32 ± 0.525**	**1.94 ± 0.453**
BrainSeg-DCANet-Last	0.862	0.914	0.925	2.11 ± 0.576	1.34 ± 0.507	1.97± 0.471
RUNet	0.857	0.909	0.917	2.22 ± 0.883	1.36± 0.526	2.08 ± 0.803
UNet	0.846	0.904	0.919	2.25 ± 0.518	1.41 ± 0.511	2.10 ± 0.425
DeepLabv3	0.810	0.871	0.876	5.09± 0.950	3.74± 1.190	5.83± 0.778

Table 2.1 show that the DeepLabv3 achieved the lowest dice and HD scores than the base UNet model and all variant of the proposed model. Similarly, the base UNet model achieved the dice score and HD scores of 0.84, 0.90 and 0.91 and 2.25, 1.41, and 2.10 for ET, TC, and WT which are less than the achieved scores of all three proposed model variants. The RUNet in comparison to its enhanced versions of BrainSeg-DCANet-Last and the proposed model BrainSeg-DCANet achieved less dice and HD scores. Lastly, the proposed BrainSeg-DCANet achieved the highest dice and HD scores in comparison with the base UNet and proposed model variants. It generated the highest dice score of 0.930 for WT segmentation and the smallest Hausdorff distance (HD) of 1.32 ± 0.525 for TC segmentation. It reveals that our proposed model outperforms the baseline UNet model and the other two models from the ablation study on brain tumor segmentation tasks.

The model hyperparameters such as width, depth, and training strategies are kept the same for a fair comparison with other deep learning models. It shows that the proposed DCAM module with residual blocks helps to produce larger Dice scores and smaller Hausdorff distances on

ET, WT, and TC segmentation and substantially improved the performance for all metrics. The different hyperparameters have been used for ablation study to measure the robust performance of the proposed model such as different loss functions. The loss function plays an important role in the optimization of training deep neural networks.

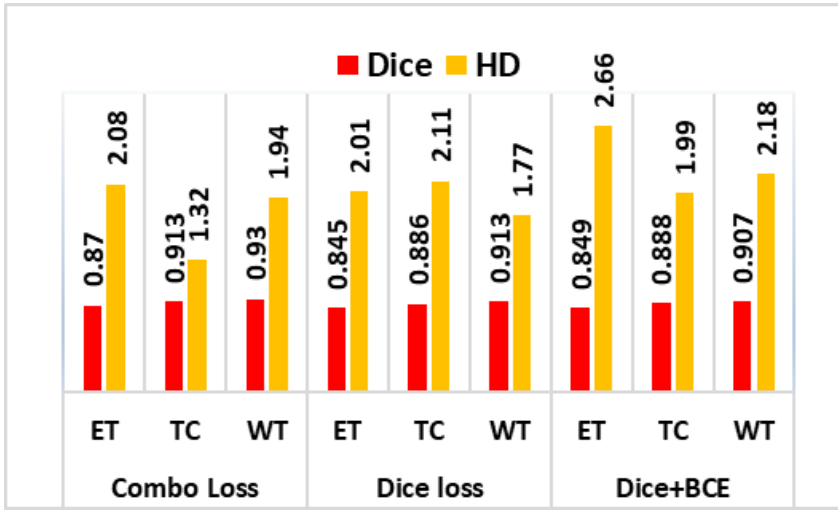


Figure 2.5: The evaluation of the BRAINSEG-DCANET with different loss functions on the test set (mean ± standard deviation). BCE stands for binary cross-entropy.

In Figure 2.5, the proposed hybrid (combo) loss function has been compared with other loss functions such as dice loss, and dice plus binary cross-entropy (Dice+BCE). The Dice Loss and the Dice+BCE loss functions achieved almost similar dice scores for ET and TC classes while it is noticed that for WT both loss functions provided variant responses. The proposed hybrid (combo) loss function produced the best dice scores of 0.87, 0.91, and 0.93 for all three tumor classes of ET, TC, and WT in comparison to the other mentioned loss functions. A similar pattern is achieved in the case of HD scores. It is seen that for both dice and HD scores the proposed combo loss function produced better results than

other common loss functions with 2.08, 1.32, and 1.94 scores for ET, TC, and WT classes.

2.3.4 Performance analysis of proposed model

The 3D visualization of the predicted segmentation map with axial slices and whole 3D volume for five cases is shown in Figure 2.6. The 2D-predicted segmentation maps are stacked to generate the 3D volume. The first column shows the GT segmentation map and the second column shows the segmentation map of the proposed BrainSeg-DCANet model. The third, fourth, and fifth columns show the predicted segmentation map of BrainSeg-DCANet-Last, RUnet, and baseline UNet models.

The proposed model predicts more pixels of necrotic/non-enhancing tumor core in 3D view of the whole volume as compared to base UNet and the other variants of the proposed BrainSeg-DCANet model. The validation from visualization results shows that our proposed model produced a better performance as compared to the other models from the ablation study. Both qualitative and quantitative results demonstrate the effectiveness of the proposed model for 2D segmentation of the brain tumor classes.

Figure 2.7 represents the feature maps using proposed BrainSeg–DCANet, BrainSeg-DCANet-Last, RUNet (baseline with residual block), and base UNet models. The feature maps produced by the BrainSeg-DCANet model that used proposed residual and DCAM modules at every stage of the encoder highlighted far better contextual information as well as positions of brain tumor sub-regions than the other presented models. The proposed DCAM module can capture the local cross-channel interaction efficiently and channel-wise convolutions layers in the proposed module would effectively capture the features of the small region tumors that can be seen in activation maps. It can highlight the distinct boundaries in the tumor region, which would be useful for achieving

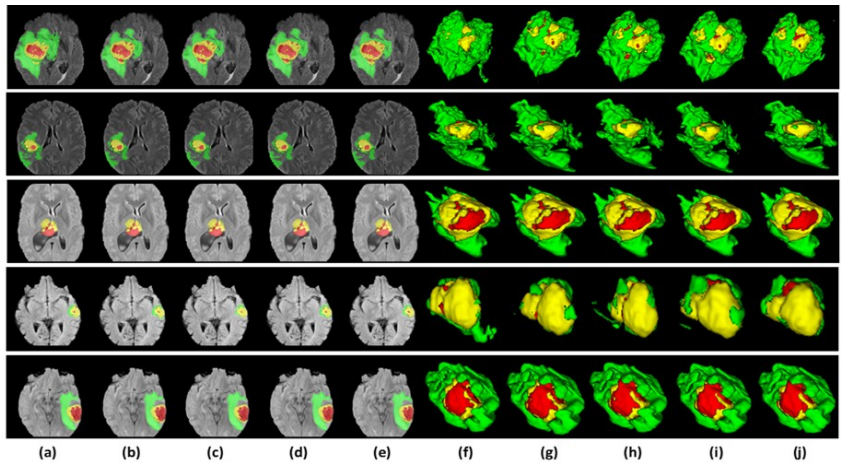


Figure 2.6: Segmentation results obtained by the proposed BrainSeg-DCANet, baseline UNet model, UNet model with the residual block (RUNet), and the proposed DCAM module at the last Conv block of the encoder with residual blocks (BrainSeg-DCANet-Last) to the five cases on the BraTS2020 dataset. From left to right: (a) 2D ground truth overlaid on FLAIR slices, (f) 3D ground truth of tumor, and the 2D and 3D segmentation results of (b,g) the proposed BrainSeg-DCANet model, (c,h) the (BrainSeg-DCANet-Last), (d, i) the RUNet, and (e,j) the baseline UNet model. Here, green color is the enhancing tumor class, yellow is the peritumoral edema while red shows the necrotic/non-enhancing tumor core.

better segmentation results. Therefore, BrainSeg-DCANet is potentially good for better performance in brain tumor segmentation.

Furthermore, descriptive statics of Dice and Hausdorff is given for the evaluation of the proposed model efficacy. Boxplots of Dice coefficients and Hausdorff distance achieved by BrainSeg-DCANet and the other presented models on BraTS2020 for each tumor class and their average for validation samples are shown in Figure 2.8. For the validation set of 74 samples, the proposed model has 2 outliers in the average dice coefficient plot in Figure 2.8(a) with high mean value than the UNet and

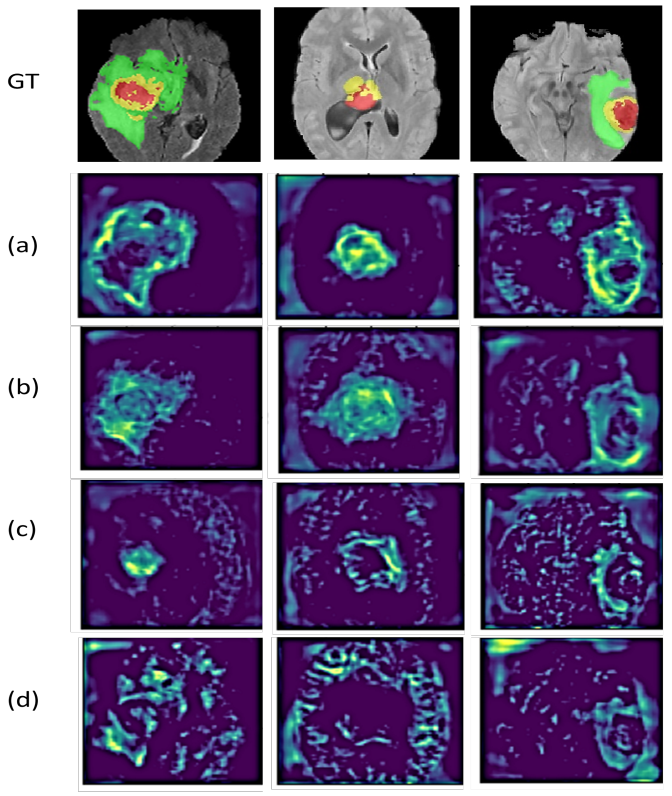


Figure 2.7: Visualization of feature maps produced by different models. All the feature maps are extracted from the output of 2nd last stage in the decoder. From top to down: (GT) Flair brain MRI slices with ground truth as input, and feature maps produced by (a) BrainSeg-DCANet (b) BrainSeg-DCANet-Last, (c) RUNet (baseline with residual block), and (d) baseline (UNet) model.

RUNET. Certainly, BrainSeg-DCANet-Last has a slightly high mean but it has a greater number of outliers and a small upper limit range than the proposed BrainSeg-DCANet. Likewise, it is observed that in Figure 2.8(b) the proposed model has the five outliers in the average Hausdorff distance like UNet but it has the lowest mean and standard deviation

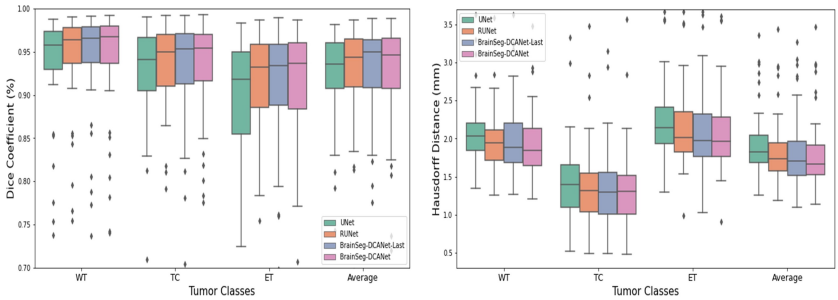


Figure 2.8: (a) and (b) show the Dice coefficients and Hausdorff distance for each tumor class and their average based on the proposed BrainSeg-DCANet model with other implemented models. The horizontal black lines inside the boxes show the mean values. The plots represent the upper and lower whiskers and the outliers. The whiskers are computed as 1.5 times the distance of the upper and lower limits of the box. The (♢) symbol represents the outliers, values outside the whiskers range.

value than the UNet, RUNET, and also the BrainSeg-DCANet-Last.

2.3.5 Comparison of proposed BrainSeg-DCANetmodel on other datasets

The proposed model evaluated on the BRATS2019 dataset is shown in Table 2.2.

Table 2.2: The performance analysis of the proposed model on the Brats2019 dataset.

Performance Metrics	**Classes**		
	ET	**TC**	**WT**
Accuracy	0.985	0.992	0.990
Dice	0.572	0.645	0.745
Hausdorff	3.50± 0.426	2.38 ± 0.667	3.08 ± 0.379
Sensitivity	0.494	0.327	0.576

The hyperparameters of the proposed model such as width, depth, and training strategies are kept the same as for BraTS 2020 dataset fair comparison.

2.3.6 Comparison of proposed BrainSeg-DCANet model with state-of-the-art methods

In BraTS 2020, the whole tumor (WT), tumor core (TC), and enhancing tumor (ET) regions Menze et al., 2014 are used for the evaluation of the segmentation instead of given class labels as enhancing tumor, non-enhancing/ necrotic tumor, and peritumoral edema. Since the optimization of these regions is more fruitful for the performance analysis Wang et al., 2018. The WT is a union of all given labels and the TC only holds enhancing and non-enhancing/necrotic tumor region while the ET is the hyperactive tumor area.

Table 2.3: The performance comparison of the proposed model with state-of-the-art methods using the brats2020 dataset.

Models	**DSC**		
	WT	**TC**	**ET**
Two-Stage Cascade Model(Lyu and Shu, 2020)	0.904	0.835	0.795
nnU-Net(Isensee et al., 2020)	0.889	0.850	0.820
Self-ensembled, deeply-supervised 3D U-net(Henry et al., 2020)	0.910	0.850	0.810
Automatic Hard mining in 3D CNN Architecture(Anand et al., 2021)	0.850	0.815	0.776
Deep Layer Aggregation(Silva et al., 2021)	0.880	0.820	0.790
Proposed Model	**0.930**	**0.913**	**0.870**

The performance comparison of the proposed model with existing models on the BRATS2020 dataset is shown in Table 2.3. The highest DSC achieved by above mentioned state-of-the-art methods for ET was 82% while our proposed model achieved an 87% score. Similarly, for the TC class, the highest score achieved by state-of-the-art methods was 85% while the proposed model achieved 91%. Lastly, the proposed model

also outperformed in the WT class by yielding the highest DCS score of 93% than other existing models.

Likewise, the comparison between proposed and existing models on the BraTS2019 dataset in terms of DSC is shown in Table 2.4. The highest DSC scores achieved by the state-of-art methods for ET, TC, and WT classes were 80%, 83.4%, and 89.4% while the proposed model outpaced the existing methods here also by achieving the high scores of 86%, 89%, and 92%.

Table 2.4: Comparison of the proposed model with existing methods on brats 2019 dataset

Models	**DSC**		
	ET	**TC**	**WT**
Wang et al. (UNet) Wang et al., 2020	0.737	0.807	0.894
Naceur et al. (DCNNs) Ben Naceur et al., 2020	0.740	0.760	0.850
Pei et al. (CANet) Pei et al., 2020	0.800	0.834	0.894
Proposed Model	**0.860**	**0.890**	**0.920**

2.3.7 Analyzing the generalization capability of the proposed model

To demonstrate the generalization capability of the proposed model, we trained and tested it on the SegTHOR2019 dataset that includes different segmentation tasks than brain tumor segmentation, namely Esophagus, Heart, Trachea, and aorta. The average DSC and HD scores of all classes are shown in Table 2.5. The proposed model achieved DSCs higher than 0.88 with Esophagus, Heart, Trachea, Aorta segmentation tasks. It also achieved HD values lower than 0.3 with all tasks. These excellent results prove the generalization capability of the proposed model. However, due to a 2D-based model, it cannot capture the spatial consistency which will be addressed in the future by replacing the proposed model with a 3D approach.

Table 2.5: The performance analysis of the proposed model on the segthor 2019 dataset

Performance Metrics	**Classes**			
	Esophagus	**Heart**	**Trachea**	**Aorta**
Dice	0.8844	0.9463	0.9366	0.9667
Hausdorff	0.2434	0.1377	0.1204	0.1168

The validation of the proposed BrainSeg-DCANet model with the other state-of-the-art methods on the SegTHOR2019 dataset is shown in Table 2.6. The proposed model produced an excellent performance as compared to existing deep learning models for SegTHOR2019 similar to the BraTS2020 and BraTS 2019 datasets. From the abovementioned results, it can be proven that the proposed BrainSeg9-DCANet is a generalized model for the brain tumor segmentation field.

Table 2.6: Dice score of proposed and existing models on segthor 2019 dataset

Models	**DSC**			
	Esophagus	**Heart**	**Trachea**	**Aorta**
He et al., 2020	0.859	0.950	0.920	0.948
Wang et al., 2019b	0.859	0.945	0.921	0.943
Vesal, Ravikumar, and Maier, 2019	0.858	0.941	0.926	0.938
Chen, P. , Xu, C. , Li, X. , Ma, Y. , Sun, 2019	0.816	0.932	0.891	0.923
(U-Net) Ronneberger, Fischer, and Brox, 2015b	0.838	0.915	0.887	0.911
(DS-Net) Hasan et al., 2020	0.851	0.927	0.900	0.908
Proposed Model	**0.884**	**0.946**	**0.936**	**0.966**

2.4 Conclusion

The segmentation of brain tumors in MR images is a complex and time-consuming task. Although various deep learning models have been proposed for brain tumor segmentation, small tumor segmentation is still challenging due to the loss of location and spatial information because of the neural network's constant convolution and transformation operation. Therefore, BrainSeg-DCANet has been proposed in this chapter to address this issue and improve the brain tumor semantic segmentation of MR images. The BrainSeg-DCANet has incorporated an efficient depth-wise attention module to preserve the important information from the feature's maps (i.e., brain tumor-relevant features) at each encoder stage. DWT has been also used to enlarge the receptive fields of the segmentation model. These modules can extract the local cross-channel interaction efficiently and generate powerful contextual feature information, and thus it can segment and localize small brain tumors accurately. The ablation study showed the effectiveness of the proposed attention module on different brain tumor datasets such as BraTS 2020, BraTS2019, and the SegTHOR 2019 multiclass medical segmentation dataset. The proposed model produced an outstanding performance as compared to the existing state-of-the-art methods. BrainSeg-DCANet has also outperformed its baseline UNet architecture. In the future, small and lightweight attention modules and 3D deep learning models will be explored to further improve the performance of the proposed model.

Chapter 3

Effective Approaches for Survival Prediction based on the 2D-3D Deep Learning Approach

To completely comprehend neurodevelopment in healthy and congenitally abnormal fetuses, quantitative analysis of the human fetal brain is essential. This analysis requires the use of automatic multi-tissue fetal brain segmentation techniques. This Chapter proposes an end-to-end automatic yet effective method for a multi-tissue fetal brain segmentation model called IRMMNET. It includes an inception residual encoder block (EB) and a dense spatial attention (DSAM) block, which facilitate the extraction of multi-scale fetal-brain-tissue-relevant information from multi-view MRI images, enhance the feature reuse, and substantially reduce the number of parameters of the segmentation model. Additionally, we propose three methods for predicting gestational age (GA), GA prediction by using a 3D autoencoder, GA prediction using radiomics features, and GA prediction using the IRMMNET segmentation model's

encoder. Our experiments were performed on a dataset of 80 pathological and non-pathological magnetic resonance fetal brain volume reconstructions across a range of gestational ages (20 to 33 weeks) that were manually segmented into seven different tissue categories. The results showed that the proposed fetal brain segmentation model achieved a Dice score of 0.791±0.18 , outperforming the state-of-the-art methods. The radiomics-based GA prediction methods achieved the best results (RMSE: 1.42). We also demonstrated the generalization capabilities of the proposed methods for tasks such as head and neck tumor segmentation and the prediction of patients' survival days.

3.1 Introduction

Congenital disorders are some of the leading causes of infant mortality worldwide Ebner et al., 2020. Recently, in-utero magnetic resonance imaging (MRI) of the fetal brain has emerged as a valuable tool for investigating the neurological development of fetuses with congenital disorders to aid in prenatal planning. Fetal MRI requires clinical and technical expertise and is a challenging imaging modality due to the ability to move freely. T2-weighted single-shot fast spin echo (ssFSE) sequences, such as ultra-fast MRI sequences, can be used to attain information in all planes.

Super-resolution (SR) reconstruction algorithms, including outlier rejection and motion correction strategies (Ebner et al., 2020), can then be applied in order to combine several low-resolution images into a single high-resolution volume that can be used for further quantitative analysis. Automated quantification of the highly complex and rapidly changing brain morphology in MRI data could improve the diagnostic and decision-making processes.

Image segmentation is an early step for the volumetric quantification of the fetal brain. Shape or volume information could be relevant

to the developing cortex, cerebellum, brainstem, white matter, and cerebrospinal fluid spaces Egaña-Ugrinovic et al., 2013; Jarvis, Finney, and Griffiths, 2019. The automatic segmentation of the developing human brain is a primary step for analysis, as manual segmentation is time-consuming and may be prone to human error. However, fetal brain segmentation based on SR fetal brain volumes is still challenging due to artifacts that are blurry or caused by motion, rapidly changing fetal brain anatomy and the effects of partial volume.

Various atlas-based methods have been developed for brain tissue segmentation Habas et al., 2010. However, these methods need an atlas, which now only exists for normally developing fetuses. Falick Michaeli et al., 2019 used single-class high-resolution fetal brain volumes for fetal brain segmentation, but multiclass segmentation still needs to be explored. Deep-learning-based segmentation models have recently been employed to segment the fetal brain into different tissue types by using low-resolution coronal-direction slices to handle fetal brain tissue segmentation problems Khalili et al., 2019.

Faghihpirayesh et al., 2022 used an encoder-decoder UNet model with multiple branches and skip connections to maintain high accuracy while devising a parallel combination of convolution and pooling operations. They used a private dataset to train their proposed model. However, they only handled the single-class segmentation problem by using 2D slices, which is not challenging and quite simple. A 2D segmentation model for volumetric 3D segmentation cannot handle temporal relationships, unlike 3D segmentation models. Moreover, they used only binary class segmentation, while the proposed model addresses the problem of multi-tissue fetal brain segmentation.

Asis-Cruz et al., 2022 used an end-to-end generative adversarial neural network (GAN) to segment the fetal brain in functional magnetic resonance images (rs-fMRI). They segmented the full fetal brain and handled binary class problems by using a private dataset. Unlike the models in these works, the proposed multi-view segmentation model

can handle the 3D segmentation of volumetric data by using a stacking approach to multi-view segmentations.

Zhao et al., 2022 trained a patch-based 3D segmentation model for fetal brain segmentation by using an in-house dataset. This 3D segmentation model required powerful computational resources. However, the 3D-CNN holds great potential for fully utilizing the 3D information from MRI data, which also contains multi-view information. However, 3D-CNN-based segmentation greatly increases the network scale and computational cost Prasoon et al., 2013. It should be noted that the major bottleneck in the development of segmentation algorithms for medical imaging is the lack of data—either the availability of atlases for atlas-based segmentation or that of training data for supervised machine learning methods. In addition, there is still a need to explore and implement deep-learning-based approaches, no clear benchmark is available for fetal brain segmentation.

In turn, the dating of the precise gestational age (GA) is essential for assessing pregnancy, fetal development, and neonatal care. Before sonography, obstetricians routinely relied on the last menstrual period for the dating of the gestational age in pre-birth life Falick Michaeli et al., 2019. The crown–rump length (CRL) method is used in the first trimester to estimate gestational age. Other methods are used in the last two trimesters, such as brain bi-parietal diameter, head circumference, femur length, and abdominal circumference. These methods were reported decades ago and are still used today Falick Michaeli et al., 2019.

Though sonographic assessment during the first trimester is the most well-known and accurate method for estimating gestational age, it shows large variations in the second and third trimesters due to the variability in organ size. According to previous studies, the assessment of gestational age by combining the above-mentioned biometric data can achieve an accuracy of ± 7 to 10 days for the second trimester and ± 21 to 30 days for the third trimester. Various methods, such as the measurement of the cerebellar length and the trans cerebellar diameter, accurately predict

gestational age in singleton and twin pregnancies Chavez et al., 2006; however, they require good visualization of the cerebellum by specialized sonographers.

In summary, estimations made with sonographic measurements are strongly affected by the inherent variability in organ size and the intrinsic signal properties of ultrasonography Wu et al., 2015. The inaccuracy of sonographic assessment has driven the need to find different approaches that can be used to accurately determine gestational age. MRI is gradually being recognized as a powerful helper for ultrasonography in the evaluation of the fetal brain. MRI-based methods provide a high resolution, soft-tissue contrast, and visibility of the whole brain independently of fetal presentation Kyriakopoulou et al., 2017; Blondiaux and Garel, 2013. As pregnancies advance, the biological variations among normal fetuses increase, and the ranges of values of each biometric measurement associated with a specific GA also increase. This means that while the predictive error at $\pm$10 days GA is considered acceptable in most clinical settings, the predictive error at $\pm$18 days is estimated to offer little clinical value Namburete et al., 2015. Therefore, when screening occurs in the second and third trimesters, the error margins produced by current methods are highly increased; thus, they are not clinically useful. Accordingly, there is a need to develop an alternative technique for estimating the GA.

Fung et al., 2020 developed a machine learning (ML) model for estimating the GA and predicting future growth. They utilized multicenter, international, and population-based project data from the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). Kojita et al., 2021 developed VGG-based transfer learning models for GA prediction. They employed an in-house (private) dataset. The deep learning model was trained with T2-weighted images from 126 training cases and 29 validation cases. The remaining 29 cases were utilized as test data, with the fetal age being estimated by the model and by using BPD (biparietal diameter) measurements.

They drew a relationship between the estimated and standard gestational ages by using Lin's concordance correlation (ρc). The model's outcome in terms of concordance was significant (ρc = 0.964).

Lu et al., 2019 developed machine learning models that could provide accurate estimations for obstetricians alongside traditional clinical practices and an efficient and effective supporting tool for pregnant women for self-monitoring. A total of 4212 intrapartum recordings were selected, of which 3370 samples were used as the training set and 842 samples were used as the test set. In addition, several simple and powerful machine learning algorithms were trained, and their performance was evaluated with real test data. The experimental results showed an intersection over union (IoU) of 0.64 between the predicted range of fetal weight at any gestational age from the ensemble model and that from ultrasound. Using their private dataset, they used simple clinical features with traditional machine learning models for the prediction of gestational age and weight. No deep-learning-based models were used as a comparison with the machine-learning models. No efficient feature engineering approaches were used to predict gestational age.

Maternal et al., 2021 developed a novel method based on machine learning models and used each subset of predictors based on an ensemble model constructed by using the Super Learner algorithm. The resulting model was a weighted average of multivariate adaptive regression splines, random forests, gradient boosting, support vector machines, and multiple linear regression. They assessed the diagnostic accuracy by using the receiver operating curve (AUC) and Bland–Altman analysis. They collected datasets from population-based cohorts in five countries (Bangladesh, Ghana, Pakistan, Tanzania, and Zambia). Women at <20 weeks of gestation according to ultrasound-based dating were used as a study case for the prediction of gestational age. A total of 7428 live-born infants were included. This dataset is not publicly available. The

resulting model was a weighted average of multivariate adaptive regression splines, random forests, gradient boosting, support vector machines, and multiple linear regression. They achieved the highest AUC of 0.96. They used only clinical features with traditional machine learning models for the age prediction. No imaging-based features were used to evaluate the performance of the machine learning or deep learning models for the prediction of gestational age.

Payette et al., 2021 employed deep learning models such as ResNet-18 and ResNet-50 with a combination of different layers for the prediction of gestational age. They collected 741 fetal brain MRIs in order to predict fetal gestational age (in days). The authors proposed a basic ResNet18-based regressor model that used a private dataset, and they did not use any other segmentation-based or 3D volumetric-based features for gestational age prediction. They used cropped 2D images covering the fetal area only in the input images to train the basic ResNet18 with overall global features, and this could efficiently help in the extraction of local image features for the prediction of fetal gestational age.

Shen et al., 2022 used attention-guided, multi-plane ResNet-50 models trained on Stanford data to predict the gestational age. They trained various CNN models based on only imaging features for the prediction of gestational age. Imaging features might not be sufficient to accurately predict gestational age. However, we used various feature extraction approaches, including imaging, radiomics, 3D latent space autoencoder-based features, and deep features extracted from the last layer of multi-view 2D image slices from segmented brain tissues, to extract more localized features for the prediction of gestational age. The fusion of multi-scale segment-based deep features achieved better performance than that of the state-of-the-art methods.

There is a further need to investigate different methods with deep learning models for GA prediction. The existing methods are based on

single-feature extraction techniques that use basic deep-learning models. Correspondingly, the datasets used for the existing methods are in-house and private. There is a need to set a benchmark on a publicly available dataset for further comparisons and enhancements in deep learning/machine learning for the prediction of gestational age and segmentation of the fetal brain.

To solve the above-mentioned issues, we propose effective yet automatic methods for the segmentation of fetal brain tissue and the prediction of gestational age. To the best of the authors' knowledge, this is the first method to propose an end-to-end solution for fetal brain segmentation in MRI images and GA prediction. Deep learning is the basis for the proposed fetal segmentation method, IRMMNET (inception residual multi-scale multi-view network). By effectively combining segmentation maps from the axial, coronal, and sagittal views to create a 3D segmentation volume, IRMMNET incorporates important insights from multi-view MRI. IRMMNET consists of several layers with the capacity to reuse features and information at several scales and depths. The inception residual encoder block (EB) and the dense spatial attention (DSAM) block are two proposed blocks that are part of IRMMNET. The EB aids in extracting information from multi-view MRI scans that are pertinent to multi-scale fetal brain tissue. The DSAM improves feature reuse while lowering the model's parameter count. The EB and DSAM help segment small lesions that have a small number of semantic pixels that are missed by traditional encoder-decoder networks. Then, we propose three methods for GA prediction—GA prediction using the IRMMNET segmentation model's encoder, GA prediction using a 3D autoencoder, and GA prediction using radiomics features.

The following is a list of this chapter's major contributions:

1. Proposal of a novel multi-view multi-scale 3D fetal brain segmentation method named IRMMNET. It combines the key insights from multi-view MRI, including the axial, coronal, and sagittal views.

IRMMNET comprises different layers with feature reuse capabilities and with various depths and multi-scale information. An efficient method for fusing segmentation maps of the axial, coronal, and sagittal views to develop a 3D segmentation volume is also presented.

2. Presentation of two effective blocks: the inception residual encoder block (EB) and the dense spatial attention (DSAM) block. The EB helps the fetal brain segmentation network extract multi-scale fetal-brain-tissue-relevant information from multi-view MRI images. The DSAM block enhances feature reuse and substantially reduces the number of parameters of the segmentation model. Extensive experiments were performed with various combinations and settings of the fetal brain segmentation model.

3. Proposal of three approaches to predicting GA: GA prediction by utilizing the IRMMNET segmentation model's encoder, GA prediction by utilizing a 3D autoencoder, and GA prediction by utilizing radiomics features. The explainability and importance of the radiomics features are also presented.

4. Demonstration of the generalization capabilities of the proposed fetal brain segmentation and GA prediction methods on two different tasks: the segmentation of head and neck tumors and the prediction of patients' survival days.

The rest of this Chapter is presented as follows: Section 3.2 introduces the proposed multi-view multi-scale 3D fetal brain segmentation method and the proposed fetal age prediction method. Section 3.3 presents the datasets used in our study, the proposed fetal brain segmentation model's results, and the GA prediction method's results. Section 3.3.5 discusses the generalization capabilities of the proposed methods. Section 3.4 discusses the findings of the study and its limitations. Section 3.5 concludes the Chapter and presents the future work.

3.2 Methodology

In this section, we explain the proposed multi-view multi-scale 3D fetal brain segmentation method (Section 3.2.1) and the proposed fetal age prediction method (Section 3.2.2).

