

Shreyansh Gupta

5Thesis_ShreyanshGupta 2

 M.TECH. THESIS 2

Document Details

Submission ID

trn:oid:::27535:141092234

Submission Date

May 31, 2026, 11:53 AM GMT+5:30

Download Date

Jul 2, 2026, 2:37 PM GMT+5:30

File Name

5Thesis_ShreyanshGupta 2.pdf

File Size

2.3 MB

63 Pages

14,957 Words

88,816 Characters

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Small Matches (less than 8 words)

Match Groups

-  **73 Not Cited or Quoted 5%**
Matches with neither in-text citation nor quotation marks
-  **3 Missing Quotations 0%**
Matches that are still very similar to source material
-  **2 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 2%  Internet sources
- 3%  Publications
- 3%  Submitted works (Student Papers)

Match Groups

- 73 Not Cited or Quoted 5%**
Matches with neither in-text citation nor quotation marks
- 3 Missing Quotations 0%**
Matches that are still very similar to source material
- 2 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 2% Internet sources
- 3% Publications
- 3% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	arxiv.org	<1%
2	Publication	"Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries", Spr...	<1%
3	Internet	www.mdpi.com	<1%
4	Publication	Shuming Lin, Sang Fyeng, Jinyi Liang, Junnan Tan. "PIDE-Net: A Heterogeneous Pr...	<1%
5	Student papers	Coventry University on 2025-12-04	<1%
6	Student papers	Indian School of Business on 2026-06-02	<1%
7	Publication	Khadija Nawaz, Atika Zanib, Yu Wang, Noor Ayesha, Saeed Ali Omar Bahaj, Tariq ...	<1%
8	Publication	Ahmadkhani, Mohsen. "Topologically-Enhanced Deep Learning for Spatial Image ...	<1%
9	Student papers	University of Surrey on 2025-09-09	<1%
10	Publication	Chenguang You, Chenzhuang Dong, Jiachen Lu, Siyuan Xie, Yijie Duan, Lili Duan, ...	<1%

11	Student papers	University of Birmingham on 2026-02-12	<1%
12	Internet	www.datatechnotes.com	<1%
13	Publication	Xueke Liu, Binglin Zhang. "Semantic-Aware Temporal Modeling for Intelligent Cla...	<1%
14	Publication	M. Karthiga, U. Karthikeyan, N. Janani, A. Saran Kumar. "SWFT-net: sparse wavele...	<1%
15	Student papers	Tilburg University on 2025-05-15	<1%
16	Internet	mdpi-res.com	<1%
17	Internet	pure.rug.nl	<1%
18	Publication	"Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brai...	<1%
19	Internet	ddd.uab.cat	<1%
20	Publication	Shuo-Peng Gong, Mohammad Khishe, Mokhtar Mohammadi. "Niching Chimp Opt...	<1%
21	Internet	spectrum.library.concordia.ca	<1%
22	Internet	ocr.oldfish.cn	<1%
23	Internet	archiv.ub.uni-heidelberg.de	<1%
24	Internet	diposit.ub.edu	<1%

25	Student papers	Skolkovo Institute of Science and Technology (Skoltech) on 2022-05-28	<1%
26	Student papers	University of Macau on 2025-02-28	<1%
27	Internet	doczz.net	<1%
28	Internet	www.researchsquare.com	<1%
29	Publication	Malak Bachri, Qusai Gazawy, Ahmad Al Shami. "Chapter 9 PH-SAM2: Persistent Ho...	<1%
30	Student papers	Queen Mary and Westfield College on 2026-04-13	<1%
31	Student papers	Trinity College Dublin on 2025-04-10	<1%
32	Student papers	University of Exeter on 2025-05-02	<1%
33	Internet	www.istaer.online	<1%
34	Internet	www.sciencegate.app	<1%
35	Publication	"Soft Computing for Problem Solving", Springer Science and Business Media LLC, ...	<1%
36	Publication	Bhabani Prasanna Pattanaik, Dharendra Kumar Shukla, Sambit Satpathy, Kamala...	<1%
37	Publication	C Amantini. "Capsaicin-induced apoptosis of glioma cells is mediated by TRPV1 va...	<1%
38	Publication	Guowei Dai, Chaoyu Wang, Duwei Dai, Qingfeng Tang, Yi Zhang, Hu Chen. "Interp...	<1%

39	Student papers	ICTS on 2026-06-20	<1%
40	Student papers	Indian School of Mines on 2023-01-11	<1%
41	Student papers	Jacobs University, Bremen on 2021-04-19	<1%
42	Publication	Sandeep Kumar Mathivanan, Saurav Mallik, Basu Dev Shivahare, S.K.B. Sangeeth...	<1%
43	Student papers	University of Malaya on 2023-06-25	<1%
44	Publication	Xiu Shu, Youqiang Xiong, Zhangli Ma, Xinming Zhang, Di Yuan. "Efficient Medical I...	<1%
45	Internet	www.biorxiv.org	<1%