3.2.1 Proposed Multi-View Multi-Scale 3D Fetal Brain Segmentation Method

Figure 3.1 presents a schematic diagram of the proposed multi-view multi-scale 3D fetal brain segmentation method. The three available views—axial, coronal, and sagittal—are inputted into the proposed segmentation model to generate a 2D segmentation mask for each view. Then, the resulting segmentation masks are combined to construct a 3D segmentation map. Specifically, we stack a 2D segmentation mask of each view to form a predicted 3D segmentation mask. This process can be expressed as follows:

$$\begin{aligned} O_A &= \text{orgmax}\left(\text{Sigmoid}\left(L_A\right)\right), \\ O_S &= \text{orgmax}\left(\text{Sigmoid}\left(L_S\right)\right), \\ O_C &= \text{orgmax}\left(\text{Sigmoid}\left(L_C\right)\right) \end{aligned} \tag{3.1}$$

$$\begin{aligned} 3DV[:,:,A] &= \sum_{i=1}^{N} O_A, \\ 3DV[:,S,:] &= \sum_{i=1}^{N} O_S, \\ 3DV[C,:,:] &= \sum_{i=1}^{N} O_C \end{aligned} \tag{3.2}$$

where L_A represents linear-layer neurons using the axial view, L_S using the sagittal view, and L_C using the coronal axis view.

$$3DV[:,:,:] = 3DV[:,:,A] + 3DV[:,S,:] + 3DV[C,:,:] \tag{3.3}$$

where O_A stands for the output prediction of the proposed model when using axial slices, O_S is the prediction of the proposed model when using the sagittal view, and O_C is the output when using coronal slices; $3DV$ is the 3D prediction volume that is reconstructed by stacking up the predicted 2D slices from each 2D view (O_A, O_s, and O_c).

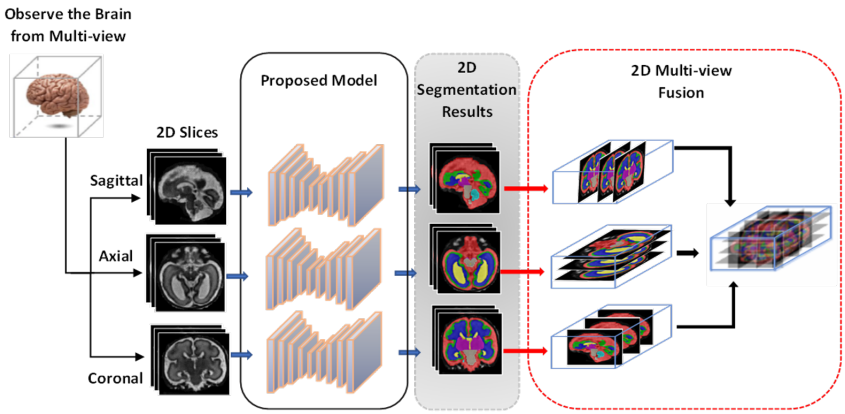


Figure 3.1: A schematic diagram of the proposed method for 3D fetal brain segmentation from the axial, coronal, and sagittal views.

3.2.1.1 Multi-View Multi-Scale Segmentation Network

Figure 3.2 presents the proposed multi-view multi-scale segmentation network. The proposed model is designed based on the concept of an encoder-decoder with skip connections. The proposed inception residual encoder block (EB) and dense spatial attention (DSAM) block are used in the encoding path. In contrast, an efficient yet simple 2D convolutional layer module with 2D upsampling layers, including regularization layers, is used on the decoding path.

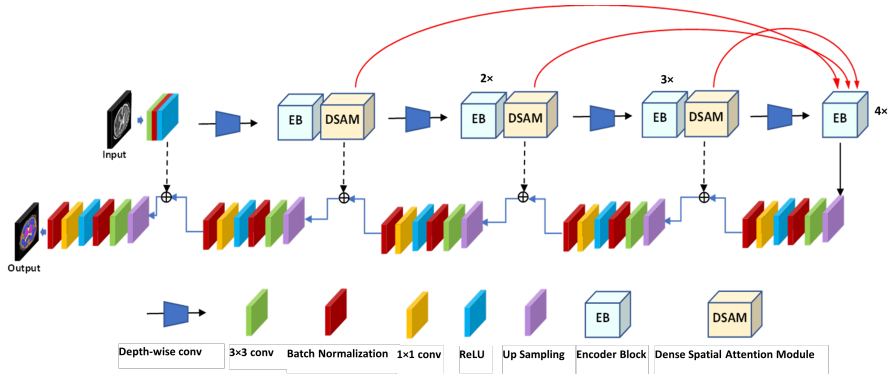


Figure 3.2: The proposed IRMMNET multi-view multi-scale segmentation network.

In the encoding path, a DSAM block is used in each encoder block, which sends information at every block from each encoder layer to the bottom layer. The number of channels is doubled at each EB block, and the input size of the feature maps is reduced by half based on the depth-wise convolutional layer in the encoding path. An increasing number of feature map blocks are used in each stage of encoder blocks; the number of EB blocks progressively increases at each stage of the encoder side. The first encoder block uses one EB block. Similarly, the second, third, and fourth blocks employ 2, 3, and 4 inception residual blocks, respectively.

The red lines in Figure 3.2 highlight the multi-scale feature maps from each encoder block fed to the model's bottom layer. This strategy increases the efficiency of the feature maps by reusing and fusing the feature information at the level of downsampling. In the proposed segmentation model, the features are extracted from three levels (red lines), which enables the model benefit from the multi-scale transformation of high-level semantic information and low-level information of the position and texture. Three downsampling layers that carry features' information are passed to the bottom-layer module to guarantee an improved

cross-level feature connection and complementarity in cross-level information.

In the decoding path, the size of the feature maps increases after each 2D upsampling layer, and the original size of the training input images returns in the output in the final layer. The first 2D upsampling layer comprises two efficient convolutional layers (2D 1×1 Conv, 2D 3×3 Conv) with a BN and ReLU layer. To reconstruct the semantic information, the feature maps are concatenated with each encoder and decoder block. The BN and ReLU regularization layers are used with 2D upsampled layers and 2D 3×3 Conv layers for smooth optimization and training of the proposed model. The 1×1 Conv layer and sigmoid activation function are used for the reconstruction of the segmentation map.

Below, we explain the architectures of the proposed EB and DSAM blocks.

3.2.1.2 Inception Residual Encoder Block (Eb)

Figure 3.3 depicts the proposed inception residual encoder block. Unlike in the Inception-Res architecture (Szegedy et al., 2017), we introduce a batch normalization (BN) layer after each convolutional layer, except for the bottleneck layers. In addition, we use 1×1 and 1×3 kernels with a 1×5 kernel branch, which was inspired by the DeepLab architecture (Chen et al., 2017).

It should be noted that the batch normalization layer produces smooth training and can avoid gradient vanishing while retaining the convolutional layers. The feature maps are aggregated by convolving them with three kernels: 1×1, 1×3, and 1×5. It is assumed that x_l is the input and x_{l+i} is the output of the ith layer. $c_{1\times n}$ is a $1 \times n$ kernel convolutional layer and c_b represents the batch normalization layer. $c_{1\times 1}$ denotes the bottleneck layer. The output of each EB module from the encoder path can be expressed as follows:

$$x_{l+1} = l_1 \times l_2 \times l_3 \times l_4 \tag{3.4}$$

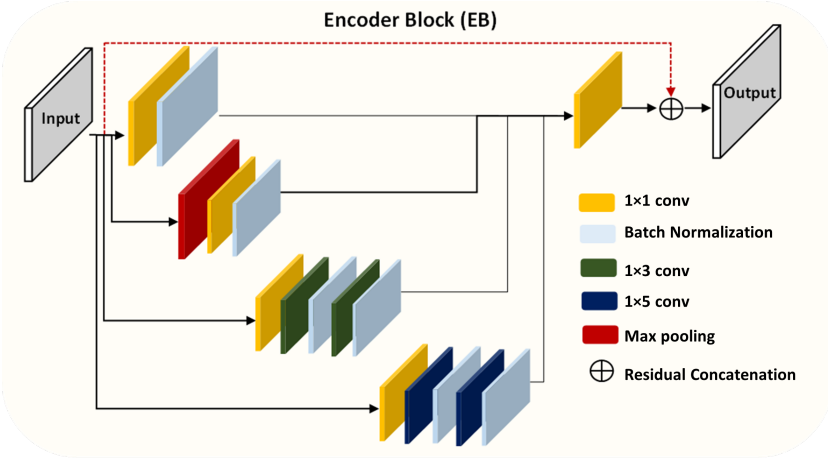


Figure 3.3: The schematic diagram of the proposed inception residual encoder block (EB).

$$l_1 = c_b(c_{1\times1}(k)) \tag{3.5}$$

$$l_2 = Maxpool(c_b(c_{1\times1}(k))) \tag{3.6}$$

$$l_3 = c_b(c_{1\times3}(c_b(c_{1\times3}(c_{1\times1}(k))))) \tag{3.7}$$

$$l_4 = c_b(c_{1\times5}(c_b(c_{1\times5}(c_{1\times1}(k))))) \tag{3.8}$$

where $k = x_l$.

3.2.1.3 The Proposed Dense Spatial Attention Module (DSAM)

It should be noted that the attention modules that are often used in image segmentation and object detection models are mainly divided into

channel-wise attention and point-wise attention modules, and the representatives of these two attention models are squeeze and excitation (SE) and the spatial attention module (SAM) (Woo et al., 2018b). A channel attention map exploits features' inter-channel relationships, and the feature map obtained from the channel attention is considered a feature detector. The spatial attention module focuses on what is meaningful given an input image through the benefit of the combination of average pooling and max pooling. Figure 3.4 shows a schematic diagram of the DSAM. As shown, we modified the SAM by replacing the point-wise convolutional layer with a basic dense layer (DL). We also used the swish activation function instead of the sigmoid function (Ramachandran, Zoph, and Le, 2017b), as swish is less prone to vanishing gradient problems. The swish activation function can be expressed as follows:

$$Swish(x) = x/(1 + e - x)) \tag{3.9}$$

where x is the input feature map.

DL is a basic dense layer. The input x is multiplied by the DL and by the swish activation function.

In the DL, the feature maps of all preceding layers are used as inputs, and their feature maps are used as inputs for all subsequent layers. The output of each dense block is concatenated with every previous dense block. The operation of the DL can be expressed as follows:

$$DL = [x, B_1(x), B_2([x, B_1(x)]), B_3(x, B_1(x), B_2([x, B_1(x)])), \dots] \tag{3.10}$$

where $B_1(x)$ is dense block 1, B_2 is dense block 2, and so on. x is the input feature map. We used 12 dense blocks in our case. The output of the proposed DSAM block, $DSAM_{out}$, can be formulated as follows:

$$DSAM_{out} = x * (Swish(DL(x))) \tag{3.11}$$

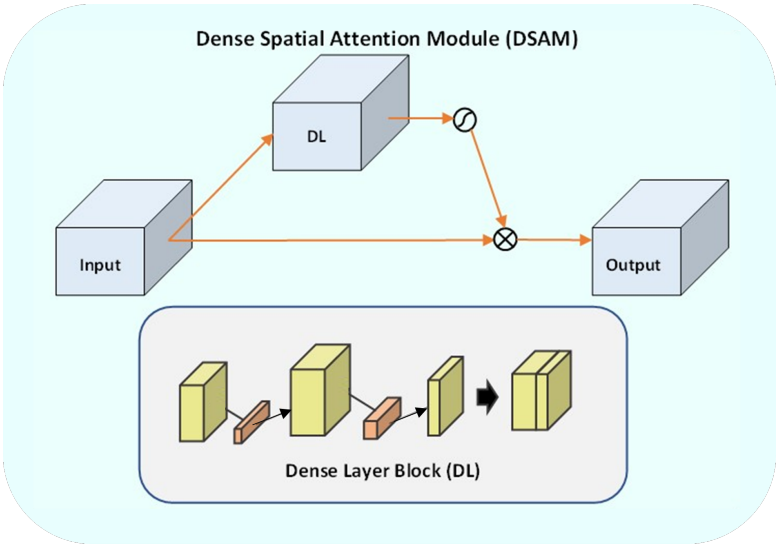


Figure 3.4: A schematic diagram of the proposed dense spatial attention module (DSAM).

The key advantages of the DL are (1) the alleviation of the vanishing gradient problem, (2) the strengthening of feature propagation, (3) the encouragement of feature reuse, and (4) the substantial reduction in the number of parameters. The feature maps are concatenated from the previous layer to the next layer to build the dense block in the proposed dense layer. The feature maps keep the relevant information from every layer and are reused in the final layer to get semantic information. The proposed DSAM uses the dense block to provide better semantic information and improves the flow of gradient information for easy training of the proposed model. The DSAM block reduces the problem of overfitting with smaller training set sizes by using dense connections. DSAM blocks also provide direct access to the gradients from the loss function and the original input signal, which leads to implicit deep supervision. It is worth noting that all layer weights in the proposed model were trained from scratch with the FeTA dataset.

3.2.1.4 Loss Function and Implementation Details

In this Chapter, we employ the Combo loss function Taghanaki et al., 2019b to train the proposed model with multi-class settings for fetal brain tissue segmentation. Combo loss can be expressed as follows:

$$L = \alpha \left(-\frac{1}{N} \sum_{i=1}^{N} \beta \left(t_i - \ln p_i \right) + (1-\beta) \left[(1-t_i) \ln (1-p_i) \right] \right) - (1 - \alpha) \sum_{i=1}^{K} \left(\frac{2 \sum_{i=1}^{N} p_i t_i + S}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} t_i + S} \right) \tag{3.12}$$

where t_i is the one-hot-encoded target or ground truth, p_i is the predicted probability, N is the number of classes multiplied by the number of samples, and the S is a small constant number that is added to prevent division by zero. α controls the amount contributed by the Dice term in the loss function L. $\beta \in [0,1]$ controls the level of penalization of the model for false positives/negatives. The S term is added to prevent division by zero. The S constant is added in both the denominator and the numerator of the Dice term.

All models were trained using an Adam optimizer with a learning rate of 0.0001, $\rho = 0.95$, $\epsilon = 1 * 10^{-8}$, and decay = 0. Based on the experimental evaluations, it was found that $\alpha = 0.5$ for the Dice and cross-entropy terms produced the best results. Different values of β were tried for all datasets, and it was found that $\beta = 0.5$ was the best value for the proposed dataset for the segmentation task.

3.2.2 Approaches to the Prediction of Gestational Age (GA)

This subsection proposes three approaches to predicting GA: (1) GA prediction by utilizing the IRMMNET segmentation model's encoder, (2) GA prediction by utilizing a 3D autoencoder, and (3) GA prediction by utilizing radiomics features.

3.2.2.1 GA Prediction by Utilizing the Encoder of IRMMNET Segmentation Model

The encoder of the proposed IRMMNET segmentation model can automatically learn multiple filters in parallel and extract low- and high-level features, such as edges, intensities, and textures. Different filters capture various characteristics of the input images that are used in GA prediction. Figure 3.5 presents the proposed framework for GA prediction by utilizing the IRMMNET segmentation model's encoder. As one can see, deep features are extracted from the trained encoder of each view's trained segmentation model. Each view's encoder generates a feature vector with dimensions of 1×256. It should be noted that the feature vectors of all slices of each volume are combined to produce one feature vector that represents the whole volume. The feature vectors of the three views are concatenated to form one single feature vector with dimensions of 1×768. Different regression algorithms are trained with the extracted feature vector in order to predict the GA.

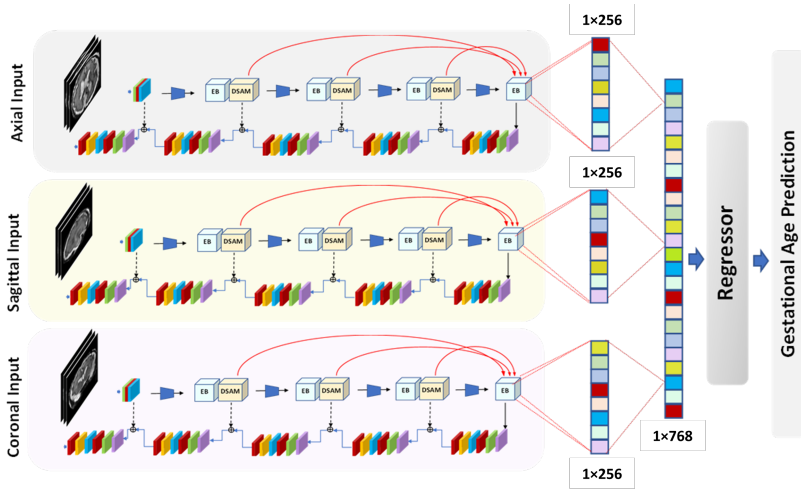


Figure 3.5: GA prediction by utilizing the IRMMNET segmentation model's encoder.

3.2.2.2 GA Prediction by Utilizing a 3D Autoencoder

Figure 3.6 presents the proposed approach to GA prediction by utilizing a 3D autoencoder. The 3D autoencoder consists of an encoder- and decoder-based model. A 3D volume is fed to the proposed 3D autoencoder to extract latent space features. As shown in Figure 3.6, the encoder part of the 3D autoencoder consists of a 3D convolutional layer, batch normalization layer, ReLU layer, and 3D maxpool layer, and the decoder side of the 3D autoencoder comprises a 3D transposed layer at each level on that side along, with a 3D convolutional layer and a batch normalization layer.

On the encoder side, the volume is reduced at each encoder block, while the volume is increased at each decoder block. The 3D transposed layer is used to upsample the feature maps at each decoder block. The bottom layer represents the 3D volume in the lower dimension. After flattening the lower dimension, a latent vector is produced. Different regression algorithms are trained with the latent features to predict the GA, as shown in Figure 3.6.

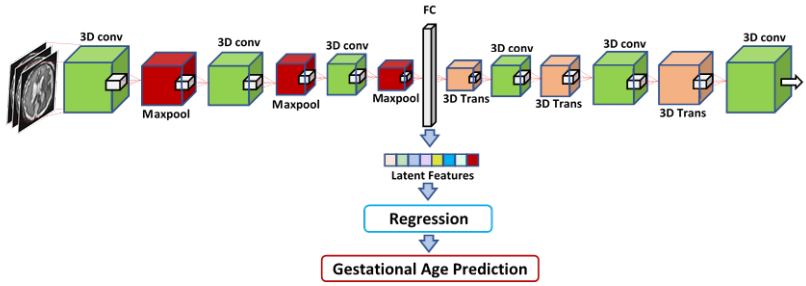


Figure 3.6: GA prediction by utilizing a 3D autoencoder. The latent features are extracted by using input image volumes for GA prediction.

3.2.2.3 GA Prediction by Utilizing Radiomics Features

Figure 3.7 shows the proposed approach to GA prediction by utilizing radiomics features. Here, we also extract 2D slices of the axial, sagittal, and coronal views from a 3D volume dataset. We train the proposed segmentation model for the dataset containing the three views and stack the output of each 2D view to reconstruct a 3D volume. The radiomics features are extracted from the input volumes of the dataset. The radiomics features used include shape-based, statistical, and wavelet features. Among the 108 radiomics features, a set was chosen based on a correlation-based feature selection technique (Gopika and ME, 2018). The selected radiomics features were elongation, flatness, major axis length, minor axis length, max 3D diameter, sphericity, surface area, energy, entropy, kurtosis, mean, skewness, coarseness, contrast, correlation, inverse-diff moment, complexity, and strength. The radiomic features were extracted from Feta MRI images by using the Pyradiomics package (Van Griethuysen et al., 2017). Different regression algorithms were trained with the extracted radiomics features to predict the GA.

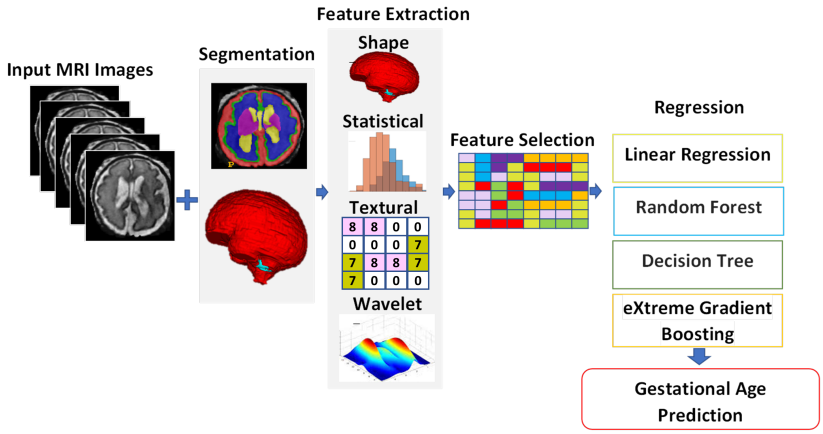


Figure 3.7: GA prediction by utilizing radiomics features.

3.2.2.4 Regression Techniques for GA Prediction

Different regression techniques were tested on the features extracted with the three approaches mentioned above for the prediction of the GA. We found that four regression techniques give acceptable results: random forest (RF) (Almalki et al., 2021), regression trees (RT), linear regression (LR), and extreme gradient boosting (XGB) (Fernández-Delgado et al., 2019). RF is supervised by traditional machine learning and is widely used for classification and regression problems. It uses a bootstrapped dataset as a subset, picks random subsets of features, and runs random trees in parallel while building the trees. We set the number of trees to be from 100 to 1000 to create the forest. LR is a linear model that builds a linear relationship between input variables and a single output variable. RT is a tree-based regression model that trains a model by observing the input object's features and generating a continuous output. The gradient-boosting regressor is a method that uses an additive forward model and allows arbitrary differentiable loss functions for optimization during training. The gradient-boosting model is an ensemble model that can be used for regression, classification, and predictive modeling problems. Extreme gradient boosting (XGB) is an open-source approach to gradient-boosting regression.

3.3 Experimental Results and Discussion

In this section, we explain the dataset used in our study (Section 3.3.1), present and discuss the results of the proposed fetal brain segmentation model (Section 3.3.3), and analyze the results of the GA prediction models (Section 3.3.4).

3.3.1 Dataset Description

The dataset included 80 T2-weighted fetal brain reconstructions with a corresponding label map that was manually segmented into seven

different tissue labels (Payette et al., 2021). The seven labels were external cerebrospinal fluid (ECF), fluid gray matter (FGM), white matter (WM), ventricles (VCs), cerebellum (CBM), deep gray matter (DGM), and brainstem (BSTM). The dataset consisted of clinically acquired fetal brain reconstructions of both neurotypical and pathological brains with a range of gestational ages. The data were acquired using $1.5T$ and $3T$ clinical GE whole-body scanners (Signa Discovery MR450 and MR750) with either an eight-channel cardiac coil or a body coil. T2-weighted single-shot fast spin echo sequences were acquired with an in-plane resolution of 0.5 mm $\times$ 0.5 mm and a slice thickness of 3 to 5 mm. The sequence parameters were the following: TR: 2000–3500 ms; TE: 120 ms (minimum); flip angle: 90°; sampling percentages: 55%.

Figure 3.8 shows the class-mapping function for the axial, sagittal, and coronal slices. The different colors show the seven classes used to predict fetal tissue segmentation.

Two different methods were used to create a super-resolution reconstruction of the fetal brain for each case from the acquired low-resolution axial, coronal, and sagittal images. Equal numbers of cases were used in the training and evaluation datasets for each reconstruction method. For each case, the gestational age in weeks and the neurotypical/pathological label was given, in addition to the label maps.

The dataset was divided into 80% for training and 20% for testing. There were totals of 64 subjects in training and 16 subjects in the testing phase.

3.3.2 Evaluation Metrics

In this study, four metrics were used to assess the performance of the fetal brain segmentation models: Dice (DSC), Hausdorff distance (HD95), sensitivity, and specificity, which is commonly used for the validation of medical volume segmentation approaches (Qayyum, Lalande, and Meriaudeau, 2020). This is also called the overlap index. It measures the

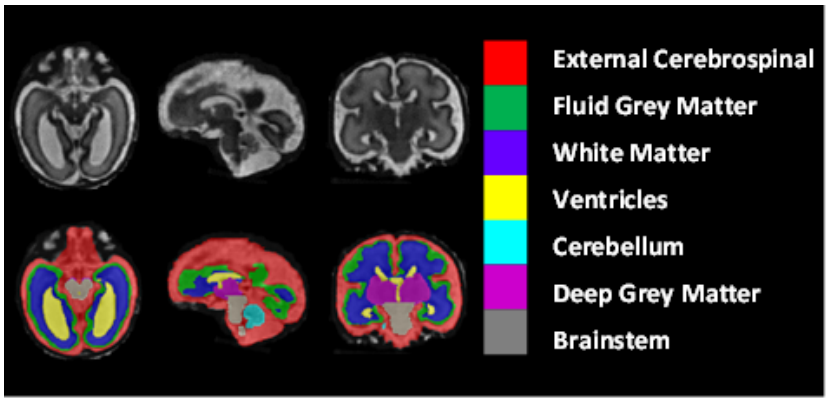


Figure 3.8: The class-mapping function for the axial, sagittal, and coronal slices.

overlap between ground truth (GT) and predicted segmentation masks. For the GT and predicted masks, DSC is defined as follows:

$$\text{Dice}(X, Y) = 2|X \cap Y|/|X \cup Y| \tag{3.13}$$

Hausdorff distance (HD95): The HD95 is calculated as the mean of two directed 95% Hausdorff distances:

$$HD95 = \frac{\vec{d}H, 95(X, Y) + \vec{d}H, 95(Y, X)}{2} \tag{3.14}$$

where X is the ground truth (GT) and Y is the predicted mask. HD is the maximum distance between the sets of points X and Y and between Y and X.

Sensitivity is used to compute the positive portion of voxels by using the ground-truth (GT) and predicted segmentation masks.

$$\text{Sensitivity } = TRP = TP/(TP + FN) \tag{3.15}$$

where TRP is the true positive rate, TP is true positive, and FN is false negative.

Specificity is also called the true negative rate (TNR), and it is used to compute performance based on the GT and predicted segmentation masks.

$$\text{Specificity } = TNR = TN/(TN + FP) \tag{3.16}$$

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

In addition, we use the root-mean-square error (RMSE) and concordance (C-index) to evaluate the GA prediction models. The C-index is used to compute the correlation between the predicted gestational age and ground-truth gestational age. The RMSE can be expressed as follows:

$$RMSE = \sqrt{\frac{\sum(GA_{pre} - GA_{GT})^2}{N}} \tag{3.17}$$

The C-index can be formulated as follows:

$$\text{C-index} = concordance - index(GA_{GT}, GA_{pre}) \tag{3.18}$$

where GA_{pre} is the predicted value, GA_{GT} is the ground-truth value for the ith observation in the dataset, and N is the sample size.

3.3.3 Performance Analysis of the Proposed Segmentation Model

3.3.3.1 Ablation Study

Table 3.1 tabulates the DSC, HD95, sensitivity, and specificity values of the proposed segmentation model, IRMMNET, with the axial, sagittal,

and coronal views (IRMMNET-Axial, IRMMNET-Sagittal, and IRMMNET-Coronal). IRMMNET-Coronal obtained better performance for all classes than IRMMNET with the axial and sagittal planes. It achieved DSC, HD95, sensitivity, and specificity scores of 0.789, 21.56, 0.818, and 0.976, respectively.

Table 3.1: Performance of the proposed segmentation model with the axial, sagittal, and coronal views.

Model	**DSC**	**HD95**	**Sensitivity**	**Specificity**
IRMMNET_Axial	0.778	24.06	0.8126	0.974
IRMMNET _Sagittal	0.781	22.80	0.817	0.974
IRMMNET _Coronal	0.789	21.56	0.818	0.976

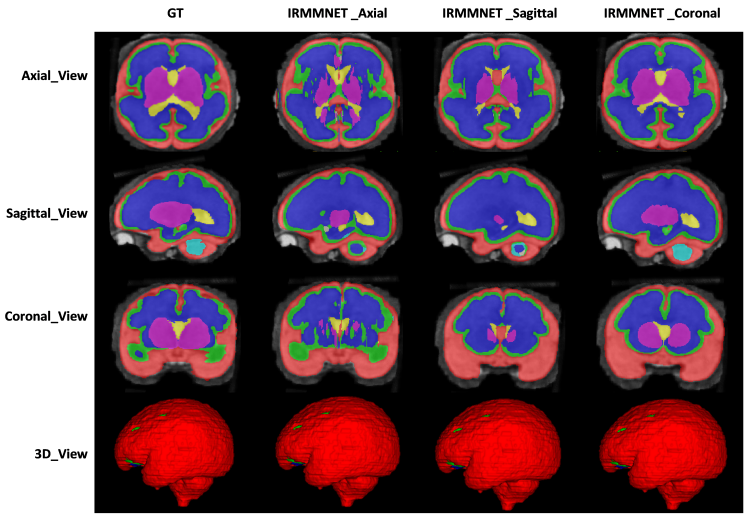


Figure 3.9: The segmentation maps of the proposed model with different views.

Figure 3.9 shows the proposed IRMMNET model's segmentation maps for the axial, sagittal, and coronal planes. In addition, 2D and 3D volumetric views of the segmentation images are shown in Figure 3.9 for

the axial, sagittal, and coronal planes. It can be visibly noticed that the coronal-view model generated accurate segmentation masks, in which the predictions for the deeper and smaller classes were close to the GT. It should be noted that the proposed model was initially tried on the 2D axial slices, but the predicted segmented images yielded a bad prediction for the deep classes. Therefore, the proposed segmentation model was applied to the three views' 2D slices (axial, sagittal, and coronal) for fetal brain segmentation. Among the predicted segmentation results for the fetal brain, the 2D coronal view produced the best results.

In Table 3.2, we compare the proposed model with different segmentation models. Specifically, the basic UNet (BaseUNet) for 2D brain tissue segmentation was trained by using the axial, coronal, and sagittal views (BaseUNet-Axial, BaseUNet-Sagittal, and BaseUNet-Coronal). As one can see, BaseUNet-Coronal yielded the highest DSC score (0.728) and the lowest HD95 score (29.042). In addition, different ResUnet models were trained by using the axial, coronal, and sagittal views (ResUnet-Axial, ResUnet-Sagittal, and ResUnet-Coronal) for 2D brain tissue segmentation. In the ResUnet model, the residual blocks were added to the base UNet model. The results of the three ResUnet models were better than those of the BaseUNet models. Finally, we added squeeze-and-excitation (SE) blocks into ResUNET (SE-ResUNet) and trained it on the three views, yielding SE-ResUNet-Axial, SE-ResUNet-Sagittal, and SE-ResUNet-Coronal. However, SE-ResUNet-Coronal had an improved DSC score compared to that of ResUnet-Coronal; its HD95 and specificity values were worse. As shown in Table 3.2, the proposed IRMMNET model comparatively produced a better performance with the axial and sagittal views. However, IRMMNET-Coronal achieved a DSC score of 0.789 and HD95 score of 21.565, which were better than those of all models used for comparison.

To enhance the proposed model's prediction, we fused the axial, sagittal, and coronal outputs of the proposed model to create a so-called multi-view model, which provided a 3D segmentation map. Later, the

Table 3.2: Comparing IRMMNET with different segmentation models.

Model	**DSC (%)**	**HD95 (mm^3)**	**Sensitivity (%)**	**Specificity (%)**
BaseUNet-Axial	0.727	29.401	0.806	0.952
BaseUNet-Sagittal	0.723	31.381	0.790	0.947
BaseUNet-Coronal	0.728	29.042	0.813	0.963
ResUNet-Axial	0.748	26.111	0.827	0.968
ResUNet-Sagittal	0.756	27.334	0.818	0.969
ResUNet-Coronal	0.752	26.014	0.828	0.977
SE-ResUNet-Axial	0.762	28.262	0.809	0.973
SE-ResUNet-Sagittal	0.769	28.888	0.808	0.978
SE-ResUNet-Coronal	0.773	27.101	0.879	0.969
IRMMNET-Axial	0.778	24.062	0.819	0.972
IRMMNET-Sagittal	0.781	22.801	0.817	0.974
IRMMNET-Coronal	0.789	21.565	0.818	0.976

performance was evaluated by using the predicted 3D segmentation map achieved with our three multi-view models and a GT segmentation map. It resulted in a better estimation in terms of the Dice, HD95, sensitivity, and specificity scores. We constructed a 3D segmentation map from the three views of the proposed model and evaluated the performance by using the predicted 3D segmentation map (achieved with our three multi-view models) and ground-truth segmentation map. Table 3.3 presents an ablation study of the proposed IRMMNET. Although the baseline Multi-view-2D-Inception+Residual model achieved optimal performance, its performance was upgraded when the DSAM module (i.e., Multi-view-2D Inception + Residual + DSAM) was systematically added. However, adding the multi-scale feature approach to the Multi-view-2D-Inception+Residual+DSAM model with the fusion of multiple views (i.e., IRMMNET) produced the highest performance scores in comparison with those of all of the state-of-art-methods and the baseline model for the fetal brain segmentation task when using the FeTA 2021 dataset. IRMMNET achieved DSC, HD95, sensitivity, and specificity scores of 0.791, 21.66, 0.819, and 0.980, respectively.

Table 3.3: Performance of various configurations of the proposed IRMMNET model.

Model	**DSC**	**HD95**	**Sensitivity**	**Specificity**
UNet	0.733	28.58	0.817	0.968
Multi-view_2D Inception + Residual	0.778	25.42	0.8178	0.967
Multi-view_2D Inception + Residual + DSAM	0.783	23.26	0.8101	0.976
IRMMNET	0.791	21.66	0.819	0.980

We also studied the efficacy of different loss functions with the proposed model. As tabulated in Table 3.4, the proposed Combo loss function achieved better performance than that of the binary cross-entropy (BCE) and Dice loss functions. No big improvements were noticed when we combined the Dice loss and BCE loss. However, the Dice loss produced the lowest scores.

Table 3.4: Performance of IRMMNET with different loss functions.

Loss Function	**DSC (%)**	**HD95 (mm^3)**	**Sensitivity (%)**	**Specificity (%)**
BCE	0.789	23.88	0.671	0.812
Dice	0.776	24.51	0.668	0.795
BCE + Dice	0.780	22.83	0.682	0.809
Combo	0.791	21.66	0.691	0.818

The training and validation times were also estimated for the proposed and state-of-the-art methods for a comparison of the computational costs. The training time of our proposed solution was 55 min, and the time taken for validation was less than 2 min. The computational times for training and validation are given in Table 3.5.

Table 3.5: Estimations of the training and validation times for the proposed and state-of-the-art methods.

Segmentation Models	**Training Time (min)**	**Validation Time (min)**
BaseUNet-Axial	45	2
BaseUNet-Sagittal	50	2.1
BaseUNet-Coronal	55	2.01
ResUNet-Axial	60	2.23
ResUNet-Sagittal	62	2.45
ResUNet-Coronal	61	2.11
SE-ResUNet-Axial	63	2.53
SE-ResUNet-Sagittal	65	1.95
SE-ResUNet-Coronal	64	1.88
IRMMNET-Axial	53	1.3
IRMMNET-Sagittal	55	1.4
IRMMNET-Coronal	52	1.5

We applied a Mann–Whitney U test or Wilcoxon Rank Sum test to compute the p-values between the predicted masks and their corresponding ground truths (Qayyum et al., 2020; Zabihollahy et al., 2020). In segmentation tasks, the p-value must be higher than 0.05 to be statistically significant, unlike in classification tasks. A comparison of the statistical analyses of the proposed IRMMNET and the state-of-the-art methods is given in Table 3.6. In the table, a p-value that is greater than 0.05 represents a more remarkable similarity between the predicted and ground-truth segmentation maps. Similarly, a higher p-value also represents an accurate segmentation. The table shows that the proposed IRMMNET had consistent results, and it statistically validates the segmentation results.

To validate the results of the proposed fused model (i.e., Multi-view-IRMMNET), we also applied the same fusion technique to the ResUNet and SE-ResUNet models, which yielded Multi-view-ResUNET and Multi-view-SE-ResUNET. The predicted 2D slices and 3D volumes of the proposed model, ResUnet, and SE-ResUNet are shown in Figure

3.10. Although ResUNet and SE-ResUNet successfully predicted outer classes, such as external cerebrospinal fluid (ECF), they failed to predict inner/deeper classes, like deep gray matter (DGM). It is conspicuous that the proposed model's predictions for all classes were close to the given GT.

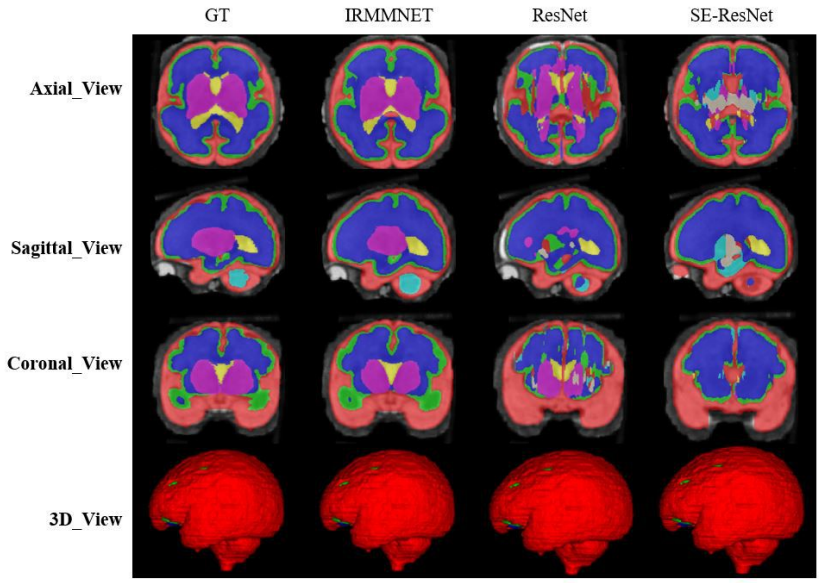


Figure 3.10: The segmentation results of the proposed Multiview IRMMNET, Multiview-ResUNET and Multiview-SE-ResUNET models.

Table 3.7 presents the Dice, HD95, sensitivity, and specificity scores of the proposed Multi-view-IRMMNET, Multi-view-ResUNET, and Multi-view-SE-ResUNET models. The proposed Multi-view-IRMMNET model achieved the best segmentation results with a DSC of 0.791, HD95 of 21.66, sensitivity of 0.819, and Specificity of 0.980. The Multi-view-ResUNET model achieved less performance scores than those of Multi-view-IRMMNET. This model achieved a DSC of 0.758, and Multi-view-SE-ResUNET achieved a maximum DSC of 0.772. We

Table 3.6: Comparison of the statistical analyses of the proposed IRMMNET and the state-of-the-art methods.

Segmentation Models	***p*-Value**
BaseUNet-Axial	0.65
BaseUNet-Sagittal	0.68
BaseUNet-Coronal	0.72
ResUNet-Axial	0.89
ResUNet-Sagittal	0.87
ResUNet-Coronal	0.97
SE-ResUNet-Axial	1.15
SE-ResUNet-Sagittal	1.10
SE-ResUNet-Coronal	1.01
IRMMNET-Axial	1.63
IRMMNET-Sagittal	1.7
IRMMNET-Coronal	1.9

Table 3.7: Comparing the segmentation results of the proposed Multi-view-IRMMNET, Multi-view-ResUNET, and Multi-view-SE-ResUNET models.

Model	**Dice**	**HD95**	**Sensitivity**	**Specificity**
Multi-view_ResUNET	0.758	27.39	0.826	0.97
Multi-view_SE-ResUNET	0.772	27.56	0.812	0.973
Multi-view_ IRMMNET	0.791	21.66	0.819	0.980

can conclude that our proposed Multi-view-IRMMNET achieved the highest DSC score in comparison with the other models, namely, Multi-view-ResUNET and Multi-view-SE-ResUNET. Similarly, Multi-view-IRMMNET achieved the lowest HD95 score of 21.66, as compared to those of Multi-view-ResUNET and Multi-view-SE-ResUNET.

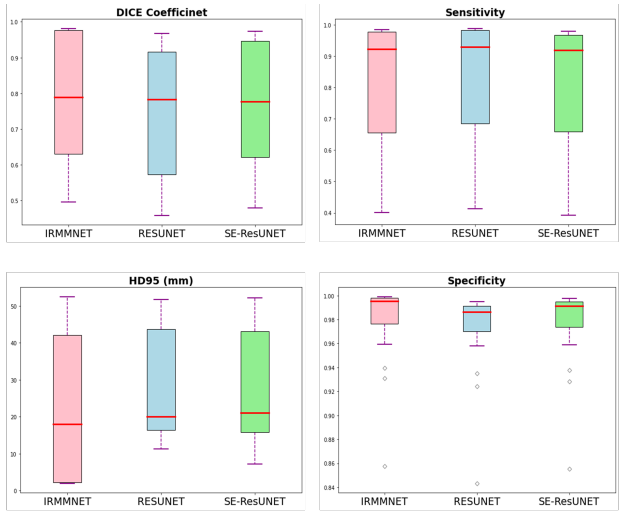


Figure 3.11: Box plots of the DSC, HD95, sensitivity, and specificity scores of the proposed Multi-view-IRMMNET, Multi-view-ResUNET, and Multi-view-SE-ResUNET models.

Figure 3.11 shows the box plots of the DSC, HD95, sensitivity, and specificity scores of the proposed Multi-view-IRMMNET, Multi-view-ResUNET, and Multi-view-SE-ResUNET models. Multi-view-IRMMNET showed the highest Q3 quartile in the DSC and specificity plots. In contrast, Multi-view-IRMMNET's median and maximum values remained higher in all plots compared to those of the Multi-view-ResUNET and Multi-view-SE-ResUNET models.

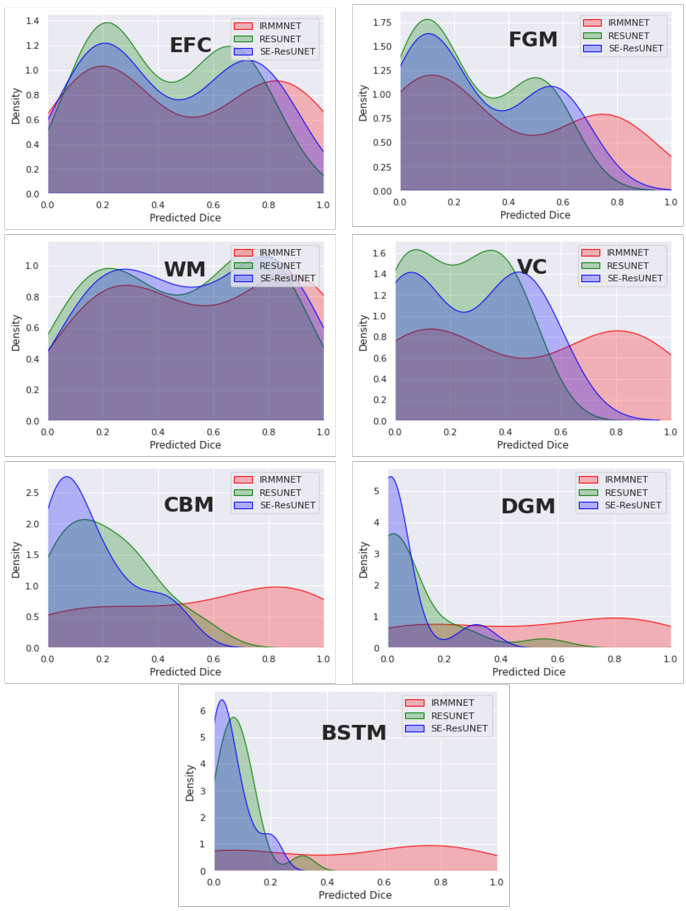


Figure 3.12: Density plots of the predicted 2D DSC for the proposed Multi-view-IRMMNET, Multi-view-ResUNET, and Multi-view-SE-ResUNET models.

Figure 3.12 overlays the density plots of the predicted 2D DSCs of the proposed Multi-view-IRMMNET, Multi-view-ResUNET, and Multi-view-SE-ResUNET models for each class. First, the distribution of the predicted 2D DSCs for Multi-view-IRMMNET was always significantly

different from those of the Multi-view-ResUNET and Multi-view-SE-ResUNET models for all classes. This was especially the case for VC, CBM, DGM, and BSTM, which were the classes opposite to EFC, FGM, and WM. It is worth noting that when a class was absent in a segmented slice, the predicted 2D DSC was zero for that slice. This showed a small trend of zero appearing for all classes, especially for deep classes, such as VC, CBM, DGM, and BSTM.

Similarly, the distribution scores of all classes with Multi-view-IRMMNET were greater than those with Multi-view-ResUNET and Multi-view-SE-ResUNET. However, regardless of the class, the 2D DSC distributions of the Multi-view-ResUNET and Multi-view-SE-ResUNET models were always significantly different, were more shifted to the left, and had a larger standard deviation compared to that of the distribution of the proposed solution. The proposed solution was always shifted toward the higher values on the right.

3.3.3.2 Comparing the Proposed Segmentation Model with Existing Methods

There has not been much research on the FeTA dataset and fetal brain segmentation, as private datasets are most commonly used in research. Table 3.8 compares the performance of the proposed Multi-view-IRMMNET with that of two state-of-the-art methods called DA-FaBiAN-Baseline presented in (Dumast et al., 2022) and TopoCP (2D) presented in (Dumast et al., 2021) in terms of DSC scores. As shown, Multi-view-IRMMNET achieved the best results, with a DSC score of 0.791. However, DA-FaBiAN-Baseline and TopoCP (2D) produced the same DSC score (0.70), but TopoCP (2D) obtained the smallest standard deviation.

Table 3.8: Comparison of the performance of the proposed Multi-view-IRMMNET model with that of two existing methods DA_FaBiAN_Baseline and TopoCP (2D).

Model	**DSC (%)**
DA_FaBiAN_Baseline	0.70 ± 0.24
TopoCP (2D)	0.70 ± 0.14
Multi-view-IRMMNET	0.791± 0.18

3.3.4 GA Prediction Results

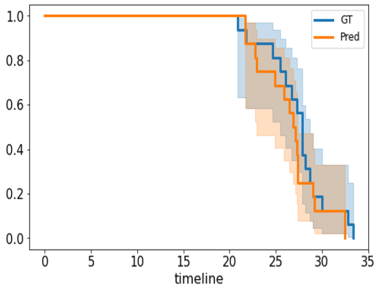
3.3.4.1 Analyzing the Performance of GA Prediction Models

As mentioned in Section 3.2.2, three different methods were used to extract features from the MRI images of the fetal brain, and then these features were fed into a regression algorithm to predict the GA. Four different regression techniques were used for the regression of the input features: LR, XGB, RF, and RT. The first feature set, which was called IRMMNET Deepfeat, was extracted from the last encoder layer of the proposed IRMMNET model, as depicted in Figure 3.5. The second feature set, the so-called 3D deep autoencoder features, was extracted from the autoencoder depicted in Figure 3.6. The third feature set included radiomics features (Figure 3.7).

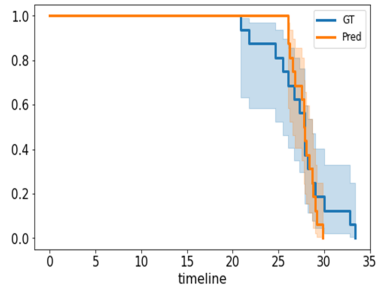
Table 3.9 shows that the RF regressor produced the lowest RMSE values with the radiomics, IRMMNET Deepfeat, and 3D deep autoencoder features. The RF regressor achieved the best GA prediction results with the radiomics features, with an RMSE score of 1.42 and a C-index score of 0.888. These results indicate that the proposed radiomics features used with the RF regressor were the most reliable method for predicting GA.

Table 3.9: Analyzing the performance of different feature types and regression techniques for GA prediction.

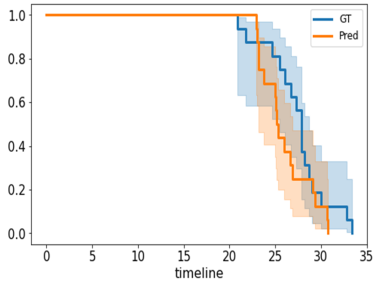
Models	Linear Regressor		XGB Regressor		Random Forest		Decision Tree	
	RMSE	C-index	RMSE	C-index	RMSE	C-index	RMSE	C-index
Radiomics	1.70	0.837	1.477	0.854	**1.42**	**0.888**	1.44	0.858
Deep Features	4.56	0.465	4.10	0.252	3.46	0.371	5.97	0.542
3D Autoencoders	4.51	0.418	3.71	0.427	3.26	0.512	4.64	0.517



(a) Radiomics features



(b) Deep features



(c) 3d Autoencoder features

Figure 3.13: Kaplan–Meier plots of the radiomics, IRMMNET Deepfeat, and 3D deep autoencoder features with RF.

Figure 3.13 presents Kaplan–Meier plots of the radiomics, IRMMNET Deepfeat, and 3D deep autoencoder features. These plots show the gestational days predicted by the proposed GA prediction model

in comparison with the ground truth for the validation datasets. Figure 3.13a shows the days predicted with the radiomics features with RF, Figure 3.13b shows the days predicted with the IRMMNET Deepfeat with RF, and Figure 3.13c shows the days predicted with the 3D deep autoencoder features with RF. As shown, the curves of the predicted days were very close to the GT curve for the radiomics-based features, unlike the curves for the IRMMNET Deepfeat and 3D deep autoencoder features. The curve of the predicted days that was produced based on the 3D deep autoencoder features was far from the curve based on the GT days. Hence, these curves prove the efficacy of the proposed GA prediction method based on radiomics features and RF.

3.3.4.2 Analyzing the Explainability of the Radiomics Features

As shown above, the radiomics features yielded the best GA prediction results. Here, we employ the SHAP explainability method (SHapley Additive exPlanations) in order to analyze the most explainable radiomics features and their importance. SHAP is a game-theoretic approach to explaining the output of any machine learning model (Antwarg et al., 2021). Figure 3.14 shows the feature importance of the radiomics features for the further analysis of the approach to using the radiomics features for GA prediction. We used the 18 best-explainable radiomics features with RF and obtained better scores than with all 108 extracted radiomics features. It was shown that the max 3D diameter and major axis length had the highest feature importance. We can say that the max 3D diameter and minor axis length were the most important features for GA prediction in comparison with the other features, which can help in future planning.

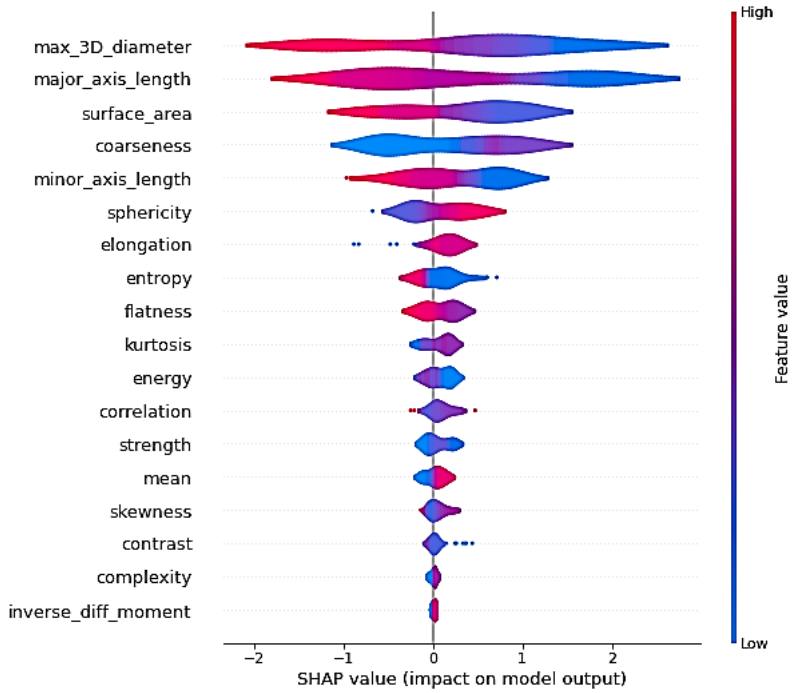


Figure 3.14: Feature importance of the radiomics features for GA prediction. The colors from red to blue represent the significance of the features in model prediction.

3.3.5 Generalization Capabilities of the Proposed Fetal Brain Segmentation and GA Prediction Models

To demonstrate the generalization capabilities of the proposed fetal brain segmentation model, we tested it on the Hecktor 2021 dataset (Andrearczyk et al., 2022). A total of 224 training samples were provided. This dataset also contained clinical values and imaging samples. The dataset was divided into 80% for training and 20% for testing. The Hecktor 2021 dataset was converted into axial, coronal, and sagittal views. We trained the proposed IRMMNET fetal brain segmentation model

with the axial, coronal, and sagittal views from a head and neck tumor dataset. Each volume in the Hecktor 2021 dataset had a spatial resolution of $144 \times 144 \times 144$. The segmentation masks of the IRMMNET model for each view were fused to construct the 3D segmentation of the tumors in the head and neck dataset. Table 3.10 compares the results of the proposed model with those of existing state-of-the-art methods in terms of the DSC and HD95 scores. IRMMNET achieved a DSC score of 0.77 and an HD95 score of 3.02, which were better than those of the methods used for the comparison.

In addition, the proposed GA prediction method was tested on the the Hecktor 2021 dataset for the prediction of the patients' survival days. The progression-free survival outcomes for all patients were provided in the CSV files with their clinical data and with various clinical variables. The head and neck tumor progression was based on the RECIST criteria: either an increased size of a known tumor (change in N and/or T) or the existence of a new tumor (change in M and/or N). Death due to a specific disease was also considered the progression of a disease that was previously considered stable.

Table 3.10: Analysis of the generalization capabilities of the proposed fetal brain segmentation model on the Hecktor 2021 dataset.

Method	**DSC (%)**	**HD95 (mm^3)**
IRMMNET	0.77	3.02
(Andrearczyk et al., 2021)	0.65	4.07
(Bourigault et al., 2021)	0.75	3.27

Clinical variables, such as the patient age, patient gender, center ID, TNM group, M-stage, N-stage, T-stage, TNM edition, and chemotherapy status, were given with different values, and some variables had missing values. We used imputation to complete the missing values for all clinical features. We mapped integer values to each the individual N-, M-, and T- staging datum as follows: T-stage (Tx: 0, T1: 1, T2 : 2, T3: 3, T4: 4, T4a: 5, T4b: 6), N-stage (N0: 0, N1: 1, N2:N2a: 3, N2b: 4, N2c: 5,

N3: 6), and M-stage (Mx: 0, M0: 0, M1:1). In addition, the TNM group was also mapped to an ordinal categorical variable, which was based on the corresponding TNM stage information (7 I: 0, 8 I: 0, 7 II: 1, 8 II: 1, 7 III: 2, 8 III: 2, 7 IV: 3, 8 IV: 3, 7 IVA: 4, 8 IVA: 4, 7 IVB: 5, 8 IVB: 5, 7 IVC: 6, 8 IVC: 6).

The min/max normalization method provided by the scikit-learn Python package was used to normalize the clinical features' values. A scaler was instantiated by using only the training data and then applied to the test set.

In these experiments, the RF regression technique was employed. As tabulated in Table 3.11, the clinical features obtained a C-index of 0.692. However, the C-index of the clinical features was lower than those of the radiomics, DeepFeat (deep features extracted from the encoder of the proposed IRMMNET model), and 3D deep autoencoder features, but it was better than those in existing studies, such as (Starke et al., 2020; Aerts et al., 2014). The Clinical + DeepFeat + Radiomics combination led to a C-index of 0.786, which was lower than the Deep-Features + Radiomics combination, meaning that we could achieve accurate predictions of patients' survival days without employing clinical data. Table 3.11 also demonstrated that the features based on radiomics and DeepFeat achieved the highest C-index scores (0.821) compared to those of methods that were specially designed for the prediction of patients' survival days, such as (Bourigault et al., 2021; Juanco-Müller et al., 2021; Starke et al., 2020; Aerts et al., 2014).

3.4 Discussion

Image segmentation is the first stage in the volumetric quantification of the developing fetal brain, which is used to examine the neurological growth of fetuses with congenital abnormalities and inform prenatal planning. The predicted delivery date is crucial for an accurate estimation of gestational age when managing any pregnancy. The timing of

Table 3.11: Analysis of the generalization capabilities of the proposed feature extraction methods on the Hecktor 2021 dataset for the prediction of patients' survival days.

Method	**C-Index**
Clinical features	0.692
Radiomics	0.791
DeepFeat	0.771
3D deep autoencoder features	0.723
Deep-Features + Radiomics	0.821
Clinical + DeepFeat + Radiomics	0.786
(Bourigault et al., 2021)	0.810
(Starke et al., 2020)	0.470
(Aerts et al., 2014)	0.690

appropriate obstetric treatment and the scheduling and interpretation of some antepartum diagnostics that assess the appropriateness of fetal growth and measures of development in order to prevent preterm births and associated morbidities depend heavily on accurate knowledge of the gestational age. We must create an automated, accurate, and precise procedure for fetal brain segmentation and gestational age calculation.

This Chapter tackled the tasks of fetal tissue segmentation and gestational age prediction. The 2D-based multi-view (axial, coronal, and sagittal) models were analyzed for fetal brain tissue segmentation. The end-to-end fetal tissue segmentation and GA prediction models were trained and tested with the FeTA 2021 MRI dataset. Initially, we trained our segmentation model on only axial slices. However, the results were not convincing, especially in brain tissues in which the number of class pixels was small. In addition, we acquired all axial, sagittal, and coronal slices from the 3D fetal brain input volume. We trained the model on those particular view slices. The coronal view model performed better than the sagittal and axial models based on the experimental results and observations. Later, the fusion of the predictions of the axial, sagittal, and coronal views was combined to enhance the model's outcome.

Various regression techniques were developed for GA prediction by using different feature extraction methods, including a 3D autoencoder, radiomics features, and features extracted from the IRMMNET encoder. The IRMMNET encoder's features were used with different regressors (LR, RF, RT, and XGB) to predict gestational age. Radiomics- and 3D-autoencoder-based features were also used for GA prediction. Extensive experiments were performed to achieve the optimal performance in optimal segmentation and gestational age prediction. Metrics such as those of the Kaplan–Meier and the SHAP explainability methods were used to study the explainability of the proposed GA models. Different comparisons and datasets were utilized to validate the generalization of the proposed models.

The main limitation of our study is that the sample size for each GA period in the held-out validation dataset was relatively small. Therefore, validating the proposed GA estimation model with data from larger populations and settings will be critical in order to extend the current use of this MRI-based biometric measurement prediction model to clinical application scenarios. Since fetal growth is influenced by each mother's previous gestational history, body condition, and composition, future studies should consider mothers' demographics as variables or covariables in their models, which might be useful for improving the precision of gestational dating.

We chose multi-view multi-scale deep learning segmentation models in order to deliver the complete volumetric information and to provide an efficient application to fetal brain segmentation, gestational age prediction, and head and neck tumor segmentation and survival analysis. Our proposed approach consisted of various modules for predicting 3D segmentation maps with a limited dataset and lower computational resources. Our proposed approach produced a more efficient and faster response on the validation dataset for the task of fetal brain segmentation.

The imaging features might not be sufficient to accurately predict gestational age. However, we used various feature extraction approaches, which included imaging, radiomics, 3D latent space autoencoder-based features, and deep features extracted from the last layer of multi-view 2D image slices from segmented brain tissues, to extract more localized features for gestational age prediction. The fusion of multi-scale segment-based deep features achieved better performance than that of state-of-the-art methods. There is a further need to investigate different methods with deep learning models for GA prediction. The existing methods are based on single-feature extraction techniques that use basic deep learning models. Correspondingly, the datasets used in existing methods are in-house and private. There is a need to set a benchmark on a publicly available dataset for further comparisons and enhancements in deep learning/machine learning for gestational age prediction and fetal brain segmentation.

We developed a couple of imaging-based features and validated them on two different medical imaging datasets for segmentation, fetal age prediction, and head and neck survival analysis. This is a comprehensive end-to-end solution for fetal brain segmentation and gestational age prediction. Three types of features radiomics features, clinical features, and latent features from the 3D autoencoder were used for the prediction of gestational age with various regression models. These regression models were applied together with various feature fusion combinations for the prediction of gestation age. Our proposed model has also used in head and neck cancer segmentation. The head and neck segmentation features were used in the prediction of the survival age, and the performance of the proposed model was compared with the performance of state-of-the-art models for the tasks of head and neck cancer segmentation and survival prediction.

3.5 Conclusion

This Chapter proposed an end-to-end, fully automated, and effective method for multi-tissue fetal brain segmentation called IRMMNET. We found that the proposed fetal brain segmentation model obtained the best results with the Combo loss function, achieving a DSC score of 0.791 and an HD95 score of 21.66, outperforming other state-of-the-art models. In addition, the density plot analysis demonstrated that with the proposed segmentation model, the distribution scores of all classes were greater than those with other models.

In the GA (Gestational Age) estimation task, different regression techniques (LR, RF, RT, and XGB) were assessed in combination with three feature extraction approaches (3D autoencoder, radiomics, and IRMMNET encoder). We found that when used with RF, the radiomics features led to the best GA prediction results, with an RMSE score of 1.42 and a C-index score of 0.888. Further, we studied the explainability and the importance of the radiomics features, and we found that the max 3D diameter and minor axis length were the most important features in GA prediction in comparison with the other features.

Finally, we studied the generalization capabilities of the proposed fetal brain segmentation and GA prediction methods for two tasks, namely, head and neck tumor segmentation and the prediction of patients' survival days. When applied to head and neck tumor segmentation, we found that the proposed segmentation model outperformed existing models that were specifically designed for this task.

In future work, we will use different patch-based 3D segmentation models and 3D transformers for the task of fetal brain segmentation in order to further enhance the system's performance. In addition, we will validate the GA estimation method with data from larger populations and settings.

Part III

3D Deep Learning Approach for Brain Tumor Segmentation & Survival Prediction

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Chapter 4

3D Brain Tumour Segmentation Based on Self-supervised Contrastive Learning and Transformers

Medical image segmentation has seen significant progress using supervised deep learning. Hereby, large annotated datasets were employed to segment anatomical structures reliably. To reduce the requirement for annotated training data, self-supervised pre-training strategies on non-annotated data were designed. Contrastive learning schemes operating on dense pixel-wise representations have been introduced as an effective tool. In this work, we expand on this strategy and leverage inherent anatomical similarities in medical imaging data. We apply our approach to the task of semantic segmentation in a self-supervised setting with limited amounts of annotated volumes. Trained alongside a segmentation loss in one single training stage, a contrastive loss aids in differentiating between salient anatomical regions that conform to the available annotations. We proposed an efficient parallel

transformer module using Multiview multiscale features and depth-wise features. The multiEncoder-based proposed transformer is trained on a self-supervised approach using contrastive loss. First, we prepared the proposed transformer with an unlabelled dataset and fine-tuned one encoder coming from the first stage and a second encoder trained with a few annotated segmentation masks. Further, we concatenated the features of both encoders for final brain tumor segmentation. MultiEncoder-based transformers achieved excellent results in various medical image segmentation tasks. We have validated our proposed solution on interdisciplinary real-time medical imaging datasets. Comprehensive experiments demonstrate the proposed model has potential as a versatile segmentation model as outperformed on brain tumor datasets as well as other interdisciplinary datasets.

4.1 Introduction

With the advent of supervised deep learning, semantic segmentation of anatomical structures became possible with high accuracy in various medical imaging modalities Isensee et al., 2021. Nonetheless, providing accurate and consistent manual annotations in sufficient amounts for algorithm training remains a challenging and tedious task Tajbakhsh et al., 2020. With an ever-increasing abundance of available imaging data, the lack of high-quality annotations impedes their automated analyses.

To decrease the necessity for annotated training data, new training schemes using the paradigm of self-supervised learning (SSL) have gained recent interest. Hereby, vector embeddings of image information are learned that are applicable for various tasks, including natural language processing Devlin et al., 2018, computer vision, or medical imaging Bai et al., 2019. Contrastive learning (CL) Bardes, Ponce, and LeCun, 2022 as a specific form of SSL has been shown to enable state-of-the-art algorithm performance in the case of substantially reduced amounts of given annotations.

Most prominent realizations of CL use differently augmented pairs of input samples (positives), that are encouraged to be encoded with similar representations, whereas the representations of distinct samples (negatives) are forced to be dissimilar. The resulting neural network parameters are suitable as initialization for subsequent tasks, whereby the learned embeddings are adapted to the respective downstream task by fine-tuning. Naturally, this approach has been expanded to dense pixel-wise imaging data, enabling semantic segmentation with limited amounts of data Xie et al., 2021. However, recent approaches of CL limit themselves to a two-stage procedure. In the first stage, a general embedding of samples is established. This is followed by a second stage, where task-dependent finetuning is performed on a downstream task. The second stage may be detrimental to the learned embedding which is progressively altered to be conducive to the enforced objective. In practice, this makes the availability of separate validation samples a necessity to avoid overfitting to a (sparsely annotated) downstream task and thereby to ensure adequate generalization to unseen data. The separate training stages hinder jointly learning the objectives of both stages, which may aid in providing a more robust embedding. Another drawback is that contrastive approaches rely on a separation into positive and negative examples. This results in embeddings that follow pre-defined or heuristic local or global separations. Therefore, this leads to rigid representations that adhere to the induced structural bias. To improve on these limitations, we make use of inherent structural similarities of the underlying human anatomy and proposed Multi transformer encoder parallel transformer-based solution in a self-supervised approach.

Transformers models have been widely used in many computer vision applications Vig, 2019. However, the transformer required more training data as compared to the convolutional neural network. Very long sequence length is another issue that needs to be addressed in the medical imaging field, especially for 3D volume for segmentation and classification tasks. Several works attempt to introduce inductive bias

into Transformer to reduce the requirement of training data such as CvT Wu et al., 2021 and CoAtNet Dai et al., 2021. Vision transformer used patches of images into tokens that would discard all structural information within the patch. The self-attention based on quadratic complexity requires a long sequence and this would be difficult to get dense prediction in the segmentation task. Various researchers proposed different directions to handle the computational complexity while using transformer architectures. SwinTransformer proposed Liu et al., 2021 a local region of convolution-based self-attention using non-overlap window-based multi-head self-attention. CCNet Huang et al., 2019 decomposed 2D attention to 1D attention. UTNet Gao, Zhou, and Metaxas, 2021 and concurrent work CvT Wu et al., 2021, and PvT Wang et al., 2021 proposed methods to reduce the tokens in key and values to improve the performance in terms of computational complexity. We have proposed a Multiview approach to use multiscale input features in the self-attention module to reduce the complexity. Furthermore, we also used a depth-wise convolutional layer block in parallel with the Multiview attention block to further improve the performance of the proposed solution. There are attempts to apply Transformer in the medical image segmentation field. TransUNet Chen et al., 2021 and UNETR Hatamizadeh et al., 2022 add a 2D or 3D convolutional decoder to a ViT-like encoder into medical image segmentation. SwinUNet Cao et al., 2022 and VTUNet Peiris et al., n.d. proposed a pure Transformer model based on Swin-Transformer for 2D or 3D segmentation. nnFormer Zhou et al., 2021, SwinUNETR Hatamizadeh et al., 2021 and HiFormer (Heidari et al., 2023) use Swin-like hybrid architectures. These models are used to train with large training datasets or to be initialized with pre-trained weights on large-scale natural image datasets or longer training epochs, otherwise they prospective achieve inferior performance on medical image datasets.

The main contributions of this chapter are:

1. We proposed an efficient parallel transformer module using Multiview multiscale features and depth-wise features.

2. The multiEncoder-based proposed transformer trained in self-supervised using contrastive loss. First, we prepared the proposed transformer with an unlabelled dataset and fine-tuned one encoder coming from the first stage and a second encoder trained with a few annotated segmentation masks. Further, we concatenated the features of both encoders for brain tumor segmentation.

3. Multiencoder-based transformer achieved excellent results in various medical image segmentation tasks.

4.2 Material & Methods

This section discusses the datasets and the methodologies used in this manuscript. The datasets include the labeled and unlabelled data for training the proposed self-supervised contrastive learning approach.

4.2.1 Dataset

There are two datasets from BraTS are used in this manuscript.

1. To handle the unlabelled part of the proposed solution the BraTS2021 training dataset Baid et al., 2021 Menze et al., 2014 Bakas et al., 2017 without labels was used for training the encoder part of the self-supervised contrastive learning approach. The total number of subjects is 1200 which is further divided into training and testing sets on 80:20 ratio. The total training set holds 960 subjects, and the testing set holds 240 subjects. Each subject has nifti volumes for Flair, T1, T1CE, and T2 MRI modalities. There are three tumor labels have been given in the ground truth (GT) masks including enhancing tumor (ET) (labeled as class-4 in GT), peritumoral edema (ED) (labeled as class-2 in GT), and non-enhancing tumor/ narcotic tumor (NCR) (labeled as class-1 in GT). All given four modalities are stacked and given to the model for training

and testing. Figure 4.1 shows the Flair, T2, T1, T1CE, with their corresponding segmentation GT of a subject from the BraTS 2021 dataset.

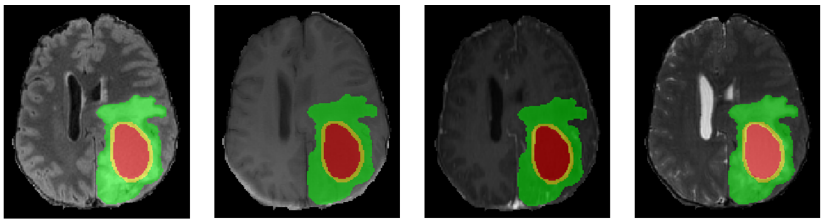


Figure 4.1: The four MRI modalities overlayed with their respective ground-truth (a) shows the Flair image, (b) shows the T1 image, (c) shows the T1CE image, and (d) shows the T2image.

2. The labeled BraTS2020 dataset has been used in this manuscript for training the second part of the proposed segmentation solution. The total number of subjects in the training set of BraTS 2020 with their masks is 369 Menze et al., 2014. Similar to the BraTS 2021 dataset each subject has nifti volumes for Flair, T1, T1CE, and T2 MRI modalities. There are three classes of tumor given in the ground truth (GT) masks including enhancing tumor (ET) (labeled as class-4 in GT), peritumoral edema (ED) (labeled as class-2 in GT), and non-enhancing tumor/ narcotic tumor (NCR) (labeled as class-1 in GT). In our experiment, all four MRI modalities are stacked for brain tumor segmentation using the proposed model. Here, the green color in the GT segmentation represents the ET class, the yellow color stands for ED, and the red color stands for the NCR tumor class.

4.2.2 Proposed Self-Supervised Learning model

The proposed two-stage self-supervised learning for brain tumor segmentation has two main parts. We named the proposed model 'Multiview Multiscale Parallel Attention Transformer UNET' (MMPATransUNet).

In the first stage of the model, augmentation has been used to mutate the data and secondly, it utilizes regularized contrastive loss Chen et al., 2020 to learn feature representations of the unlabeled data. The multiple augmentations are applied on a randomly selected 3D foreground patch from a 3D volume. Two augmented views of the same 3D patch are generated for the contrastive loss as it functions by drawing the two augmented views closer to each other if the views are generated from the same patch, if not then it tries to maximize the disagreement. Contrastive learning offers this functionality on a mini-batch. We have used masked volume inpainting, contrastive learning, and rotation prediction as proxy tasks for learning contextual representations of input images.

The primary task of the network is to reconstruct the original image. The different augmentations used are classical techniques such as in-painting Pathak et al., 2016, out-painting, and noise augmentation to the idea by local pixel shuffling Chen et al., 2019. The secondary task of the network is to simultaneously reconstruct the two augmented views as similar to each other as possible via regularized contrastive loss Chen et al., 2020 as its objective is to maximize the agreement. The term regularized has been used here because contrastive loss is adjusted by the reconstruction loss as a dynamic weight itself. Multiple patches having size 64x64x64 are generated and used in different views based on the augmentation via the transforms on the same cubic patch. The objective of the self-supervised learning (SSL) network is to reconstruct the original image. The contrastive loss is driven by maximizing agreement of the reconstruction based on the input of the two augmented views. For the first stage proposed model, 960 subjects from Brats 2021 dataset are

used.

In the second stage, the pre-trained encoder weights from the first stage are used in the proposed 3D Multiview and multiscale parallel attention transformer (MMPTtransNet) for brain tumor segmentation on BraTS 2021 dataset. We trained our proposed model using a training-supervised dataset based on patches. We randomly generated patches with size 64x64x64 from input volume and used different augmentation to train the proposed model. We used a sliding window with 64 strides to generate the prediction volume. The training transforms such as RandCrop, RandGaussianNoise, RandGaussianSmooth, RandShiftIntensity, RandAdjustContrast, and RandZoomd was used to train the second stage proposed model. For the second stage proposed model, 369 subjects from Brats 2020 dataset are used. The complete architecture of the Two-stage self-supervised contrastive learning for brain tumor segmentation ((MMPATransUNet) is given in Figure 4.2.

Multiview and multiscale parallel attention block This manuscript proposes an efficient multiscale Multiview parallel attention transformer block for 3D volumetric segmentation.

The inp_v feature maps from the decoder side of the proposed module are passed to the DConv layer before the proposed attention module. Therefore, the proposed attention transformer block used a full token map and passed this token map to the depth-wise convolutional layer. We use the depth-wise separable convolution as an efficient version of convolution implemented by: depthwise conv $\rightarrow$ pointwise conv, where the depth-wise convolution gathers the spatial information. In contrast, the pointwise convolution gathers along the channel dimension. Given token sequence Att_f after the attention module, we first reshape them back to 2D and transform them with convolutional blocks as shown in the equation. The proposed convolutional projection and feed-forward network are a generalized version of the origin Transformer design that is implemented using a 1×1 convolution layer in both modules.

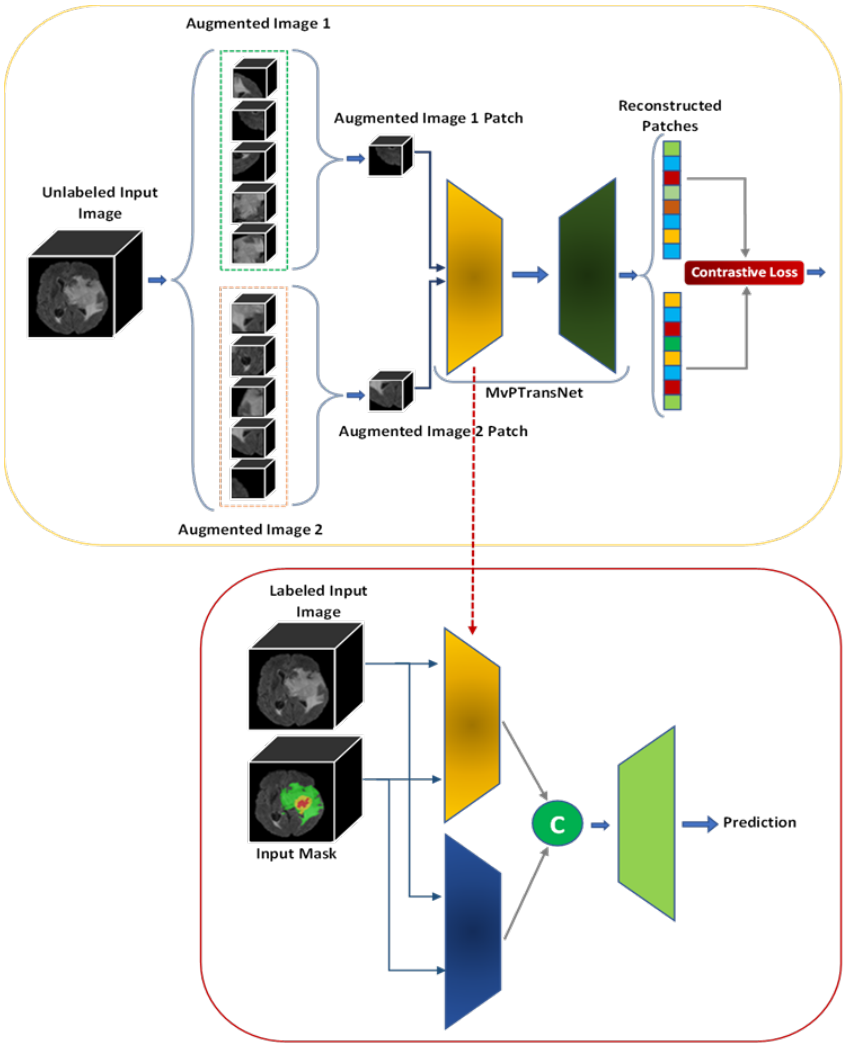


Figure 4.2: Two stages Self-supervised learning for BraTS segmentation.

$$F = DConv(inp_v) \tag{4.1}$$

$$V_f = Fv \tag{4.2}$$

$$Q_f = Fq \tag{4.3}$$

$$K_f = Fk \tag{4.4}$$

$$Att_f = \frac{Q_f^T \otimes K_f}{\sqrt{d_k}} \bigotimes V_f \tag{4.5}$$

$$\tilde{F} = F + DConv\left(Reshpe_{2D}\left(\frac{Q_f^T \otimes K_f}{\sqrt{d_k}} \bigotimes V_f\right)\right) \tag{4.6}$$

In the second branch, The feature obtained from three views (sagittal, coronal, and axial) is passed to three Conv layers with BN and ReLU using different kernel sizes such as $3 \times 1 \times 1$, $1 \times 3 \times 1$, and $1 \times 1 \times 3$ in each convolutional layers to produced three feature maps. The activation maps from Q(sagittal view), K(coronal view), and V are passed to the softmax layer. The importance of each feature is performed and these multi-scale feature maps to get the output of the proposed module as shown in Eq. 4.11

$$Q_{sagittal} = Conv(:,:,k_{size} = (1,3,1), pad = (0,1,0)) \tag{4.7}$$

$$Q_{Coronal} = Conv(:,:,k_{size} = (3,1,1), pad = (1,0,0)) \tag{4.8}$$

$$Q_{axial} = Conv(:,:,k_{size} = (1,1,3), pad = (0,0,1)) \tag{4.9}$$

$$MV_f = HF + softmax(Q_{sagittal^T} \bigotimes K_{Coronal}) \bigotimes V_{axial} \tag{4.10}$$

$$PMVF = \tilde{F} + MV_f \tag{4.11}$$

where $\otimes$ represents matrix multiplication and MV_f represented Multiview feature maps. By adopting the proposed module using MV_f, $\tilde{F}$ depth-wise attention-based features used in the proposed model can enhance the capability and efficiency of the proposed solution. The proposed MMPA transformer block is shown in Figure 4.3.

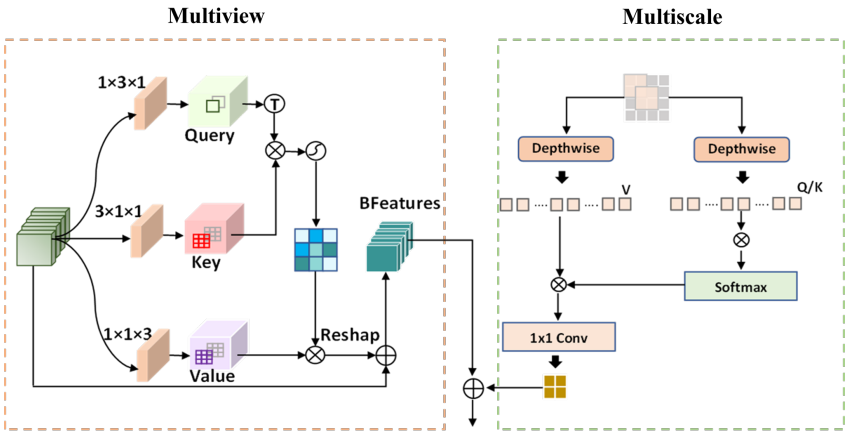


Figure 4.3: The efficient parallel Multiview Multiview multiscale and depth-wise features attention block.

A framework of the second stage proposed model is presented as an encoder, decoder, and a baseline module is shown in Figure 4.4. The 1x1 convolutional layer with softmax function has been used at the end of the model. The 3D MMPTtranNet from each block on the encoder side. In the encoder block, the spatial input size has been reduced with an increasing number of feature maps, and on the decoder side, the input image spatial size has been increased using a 3D Conv-Transpose layer. The input features' maps that are obtained from every encoder block are concatenated with every decoder block feature map to reconstruct the semantic information. The convolutional (3x3x3conv-BN-ReLu) layer

used the input feature maps extracted from every convolutional block in the encoder side and further passed these feature maps into the proposed residual module. The spatial size doubled at every encoder block and feature maps are halved at each decoder stage of the proposed model. The feature concatenation has been done at every encoder and decoder block except the last 1x1 convolutional layer. The three-level deep-supervision techniques are applied to get the aggregated loss between ground truth and prediction.

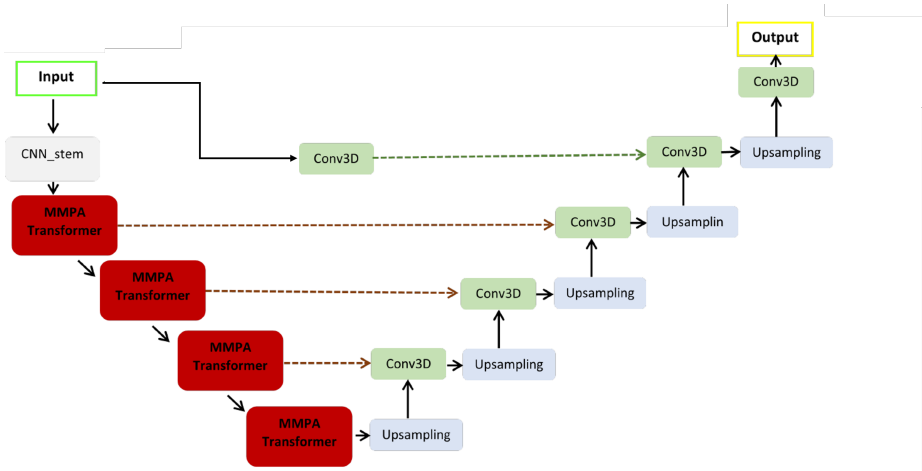


Figure 4.4: The proposed model used for Brain tumor segmentation. MMPATransformer module used at encoder and 3D conv with 3D upsampling layer used in decoder.

Training and optimization of the proposed solution: The learning rate of 0.0001 with the Adam optimizer has been used for training the proposed model. The cross-entropy and dice function is used as a loss function between the output of the model and the ground-truth sample. 2 batch-size with 1000 epochs has been used with 20 early stopping steps. The best model weights have been saved for prediction in the

validation phase. The 64x64x64 input image patch was used for training and a sliding window with stride 64 was used as prediction. The Pytorch library is used for model development, training, optimization, and testing. The V100 Tesla NVidia-GPU machine is used for training and testing the proposed model. The data augmentation methods mentioned in Table 1 are used for self-supervision and proposed model training and optimization. The dataset cases have different intensity ranges. The dataset is normalized between 0 and 1 using the max and min intensity normalization method. The detail of the training protocol is shown in Table 4.1

Table 4.1: Training protocols.

Data augmentation methods	RandCrop, RandGaussianNoise, RandGaussianSmooth, RandShiftIntensity, RandAdjustContrast, RandZoomd
Initialization of the network	"he" normal initialization
Batch size	2
Patch size	64x64x64
Total epochs	1000
Optimizer	Adam
Initial learning rate	0.0001
Stopping criteria, and optimal model selection criteria	The stopping criterion is reaching the maximum number of epochs (20).
Training time	16 hours

4.3 Experimental Results & Discussion

4.3.1 Results

This section of the manuscript covers the ablation study of the proposed MMPATransUNet (Multiview Multiscale Parallel Attention Transformer UNET) with its variants also its comparison with state-of-the-art methods. Moreover, to evaluate the generalization capability of the model

we have given the comparison of two other interdisciplinary medical imaging datasets with our proposed methods.

4.3.1.1 Results on Brain Tumor Segmentation (BraTS)

We have compared the performance of our proposed approach using the testing dataset of BraTS 2021 on 240 subjects with its variants and state-of-the-art deep neural networks. We have used an ablation study in this chapter and used a systematic approach to assess our method. Firstly, we have used the only Multiview approach with transformer UNET module and named it as MVATransNet. Secondly, we used a Multiscale attention module with a single encoder and named it as MMSATransNet and then we proposed our final Multiview Multiscale Parallel attention using the multi-encoder transformer Unet (MMPATransUNet) approach. Table 4.2 shows the ablation study scores in terms of dice and HD coefficients for all three tumor regions ET, TC, and WT. The proposed MMPATransUNet produced the best performance scores in comparison to the MVATransNet and MMSATransNet.

Table 4.2: Comparison of proposed brain tumor segmentation model with its variants on BraTS 2021 dataset.

Model	Dice-ET	Dice-TC	Dice-WT	H95-WT	H95-TC	H95-ET
MVATransNet	83.42	86.67	90.08	5.04	6.12	8.10
MMSATransNet	85.24	88.12	91.92	7.33	7.89	9.01
Proposed self-supervised MMPATransUNet	**86.08**	**88.98**	**92.95**	**4.88**	**5.46**	**8.88**

Table 4.3: Comparison of proposed and existing state-of-the-art models for Segmentation of Brain tumor on BraTS 2021 Data.

Model	Dice-ET	Dice-TC	Dice-WT	H95-WT	H95-TC	H95-ET
3D U-Net [6]	78.94	84.31	85.52	10.30	11.32	14.13
SwinUNet [5]	83.15	84.53	89.86	8.51	9.35	13.09
TransBTS [10]	80.35	85.35	89.25	7.83	8.21	15.12
UNETR [11]	80.77	84.61	86.06	9.37	10.29	12.83
nnFormer [12]	84.24	85.72	89.54	7.74	8.51	10.07
VT-UNet-S [13]	83.14	86.86	91.02	8.25	8.03	11.46
VT-UNet-B [13]	85.59	87.41	91.20	6.23	6.29	10.03
Proposed Model	86.08	88.98	92.95	4.88	5.46	8.88

The comparison of the proposed self-supervised contrastive learning approach MMPATransUnet with the state-of-the-art methods is given in Table 4.3. The proposed method has given the highest dice scores of 86.08, 88.98, and 92.95 for ET, TC, and WT tumor regions in comparison with the existing methods. Likewise, the lowest HD scores of 4.88, 5.46, and 8.88 for WT, TC, and ET labels are produced by the proposed brain tumor segmentation model in conjunction with the state-of-the-art approaches.

4.3.1.2 Results on Interdisciplinary Medical Imaging Datasets

We have tested our proposed segmentation solution on two different Medical Image Computing and Computer Assisted Interventions (MICCAI) challenge datasets to evaluate the generalization capability of our proposed solution. The proposed method has shown the above state-of-the-art performance in all MICCAI datasets. The results are shown in Table 4.4, and Table 4.5.

1. Results on HEad and neCK TumOR (HECKTOR) segmentation on PET/CT images 2021:

HECKTOR2021 CT/PET dataset of (254+71) H&N cancer patients collected from 6 different centers consisting of PET, CT images, and patient information including gender, age, sex, and acquisition center.

Table 4.4: The comparison of proposed MMPATransUNet with state-of-the-art methods for HECKTOR 2021 dataset.

Models	DSC	HD
Proposed self-supervised MMPATransUNet with Multiencoder contrastive learning	0.79	2.97
Base 3DUNet	0.70	3.88
(Andrearczyk et al., 2021)	0.65	4.07
(Bourigault et al., 2021)	0.75	3.27

In comparison to HECKTOR2020, 71 patients were added in HECKTOR2021 with an addition of 101 test cases and 224 training cases. The patient included in the dataset are histologically proven H&N cancer-diagnosed patients. The test set belongs to two different CHUP and ta CHUP centers. For consistency, expert radiation oncologists annotated gTVt (primary gross tumor volume) from multiple centers. The dataset is provided in NIFIT volume and binary contour of GTVt for segmentation of GTVt. PET and CT images are stored in SUVs and Hounsfield Units, respectively. Besides, a bounding box of testing and training sets is also provided in CSV. The evaluation is performed based on the DSC score computed within the bounding box at the original resolution of CT images. Further detail of the dataset can be found at Andrearczyk et al., 2022.

Table 4.4 shows the performance of the proposed method on the HECKTOR2022 dataset. The proposed MMPATransUNET has performed better in comparison with the state-of-the-art approaches. The highest mean dice score value of 0.76 and the lowest HD score of 3.03 have been achieved with a proposed solution.

2. Results on Myocardial Segmentation with Automated Infarct Quantification Challenge (MYOSAIQ) 2023:

Similar to the HECTOR, the proposed MMPATransUNET is tested on MYOSAIQ 2023 dataset to evaluate the efficacy of the proposed solution. This dataset holds the four cardiac classes including left ventricle (LV), Left ventricle Myocardium (MYO), Left ventricle Myocardium

Infarction (MI), and Microvascular obstruction (MVO). The full dataset is composed of 467 LGE exams from two different cohorts to quantify MI lesions at different phases of the longitudinal evolution of the disease: 4-8 days post-MI (MIMI cohort) and 1 month / 12 months post MI & reperfusion (HIBISCUS cohort). This dataset is divided into 3 subgroups. The first group consisted of 123 patients with LGE images of acute myocardial infarction up to 8 days post-MI (MIMI cohort). These patients were treated by two different procedures: i) with Percutaneous Coronary Intervention (PCI) and immediate stenting or ii) with the Minimalist Immediate Mechanical Intervention approach. The second group has 204 patients with LGE images obtained at 1 month post-PCI and reperfusion (HIBISCUS-STEMI cohort), and the third group contains 140 patients with LGE images obtained at 12 months post-PCI and reperfusion (HIBISCUS-STEMI cohort). The 374 training samples were used for training and 93 for testing the proposed model. The full dataset description is given as 'https://www.creatis.insa-lyon.fr/Challenge/myosaiq/databases.html'.

Table 4.5 shows the performance results of the proposed method on the MYOSAIQ dataset in terms of dice and HD scores for all given labels.

Table 4.5: The Average Dice & HD coefficients for MYOSAIQ 2023 dataset using proposed MMPATransUNet approach

Classes	**DICE**	**HD**
Left ventricle (LV)	0.953	6.021
Left ventricle Myocardium (MYO)	0.893	11.207
Left ventricle Myocardium Infarction (MI)	0.702	15.871
Microvascular obstruction (MVO)	0.801	13.994

Table 4.6: HD comparison of proposed MMPATransUNet approach with other teams' scores on leaderboard for LV, MYO, MI and MVO class labels

Leaderboard positions	Left ventricle (LV)	Left ventricle Myocardium (MYO)	Left ventricle Myocardium Infarction (MI)	Microvascular obstruction (MVO)
Team1	0.937	0.822	0.684	0.720
Team2	0.935	0.819	0.660	0.629
Team3	0.935	0.817	0.658	0.615
Team4	0.934	0.817	0.656	0.599
Our Results	0.953	0.893	0.702	0.801

Table 4.7: HD comparison of proposed MMPATransUNet approach with other teams' scores on leaderboard for LV, MYO, MI and MVO class labels

Leaderboard positions	Left ventricle (LV)	Left ventricle Myocardium (MYO)	Left ventricle Myocardium Infarction (MI)	Microvascular obstruction (MVO)
Team1	6.406	12.214	16.746	16.548
Team2	6.471	12.301	18.201	21.871
Team3	6.667	11.787	19.790	14.539
Team4	6.641	11.794	18.770	16.705
Our Results	**6.021**	**11.207**	**15.871**	**13.994**

The performance analysis of the proposed model and other team solutions is shown in Table 4.6 and Table 4.7. The challenge organizer announced ranking based on the method's novelty and achieved better performance. Our proposed model achieved comparable performance as compared to the top four teams. These are leaderboard scores using 93 test samples. For the Blood pool/ left ventricle, our proposed solution almost achieved better results as compared to other teams for all classes.

4.3.1.3 Qualitative Results based on proposed model

In this section, we have provided the results of the brain tumor segmentation 2021 dataset and recently participated MYOSAIQ challenge

to show the qualitative performance of the proposed solution.

1. 3D visualization of proposed and existing models for BraTs segmentation:

Figure 4.5 shows a 2D and 3D visualization of the input image with its ground-truth and prediction masks on proposed MMPATransUNet and other state-of-the-art methods for brain tumor segmentation. Here, the green color in the segmentation represents the ET class, the yellow color signifies ED, while the red color denotes the NCR tumor class. The 3D visualization of GT and the proposed MMPATransUNet are very close to each other. The enhancing tumor class is always comparatively bigger than the peritumoral edema and necrotic tumor class while at the time of diagnosis, these tumors show heterogeneous contrast enhancement, often with a necrotic center, and peritumoral edema. The proposed model effectively segmented all three classes also as presented in terms of dice and HD scores in Table 4.3.

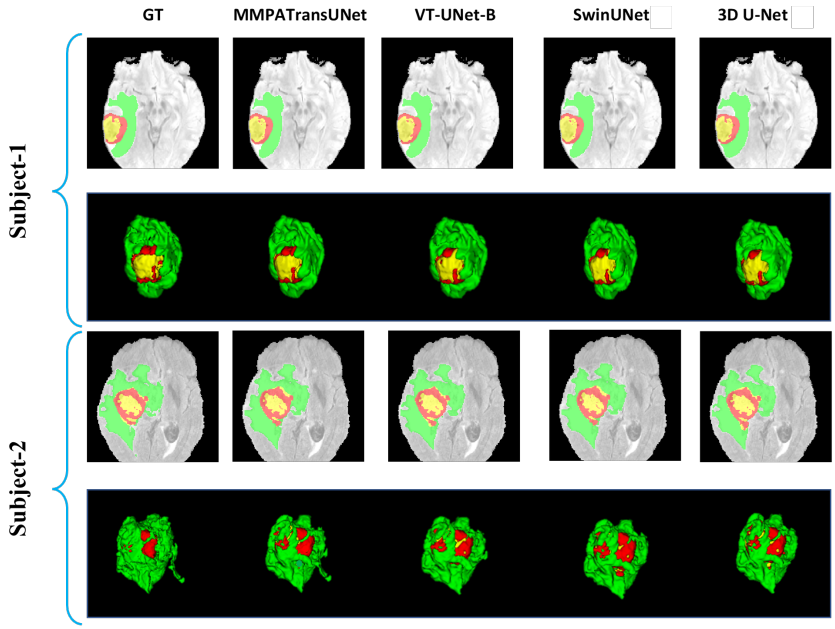


Figure 4.5: Segmentation results obtained by the proposed MMPATransUNet and existing state-of-the-art methods on the BraTS2021 dataset for two subjects. First row from left to right: (a) ground truth, (b) prediction MMPATransUNet, (c) prediction VT-UNet-B, (d) SwinUNet, and (e) 3D UNet overlayed on corresponding T1 images. The second row represents the 3D masks of the corresponding 2D views from the first row.

The performance analysis of the proposed model is compared with other state-of-the-art models given in Figure 4.6 and Figure 4.7. Our proposed solution achieved the highest dice as compared to existing methods as shown in Figure 4.6. Figure 4.7 shows the HD of proposed and existing state-of-the-art models.

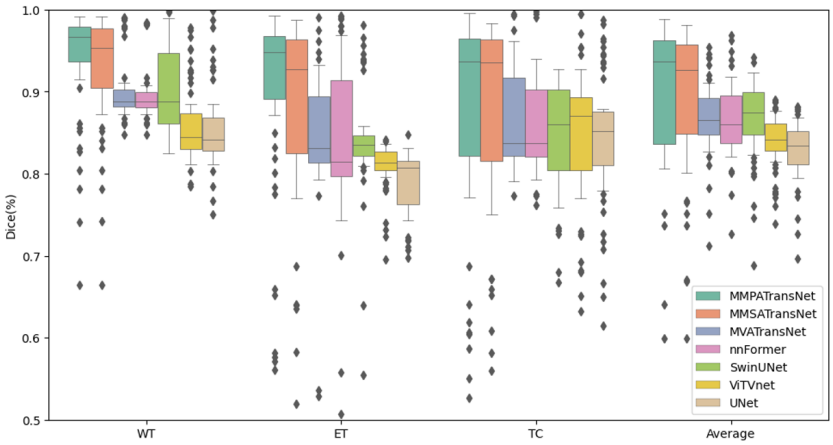


Figure 4.6: The Dice coefficient comparison of proposed MMPATransUNet with its variants and existing state-of-the-art methods on Brats 2021 dataset.

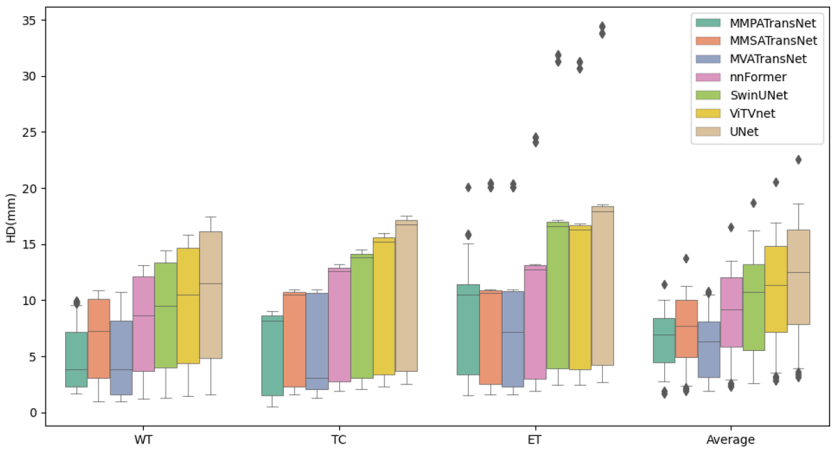


Figure 4.7: The HD comparison of proposed MMPATransUNet with its variants and existing state-of-the-art methods on the BraTS 2021 dataset.

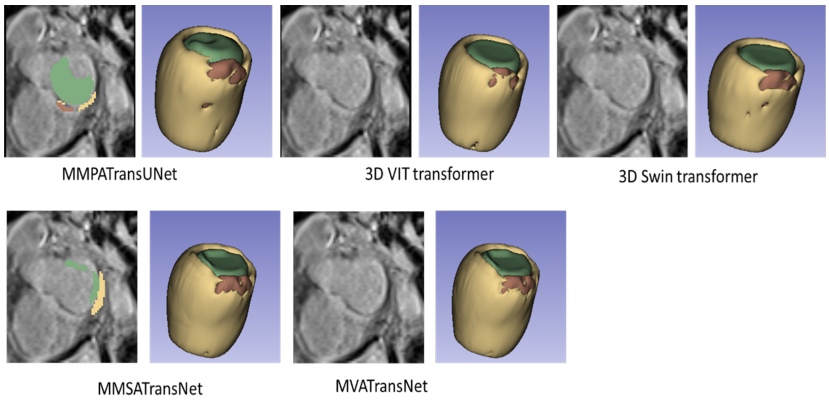
Furthermore, descriptive statics of Dice and Hausdorff are given for

the evaluation of the proposed model efficacy. Figure 2 and Figure 3 show the boxplots for Dice and HD comparison of the proposed model with its variants and state-of-the-art approaches of ET, TC, WT, and average tumor regions. The horizontal black lines inside the boxes show the mean values. The plots represent the upper and lower whiskers and the outliers. The whiskers are computed as 1.5 times the distance of the upper and lower limits of the box. The ($\diamond$) symbol represents the outliers, values outside the whiskers range. These plots are presented on the test set of 240 samples from the BraTS 2021 dataset. In Figure 4.7, the proposed MMPATransUNet has a higher mean value for all tumor regions and the average tumor label in comparison with other methods. Similarly, it can be clearly seen in Figure.. that the proposed model has the lowest mean values for all tumor regions as well as their average in compassion with other methods.

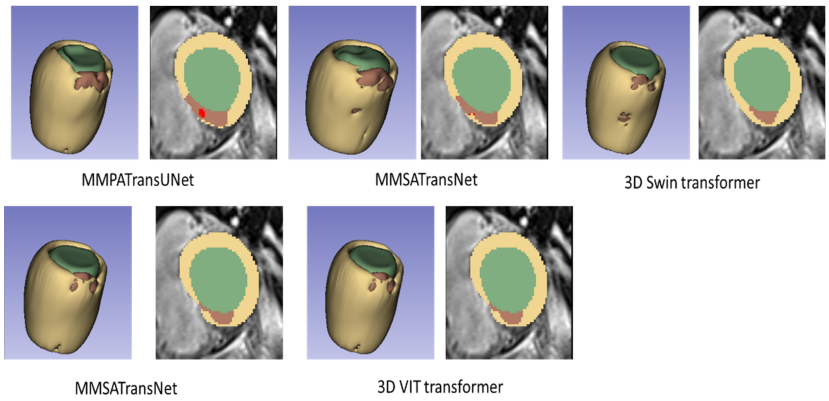
2. 3D visualization of the proposed model on MySEQ challenge:

The prediction visualization of the proposed model with the existing vision transformers is shown in Figure 4.8. The green color shows the LV class, the yellow label characterizes the MYO class, the brown color represents the MI label and the red color embodies the MVO class. The proposed MMPATransUNet achieved better performance and well-segmented the left ventricle, MYO, and infarction labels as compared to the other 3D Swin and Vision Transformers. Similarly, we can see that our proposed model efficiently segmented the MVO class where other models were not able to do so as shown 4.8 (b).

The performance of the proposed model on the interdisciplinary datasets shows that it is a well-generalized model. Therefore, it is a suitable versatile approach for medical image segmentation tasks with the capability of efficiently addressing small lesions issues.



(a) Segmentation of blood pool and MOV



(b) Segmentation of Myocardial Infarction

Figure 4.8: Comparison of proposed MMPATransUNet model with other vision transformer-based models for test subject 32.

4.3.2 Discussion

The proposed model has been validated on different challenging datasets such as heart, brain, head and neck cancer, and infant segmentation, and the proposed solution achieved state-of-the-art results

in a self-supervised setting. After systematically comparing the performance of Transformers and CNN-based models in a broad range of medical image segmentation tasks, we found that previously proposed Transformer models do not consistently outperform CNN models. The reasons behind this finding are multifaceted. Although Transformers have a greater capacity for modeling complex structures, ViT-like Transformers face computational complexity constraints that lead to excessive downsampling of the input, rendering them unsuitable for dense prediction tasks such as segmentation. Swin-based Transformers, lacking inductive bias, struggle to effectively train on limited medical image data.

Furthermore, we verified that large-scale pre-trained weights on natural images do not transfer well to medical images, especially in 3D settings. On the other hand, CNNs excel at capturing local textures and can learn with a small amount of data due to their inherent inductive bias. As a result, hybrid architectures like MedFormer successfully combine the strengths of both approaches, leveraging their advantages and demonstrating improved segmentation performance across various medical imaging scenarios. It is essential to carefully assess the individual contributions of the Transformer and CNN components in hybrid models. For example, although TransUNet introduces ViT to the ResNet backbone, adding numerous parameters, its performance is comparable to R50-UNet. Second, establishing solid core-model baselines is crucial for unbiased model evaluations. Reporting results for Transformers with auxiliary techniques (e.g., data augmentation, advanced optimizers, or model ensembles) does not faithfully assess the fairness of core architectures.

Our results, obtained through a consistent evaluation framework, indicate that the core-architecture performance of previous Transformer models does not consistently outperform CNN baselines. Third, evaluating task-agnostic datasets is vital for measuring the generalization ability of segmentation models. Current Transformer-related studies are

highly task-dependent, whereas CNN-based U-Net models have been verified across various tasks.

Our study highlights the importance of using multi-dataset evaluations with diverse anatomies for assessing generalization. Additionally, incorporating datasets from multi-center, multi-vendor, or different scanning protocols is crucial for a robust evaluation against medical domain shifts. We have designed models and investigated the utility of large-scale foundation models for medical imaging that can learn through a multi-view, multi-encoder, and multi-body training paradigm. An effective model should possess strong transfer learning and few-shot learning capabilities. We will explore human-in-the-loop learning, refining the model based on doctors' feedback on the model's prediction. This approach will help to improve the model's performance while ensuring that the model's prediction is consistent with medical professionals in real-world settings.

4.4 Conclusion

In this study, we presented a hybrid Transformer segmentation model that is scalable across data amounts, ranging from small-scale to large-scale data without pretraining. The proposed model showcases its potential for generalization by establishing new state-of-the-art baselines across seven widely used datasets and targeting anatomies (e.g., healthy organs, diseased tissue, and tumors). The strong performance of the proposed can be attributed to the desirable inductive bias introduced by the unified architecture and the efficient parallel Multiview attention and depth-wise based attention module, which learns global semantics information. Further, we designed two encoders in a self-supervised learning approach and fused the features from a pre-trained encoder and an encoder trained from scratch. These key designs enable the proposed model to capture boundary details, fuse global information

in a hierarchical manner, and exhibit robustness against data distribution shifts. Furthermore, we provide a comprehensive analysis for a fair comparison of different architectures across a wide range of medical image segmentation tasks.

Chapter 5

Survival Prediction of Brain Tumor Patients Using Attention-Guided 3D Deep Learning with Radiomics

Automatic survival prediction of gliomas from brain Magnetic Resonance Imaging (MRI) volumes is an essential step for a patient's prognosis analysis. Radiomics research delivers beneficial feature information from MRI imaging which is substantially required by clinicians and oncologists for predicting disease prognosis for precise surgical treatment and planning. In recent years, the success of deep learning has been vast in the field of medical imaging, it shows state-of-the-art performance in applications like segmentation, classification, regression, and detection. Therefore, in this chapter, we proposed a collective method using deep learning and radiomics techniques for the survival prediction of brain tumor patients. We first propose a hierarchical channel attention (HAM) module and a multi-scale-aware feature enhancement (MSAFE) to efficiently fuse adjacent hierarchical features in the proposed segmentation model. After segmentation, deep/latent features (LCNN) are extracted

from the bottom layer of the proposed segmentation model. Later, we extracted selected radiomics features (histogram, location, and shape) using input images and segmented masks from the proposed segmentation model. Further, the 3D deep learning regressor has been trained for 3D regressor-based deep feature extraction. We proposed the method of overall survival prediction for the brain tumor patients by combining all the meaningful features including clinical features (age) that also favorably contribute to the survival days prediction for the glioma patients. To predict the survival days for each patient, the selected features are trained to analyze the performance of various regression techniques like Random Forest (RF), Decision Tree (DT), and XGBoost.

Our proposed combined feature-based method achieved the highest performance for survival days prediction over the state-of-the-art methods. We also perform extensive experiments to show the effectiveness of each feature extraction method. The experimental results infer that deep learning-based features along with radiomic features and clinical features are truly vital paradigms to estimate survival days.

5.1 Introduction

Gliomas are the most common brain tumor disease developed from glial cells with the greatest mortality rate. There are 190,000 glioma cases annually occurring worldwide Castells et al., 2009. Approximately 12 months Furnari et al., 2007 is the average survival time of glioma patients and 24 months after surgical resection Louis et al., 2007, roughly 90% of patients unluckily died due to this disease. For automatic survival prediction and treatment planning, automatic delineation, early detection, and volume estimation are important tasks to detect gliomas. Due to the high variation of shape, appearance, and location of gliomas, it is a challenging task to localize and delineate the gliomas using conventional segmentation methods. Moreover, human experts need to closely monitor the manual segmentation annotation of tumor tissue

that is a tedious and time-consuming task. Consequently, there is always a need for automatic, accurate and faster methods for segmentation and survival rate prediction that could be helpful for the diagnosis and treatment of the gliomas.

Magnetic resonance imaging (MRI) is used as an effective non-invasive predictive tool for assessment and initial diagnosis of treatment response in neuro-oncology. Based on different research, MRI provides distinctive information that would be a better choice to predict survival autonomously based on pathologic and clinical data. The volume, shape, textural and intensity extracted from radiographic images is known as radiomics Chaddad et al., 2019.

Radiomics includes numerous important disciplines, incorporating computer vision for quantitative feature extraction, radiology for imaging interpretation, and machine learning for classifier assessment and regression Seow et al., 2018. Various radionics-based models have been presented for survival prediction Yao et al., 2016, and distant metastasis prediction Coroller et al., 2015. Shboul et al., 2019 proposed texture, area, volume, and Euler characteristics-based radiomics features from different intra-tumor parts using Extreme Gradient Boosting (XGBoost) regressor for survival prediction. They achieved 0.519 accuracies in survival prediction tasks in Brain Tumor Segmentation (BraTS) 2018 test data. Feng et al., 2020 proposed basic machine learning-based algorithms such as simple linear regressors using some volume, surface area, and directional gradient features for survival prediction. Their method was overfitted due to a small set of feature extraction and did not optimize well for survival prediction.

Chaddad et al., 2019 used a gradient boosting algorithm and random forest machine learning methods based on multi-scale texture features for survival prediction. Furthermore, they combined clinical factors with different seven radiomics features for survival prediction. Osman, 2019 presented 147 sets of radiomics image features using three tumor subregions and used the Least Absolute Shrinkage and Selection

Operator (LASSO) regression model for survival prediction. Further, they used the Cox model to select features to enhance the survival prediction performance. Sun et al., 2019 presented 4,524 radiomic features from the segmented area of the tumor and used a decision tree machine learning model for survival prediction. Baid, Shah, and Talbar, 2020 proposed a gray-level co-occurrence matrix, shape features, and first-order statistics features, they used Artificial Neural Network (ANN) model for survival prediction. Baid et al., 2019 introduced radiological features from MR images and used the LASSO model for survival days prediction. Weninger, Haarburger, and Merhof, 2019 proposed volume, distance, the center of the mass tumor, and age radiomics feature using linear regression for survival prediction. Bae et al., 2018 proposed radiomics analysis software to extract 43 unique quantitative features and Support Vector Machine (SVM) machine learning model used for survival prediction.

Recently deep learning models achieved state-of-the-art performance in medical image analysis for regression Mazher et al., 2022, classification Noreen et al., 2020, segmentation Payette et al., 2023,Lalande et al., 2022, and detection application Bano et al., 2022. The deep learning-based models produced an excellent performance to recognize objects and diagnose diseases from medical images Vente et al., 2023. The key advantage of deep learning models is that they extract the features automatically from images without manual intervention. However, prior approaches like radiomics always use manually engineered features like tumor size/shape. Though, one downside of deep learning is that it generally requires large, annotated datasets to work well.

Iqbal et al., 2019 proposed deep learning-based segmentation models to extract CNN features and combined handcrafted features such as texture, histogram, volume, area, and run length. They used a random forest regression model for survival prediction. Banerjee, Mitra, and Shankar, 2019 presented a deep learning-based method for survival prediction using BraTs 2019 dataset. They extracted semantic and agnostic

features using multi-planner spatial convolutional neural networks and trained multilayer perceptron (MLP) neural networks for survival prediction. Their approach did not achieve better performance in overall survival prediction task. Jungo et al., 2018 proposed a residual convolutional neural network segmentation model to extract volume, heterogeneity, rim width, and surface irregularity features from the segmented tumor. They used an ANN using segmented features for survival prediction. Islam et al., 2020 proposed a CNN model and extracted volume, heterogeneity, rim width, and surface irregularity from the segmented tumor. Further, they combined imaging and clinical features for survival prediction. They used various machine learning regressors such as SVM, ANN), Random Forest, and XGBoost for survival prediction.

Huang et al., 2021 proposed a non-Local module-based VNet model to segment brain tumors. Further, they have extracted CNN based deep features and imaging radiomics features. They combined clinical features with radiomics and deep features for survival prediction. They tested their method using BraTS 2020 dataset and achieved a 79% dice score for segmentation and 311.5 RMSE values in survival days prediction.

Pálsson et al., 2022 proposed imaging features and used a machine-learning model for survival prediction. They used an atlas-based segmentation model with an autoencoder to extract segmented features. They proposed an automatic method that is contrast-adaptive and robust to missing modalities, making the features generalizable across scanners and imaging protocols. They used a random forest machine learning model on segmented features for survival prediction. They added other clinical features with segmented features for survival days prediction. They tested their method on BRaTs 2020 dataset and achieved a 0.61 C score between predicted and ground-truth survival days.

Regardless of the different correlations between survival, genomic

expression and imaging features proven in the literature, no single analysis is significant enough to deploy in clinical practice. Therefore, various radiomic features depending on the nature of the disease are hard to connect in a meaningful way. Interpretation of features is also important and sometime clinicians may not trust the model prediction even though the model produced accurate results because the model does not give interpretable explanations of its prediction. Interpretable models may also reveal patterns in the data that provide effective new intuition into the disease and encourage future research.

The studies based on radiomics are not ideal to reproduce because of failing to generalize the solution across different scanners and implementation of software. Therefore, radiomics features use raw image characteristics that would sensitively depends on image acquisition environment and different scanning equipment. Other issues remain active though computer technology has been used in the field of brain tumor survival prediction and segmentation, there are still challenges exists to fully automate the solution in clinical practice. First, brain tumors vary from patient to patient in size, shape, and location. Second, in most cases, lesion of brain MRI is very small that creates the imbalance voxel between background and lesion area. The aforesaid challenges can create hurdles in survival prediction. Lastly, the other information such as genes, age not used during radiomic feature extraction that would restrict the ability to identify predictions.

Therefore, radiomics features are not solely enough for survival prediction tasks. CNN-based features are robust and automatically extracted from images without manual intervention. In contrast, the traditional methods used handcrafted features like radiomics, the data-driven based approach extract imaging features automatically like for complex segmentation tasks. It validated from literature that high-level features like shapes extract global semantic information and low-level features like contours extract details spatial structural information. Hence, the features from different neural networks layers are used to

combine effective features from previous cascaded layers intuitively.

Deep Learning based models can synergistically and comprehensively explore efficient features from high-level features to low-level perspectives. Combining the CNN-based deep features and meaningful radiomics features along with the clinical feature improves the accuracy of survival prediction.

In this article, we proposed an attention-based convolutional neural network (CNN). We propose a multi scale-aware feature enhancement (MSAFE) module that adjusts the receptive fields dynamically to produce efficient features effectively, so improving the feature representation capability of the network. Later, we extracted the deep/latent CNN features (LCNN) from the trained proposed CNN model along with the specifically selected radiomics feature (histogram, location, and shape) using input images and segmented masks. These features enhanced the performance of our proposed solution. A better segmentation mask produced better radiomics features that boost the performance of the proposed solution for survival days prediction.

Furthermore, we proposed our final method of overall survival prediction by combining all the meaningful features including clinical features (age) that contribute to the number of days of survival left for the patient. To predict the survival days for each patient, the selected features are trained to analyze the performance of various regression techniques. Details about how the features are selected and various experiments that are run for finding the best regression technique have been discussed in the next sections. To assess the generalization capability of our proposed solution is also given on Head Neck Tumor dataset (HECTOR2021).

5.2 Proposed tumor segmentation model

This section presents the dataset and architectural details of the proposed deep learning models and radiomics features for brain tumor segmentation and survival days prediction.

5.2.1 Dataset

The brain tumor segmentation(BraTS) has been published since 2012 at Medical Image Computing and Computer-Assisted Intervention (MICCAI) conference Menze et al., 2014. The overall survival prediction task is included from 2017 until 2020 BraTS challenges. In this article, BraTS 2020 dataset was used for the initial segmentation of brain tumor classes for later overall survival prediction of the patients. The total number of subjects in the training set of BraTS 2020 with their masks is 369 and dataset was taken from different institutions using different scanners and clinical protocols Menze et al., 2014.

This dataset is divided into two sets, one for training and one for testing the proposed model. A ratio of 80:20 is used for training and testing set division. The total training set holds 295 subjects, and the testing set holds 74 subjects. Each subject has nifti volumes for Flair, T1, T1CE, and T2 MRI modalities of size $240 \times 240 \times 155$. The manual annotation provided by one to four raters and validated by experienced neurologists having 15 years' experience. There are three classes of tumor given in the ground truth (GT) masks including enhancing tumor (ET) (labeled as class-4 in GT), peritumoral edema (ED) (labeled as class-2 in GT), and non-enhancing tumor/ narcotic tumor (NCR) (labeled as class-1 in GT). In our experiment, all four MRI modalities are stacked for brain tumor segmentation using the proposed model. Further, the resection status, age, and survival in days were also provided for overall survival prediction. Figure 5.1 shows the Flair, T2, T1, T1CE, and the corresponding segmentation GT of a subject from the BraTS dataset (2020).

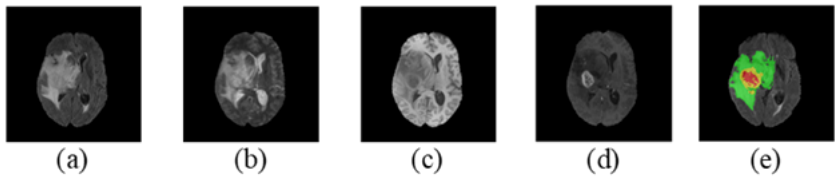


Figure 5.1: The four MRI modalities were used in this study. (a) shows the Flair image, (b) shows the T2 image, (c) shows the T1 image, (d) shows the T1CE image, and (e) shows the ground truth mask.

Here, the green color in the GT segmentation represents the ET class, the yellow color stands for ED, and the red color holds for the NCR tumor class. The segmentation labels in this dataset are enhancing tumor region (ET), tumor core (TC) and whole tumor (WT).

5.2.2 Proposed Method for Survival Prediction

The training and validation datasets use stacked images of four MRI modalities. A 3D deep learning segmentation model has been proposed to segment brain tumors and is further used in survival prediction. The deep features (LCNN) are extracted from the bottom layer of the encoder from trained segmentation masks. Furthermore, the input volume and predicted segmentation masks have been used to extract radiomics features. We also have used some feature selection techniques to select the best radiomics and LCNN for survival days prediction.

The variance-based feature selection technique is used to select the best features. Moreover, the features are normalized and fed into traditional machine learning-based classifiers such as random forest, gradient boosting, etc. Furthermore, the 3D deep learning-based CNN regressor (3DReg) model has been trained for deep feature selection. Thereafter, a feature selection is also applied to select the useful 3DReg features.

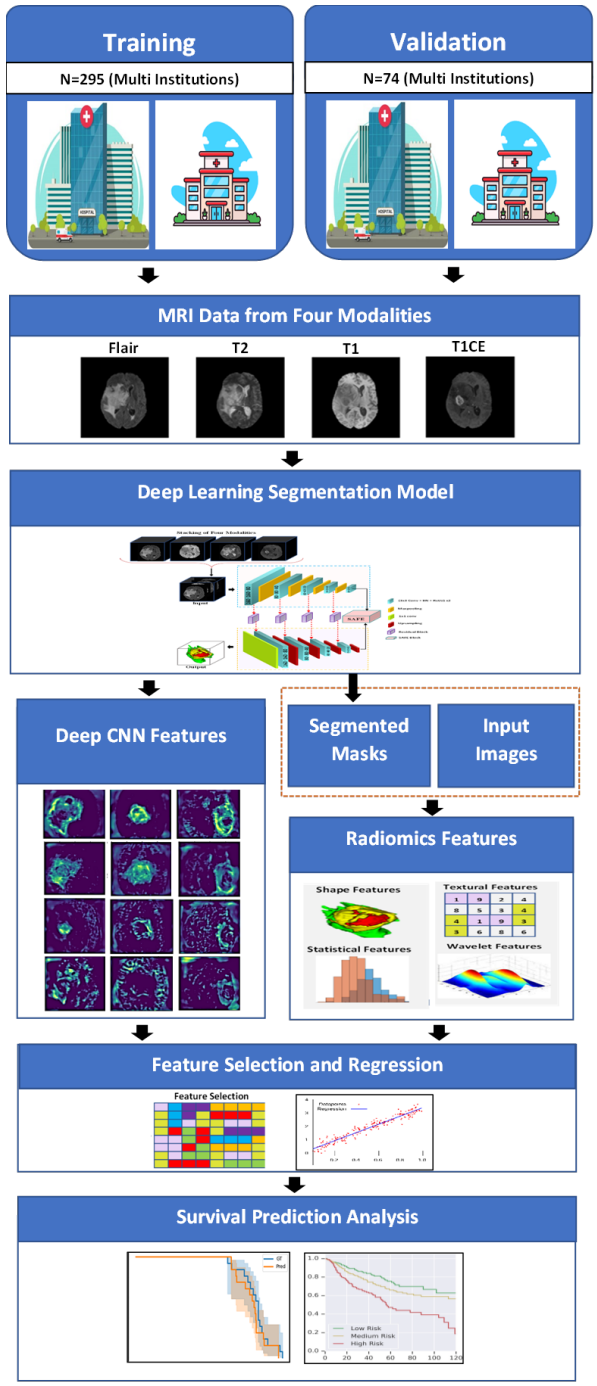


Figure 5.2: Overall proposed technique for survival days prediction based on deep and non-deep learning feature s extraction methods.

Different performance metrics have been used for survival days prediction. The overall schematic diagram of the proposed methodology is given in Figure 5.2. The detailed description of each module is explained in the following subsections.

5.2.3 The proposed Segmentation Framework

Figure 5.3 illustrates the proposed CNN model for brain tumor segmentation, which includes encoder and decoder networks.

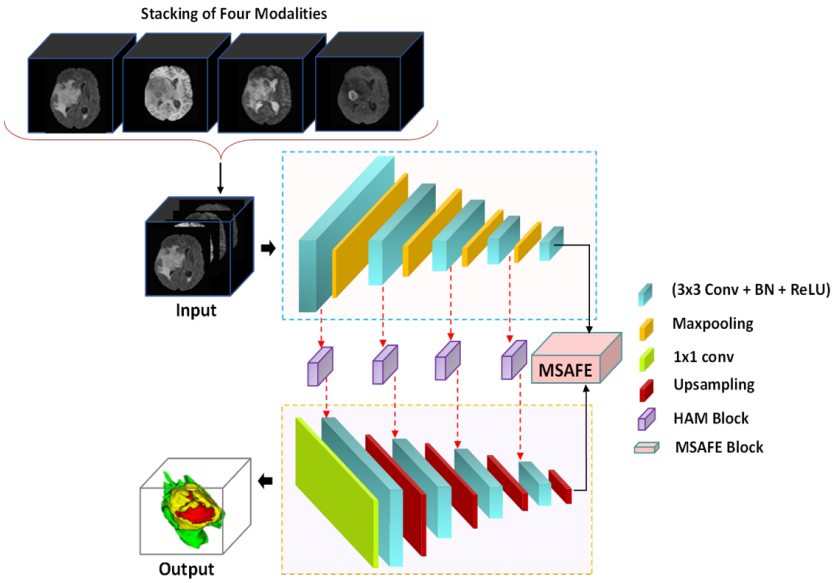


Figure 5.3: Diagram of the proposed Model with the proposed MSAFE module and the HAM block.

Four MRI modalities (T1, T1CE, T2, Flair) have been stacked as the

input, and predicted masks are produced for each tumor class (enhancing, non-enhancing/necrotic, peritumoral edema) highlighting the tumor regions. The convolutional block consists of 3D convolutional layers with Batch-Normalization and ReLU activation functions to extract the different feature maps from each block on the encoder side. The 3D max-pooling layer has been used to reduce the input image spatial size. In the encoder block, the spatial input size is reduced with increasing the number of layers, while on the decoder side, the input image spatial resolution is recovered via a 3D upsampling layer using a bilinear upsampling method. Each MRI modality has a 160x160x80 input size. The number of input channels is 4 (stacked four modalities of input dataset). The number of feature maps for each encoder block is 32, 64, 128, and 256. The kernel size of 3x3x3 is used for each Conv layer in the encoder and decoder block. The kernel size 2x2x2 for the 3DmaxPool layer is used to downsample the spatial resolution on the encoder side. The transpose3D convolutional layer with 2x2x2 kernel size with stride 2 is used to upsample each decoder size.

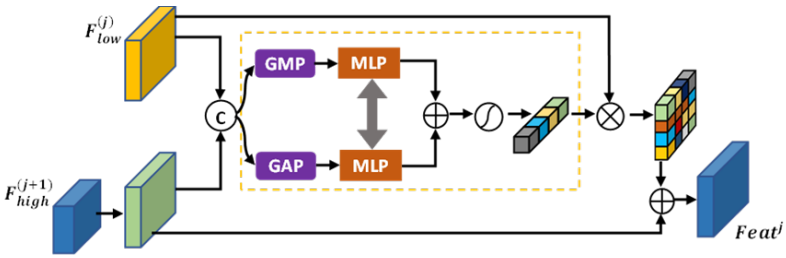


Figure 5.4: Schematic diagram of hierarchical attention module (HAM).

We introduced two special modules for our 3D CNN model. First, we have proposed a 3D hierarchical attention module (HAM) module for the proposed brain tumor segmentation model as shown in Figure 5.4. The HAM block has been used after each encoder block to concatenate with their corresponding decoder blocks. The 1x1 convolutional

layer with softmax function has been used at the end of the proposed model to generate the final output.

We add the reweighted low-level features to the high-level features to yield the result $Feat^j$ The mathematical representation of the proposed module is given below:

$$Feat_c^j = Concat(F_{low}^j, upsample(F_{high}^{j+1}) \tag{5.1}$$

$$Feat^j = sigmoid\left\{L_{mlp}\left(L_{gap}\left(Feat_c^j\right)\right) + L_{mlp}\left(L_{gmp}\left(Feat_c^j\right)\right)\right\} \bigotimes F_{low}^j + F_{high}^{j+1} \tag{5.2}$$

where Concat represents the concatenation operation, $Feat_{mlp}$ denoted as MLP operator, $Feat_{gap}$ denoted the global average pooling, $Feat_{gmp}$ represents the global max pooling, and $\bigotimes$ represents element-wise multiplication. The proposed module is used to progressively guide the fusion between high-level and low-level features that could help to suppress irrelevant background noise and preserve more semantic information.

The proposed HAM module is used to progressively guide the fusion between high-level and low-level features that could help to suppress irrelevant background noise and preserve more semantic information. High- and low-level features from the encoder and decoder have rich semantic and spatial information that would be suitable for accurate segmentation. We designed an attention module with weighting vectors and used a channel attention module to capture rich semantic information from low-level features and high spatial information from high-level features. The features are concatenated between low-level and up-samples of the high-level features. This HAM module would

be useful to exploit the useful features between high- and low-level features. Further, the combined features are passed to the channel attention with a weighting function to suppress the irrelevant background noise. The channel attention coefficient with sigmoid is used with multi-layer perceptron (MLP) and pooling layers to preserve relevant features as shown in Figure 5.4.

Second, we propose a multi-scale-aware feature enhancement (MSAFE) module. We embed the MSAFE module at the bottom of the framework which makes the model capable of extracting hidden multi-scale contextual information as well as able to aggregate multi-scale features efficiently. In a result, MSAFE) considers ample feature maps and can handle scale variations among different image instances. Specifically, the feature map produced by the bottom encoder module is equally divided into four parallel feature groups, named by fi, i ϵ 1, 2, 3, 4. The size of features in each group fi is the same as of input features, however, the number of channels is 1/4th of the original. To obtain a more abundant feature map consisting of various receptive fields, we applied dilated convolutions with different dilation rates to all four parallel groups. As a result, the proposed network can extract adequately precise features at different scales. The detailed structure configuration of the MSAFE module is shown in Figure 5.5.

$$fm_i = sigmoid(D_c^{ratei}(f_i)) \bigotimes D_c^{ratei}(f_i) \tag{5.3}$$

$$HF = Concat(fm_1, fm_2, fm_3, fm_4) \tag{5.4}$$

where, D_c^{ratei} represents the dilated convolution layer with dilated rate of $rate^i$ ($r_1 = 1, r_2 = 2, r_3 = 3$, and $r_4 = 4$)

Finally, these hierarchical features HF are concatenated and delivered to the adaptive feature aggregation component to further model the importance of each feature.

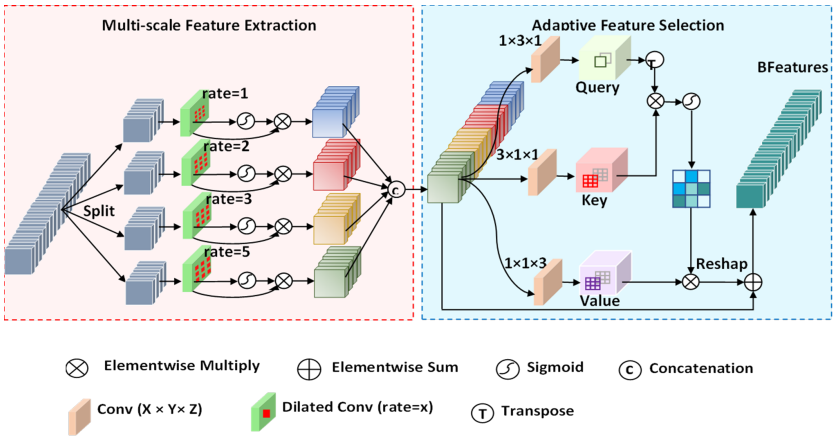


Figure 5.5: Multi-Scale aware feature enhancement (MSAFE) module for brain tumor segmentation.

The feature obtained from three view (sagittal, coronal, and axial) are passed to three Conv layers with BN and ReLU using different kernel sizes such as $3 \times 1 \times 1$, $1 \times 3 \times 1$, and $1 \times 1 \times 3$ in each convolutional layers to produced three feature maps. The activation maps from Q, K and V are passed to the softmax layer. The importance of each feature is performed and aggregate these multi-scale feature maps to get the output of the proposed module is shown in Eq. (5.5)

$$BFeatures = HF + softmax(Query^T \bigotimes Key) \bigotimes Value \tag{5.5}$$

where $\bigotimes$ represents matrix multiplication and BFeatures represented bottom feature maps. Adopting the proposed module at the bottom layer after the last layer of the encoder can enhance the capability of the proposed model.

5.2.4 Radiomics Feature and Imaging Feature selection for survival prediction

In this work, we have extracted different multiscale features from input MRI 3D volumes and segmented tumor masks. The radiomics features are categorized as volume features, intensity-based features, and Geometrical features.

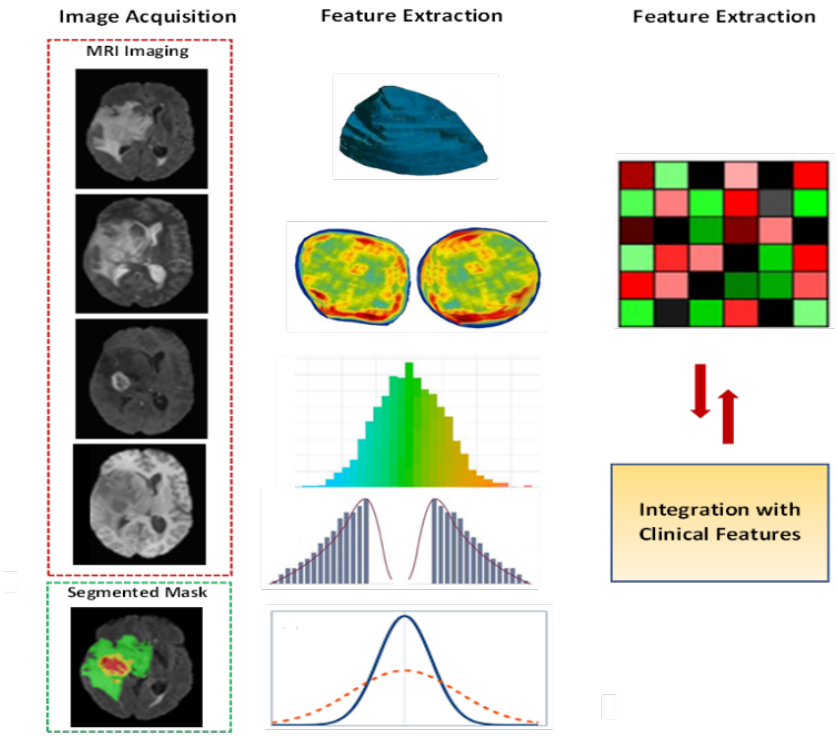


Figure 5.6: The radiomics features extraction from deep learning segmentation masks for survival rate prediction.

The intensity-based features are Kurtosis, Entropy, and Histogram. The Geometrical features are Length Coordinates, First axis, Second

axis, Third axis, Centroid coordinates, Eigenvalues, Equatorial eccentricity, and Meridional eccentricity. The further textural radiomics feature extracted from 3D MRI modalities input volume are (1) first-order statistics/statistical features (FOS/SF), gray level co-occurrence matrix (GLCM/SGLDM), gray level difference statistics (GLDS), neighborhood gray-tone difference matrix (NGTDM), statistical feature matrix (SFM), Laws texture energy measures (LTE/TEM), fractal dimension texture analysis (FDTA), gray level run length matrix (GLRLM), Fourier power spectrum (FPS), gray level size zone matrix (GLSZM), higher order spectra (HOS), and local binary pattern (LPB) Mazher et al., 2022. These radiomics features approach is shown in Figure 5.6.

5.2.5 Deep Features Extracted from the Bottom Layers of the Segmentation Model

Deep CNN features (LCNN) were extracted from the bottom layer of the trained encoder of the proposed segmentation model shown in Fig 3. The extracted feature vector has a 1x256 dimension extracted for each input 3D volume from the trained segmentation model. The performance of deep features is then analyzed using classical machine learning models.

5.2.6 3D Deep Regressor model

We have built a 3D regressor model based on the 3D convolutional, 3D Batch-Norm, and 3D ReLU layers. These three Layers (3DConv-BN-ReLU) formed one block. The proposed 3D-CNN network is a combination of these repeated layers followed by max pooling layers. After each convolution layer with the filter size of $3 \times 3 \times 3$, there is a max-pooling layer of size $2 \times 2 \times 2$. The 3D maxpooling layer is inserted after each block to downsample the spatial resolution of input feature maps. The first convolutional layer has a filter size of 16, and the number of filters

in every next block is doubled to gain a rich feature vector of images. The number of feature maps increased by 32, 64, 128, and 256, and spatial resolution decreased by 160, 80, 40, 20, and 15. One flattened layer and two fully connected layers have been used to get the final feature output. Like the LCNN feature vectors, a total of 1 x 256 size of feature vector has been extracted for each input 3D volume. The proposed model is shown in Figure 5.7. L1 loss function has been used to compute the loss between predicted and ground-truth survival days. All model layers have been optimized from scratch using PyTorch.

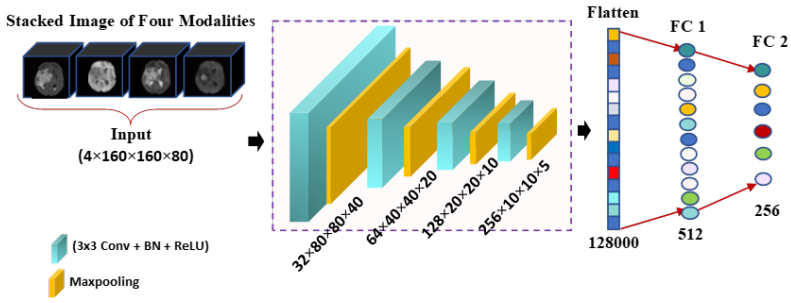


Figure 5.7: The proposed 3D deep learning regressor for survival days prediction using a brain tumor dataset.

5.2.7 Regression Models for Survival Days Prediction

The various regression models used for survival days prediction. From experimental evaluation, the four regression models like random forest (RF) Taghanaki et al., 2019b, regression trees (RT), linear regression (LR), and extreme gradient boosting (XGB) produced acceptable performance. RF is widely used in machine learning based classification, regression problems and it used bootstrapped dataset to build various number of trees in parallel. LR is another famous regression model that is used to find linear behavior between input and out variables.

RT and Extreme gradient boosting (XGB) are also famous classical machine learning models, and these models optimize the parameters during training and find the best parameters for better decision masking of the model.

All regressors have been implemented using the scikit-learn tool with default setting ($https : //scikitlearn.org/stable/supervised_learning.html$).

5.2.8 Loss Function

The Combo loss function proposed by Taghanaki et al., 2019b has been used for optimization and training the proposed segmentation models. The loss function is defined as:

$$L = \alpha \left(-\frac{1}{i=1} \sum_{N}^{N} \beta \left(t_i - lnp_i \right) + (1-\beta) \left[(1-t_i)\, ln\, (1-p_i) \right] \right) - (1-\alpha) \sum_{i=1}^{K} \left(\frac{2 \sum_{i=1}^{N} p_i t_i + S}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} t_i + S} \right) \tag{5.6}$$

where t_i represented target labels, p_i denoted as prediction mask, α , $\beta \epsilon [0, 1]$ and S are used as hyperparameters. All models are trained using Adam optimizer with a learning rate of 0.000116, $\rho = 0.95$, $\epsilon = 1 * 10^{-8}$, and $decay = 0$. Experimentally, we found that Dice and cross-entropy terms lead to better segmentation results. We tried different β values with all used brain tumor segmentation datasets, finding that $\beta = 0.5$ is the best value for our proposed model. For the survival prediction task, we have chosen root mean squared error (RMSE) as the evaluation indicator. The RMSE is shown in Eq. (5.7).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(X_{obs,i} - X_{pre,i}\right)^2}{n}} \tag{5.7}$$

Using the Adam optimizer, the proposed 3D segmentation model is optimized at the learning rate of 0.0001. The loss function between the prediction mask and the ground truth was created using the combo loss function. We have set epochs to 1000 and used a batch size of 2. The proposed model is trained using the early stopping criteria, and the training was ended after a maximum of 20 epochs on similar validation dice. The nifty data volume is read and written using SimpleITK. The data visualization is accomplished using the ITK-SNAP, Since the intensity ranges of the dataset cases vary, we used the max and min intensity normalization methods to normalize the dataset between 0 and 1. Each input image size is set as 80x160x160, and for each validation volume, the prediction mask is resampled to its original shape (224x224x155) using the nearest-neighbor interpolation approach. The prediction mask generated by the proposed model has been resampled to have the same size and spacing as the original image and copies all the meta-data, including origin, direction, and orientation.

The proposed models have been developed in the PyTorch library and trained from scratch. An NVIDIA GTX 3070 GPU having 12GM memory is used for the training and optimization of the proposed model.

5.3 Experimental Results and Discussion

This section covers in detail the results and their discussion on our proposed survival prediction techniques for brain tumor patients.

5.3.1 Results

We described various feature extraction techniques and utilized multiple regression models for survival days prediction using BraTS 2020 medical imaging dataset. Initially, we designed a segmentation model and later we used this model for deep/latent CNN (LCNN) feature extraction as well as the prediction masks of the model are used for the radiomics (RD) feature extraction method. From the trained segmentation model the LCNN features are extracted from the bottom layer of the model encoder block. Moreover, we also trained a separate 3D deep learning model called 3D regressor (3Dreg) using L1 loss for survival days prediction.

We have compared our proposed model with base UNet and ResNet-based UNet models. The base UNet is a simple encoder and decode-based model and RUNet is the ResUNet base model. The 3D HAM modules are inserted in each decoder side of the base UNet. The dice scores coefficients (DSCs) and Hausdorff distance (HD) of the proposed model as well as RUNet and UNet models when segmenting the three-brain tumor sub-regions (i.e., ET, WT, and TC) are shown in Table 5.1. The higher DSCs as well as the lower the HD in segmentation means elevated model performance.

Table 5.1 shows that the base UNet model achieved the dice score and HD scores of 0.83, 0.86, and 0.87 and 8.78, 7.90, and 3.88 for ET, TC, and WT which are less than the achieved scores among all models. Similarly, the RUNet in comparison achieved fewer dice and high HD scores to its enhanced version of the proposed model. Therefore, the proposed model benefiting from the proposed MSAFE module and additional HAM blocks reached the ultimate performance scores by generating the highest dice score and smallest Hausdorff distance (HD) scores. It reveals that our proposed segmentation model outperforms the baseline UNet model and the other two comparison models on brain tumor segmentation tasks. The higher segmentation performance scores are key

to better survival predictions as they produce good-quality prediction masks.

For performance analysis, various machine learning regressors have been trained for survival predictions as shown in Table 5.2, the performance is given in c-index. The random forest regressor produced a better performance as compared to other regressor models like Decision Tree (DTR), X-Gradient Boosting (XGBR), Gradient Bagging Regressor (GBR), extended Tree Regressor (ETR), Bagging Regressor (BR), etc.

Table 5.1: The performance comparison of proposed and existing deep learning models for brain tumor segmentation

Models	**DSCs**			**HD**		
	ET	**TC**	**WT**	**ET**	**TC**	**WT**
Proposed Model	0.883	0.891	0.902	4.18 ± 0.822	3.62 ± 0.689	2.89 ± 0.253
RUNet	0.868	0.872	0.886	6.33 ± 0.995	5.45 ± 0.887	4.35 ± 0.911
UNet	0.838	0.864	0.874	8.78 ± 0.989	7.90 ± 0.945	3.88 ± 0.673

Table 5.2: The performance comparison of different regression methods for survival days prediction.

Algorithm	**3D Regressor features**	**Clinical features**	**Latent 3D CNN features**	**Radiomics + Clinical features**	**Combined features**
RF	0.49	0.56	0.59	0.61	0.63
XGBR	0.54	0.56	0.6	0.57	0.62
DTR	0.51	0.55	0.57	0.56	0.61
GBR	0.53	0.53	0.53	0.55	0.6
ETR	0.55	0.51	0.58	0.53	0.62
BR	0.54	0.54	0.52	0.56	0.59

Table 5.3 displays the results of the random forest model for the proposed feature extraction techniques in comparison with state-of-the-art methods in terms of the c-index. It is seen that when the radiomics features are combined with the clinical feature (Age) it achieves a state-of-the-art performance score. The idea of combining the extracted features from all methods produced generous results and yields a higher performance score than state-of-the-art methods. It could be because combining many features into a single feature led to generalizability that is

Table 5.3: C-index-based performance measured for survival days prediction.

Algorithms	**C-index**
3D Deep Regressor (3DReg)	0.49
Clinical	0.56
Radiomics	0.57
3D Deep CNN Features (LCNN)	0.59
Radiomics + Clinical (RDCL)	0.61
Clinical + Radiomics + 3D Latent CNN + 3D Deep Regressor	**0.63**
Hd95 + Clinical + CoM + CEV (Pálsson et al., 2022)	0.61

derived via integrated inference skills shared across numerous feature sets.

Further, we estimated the RMSE scores for our proposed feature extraction methods using random forest regressor results and draw a performance comparison with existing state-of-the-art survival prediction methods as shown in Table 5.4. The proposed combined feature solution produced the lowest RMSE values as compared to state-of-the-art methods.

Table 5.4: Performance of proposed and existing methods for survival days prediction.

	Method	**RMSE**	**MAE**
Existing method \hl{[28]}	CNN+ Radiology+ Clinical	316.31	240.05
	Radiology+ Clinical	392.51	284.54
	CNN (RF)	378.85	284.13
	CNN(DL)	358.92	269.37
Proposed method	3D Deep Regressor (3DReg)	400.11	300.10
	Clinical	390.10	290.70
	3D Deep CNN (LCNN)	377.89	255.66
	Radiomics	321.23	247.56
	Combined + Clinical	310.11	243.99

5.3.2 Performance Analysis of the Proposed Model

The Kaplan–Meier curves based on ground truth and predicted survival days using proposed feature extraction techniques are shown in Figure 5.8. We have used a random forest regressor to predict survival days using clinical, radiomics, and latent CNN features from the proposed segmentation model and deep CNN features from the trained 3D regressor. The curves show survival days prediction based on combined extracted features with random forest algorithms produced more accurately as compared to individual feature techniques for survival days prediction.

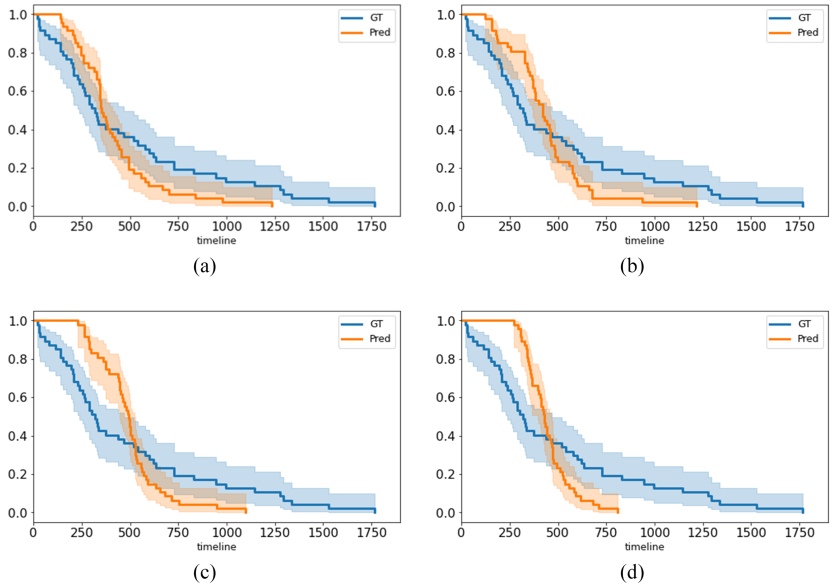


Figure 5.8: Kaplan–Meier plots of the (a) combined features radiomics, (b) radiomics features using RF classifier, (c) 3D deep extracted from segmentation model features with RF, (d) 3D deep learning regressor.

Figure 5.8. shows the survival days prediction curve with ground truth survival days using combined features (clinical, radiomics, deep

features). The random forest machine learning model based on radiomics features shows better performance after the combined feature approach is shown in Figure 5.8 (b). Figure 5.8 (c) shows results based on latent deep 3D CNN features using random forest. The 3D Regressor model using random forest is shown in Figure 5.8 (d).

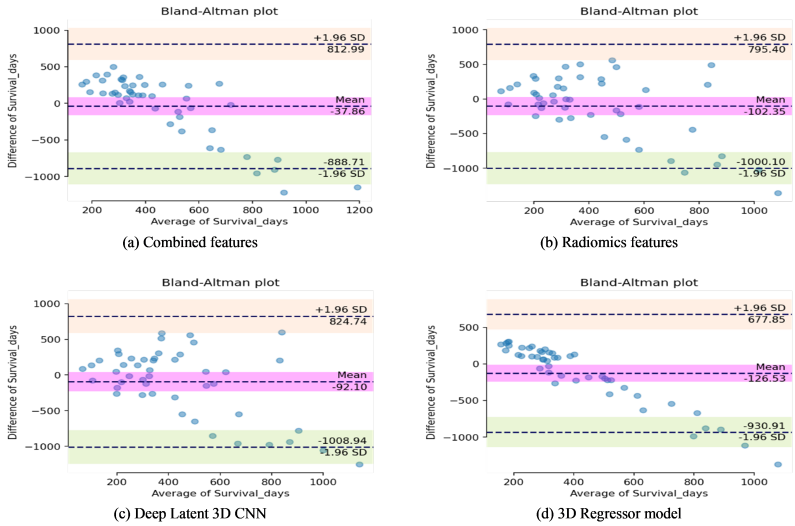


Figure 5.9: Bland Altman plot between predicted survival days and ground-truth survival days. (a) Combined (b) Radiomics + Clinical (c) Latent 3D CNN (d) 3D Regressor.

The combined feature approach produced better performance and predicted curves are very close to the target curves in terms of survival days as compared to individual based features. Therefore, the curves produced by proposed models validated the effectiveness of the survival days prediction method based on combined deep learning-based and radiomics features along with the clinical feature.

To investigate agreement between algorithm-generated and manually determined survival days, we used a Bland–Altman plot, which

graphs the mean difference of measured survival versus manual survival days and constructs limits of agreement. Bland Altman represents the distribution of survival days output for clinical, radiomics, deep learning features, and deep learning 3D regressor features. Bland Altman's plot shows the agreement between predicted and ground truth survival days. Figure 5.9 (a) based on the combined feature extraction approach represented the most agreement of predicted and ground truth survival days as compared to other bland Altman plots. Similarly, the other plots are based on radiomics, and deep features are shown in Figure 5.9 (b, c, d).

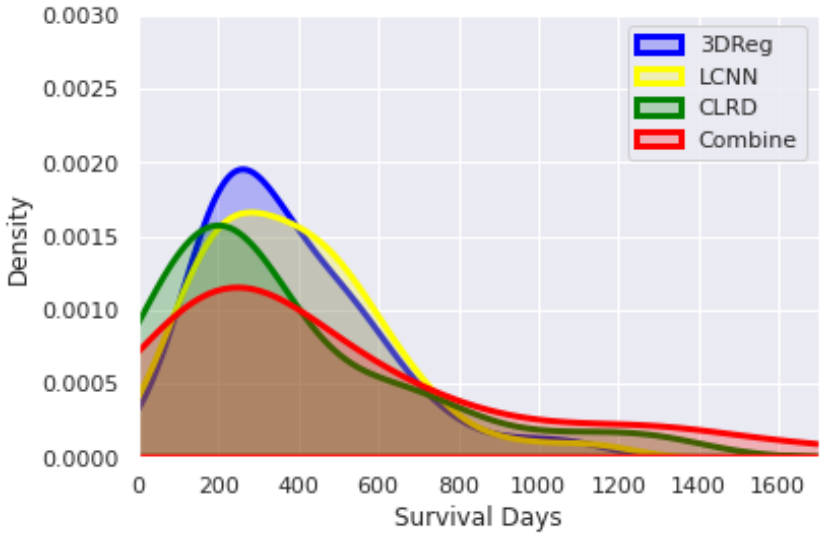
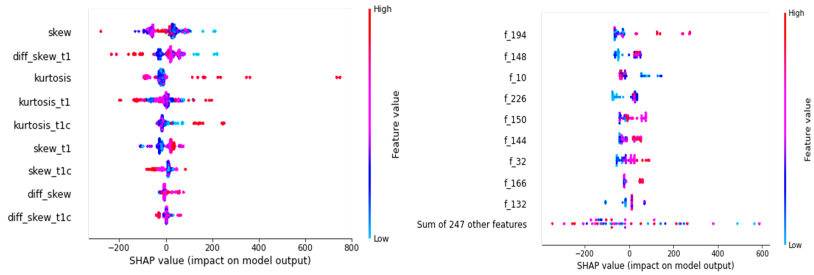


Figure 5.10: Density plots of the survival days prediction for the extracted features from 3D Regressor (3DReg), latent 3D CNN (LCNN), Radiomics + Clinical (CLRD), and their combined features (Combine).

Figure 5.10 shows the density plot for all survival prediction methods. Firstly, the distribution of the predicted survival days using combined features from all extracted feature techniques is significantly different from their separate outcomes. Similarly, the distribution score of predicted survival days with the combined feature approach is greater than their separate execution methods. However, prediction distributions of the 3D regressor distributions and latent 3D CNN techniques have larger standard deviations and more shifts towards left as compared to the distribution of the Radiomics + Clinical feature approach.

The proposed combined feature solution is always shifted toward the higher values on the right as shown in Fig 11. The SHAP explainability method (SHapley Additive exPlanations) Juanco-Müller et al., 2021 has been used to measure the importance of deep CNN and radiomics features. SHAP is used game theoretic approach and now widely used to explain the machine learning model feature importance Starke et al., 2020. The red color shows the high features and blue color presented the low feature values.

Figure 5.11 (a, b) shows the radiomics and deep features importance. It was shown that the skew, skew difference, and kurtosis achieved higher feature importance than the other features. We can say that the skew and kurtosis achieved higher feature importance as compared to other radiomics features. Out of 256 extracted features from the deep CNN method which produced the highest feature importance as compared to others are shown in Figure 5.11 (b).



(a) Radiomics Features (b) Deep Latent 3D CNN

Figure 5.11: Radiomics feature importance for survival prediction. The red color presented high and blue color presented low features in prediction of model (a) radiomics feature generation, (b) Deep Latent 3D CNN.

5.3.3 Generalization Capability Analysis

To see the generalization capability of our proposed solution for survival days prediction we used the Hecktor2021 dataset.

I. Dataset: The head and neck tumor dataset proposed in MICCAI2021 has been used to validate our proposed solution. It consisted of 224 total training patients and 101 total testing acquired from 5 different centers. The primary gross tumor volume (GTVt) along with survival days was annotated by experts. The dataset was provided in NIfTI format with binary segmentation masks. The bounding boxes at the original CT resolution were also provided to compute DSC scores for evaluation. A detailed description of the dataset can be found Bourigault et al., 2021.

II. Results: To estimate the number of survival days, we considered four previously mentioned feature extraction methods for the internal validation dataset: including clinical, CT/PET with radiomics, 3D deep segmentation-based features, and 3D regressor characteristics.

Results showed that the combined feature set (radiomics, clinical, and deep) gave the best performance when compared to the method using separate features. The combined features extraction strategy yields the best results for the RF model, as shown in Table 5.5 in terms of the c-index.

Table 5.5: The various regressor models for survival prediction in terms of C-index.

Algorithms	**C-index**
Clinical features	0.692
Deep segmentation features (LCNN)	0.788
Radiomics Features (RD)	0.757
Deep 3D regressor features (3DReg)	0.794
Combined (Clinical + RD + LCNN + 3DReg) features	**0.845**

We compare the performance of our given method in terms of c-index with existing methods for survival days prediction using the HECTOR dataset as shown in Table 5.6. It is observed that our proposed combined feature approach outperformed the other state-of-the-art methods by yielding a c-index score of 0.84.

Table 5.6: Performance comparison with state-of-the-art methods Algorithms C-index.

Algorithms	**C-index**
Proposed Method	0.84
(Bourigault et al., 2021)	0.82
(Juanco-Müller et al., 2021)	0.59
(Starke et al., 2020)	0.47
(Aerts et al., 2014)	0.69

The diagnosis of squamous cell carcinoma requires the discovery of malignancy and the evaluation of prognostic outcomes. Effective detection may aid in better decision-making. Even if the segmentation and quantitative analysis of head and neck cancer is a notably complex task,

the performance of the proposed framework is sufficient to be considered for automated diagnosis and survival rate prediction. Our combined radiomics and deep learning extracting features from the trained segmentation and regressor model for survival days prediction, which performs noticeably better than state-of-the-art techniques in terms of all assessment metrics for the tested dataset.

5.4 Discussion

In this chapter, we have proposed radiomics and deep learning-based methods for survival days prediction for brain tumor survival days and further validate our method on the Hecktor2021 dataset to compare and validate the performance of our proposed model.

We have trained a 3D segmentation model with different proposed modules to first segment the tumor regions and generate the predicted masks. The proposed model yielded the best scores for DSCs and HD as the proposed MSAFE module in the segmentation model achieved better feature maps based on parallel dilated convolutional layers with different dilation rate by focusing different receptive fields. The proposed block extracts feature information dynamically based on different spatial scales and covered situation where brain tumor widely varying the size and shape. Later, the input 3D volumes and segmented masks are passed for the radiomics feature extraction and selection. Also, we extracted the latent CNN features from the last encoder block of the trained segmentation model. Furthermore, we trained a 3D deep learning-based regressor model and extracted deep features from the last layer of the trained regressor model to predict survival days. We also used clinical features such as age and recission time provided in BraTS 2020 training data to train the random forest and other machine learning regressors.

Interestingly, adding a clinical training set did not significantly alter the results, which implies that the gap may be explained by clinical characteristics that are more pertinent when combined with different feature concatenation-based models. For the analysis of complicated tasks, feature ensembling or features concatenation approaches are well recognized to be quite effective. Contrary to most of the previous research on medical image diagnosis, in which clinical features play a significant role in prediction and diagnosis, we have discovered the fascinating fact that clinical features do not much contribute to performance improvement, indicating that most of the data is based on images. The geometric feature based on intensity, geometry and location contributed more to predict the survival days prediction. This study validated that the smaller the shape of the tumor produced the higher overall survival and the immediacy of the tumor to the center of the brain achieved lesser performance for overall survival. The radiomics using images and masks produced higher explainable features for survival days prediction. Kurtosis and skewness achieved overall better performance and played a better role for predicting the overall survival days showes in the explainable feature plots. The performance of latent CNN deep learning-based features is almost comparable to the radiomics survival days prediction by observing the c-index scores.

The overall performance of the 3D deep learning regressor is not good as compared to radiomics and LCNN feature results. A deep learning-based regression model needs more input data for training, and we have used only 295 input samples. Therefore, it is observed we need more data to fully obtain the robustness in the results based on a 3D CNN-based regressor. Radiomics-based features may have better interpretable advantages and generally more robust results can be obtained as compared to deep learning-based features.

However, our main objective is to propose imaging features that can

be replicated based on various MR preprocessing, contrasts, and scanning equipment. To attain efficient and robust solution using MR modalities based on variations in scanners or acquisition protocols deep LCNN features are obtained with a method that was designed to have these properties. Automatically computing detailed LCNN features make it easy to analyze the relative properties of tumor and their classes for survival prediction, as we have done in our experiments.

Our combined based feature method produced better performance and accurately measured the survival days on two different medical imaging datasets. It could be because combining many features into a single feature led to generalizability that is derived via integrated inference skills shared across numerous feature sets.

As demonstrated in our experiments, the features proposed in our chapter readily generalize across different datasets like BratS 2020 and Hecktor 2021. The proposed approach produced an excellent performance in Hecktor 2021 dataset within the same feature set. However, BraTS, it didn't outperform that well. Though, one possible reason could be the BraTS20 dataset consisted of anaplastic astrocytoma's, glioblastoma or both that have various survival characteristics. The circumstance that BraTS20 has pre-operative that may play an essential role as well as the influences of surgery cannot be considered.

The segmentations model caters to the information from directly multi-contrast MRI scans using f tumor region and the surrounding brain. In this work, the focus is to extract radiomics features from the tumor region itself and rest of the brain. We analyzed that the skewness, and kurtosis of deformation field amplitudes around the tumor together with conventional clinical and deep learning-based auto-generated features significantly improves survival models. Future work may therefore explore transformer-based deep learning methods along with radiomic features to further improve model accuracy. Since the transformers adopt the mechanism of self-attention, differentially weighing the significance of each part of the input data. Conclusion In this chapter, we

have proposed a hybrid deep learning and radiomics-based method for survival days prediction of brain tumor patients. Further, the proposed hybrid solution is validated on the Hecktor2021 dataset for head and neck tumor segmentation to compare and substantiate the performance of our proposed approach. We trained the proposed 3D segmentation model with two distinct proposed modules (HAM, MSAFE). Later, we extracted radiomics and deep features from the trained segmentation deep learning model. Further, we trained a 3D deep learning-based regressor model and extracted the deep features from the last layer of the trained regressor model to predict survival days of the brain tumor patients. A variety of classical machine learning models have been trained using several sets of extracted imaging features while achieving the better performance with the combined imaging features with non-imaging (Age) features. We have also presented the explainability of the given set of features. It has been observed that combined radiomics and deep features can also have strong interpretability and clinical applicability for survival days prediction in brain tumors. In the future, transformer and generative learning-based deep learning models will be explored to further improve the performance of the proposed approach for survival days prediction in brain tumor prognosis.

5.5 Conclusion

In this Chapter, we have proposed a hybrid deep learning and radiomics-based method for survival days prediction of brain tumor patients. Further, the proposed hybrid solution is validated on the Hecktor2021 dataset for head and neck tumor segmentation to compare and substantiate the performance of our proposed approach. A variety of classical machine learning models have been trained using several sets of extracted imaging features while achieving better performance with the combined imaging features with non-imaging (Age) features. We have also presented the explain ability of the given set of features. It has

been observed that combined radiomics and deep features can also have strong interpretability and clinical applicability for survival days prediction in brain tumors. In the future, transformer and generative learning-based deep learning models will be explored to further improve the performance of the proposed approach for survival days prediction in brain tumor prognosis.

Part IV

Validation of Proposed Methods

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Chapter 6

Validation of the proposed solutions on real-time medical imaging challenges

6.1 Introduction

In this chapter, we will explain the validation or assessment of our proposed solution for this dissertation in various MICCAI (Medical Image Computing and Computer Assisted Intervention) challenges. We have tested our proposed solution on real-time MICCAI datasets and compared the performance of our proposed solution with other international teams in a real-time environment.

Furthermore, most of the MICCAI challenges require solutions in docker containers for deployment in remote systems to test the models on hidden test sets. We have prepared docker containers for most of the challenges to be deployed in real time. MICCAI conference provided a dataset for various challenges every year and top world centers or deep learning engineers and researchers participated in different challenges and provided their solutions to compare the performance of their proposed solutions. These challenges are posted on the respective

websites and on MICCAI's main website. We have participated in MICCAI 2021, 2022, and 2023. The detailed description and ranking of each challenge that we have obtained are explained in this chapter. We have presented one of our MICCAI challenge CuRIOUS2022 in this chapter that has been published as a book chapter in MICCAI proceedings.

In the CuRIOUS2022 challenge, we have proposed a self-supervised contrastive learning two-stage model for the Intraoperative ultrasound (iUS) task. In the first stage, we trained the encoder of our proposed 3DResUNet model using self-supervised contrastive learning. The unlabeled dataset was used in self-supervision to train the encoder of the proposed 3DResUNet model and utilized this encoder as a pre-trained weight. In the second stage, the pre-trained weighted-based 3DResUNet proposed model was used to train on the labeled training dataset for iUS segmentation.

We achieved 1st position in CuRIOUS2022 and this solution was further extended using 3D MultiEncoder shared 3D transformer-CNN models in a self-supervised contrastive learning framework for this Ph.D. work on brain tumor segmentation task. We have explained a detailed analysis of our proposed approach in Chapter 4.

6.1.1 Position ranking of challenges

The position achieved in participating in MICCAI competitions is shown in Figure 6.1 and the position ranking for ISBI competition is shown in Figure 6.2. We have divided the ranking into three portions. The red color represents the ranking in the first 5 positions, mustard shows the top 6 to 10 positions, and the diagonal green color represents more than 10 rankings that we have achieved.

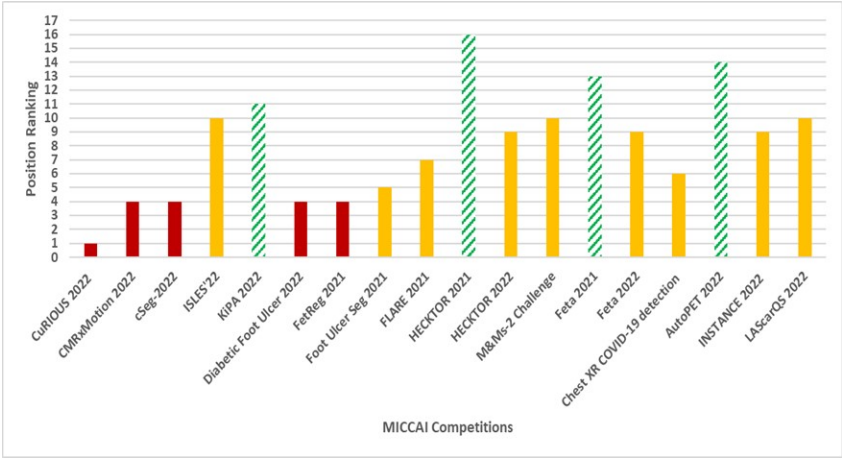


Figure 6.1: Proposed Deep Learning models validated on various MICCAI Competitions.

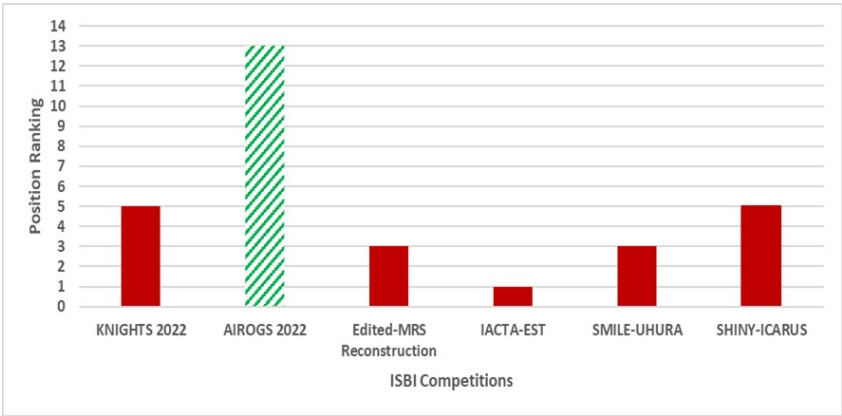


Figure 6.2: Proposed Deep Learning models validated on various ISBI (2022-2023)Competitions.

6.2 Challenges

6.2.1 Segmentation of Intra-operative Ultrasound using Self-Supervised Learning based 3D-ResUnet Model with Deep Supervision

The work presented in this section is published in Qayyum et al., 2023. Intra-operative ultrasound (iUS) is a robust and relatively inexpensive technique to track intra-operative tissue shift and surgical tools. Automatic algorithms for brain tissue segmentation in iUS, especially brain tumors and resection cavity can greatly facilitate the robustness and accuracy of brain shift correction through image registration, and allow easy interpretation of the iUS. This has the potential to improve surgical outcomes and patient survival rates. In this paper, we have proposed a self-supervised two-stage model for the Intraoperative ultrasound (iUS) task. In the first stage, we trained the encoder of our proposed 3DResUNet model using self-supervised contrastive learning. The self-supervised learning offers the promise of utilizing unlabeled data. The training samples are used in self-supervision to train the encoder of the proposed 3DResUNet model and utilized this encoder as a pre-trained weight for the Intra-operative ultrasound (iUS) segmentation. In the second stage, the pre-trained weighted-based 3DResUNet proposed model was used to train on the training dataset for iUS segmentation. Experiment on the CuRIOUS -22 challenge showed that our proposed solution showed significantly better performance before, during, and after Intra-operative ultrasound (iUS) segmentation. The code is publicly available.

6.2.1.1 Introduction

Intra-operative ultrasound is a high-energy sound wave surgical procedure that is bounced off internal tissues and organs. It is a low-cost dynamic imaging modality that provides provides interactive and timely

information during surgery which helps the surgeon to find tumors or other problems during the procedure. As the transducer is in direct contact with the organ being examined, hence, it can guide the surface incisions for deep lesion resection, limit the extent of surgical resection, accurately guide intraoperative biopsies, localize the pathology and we can obtain high-resolution images which are not degraded by air, bone, or overlying soft tissues.

For most cancers, the survival at one and five years is much higher if it is detected early (stage 1) than at a later stage. When a cancer diagnosis is delayed or inaccessible, the survival chances decrease significantly, and it may have greater problems associated with treatment and much higher costs of care i.e., 90% of patients have a 10+ year survival rate after being diagnosed in early stages in comparison to 5% for those who are diagnosed at a later stage (stage 4).

Automatic algorithms for brain tissue segmentation in iUS, especially brain tumors and resection cavities can greatly facilitate the robustness and accuracy of brain shift correction through image registration and allow easy interpretation of the iUS. The resection quality and safety are often affected by the intra-operative brain tissue shift due to several factors i.e. intracranial pressure change, drug administration, gravity, and tissue removal. Such shift in tissue may results in the displacement of the target and vital structures during the surgical procedure while the displacements may not be directly visible to the surgeon, however, it renders the surgical plan invalid. Live ultrasound overlaid onto preoperative data which allows for assessment and visualization of brain shift. The images from intra-operative ultrasound contain biological information possibly correlated to the tumor's behavior, aggressiveness, and oncological outcomes. Deep Learning has been widely used in overall medical image segmentation tasks Payette et al., 2023, Ma et al., 2022.

To tackle the discrepancies of iUS at different surgical stages, in

this work, we present an efficient segmentation of intra-operative ultrasound using self-supervised learning-based 3D-ResUnet with deep supervision. We trained the encoder of our proposed 3DResUNet model using the self-supervised contrastive learning method. The self-supervised learning offers the promise of utilizing unlabeled data. The training sample is used in self-supervision to train the encoder of the proposed 3DResUNet model and utilized this encoder as a pre-trained weight for the Intra-operative ultrasound (iUS) segmentation task. In the second stage, the pre-trained weighted-based 3DResUNet proposed model was used to train on the training dataset for iUS segmentation. Our proposed solution produced optimal performance on the validation dataset for three tasks (before, during, and after Intra-operative ultrasound (iUS) segmentation).

6.2.1.2 Methods

In self-supervised learning sitting, First, it uses augmentation to mutate the data, and second, it utilizes regularized contrastive loss Chen et al., 2020 to learn feature representations of the unlabelled data. The multiple augmentations are applied on a randomly selected 3D foreground patch from a 3D volume. Two augmented views of the same 3D patch are generated for the contrastive loss as it functions by drawing the two augmented views closer to each other if the views are generated from the same patch, if not then it tries to maximize the disagreement. We have to use masked volume inpainting, contrastive learning, and rotation prediction as proxy tasks for learning contextual representations of input images. the primary task of the network is to reconstruct the original image. The different augmentations used are classical techniques such as in-painting Pathak et al., 2016, out-painting, and noise augmentation to the image by local pixel shuffling Chen et al., 2019. The secondary task of the network is to simultaneously reconstruct the two augmented views as similar to each other as possible via regularized contrastive

loss Chen et al., 2020 as its objective is to maximize the agreement. The term regularized has been used here because contrastive loss is adjusted by the reconstruction loss as a dynamic weight itself. Multiple patches having sizes 128x128x128 are generated and used different views based on the augmentation via the transforms on the same cubic patch. The objective of the SSL network is to reconstruct the original image. The contrastive loss is driven by maximizing the agreement of the reconstruction based on the input of the two augmented views.

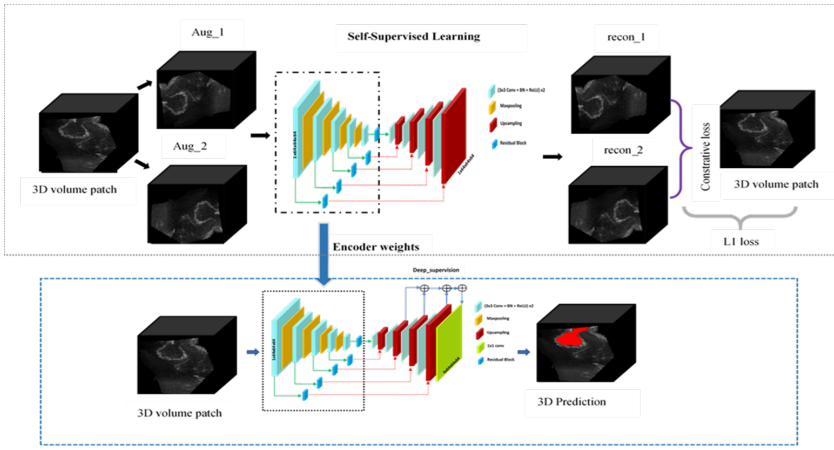


Figure 6.3: Self-supervised learning based 3DResUNet model for iUS segmentation.

The pre-trained encoder weights are used in the proposed 3DResUNet with deep supervision before, during, and after the iUS segmentation task. We trained our proposed model using a training-supervised dataset based on patches. We randomly generated patches

with size 128x128x128 from the input volume and used different augmentations to train the proposed model. We used a sliding window with 128 strides to generate the prediction volume on the validation dataset. The training transforms such as RandCrop, RandGaussianNoise, RandGaussianSmooth, RandShiftIntensity, RandAdjustContrast, and RandZoomd was used to train the proposed 3DResUNet model. All transformations were used from MONAI Library $'https://github.com/Project-MONAI/MONAI'$.

3D-ResUnet with Deep Supervision: A framework of the proposed model is presented as an encoder, a decoder, and a baseline module. The 1x1 convolutional layer with softmax function has been used at the end of the proposed model. The 3D strides convolutional layer has been used to reduce the input image spatial size.

The convolutional block consists of convolutional layers with Batch-Normalization and ReLU activation function to extract the different feature maps from each block on the encoder side. In the encoder block, the spatial input size has been reduced with an increasing number of feature maps, and on the decoder side, the input image spatial size will increase using a 3D Conv-Transpose layer. The input features' maps that are obtained from every encoder block are concatenated with every decoder block feature map to reconstruct the semantic information.

The convolutional (3x3x3conv-BN-ReLu) layer used the input feature maps extracted from every convolutional block on the encoder side and further passed these feature maps into the proposed residual module. The Residual block has been inserted at every encoder block. Each 3d residual module consisted of a 3x3 convolutional, batch norm, and Relu layer with identity skip connection. The spatial size doubled at every decoder block and feature maps are halved at each decoder stage of the proposed model. The feature concatenation has been done at every encoder and decoder block except the last 1x1 convolutional layer. The three-level deep-supervision techniques are applied to get the aggregate loss between ground truth and prediction. The proposed method is

shown in Figure 6.3.

6.2.1.3 Dataset

The challenge organizers provided 23 subjects for training using Intra-operative Ultrasound (iUS) images. They organized a challenge for three tasks (pre-resection, during, and after). All medical images used for the challenge were acquired for routine clinical care at St Olavs University Hospital (Trondheim, Norway) after patients gave their informed consent. The imaging data are available in both MINC-2 and NIFTI-1 formats and the segmentations are available in NIFTI-1 format. All images, MRI, iUS, and segmentations are in the same referential space. A detailed description can be found Xiao et al., 2017. We have used the following preprocessing steps for data cleaning:

- Cropping strategy: Yes
- Resampling Method for anisotropic data: The nearest neighbor interpolation method has been applied for resampling.
- Intensity Normalization method: The dataset has been normalized using a z-score method based on mean and standard deviation.

6.2.1.4 Implementation Details

The learning rate of 0.0004 with Adam optimizer has been for training the proposed model. The cross-entropy and dice function is used as a loss function between the output of the model and the ground-truth sample. 2 batch-size with 1000 epochs has been used with 20 early stopping steps. The best model weights have been saved for prediction in the validation phase. The 128x128x128 and other input image patches were used for training and the sliding window with stride 8 was used as the prediction. The Pytorch library is used for model development, training, optimization, and testing. The V100 Tesla NVidia-GPU machine is used for training and testing the proposed model. The data augmentation methods mentioned in Table 6.1 are used for self-supervision stage

Table 6.1: Training protocols.

Data augmentation methods	RandCrop, RandGaussianNoise, RandGaussianSmooth, RandShiftIntensity, RandAdjustContrast, RandZoom
Initialization of the network	"he" normal initialization
Patch sampling strategy	None
Batch size	2
Patch size	128x128x128
Total epochs	300
Optimizer	Adam
Initial learning rate	0.0001
Learning rate decay schedule	None
Stopping criteria, and optimal model selection criteria	The stopping criterion is reaching the maximum number of epochs (300).
Training time	8 hours

1 and the proposed model stage 2 for training and optimization. The dataset cases have different intensity ranges. The dataset is normalized between 0 and 1 using the max and min intensity normalization method. The detail of the training protocol is shown in Table 6.1.

The same preprocessing has been applied at testing time. The training size of each image is fixed (128x128x128) and used linear interpolation method to resample the prediction mask to the original shape for each validation volume. The sliding window with has been used to get the prediction. The prediction mask produced by our proposed model has been resampled such that it has the same size and spacing as the original image and copies all of the meta-data, i.e., origin, direction, orientation, etc.

6.2.1.5 Results of a Validation dataset

Figure 6.4 results in visualization of some validation cases. One bad and one good example for different subjects has been shown in Figure

6.4. The proposed model achieved good performance in some cases and predicted a good segmentation map. The Dice value for before, during, and after tasks based on the proposed model is shown in Table 6.2. The dice score in a few cases is very bad, especially during and after the stage. We will try to optimize the model in the future to get a better Dice score. The results on the test dataset are shown in Table 6.3 and Table 6.4. We have evaluated our proposed model using different patch sizes. The proposed model with a self-supervised module achieved better performance on task1 and task2.

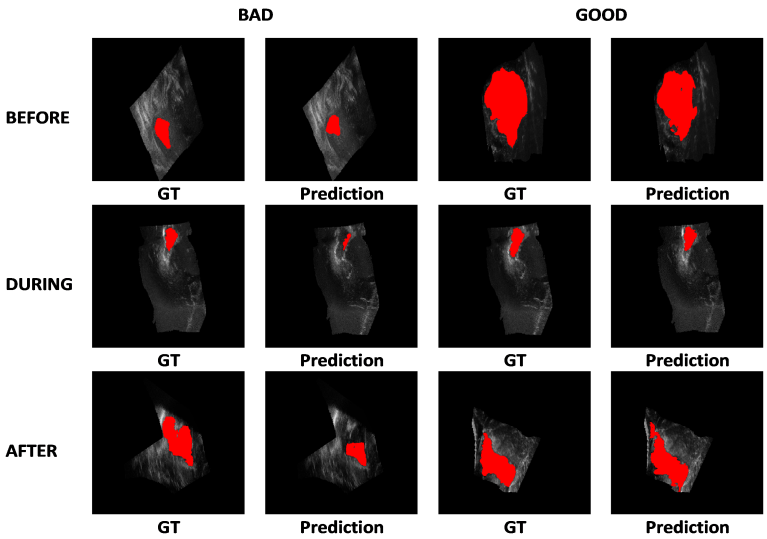


Figure 6.4: 2D visualization using our proposed model.

Table 6.2: The performance of the proposed model on five validation cases for before, during, and after tasks,

	S1	S2	S3	S4	S5	Average
Before	0.825905	0.898576	0.882049	0.310432	0.730627	0.729518
During	0.918594	0.769752	0.237226	0.505755	0.53811	0.593887
After	0.896442	0.895569	0.811527	0.055052	0.532311	0.63818

Table 6.3: The performance of the proposed model using Task 1(Brain tumor segmentation in intra-operative ultrasound) dataset.

Models	Dice	Hd95	Recall	Precision
3DResUNet (128x128x128)	0.4769	99.69	0.5409	0.6230
3DResUNet+SSL (128x128x128)	0.5746	35.26	0.5845	0.7958
3DResUNet+SSL (64x64x64)	0.5430	42.40	0.5929	0.7829
3DResUNet+SSL (256x256x256)	0.5618	86.59	0.6275	0.5814

Table 6.4: Performance of proposed models using Task 2 (Resection cavity segmentation in intra-operative ultrasound) dataset.

Models	Dice	Hd95	Recall	Precision
3DResUNet+SSL (128x128x128)	0.7027	24.54	0.6568	0.7829
3DResUNet+SSL (64x64x64)	0.6859	31.42	0.6213	0.8604
3DResUNet+SSL (256x256x256)	0.6791	22.66	0.6033	0.8686

6.2.1.6 Conclusion

In this work, we presented self-supervision segmentation of Intra-operative Ultrasound and presented a 3D-ResUnet model with deep supervision. Experiments conducted on s CuRIOUS-2022 showed significantly better performance at 0.729, 0.594, and 0.638 ((before, during, and after Intra-operative ultrasound (iUS) segmentation). The overall results on the test dataset showed better performance and our proposed solution could be used as the first step towards correct diagnoses and prediction of iUS segmentation. In the Future, we will develop the 3D transformer-based model.

6.3 Chapter summary

In this chapter, we have explained the positions and descriptions of our proposed solutions in different MICCAI challenges to validate our proposed solution on interdisciplinary datasets for segmentation tasks. We proposed different 2D and 3D segmentation models using various tricks and achieved optimal performance in various segmentation tasks. One of our MICCAI challenges CuRIOUS2022 is explained in detail as this method had been further extended for the brain tumor segmentation task for this Ph.D. dissertation.

We have participated in various MICCAI as well as ISBI challenges other than those explained in this chapter that are mentioned in Figure 6.1 and 6.2. In most of the challenges, we achieved top ranking position as compared to the performance with the top world teams. The enhanced version of the CuRIOUS2022 challenge using 3D MultiEncoder shared 3D transformer-CNN in a self-supervised contrastive learning framework was used in Chapter 4 for the brain tumor segmentation task.

We have also worked on the docker deployment in those challenges as explained previously. This docker experience will be used in the future for deploying this Ph.D. work into a real-time brain tumor detection and survival prediction application.

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Part V

Concluding Remarks and Future works

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Chapter 7

Conclusion and Future works

This dissertation covers the segmentation of the brain tumor and the survival prediction of tumor patients in terms of survival days. A common and most perilous form of brain tumor is glioma which forms when glial cells grow out of control. Gliomas are malignant (cancerous), but some can be very slow growing. They're primary brain tumors, meaning they originate in the brain tissue. Normally, these cells support nerves and help your central nervous system work.

Gliomas don't usually spread outside of the brain or spine but are life-threatening because they can be hard to reach and treat with surgery and grow into other areas of the brain. Therefore, early detection, automatic delineation, and volume estimation are vital tasks for survival prediction and treatment planning. Moreover, manual mark delineation is laborious and time-consuming work for a neurosurgeon. The accurate segmentation of brain tumors can facilitate diagnosis and help assess the prognosis and severity of the disease.

7.1 Summary of Contributions

The main goal of this doctoral dissertation project was to develop fully automated segmentation and survival prediction schemes for MRI-based brain tumor images using deep learning-based methods to overcome the challenges of manual mark delineation, accuracy, cost, and time consumption. The findings presented in this doctoral dissertation are promising and show that the proposed techniques can produce a clinically helpful computer-aided digital tool for brain tumor segmentation and survival prediction analysis by using MRI modality. The present thesis makes several noteworthy contributions to the literature on developing innovative algorithms based on machine learning and deep learning techniques to improve and increase the accuracy of fully automated MRI-based computer-aided design (CAD). In addition, this research has several practical applications, which are presented in five chapters and summarized below. The first chapters cover the fully automated 2D, 2D to 3D, and 3D deep learning-based methods for brain tumor segmentation and survival days prediction. Moreover, a brief set of validation is presented in the sixth chapter on all proposed solutions. This chapter covers the results and publications from real-time medical imaging challenges on interdisciplinary datasets to elucidate the generalization capability and the worth of the proposed solutions presented in this dissertation.

1. Chapter 2 proposes a new methodology based on a fully automated and robust 2D deep learning-based method named Depth-wise Channel Attention Deep Learning (BrainSeg-DCANet) framework presented to accurately segment the Whole Tumor, Tumor Core, and Enhancing Tumor (i.e., WT, TC, and ET) structures in brain tumor MRI images. A special depth-wise attention module has been introduced that helps enhance the extraction of small brain tumor-relevant features and improves efficiently locating the small tumor regions. The proposed

depth-wise attention module can capture the local cross-channel interaction in an efficient way to produce more powerful contextual feature information using a dynamic 1D convolution layer. Furthermore, a discrete wavelet transformed (DWT) is used to enlarge the receptive fields of the segmentation model. The proposed brain tumor segmentation model can extract the prevailing contextual features and produce a better performance as compared to state-of-the-art brain tumor segmentation methods with low computational costs.

2. In Chapter 3, we proposed a novel 2D to 3D multi-view multi-scale segmentation method, the so-called IRMMNET. It combines the key insights from multi-view MRI, including axial, coronal, and sagittal views. IRMMNET comprises different layers with feature reuse capability with various depths and multi-scale information. An efficient method for fusing segmentation maps of the axial, coronal, and sagittal views to develop a 3D segmentation volume presentation. In this chapter, there have been two effective blocks were presented named Inception-Residual encoder block (EB) and dense-spatial attention (DSAM) block. EB helps the brain segmentation network to extract multi-scale brain tissue-relevant information from the multi-view MRI images. DSAM block enhances feature reuse and substantially reduces the number of parameters of the segmentation model.

Extensive experiments are performed using various combinations and settings of the proposed segmentation model. Moreover, we proposed three approaches for survival prediction tasks including prediction utilizing IRMMNET segmentation model encoder, utilizing 3D autoencoder, and utilizing radiomics features. The explainability and importance of the radiomics features are also presented.

3. In Chapter 4, we proposed a new fully 3D CNN method for computing the brain tumor segmentation for all tumor regions including TC, WT, and ET on BraTS 2020 and BraTS2021 datasets. A novel In Chapter 4, we proposed a new fully 3D CNN method for computing the brain tumor segmentation for all tumor regions including TC,

WT, and ET on BraTS 2020 and BraTS2021 datasets. A novel two-stage self-supervised contrastive learning using parallel multiview multiscale attention-based CNN Transformers for 3D brain tumor volumetric segmentation approach was introduced. This method works efficiently for small labeled and large unlabelled datasets. The proposed model results were compared with the various state-of-the-art methods for 3D brain tumor segmentation on MRI images.

4. Chapter 5 is based on the survival prediction analysis for brain tumor patients. A 3D MR image-based survival prediction task was performed. Multiple feature extraction techniques including 3D radiomics features, 3D regressor, and 3D deep CNN features called latent feature technique were used to extract the features from the volumetric MRI images, and then different machine learning-based regression methods were applied to the extracted features. This method was applied to BraTS 2020 dataset. Also, it was tested on the Head Neck Tumor dataset (HECTOR2021) to test its generalization capability.

5. Chapter 6 covers the validation of our proposed solutions on real-time medical imaging challenges (i.e. MICCAI) and their outcomes on the proposed and designed models for this dissertation to evaluate the generalization capacity and efficacy of proposed solutions. The best method from the tested solutions in these challenges was further enhanced and used for the brain tumor segmentation task in this dissertation mentioned in Chapter 4.

7.2 Future Directions

The work presented in this dissertation addresses the interpretation of brain tumor segmentation and survival prediction of tumor patients for their prognosis analysis. We believe this is a crucial field of AI research. Several directions of future work have been identified during this work We couldn't investigate some of them due to time limitations.

1. The deep neural network could not detect small tumors due to small tumor size. The deep neural network would introduce some uncertainty, especially at the boundary of the tumor with surrounding tissues. We need to explore probabilistic deep neural networks to measure the uncertainties in deep neural networks.

2. We did not explore joint registration and segmentation models for brain tumor segmentation. Registration of multimodality could enhance the performance in segmentation tasks.

3. We could not explore the multiple models using multiple modalities. Each modality would be used in an individual model and trained in multiple models with joint loss function in a single training function. Multiple models could potentially be the future direction for brain tumor segmentation.

4. The current state-of-the-art brain tumor-based CAD systems only accept images without considering the patient's clinical data. Yet, in practice, the patient's clinical history and laboratory data enable medical doctors or neurosurgeons to interpret imaging findings in the appropriate clinical context leading to higher and more accurate diagnostics. Different data fusion techniques can be applied to combine the brain tumor images with patients' clinical and laboratory data to achieve this goal and improve the accuracy of the diagnostics.

5. Joint training and optimization of segmentation and survival prediction models for survival time prediction and analysis could be another future direction for the diagnosis and prognosis of brain tumors. We need to deploy proposed models using docker technology for real-time analysis, diagnosis, and prognosis of brain tumors in a real-time clinical setting.

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

Bibliography

Aerts, Hugo JWL, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al. (2014). "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach". In: *Nature communications* 5.1, pp. 1–9.

Almalki, Yassir Edrees, Abdul Qayyum, Muhammad Irfan, Noman Haider, Adam Glowacz, Fahad Mohammed Alshehri, Sharifa K Alduraibi, Khalaf Alshamrani, Mohammad Abd Alkhalik Basha, Alaa Alduraibi, et al. (2021). "A novel method for COVID-19 diagnosis using artificial intelligence in chest X-ray images". In: *Healthcare*. Vol. 9. 5. Multidisciplinary Digital Publishing Institute, p. 522.

Anand, Vikas Kumar, Sanjeev Grampurohit, Pranav Aurangabadkar, Avinash Kori, Mahendra Khened, Raghavendra S Bhat, and Ganapathy Krishnamurthi (2021). "Brain Tumor Segmentation and Survival Prediction using Automatic Hard mining in 3D CNN Architecture". In: arXiv: 2101.01546. URL: http://arxiv.org/abs/2101.01546.

Andrearczyk, Vincent, Valentin Oreiller, Sarah Boughdad, Catherine Chez Le Rest, Hesham Elhalawani, Mario Jreige, John O Prior, Martin Vallières, Dimitris Visvikis, Mathieu Hatt, et al. (2022). "Overview of the HECKTOR Challenge at MICCAI 2021: Automatic Head and Neck Tumor Segmentation and Outcome Prediction in PET/CT Images". In: *arXiv preprint arXiv:2201.04138*.

Andrearczyk, Vincent, Valentin Oreiller, Sarah Boughdad, Catherine Cheze Le Rest, Hesham Elhalawani, Mario Jreige, John O Prior, Martin Vallières, Dimitris Visvikis, Mathieu Hatt, et al. (2021). "Overview of the HECKTOR challenge at MICCAI 2021: automatic head and neck tumor segmentation and outcome prediction in PET/CT images". In: *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Springer, pp. 1–37.

Antwarg, Liat, Ronnie Mindlin Miller, Bracha Shapira, and Lior Rokach (2021). "Explaining anomalies detected by autoencoders using Shapley Additive Explanations". In: *Expert Systems with Applications* 186, p. 115736.

Asis-Cruz, De, Dhineshvikram Krishnamurthy, Chris Jose, Kevin M Cook, Catherine Limperopoulos, et al. (2022). "FetalGAN: Automated Segmentation of Fetal Functional Brain MRI Using Deep Generative Adversarial Learning and Multi-Scale 3D U-Net". In: *Frontiers in Neuroscience*, p. 852.

Atlas, Scott W (2009). *Magnetic resonance imaging of the brain and spine*. Vol. 1. Lippincott Williams & Wilkins.

Bae, Sohi, Yoon Seong Choi, Sung Soo Ahn, Jong Hee Chang, Seok-Gu Kang, Eui Hyun Kim, Se Hoon Kim, and Seung-Koo Lee (2018). "Radiomic MRI phenotyping of glioblastoma: improving survival prediction". In: *Radiology* 289.3, pp. 797–806.

Bai, Wenjia, Chen Chen, Giacomo Tarroni, Jinming Duan, Florian Guitton, Steffen E Petersen, Yike Guo, Paul M Matthews, and Daniel Rueckert (2019). "Self-supervised learning for cardiac mr image segmentation by anatomical position prediction". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, pp. 541–549.

Baid, Ujjwal, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C Kitamura, Sarthak Pati, et al. (2021). "The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification". In: *arXiv preprint arXiv:2107.02314*.

Baid, Ujjwal, Nisarg A Shah, and Sanjay Talbar (2020). "Brain tumor segmentation with cascaded deep convolutional neural network". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part II 5*. Springer, pp. 90–98.

Baid, Ujjwal, Sanjay Talbar, Swapnil Rane, Sudeep Gupta, Meenakshi H Thakur, Aliasgar Moiyadi, Siddhesh Thakur, and Abhishek Mahajan (2019). "Deep learning radiomics algorithm for gliomas (drag) model: a novel approach using 3d unet based deep convolutional neural network for predicting survival in gliomas". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*. Springer, pp. 369–379.

Bakas, Spyridon, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos (2017). "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features". In: *Scientific data* 4.1, pp. 1–13.

Banerjee, Subhashis, Sushmita Mitra, and B Uma Shankar (2019). "Multi-planar spatial-ConvNet for segmentation and survival prediction in brain cancer". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*. Springer, pp. 94–104.

Bano, Sophia, Alessandro Casella, Francisco Vasconcelos, Abdul Qayyum, Abdesslam Benzinou, Moona Mazher, Fabrice Meriaudeau, Chiara Lena, Ilaria Anita Cintorrino, Gaia Romana De Paolis, et al. (2022). "Fetreg2021: a challenge on placental vessel segmentation and registration in fetoscopy". In: *arXiv preprint arXiv:2206.12512*.

Bardes, Adrien, Jean Ponce, and Yann LeCun (2022). "Variance-invariance-covariance regularization for self-supervised learning". In: *ICLR, Vicreg*.

Ben Naceur, Mostefa, Mohamed Akil, Rachida Saouli, and Rostom Kachouri (2020). "Deep Convolutional Neural Networks for Brain Tumor Segmentation: Boosting Performance Using Deep Transfer Learning: Preliminary Results". In: pp. 303–315. DOI: 10.1007/978-3-030-46643-5_30. URL: http://link.springer.com/10.1007/978-3-030-46643-5_30.

Blondiaux, Eléonore and Catherine Garel (2013). "Fetal cerebral imaging–ultrasound vs. MRI: an update". In: *Acta Radiologica* 54.9, pp. 1046–1054.

Bourigault, Emmanuelle, Daniel R McGowan, Abolfazl Mehranian, and Bartłomiej W Papież (2021). "Multimodal PET/CT tumour segmentation and prediction of progression-free survival using a full-scale UNet with attention". In: *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Springer, pp. 189–201.

Cao, Hu, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang (2022). "Swin-unet: Unet-like pure transformer for medical image segmentation". In: *European conference on computer vision*. Springer, pp. 205–218.

Castells, Xavier, Juan Miguel García-Gómez, Alfredo Navarro, Juan José Acebes, Óscar Godino, Susana Boluda, Anna Barceló, Montserrat Robles, Joaquín Ariño, and Carles Arús (2009). "Automated brain

tumor biopsy prediction using single-labeling cDNA microarrays-based gene expression profiling". In: *Diagnostic Molecular Pathology* 18.4, pp. 206–218.

Chaddad, Ahmad, Michael Jonathan Kucharczyk, Paul Daniel, Siham Sabri, Bertrand J Jean-Claude, Tamim Niazi, and Bassam Abdulkarim (2019). "Radiomics in glioblastoma: current status and challenges facing clinical implementation". In: *Frontiers in oncology* 9, p. 374.

Chavez, Martin R, Cande V Ananth, Lillian M Kaminsky, John C Smulian, Lami Yeo, and Anthony M Vintzileos (2006). "Fetal transcerebellar diameter measurement for prediction of gestational age in twins". In: *American journal of obstetrics and gynecology* 195.6, pp. 1596–1600.

Chen, Jieneng, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou (2021). "Transunet: Transformers make strong encoders for medical image segmentation". In: *arXiv preprint arXiv:2102.04306*.

Chen, Liang, Paul Bentley, Kensaku Mori, Kazunari Misawa, Michitaka Fujiwara, and Daniel Rueckert (2019). "Self-supervised learning for medical image analysis using image context restoration". In: *Medical image analysis* 58, p. 101539.

Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille (2017). "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs". In: *IEEE transactions on pattern analysis and machine intelligence* 40.4, pp. 834–848.

Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton (2020). "A simple framework for contrastive learning of visual representations". In: *International conference on machine learning*. PMLR, pp. 1597–1607.

Chen, Yunpeng, Yannis Kalantidis, Jianshu Li, Shuicheng Yan, and Jiashi Feng (2018). "A^2-Nets: Double Attention Networks". In: arXiv: 1810.11579. URL: http://arxiv.org/abs/1810.11579.

Chen, P. , Xu, C. , Li, X. , Ma, Y. , Sun, F. (2019). "Two-stage network for OAR segmentation". In: *Proceedings of the Challenge on Segmentation of THoracic Organs at Risk in CT Images*.

Çiçek, Özgün, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger (2016). "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". In: pp. 424–432. DOI: 10.1007/978-3-319-46723-8_49. URL: http://link.springer.com/10.1007/978-3-319-46723-8_49.

Coroller, Thibaud P, Patrick Grossmann, Ying Hou, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Gretchen Hermann, Philippe Lambin, Benjamin Haibe-Kains, Raymond H Mak, and Hugo JWL Aerts (2015). "CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma". In: *Radiotherapy and Oncology* 114.3, pp. 345–350.

Dai, Zihang, Hanxiao Liu, Quoc V Le, and Mingxing Tan (2021). "Coatnet: Marrying convolution and attention for all data sizes". In: *Advances in Neural Information Processing Systems* 34, pp. 3965–3977.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805*.

Dong, Hao, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo (2017). "Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks". In: pp. 506–517. DOI: 10.1007/978-3-319-60964-5_44. URL: http://link.springer.com/10.1007/978-3-319-60964-5_44.

Dumast, Priscille de, Hamza Kebiri, Kelly Payette, Andras Jakab, Hélène Lajous, and Meritxell Bach Cuadra (2022). "Synthetic magnetic resonance images for domain adaptation: Application to fetal brain tissue segmentation". In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE, pp. 1–5.

Dumast, Priscille de, Hamza Kebiri, Chirine Atat, Vincent Dunet, Mériam Koob, and Meritxell Bach Cuadra (2021). "Segmentation of

the cortical plate in fetal brain MRI with a topological loss". In: *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis*. Springer, pp. 200–209.

Ebner, Michael, Guotai Wang, Wenqi Li, Michael Aertsen, Premal A Patel, Rosalind Aughwane, Andrew Melbourne, Tom Doel, Steven Dymarkowski, Paolo De Coppi, et al. (2020). "An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI". In: *NeuroImage* 206, p. 116324.

Egaña-Ugrinovic, Gabriela, Magdalena Sanz-Cortes, Francesc Figueras, Nuria Bargalló, and Eduard Gratacós (2013). "Differences in cortical development assessed by fetal MRI in late-onset intrauterine growth restriction". In: *American journal of obstetrics and gynecology* 209.2, 126–e1.

Faghihpirayesh, Razieh, Davood Karimi, Deniz Erdogmus, and Ali Gholipour (2022). "Deep Learning Framework for Real-time Fetal Brain Segmentation in MRI". In: *arXiv preprint arXiv:2205.01675*.

Falick Michaeli, Tal, Adam Spiro, Ofra Sabag, Gilad Karavani, Simha Yagel, Smadar Eventov-Friedman, Howard Cedar, Yehudit Bergman, and Yuval Gielchinsky (2019). "Determining gestational age using genome methylation profile: a novel approach for fetal medicine". In: *Prenatal diagnosis* 39.11, pp. 1005–1010.

Feng, Xue, Nicholas J Tustison, Sohil H Patel, and Craig H Meyer (2020). "Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features". In: *Frontiers in computational neuroscience* 14, p. 25.

Fernández-Delgado, Manuel, Manisha Sanjay Sirsat, Eva Cernadas, Sadi Alawadi, Senén Barro, and Manuel Febrero-Bande (2019). "An extensive experimental survey of regression methods". In: *Neural Networks* 111, pp. 11–34.

Fu, Jun, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu (2018). "Dual Attention Network for Scene Segmentation". In: arXiv: 1809.02983. URL: http://arxiv.org/abs/1809.02983.

Fung, Russell, Jose Villar, Ali Dashti, Leila Cheikh Ismail, Eleonora Staines-Urias, Eric O Ohuma, Laurent J Salomon, Cesar G Victora, Fernando C Barros, Ann Lambert, et al. (2020). "Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study". In: *The Lancet Digital Health* 2.7, e368–e375.

Furnari, Frank B, Tim Fenton, Robert M Bachoo, Akitake Mukasa, Jayne M Stommel, Alexander Stegh, William C Hahn, Keith L Ligon, David N Louis, Cameron Brennan, et al. (2007). "Malignant astrocytic glioma: genetics, biology, and paths to treatment". In: *Genes & development* 21.21, pp. 2683–2710.

Gambhir, Sanjiv Sam (2002). "Molecular imaging of cancer with positron emission tomography". In: *Nature Reviews Cancer* 2.9, pp. 683–693.

Gao, Yunhe, Mu Zhou, and Dimitris N Metaxas (2021). "UTNet: a hybrid transformer architecture for medical image segmentation". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*. Springer, pp. 61–71.

Gopika, N and A Meena Kowshalaya ME (2018). "Correlation based feature selection algorithm for machine learning". In: *2018 3rd international conference on communication and electronics systems (ICCES)*. IEEE, pp. 692–695.

Habas, Piotr A, Kio Kim, James M Corbett-Detig, Francois Rousseau, Orit A Glenn, A James Barkovich, and Colin Studholme (2010). "A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation". In: *Neuroimage* 53.2, pp. 460–470.

Hanahan, Douglas and Robert A Weinberg (2011). "Hallmarks of cancer: the next generation". In: *cell* 144.5, pp. 646–674.

Hasan, Md. Kamrul, Lavsen Dahal, Prasad N. Samarakoon, Fakrul Islam Tushar, and Robert Martí (2020). "DSNet: Automatic dermoscopic skin lesion segmentation". In: *Computers in Biology and Medicine* 120, p. 103738. ISSN: 00104825. DOI: 10.1016/j.compbiomed.2020.103738. URL: https://linkinghub.elsevier.com/retrieve/pii/S0010482520301190.

Hatamizadeh, Ali, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu (2021). "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images". In: *International MICCAI Brainlesion Workshop*. Springer, pp. 272–284.

Hatamizadeh, Ali, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu (2022). "Unetr: Transformers for 3d medical image segmentation". In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 574–584.

Havaei, Mohammad, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle (2017). "Brain tumor segmentation with Deep Neural Networks". In: *Medical Image Analysis* 35, pp. 18–31. ISSN: 13618415. DOI: 10.1016/j.media.2016.05.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S1361841516300330.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2015). "Deep Residual Learning for Image Recognition". In: arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385.

He, Tao, Junjie Hu, Ying Song, Jixiang Guo, and Zhang Yi (2020). "Multi-task learning for the segmentation of organs at risk with label dependence". In: *Medical Image Analysis* 61, p. 101666. ISSN: 13618415. DOI: 10.1016/j.media.2020.101666. URL: https://linkinghub.elsevier.com/retrieve/pii/S1361841520300323.

Heidari, Moein, Amirhossein Kazerouni, Milad Soltany, Reza Azad, Ehsan Khodapanah Aghdam, Julien Cohen-Adad, and Dorit Merhof (2023). "Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 6202–6212.

Henry, Theophraste, Alexandre Carre, Marvin Lerousseau, Theo Estienne, Charlotte Robert, Nikos Paragios, and Eric Deutsch (2020). "Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a BraTS 2020 challenge solution". In: arXiv: 2011.01045. URL: http://arxiv.org/abs/2011.01045.

Hu, Jie, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu (2017). "Squeeze-and-Excitation Networks". In: arXiv: 1709.01507. URL: http://arxiv.org/abs/1709.01507.

Huang, He, Wenbo Zhang, Ying Fang, Jialing Hong, Shuaixi Su, and Xiaobo Lai (2021). "Overall survival prediction for gliomas using a novel compound approach". In: *Frontiers in Oncology* 11, p. 724191.

Huang, Zilong, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu (2019). "Ccnet: Criss-cross attention for semantic segmentation". In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 603–612.

Iqbal, Sajid, Muhammad U Ghani Khan, Tanzila Saba, Zahid Mehmood, Nadeem Javaid, Amjad Rehman, and Rashid Abbasi (2019). "Deep learning model integrating features and novel classifiers fusion for brain tumor segmentation". In: *Microscopy research and technique* 82.8, pp. 1302–1315.

Isensee, Fabian, Paul F. Jaeger, Peter M. Full, Philipp Vollmuth, and Klaus H. Maier-Hein (2020). "nnU-Net for Brain Tumor Segmentation". In: arXiv: 2011.00848. URL: http://arxiv.org/abs/2011.00848.

Isensee, Fabian, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein (2021). "nnU-Net: a self-configuring method for deep

learning-based biomedical image segmentation". In: *Nature methods* 18.2, pp. 203–211.

Isensee, Fabian, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H. Maier-Hein (2018). "Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge". In: pp. 287–297. DOI: 10.1007/978-3-319-75238-9_25. URL: http://link.springer.com/10.1007/978-3-319-75238-9_25.

Islam, Mobarakol, VS Vibashan, V Jeya Maria Jose, Navodini Wijethilake, Uppal Utkarsh, and Hongliang Ren (2020). "Brain tumor segmentation and survival prediction using 3D attention UNet". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5*. Springer, pp. 262–272.

Jarvis, Deborah A, Chloe R Finney, and Paul D Griffiths (2019). "Normative volume measurements of the fetal intra-cranial compartments using 3D volume in utero MR imaging". In: *European radiology* 29.7, pp. 3488–3495.

Juanco-Müller, Ángel Víctor, João FC Mota, Keith Goatman, and Corné Hoogendoorn (2021). "Deep supervoxel segmentation for survival analysis in head and neck cancer patients". In: *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Springer, pp. 257–265.

Jungo, Alain, Richard McKinley, Raphael Meier, Urspeter Knecht, Luis Vera, Julián Pérez-Beteta, David Molina-García, Víctor M Pérez-García, Roland Wiest, and Mauricio Reyes (2018). "Towards uncertainty-assisted brain tumor segmentation and survival prediction". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3*. Springer, pp. 474–485.

Kamnitsas, Konstantinos, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker (2017a). "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *Medical image analysis* 36, pp. 61–78.

Kamnitsas, Konstantinos, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker (2017b). "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *Medical Image Analysis* 36, pp. 61–78. ISSN: 13618415. DOI: 10.1016/j.media.2016.10.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S1361841516301839.

Ke, Rihuan, Aurelie Bugeau, Nicolas Papadakis, Mark Kirkland, Peter Schuetz, and Carola-Bibiane Schonlieb (2021). "Multi-Task Deep Learning for Image Segmentation Using Recursive Approximation Tasks". In: *IEEE Transactions on Image Processing* 30, pp. 3555–3567. ISSN: 1057-7149. DOI: 10.1109/TIP.2021.3062726. URL: https://ieeexplore.ieee.org/document/9371403/.

Khalili, Nadieh, Nikolas Lessmann, Elise Turk, N Claessens, Roel de Heus, Tessel Kolk, Max A Viergever, Manon JNL Benders, and Ivana Išgum (2019). "Automatic brain tissue segmentation in fetal MRI using convolutional neural networks". In: *Magnetic resonance imaging* 64, pp. 77–89.

Kojita, Yasuyuki, Hidetoshi Matsuo, Tomonori Kanda, Mizuho Nishio, Keitaro Sofue, Munenobu Nogami, Atsushi K Kono, Masatoshi Hori, and Takamichi Murakami (2021). "Deep learning model for predicting gestational age after the first trimester using fetal MRI". In: *European Radiology* 31.6, pp. 3775–3782.

Kong, Xiangmao, Guoxia Sun, Qiang Wu, Ju Liu, and Fengming Lin (2018). "Hybrid Pyramid U-Net Model for Brain Tumor Segmentation". In: pp. 346–355. DOI: 10.1007/978-3-030-00828-4_35. URL: http://link.springer.com/10.1007/978-3-030-00828-4_35.

Kyriakopoulou, Vanessa, Deniz Vatansever, Alice Davidson, Prachi Patkee, Samia Elkommos, Andrew Chew, Miriam Martinez-Biarge, Bibbi Hagberg, Mellisa Damodaram, Joanna Allsop, et al. (2017). "Normative biometry of the fetal brain using magnetic resonance imaging". In: *Brain Structure and Function* 222.5, pp. 2295–2307.

Lalande, Alain, Zhihao Chen, Thibaut Pommier, Thomas Decourselle, Abdul Qayyum, Michel Salomon, Dominique Ginhac, Youssef Skandarani, Arnaud Boucher, Khawla Brahim, et al. (2022). "Deep learning methods for automatic evaluation of delayed enhancement-MRI. The results of the EMIDEC challenge". In: *Medical Image Analysis* 79, p. 102428.

LeCun, Yann, Yoshua Bengio, et al. (1995). "Convolutional networks for images, speech, and time series". In: *The handbook of brain theory and neural networks* 3361.10, p. 1995.

Li, Haichun, Ao Li, and Minghui Wang (2019). "A novel end-to-end brain tumor segmentation method using improved fully convolutional networks". In: *Computers in Biology and Medicine* 108, pp. 150–160. ISSN: 00104825. DOI: 10.1016/j.compbiomed.2019.03.014. URL: https://linkinghub.elsevier.com/retrieve/pii/S0010482519300873.

Liu, Pengju, Hongzhi Zhang, Wei Lian, and Wangmeng Zuo (2019). "Multi-Level Wavelet Convolutional Neural Networks". In: *IEEE Access* 7, pp. 74973–74985. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2921451. URL: https://ieeexplore.ieee.org/document/8732332/.

Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo (2021). "Swin transformer: Hierarchical vision transformer using shifted windows". In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell (2014). "Fully Convolutional Networks for Semantic Segmentation". In: arXiv: 1411.4038. URL: http://arxiv.org/abs/1411.4038.

Louis, David N, Hiroko Ohgaki, Otmar D Wiestler, Webster K Cavenee, Peter C Burger, Anne Jouvet, Bernd W Scheithauer, and Paul Kleihues (2007). "The 2007 WHO classification of tumours of the central nervous system". In: *Acta neuropathologica* 114, pp. 97–109.

Lu, Yu, Xi Zhang, Xianghua Fu, Fangxiong Chen, and Kelvin KL Wong (2019). "Ensemble machine learning for estimating fetal weight at varying gestational age". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01, pp. 9522–9527.

Lyu, Chenggang and Hai Shu (2020). "A Two-Stage Cascade Model with Variational Autoencoders and Attention Gates for MRI Brain Tumor Segmentation". In: arXiv: 2011.02881. URL: http://arxiv.org/abs/2011.02881.

Ma, Jun, Yao Zhang, Song Gu, Xingle An, Zhihe Wang, Cheng Ge, Congcong Wang, Fan Zhang, Yu Wang, Yinan Xu, et al. (2022). "Fast and low-GPU-memory abdomen CT organ segmentation: the flare challenge". In: *Medical Image Analysis* 82, p. 102616.

Ma, Ningning, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun (2018). "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design". In: arXiv: 1807.11164. URL: http://arxiv.org/abs/1807.11164.

Maternal, Alliance for, Newborn Health Improvement (AMANHI) Gestational Age Study Group, Alliance for Maternal, and Newborn Health Improvement (AMANHI) GA Study Group (2021). "Simplified models to assess newborn gestational age in low-middle income countries: findings from a multicountry, prospective cohort study". In: *BMJ global health* 6.9, e005688.

Mazher, Moona, Abdul Qayyum, Domenec Puig, and Mohamed Abdel-Nasser (2022). "Effective Approaches to Fetal Brain Segmentation in MRI and Gestational Age Estimation by Utilizing a Multiview Deep Inception Residual Network and Radiomics". In: *Entropy* 24.12, p. 1708.

Mehta, Neeta (2011). "Mind-body dualism: A critique from a health perspective". In: *Mens sana monographs* 9.1, p. 202.

Menze, Bjoern H, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. (2014). "The multimodal brain tumor image segmentation benchmark (BRATS)". In: *IEEE transactions on medical imaging* 34.10, pp. 1993–2024.

Namburete, Ana IL, Richard V Stebbing, Bryn Kemp, Mohammad Yaqub, Aris T Papageorghiou, and J Alison Noble (2015). "Learning-based prediction of gestational age from ultrasound images of the fetal brain". In: *Medical image analysis* 21.1, pp. 72–86.

Noori, Mehrdad, Ali Bahri, and Karim Mohammadi (2019). "Attention-Guided Version of 2D UNet for Automatic Brain Tumor Segmentation". In: *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*. IEEE, pp. 269–275. ISBN: 978-1-7281-5075-8. DOI: 10.1109/ICCKE48569.2019.8964956. URL: https://ieeexplore.ieee.org/document/8964956/.

Noreen, Neelum, Sellappan Palaniappan, Abdul Qayyum, Iftikhar Ahmad, Muhammad Imran, and Muhammad Shoaib (2020). "A deep learning model based on concatenation approach for the diagnosis of brain tumor". In: *IEEE Access* 8, pp. 55135–55144.

Osman, Alexander FI (2019). "A multi-parametric MRI-based radiomics signature and a practical ML model for stratifying glioblastoma patients based on survival toward precision oncology". In: *Frontiers in Computational Neuroscience* 13, p. 58.

Pálsson, Sveinn, Stefano Cerri, Hans Skovgaard Poulsen, Thomas Urup, Ian Law, and Koen Van Leemput (2022). "Predicting survival of glioblastoma from automatic whole-brain and tumor segmentation of MR images". In: *Scientific Reports* 12.1, p. 19744.

Pathak, Deepak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros (2016). "Context encoders: Feature learning by inpainting". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2536–2544.

Payette, Kelly, Priscille de Dumast, Hamza Kebiri, Ivan Ezhov, Johannes C Paetzold, Suprosanna Shit, Asim Iqbal, Romesa Khan, Raimund Kottke, Patrice Grehten, et al. (2021). "An automatic multi-tissue human fetal brain segmentation benchmark using the Fetal Tissue Annotation Dataset". In: *Scientific Data* 8.1, pp. 1–14.

Payette, Kelly, Hongwei Li, Priscille de Dumast, Roxane Licandro, Hui Ji, Md Mahfuzur Rahman Siddiquee, Daguang Xu, Andriy Myronenko, Hao Liu, Yuchen Pei, et al. (2023). "Fetal brain tissue annotation and segmentation challenge results". In: *Medical Image Analysis*, p. 102833.

Pei, Linmin, Lasitha Vidyaratne, Md Monibor Rahman, and Khan M. Iftekharuddin (2020). "Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images". In: *Scientific Reports* 10.1, p. 19726. ISSN: 2045-2322. DOI: 10.1038/s41598-020-74419-9. URL: http://www.nature.com/articles/s41598-020-74419-9.

Peiris, H, M Hayat, Z Chen, G Egan, and M Harandi (n.d.). "A volumetric transformer for accurate 3d tumor segmentation. arXiv 2021". In: *arXiv preprint arXiv:2111.13300* ().

Pereira, Sergio, Adriano Pinto, Victor Alves, and Carlos A. Silva (2016). "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images". In: *IEEE Transactions on Medical Imaging* 35.5, pp. 1240–1251. ISSN: 0278-0062. DOI: 10.1109/TMI.2016.2538465. URL: http://ieeexplore.ieee.org/document/7426413/.

Prasoon, Adhish, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, and Mads Nielsen (2013). "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network".

In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp. 246–253.

Qayyum, Abdul, Alain Lalande, Thomas Decourselle, Thibaut Pommier, Alexandre Cochet, and Fabrice Meriaudeau (2020). "Segmentation of the myocardium on late-gadolinium enhanced MRI based on 2.5 D residual squeeze and excitation deep learning model". In: *arXiv preprint arXiv:2005.13643*.

Qayyum, Abdul, Alain Lalande, and Fabrice Meriaudeau (2020). "Automatic segmentation of tumors and affected organs in the abdomen using a 3D hybrid model for computed tomography imaging". In: *Computers in Biology and Medicine* 127, p. 104097.

Qayyum, Abdul, Moona Mazher, Steven Niederer, and Imran Razzak (2023). "Segmentation of Intra-operative Ultrasound Using Self-supervised Learning Based 3D-ResUnet Model with Deep Supervision". In: *Lesion Segmentation in Surgical and Diagnostic Applications: MICCAI 2022 Challenges, CuRIOUS 2022, KiPA 2022 and MELA 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18–22, 2022, Proceedings*. Springer, pp. 55–62.

Ramachandran, Prajit, Barret Zoph, and Quoc V. Le (2017a). "Searching for Activation Functions". In: arXiv: 1710.05941. URL: http://arxiv.org/abs/1710.05941.

Ramachandran, Prajit, Barret Zoph, and Quoc V Le (2017b). "Searching for activation functions". In: *arXiv preprint arXiv:1710.05941*.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015a). "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, pp. 234–241.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015b). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28. URL: http://link.springer.com/10.1007/978-3-319-24574-4_28.

Roy, Abhijit Guha, Nassir Navab, and Christian Wachinger (2018). "Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer, pp. 421–429.

Saut, Olivier, Jean-Baptiste Lagaert, Thierry Colin, and Hassan M. Fathallah-Shaykh (2014). "A Multilayer Grow-or-Go Model for GBM: Effects of Invasive Cells and Anti-Angiogenesis on Growth". In: *Bulletin of Mathematical Biology* 76.9, pp. 2306–2333. ISSN: 0092-8240. DOI: 10.1007/s11538-014-0007-y. URL: http://link.springer.com/10.1007/s11538-014-0007-y.

Schlemper, Jo, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert (2019). "Attention gated networks: Learning to leverage salient regions in medical images". In: *Medical Image Analysis* 53, pp. 197–207. ISSN: 13618415. DOI: 10.1016/j.media.2019.01.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S1361841518306133.

Schwartzbaum, Judith A, James L Fisher, Kenneth D Aldape, and Margaret Wrensch (2006). "Epidemiology and molecular pathology of glioma". In: *Nature clinical practice Neurology* 2.9, pp. 494–503.

Seow, Pohchoo, Jeannie Hsiu Ding Wong, Azlina Ahmad-Annuar, Abhishek Mahajan, Nor Aniza Abdullah, and Norlisah Ramli (2018). "Quantitative magnetic resonance imaging and radiogenomic biomarkers for glioma characterisation: a systematic review". In: *The British journal of radiology* 91.1092, p. 20170930.

Shboul, Zeina A, Mahbubul Alam, Lasitha Vidyaratne, Linmin Pei, Mohamed I Elbakary, and Khan M Iftekharuddin (2019). "Feature-guided deep radiomics for glioblastoma patient survival prediction". In: *Frontiers in neuroscience* 13, p. 966.

Shen, Liyue, Jimmy Zheng, Edward H Lee, Katie Shpanskaya, Emily S McKenna, Mahesh G Atluri, Dinko Plasto, Courtney Mitchell, Lillian M Lai, Carolina V Guimaraes, et al. (2022). "Attention-guided deep learning for gestational age prediction using fetal brain MRI". In: *Scientific reports* 12.1, pp. 1–10.

Silva, Carlos A., Adriano Pinto, Sérgio Pereira, and Ana Lopes (2021). "Multi-stage Deep Layer Aggregation for Brain Tumor Segmentation". In: arXiv: 2101.00490. URL: http://arxiv.org/abs/2101.00490.

Starke, Sebastian, Stefan Leger, Alex Zwanenburg, Karoline Leger, Fabian Lohaus, Annett Linge, Andreas Schreiber, Goda Kalinauskaite, Inge Tinhofer, Nika Guberina, et al. (2020). "2D and 3D convolutional neural networks for outcome modelling of locally advanced head and neck squamous cell carcinoma". In: *Scientific reports* 10.1, pp. 1–13.

Sun, Li, Songtao Zhang, Hang Chen, and Lin Luo (2019). "Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning". In: *Frontiers in neuroscience* 13, p. 810.

Szegedy, Christian, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning". In: *Thirty-first AAAI conference on artificial intelligence*.

Taghanaki, Saeid Asgari, Yefeng Zheng, S. Kevin Zhou, Bogdan Georgescu, Puneet Sharma, Daguang Xu, Dorin Comaniciu, and Ghassan Hamarneh (2019a). "Combo loss: Handling input and output imbalance in multi-organ segmentation". In: *Computerized Medical Imaging and Graphics* 75, pp. 24–33. ISSN: 08956111. DOI: 10.1016/j.compmedimag.2019.04.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0895611118305688.

Taghanaki, Saeid Asgari, Yefeng Zheng, S Kevin Zhou, Bogdan Georgescu, Puneet Sharma, Daguang Xu, Dorin Comaniciu, and

Ghassan Hamarneh (2019b). "Combo loss: Handling input and output imbalance in multi-organ segmentation". In: *Computerized Medical Imaging and Graphics* 75, pp. 24–33.

Tajbakhsh, Nima, Laura Jeyaseelan, Qian Li, Jeffrey N Chiang, Zhihao Wu, and Xiaowei Ding (2020). "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation". In: *Medical Image Analysis* 63, p. 101693.

Van Griethuysen, Joost JM, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts (2017). "Computational radiomics system to decode the radiographic phenotype". In: *Cancer research* 77.21, e104–e107.

Vente, Coen de, Koenraad A Vermeer, Nicolas Jaccard, He Wang, Hongyi Sun, Firas Khader, Daniel Truhn, Temirgali Aimyshev, Yerkebulan Zhanibekuly, Tien-Dung Le, et al. (2023). "AIROGS: Artificial Intelligence for RObust Glaucoma Screening Challenge". In: *arXiv preprint arXiv:2302.01738*.

Vesal, Sulaiman, Nishant Ravikumar, and Andreas Maier (2019). "A 2D dilated residual U-Net for multi-organ segmentation in thoracic CT". In: arXiv: 1905.07710. URL: http://arxiv.org/abs/1905.07710.

Vig, Jesse (2019). "A multiscale visualization of attention in the transformer model". In: *arXiv preprint arXiv:1906.05714*.

Wang, Feifan, Runzhou Jiang, Liqin Zheng, Chun Meng, and Bharat Biswal (2020). "3D U-Net Based Brain Tumor Segmentation and Survival Days Prediction". In: pp. 131–141. DOI: 10.1007/978-3-030-46640-4_13. URL: http://link.springer.com/10.1007/978-3-030-46640-4_13.

Wang, Guotai, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren (2018). "Automatic Brain Tumor Segmentation Using Cascaded Anisotropic Convolutional Neural Networks". In: pp. 178–190. DOI: 10.1007/978-3-319-75238-9_16. URL: http://link.springer.com/10.1007/978-3-319-75238-9_16.

Wang, Guotai, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren (2019a). "Automatic Brain Tumor Segmentation Using Convolutional Neural Networks with Test-Time Augmentation". In: pp. 61–72. DOI: 10.1007/978-3-030-11726-9_6. URL: http://link.springer.com/10.1007/978-3-030-11726-9_6.

Wang, Qin, Weibing Zhao, Chun-Hui Zhang, Liyue Zhang, Changmiao Wang, Zhen Li, Shuguang Cui, and Guanbin Li (2019b). "3D Enhanced Multi-scale Network for Thoracic Organs Segmentation." In: *SegTHOR@ ISBI* 3, pp. 1–5.

Wang, Wenhai, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao (2021). "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions". In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 568–578.

Weninger, Leon, Christoph Haarburger, and Dorit Merhof (2019). "Robustness of radiomics for survival prediction of brain tumor patients depending on resection status". In: *Frontiers in computational neuroscience* 13, p. 73.

Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon (2018a). "CBAM: Convolutional Block Attention Module". In: arXiv: 1807.06521. URL: http://arxiv.org/abs/1807.06521.

Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon (2018b). "Cbam: Convolutional block attention module". In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19.

Wu, Haiping, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang (2021). "Cvt: Introducing convolutions to vision transformers". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22–31.

Wu, J, SP Awate, DJ Licht, C Clouchoux, Adre J Du Plessis, BB Avants, A Vossough, JC Gee, and Catherine Limperopoulos (2015). "Assessment of MRI-based automated fetal cerebral cortical folding measures in prediction of gestational age in the third trimester". In: *American Journal of Neuroradiology* 36.7, pp. 1369–1374.

Xiao, Yiming, Maryse Fortin, Geirmund Unsgård, Hassan Rivaz, and Ingerid Reinertsen (2017). "RE troSpective Evaluation of Cerebral Tumors (RESECT): A clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries". In: *Medical physics* 44.7, pp. 3875–3882.

Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He (2016). "Aggregated Residual Transformations for Deep Neural Networks". In: arXiv: 1611.05431. URL: http://arxiv.org/abs/1611.05431.

Xie, Zhenda, Yutong Lin, Zheng Zhang, Yue Cao, Stephen Lin, and Han Hu (2021). "Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16684–16693.

Yao, Jiawen, Sheng Wang, Xinliang Zhu, and Junzhou Huang (2016). "Imaging biomarker discovery for lung cancer survival prediction". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer, pp. 649–657.

Zabihollahy, Fatemeh, Martin Rajchl, James A White, and Eranga Ukwatta (2020). "Fully automated segmentation of left ventricular scar from 3D late gadolinium enhancement magnetic resonance imaging using a cascaded multi-planar U-Net (CMPU-Net)". In: *Medical physics* 47.4, pp. 1645–1655.

Zhao, L, JD Asis-Cruz, X Feng, Y Wu, K Kapse, A Largent, J Quistorff, C Lopez, D Wu, K Qing, et al. (2022). "Automated 3D Fetal Brain Segmentation Using an Optimized Deep Learning Approach". In: *American Journal of Neuroradiology* 43.3, pp. 448–454.

Zhou, Hong-Yu, Jiansen Guo, Yinghao Zhang, Lequan Yu, Liansheng Wang, and Yizhou Yu (2021). "nnformer: Interleaved transformer for volumetric segmentation". In: *arXiv preprint arXiv:2109.03201*.

Zhou, Zhuxian, Xueming Wu, Adam Kresak, Mark Griswold, and Zheng-Rong Lu (2013). "Peptide targeted tripod macrocyclic Gd (III) chelates for cancer molecular MRI". In: *Biomaterials* 34.31, pp. 7683–7693.

UNIVERSITAT ROVIRA I VIRGILI
SELF-SUPERVISED ADVANCED DEEP LEARNING FOR CHARACTERIZATION OF BRAIN TUMOR AGGRESSIVENESS AND PROGNOSIS ANALYSIS THROUGH MULTIMODALITY MRI IMAGING
Moona Mazher

UNIVERSITAT ROVIRA i VIRGILI