ABSTRACT

The accurate segmentation of brain tumour sub-regions: necrotic core (NCR), peritumoral edema (ED), and gadolinium-enhancing tumour (ET), using multi-parametric Magnetic Resonance Imaging (mpMRI) is a clinically essential

problem for pre-operative surgical planning, radiotherapy contouring, and treatment response assessment. While there have been considerable advances in this field, two major families of Deep Learning architectures face significant intrinsic limitations. CNNs models local spatial dependencies but fail to represent long-range volumetric dependencies; ViTs tackle this problem but make them computationally infeasible due to their quadratic complexity on high-resolution 3D volumes. This thesis proposes a clinically-oriented deep learning approach that integrates three different technologies. Firstly, the selective state-space models (SSMs) – more specifically, the Mamba architecture – allow to efficiently learn the global spatial context on volumetric MRI with linear complexity $O(N)$. Secondly, the integration of the Topological Data Analysis (TDA) technique, provides a means of imposing an anatomically plausible regularisation on the tumour segmentation process. In addition to the segmentation component, the proposed system architecture employs radiogenomic multi-task learning, allowing for non-invasive virtual biopsy by predicting clinically relevant molecular biomarkers, including IDH mutations and MGMT promoter methylation, using image-based features. Domain adaptation techniques with GRL and federated learning approaches are included to tackle the problem of domain shift, especially the scarcity of paediatric glioma cases. Interpretability in clinical settings is ensured through the application of CBMs and Grad-CAM saliency maps, making it possible to audit and trust the model's decision-making process. Comparing with previous state-of-the-art frameworks on the BraTS dataset, our hybrid CNN-Mamba network architecture achieves superior Dice Similarity Coefficient (DSC) scores, with the IChOA optimised CNN model reaching 0.97 mean DSC, and the topology-aware loss minimises anatomically inconsistent outputs. Our framework provides a comprehensive, explainable, and deployable pipeline for automatic analysis of neuro-oncological tumours in both adult and paediatric population.

Chapter 1

Introduction

1.1 Overview

Brain cancer is an umbrella term for the malignant tumor that arises due to uncontrolled cell growth in the brain and constitutes one of the deadliest conditions in human medicine. Brain tumors can originate either from the glial cells of the brain parenchyma (brain glioma) or by the metastatic dissemination of the peripheral malignancies. The latter accounts for the majority of the brain tumor cases (primary brain tumors) and demonstrates significant variability in the tumoral growth and infiltrating pattern, thus complicating their complete removal.

In clinical practice, accurate localization and segmentation of the tumor subregions within the acquired three-dimensional volume of the brain MRI images play an essential role in diagnostics, surgical planning, radiation therapy planning, and post-treatment follow-ups. The traditional method of diagnosis includes the use of mpMRI technique that encompasses four different MRI sequences, including T1, T1CE, T2, and FLAIR. These techniques yield specific information regarding the subregions of the tumor, including ET from T1CE sequence, ED from T2/FLAIR

sequence, and NCR from T1 sequence. In combination, they form the Whole Tumor (WT) structure.

Conventionally, tumour segmentation involved time-intensive manual delineation by specialist neuroradiologists, which is prone to considerable inter- and intra-rater variability and impractical at scale (typically taking up to 1-4 hours per patient). Deep learning has brought about a paradigm shift, where fully automated tumour segmentation pipelines can now segment the entire mpMRI data set in seconds, with accuracy comparable to expert performance. Over time, the MICCAI community-driven Brain Tumour Segmentation (BraTS) challenge became the benchmark of choice for validating new advances in automatic tumour segmentation, culminating in the nnU-Net self-configuring framework setting the bar with a CNN baseline achieving 0.89 mean Dice coefficient.

Despite impressive benchmark performance, the current state-of-the-art in automated segmentation systems suffers from four main translational barriers preventing the transition from bench to bedside: heavy computational burden due to global modelling by transformers, topological failures resulting in unrealistic anatomical tumour shapes, separation between automated segmentation and molecular classification of brain tumours, and poor out-of-domain performance when moving from adults to children.

1.2 Motivation

The major motivation behind this study can be represented as follows:

Clinical Need: Since the addition of molecular markers, such as IDH mutations and MGMT promoter methylation, to the 2021 WHO classification system, treatment paradigms have been entirely modified. IDH-mutated gliomas have a more favourable prognosis and react to alkylating chemotherapies in a different way from IDH-wildtype gliomas.

CNN Upgrades : Existing CNN-based methods perform well on learning local spatial patterns, but cannot learn the extended spatial relationships inherent in infiltration zones of glioma. Although Vision Transformers could capture large spatial dependencies through global self-attention, they end up being an inefficient solution on high-resolution volumetric scans, due to their lack of communication and memory capabilities, both of which scale as $O(L^2)$. Furthermore, usual voxel-level loss functions such as Dice and cross entropy maximize overlap metrics, and although the sub-regions of a tumour are structurally consistent, the network may make senseless predictions without this consistency; for example: producing physically impossible results in which the segments do not overlap, there are holes, and the hierarchies within sub-regions are ill-suited.

Technology Advancement: It is particularly suitable for volumetric medical images where long-distance dependency and computational efficiency are essential requirements. Leveraging the synergy between Mamba's linear complexity, the inductive bias of CNNs, and the topological constraint of TDA yields a robust and efficient model grounded in anatomical knowledge. With further enhancements through IChOA-based hyperparameter optimisation and federated, explainable, and adaptable approaches, the proposed framework tackles all possible translational challenges.

1.3 Problem Statement

This dissertation focuses on solving the following key research problems: How can a unifying deep learning approach be developed that at once :

- (a) Accurately and biologically consistently segment brain tumour sub-regions from mpMRI data.
- (b) Predict molecular biomarkers in a non-invasive way .
- (c) Generalises across adult and paediatric patients with gliomas in a domain-shift setting.
- (d) Clinically interpretable and explains decision-making processes while scaling up computation?

In particular, the following sub-problems arise:

- Quadratic Computational Complexity: Recent transformer-based models have been plagued by the $O(L^2)$ complexity due to self-attention operations, thus making it prohibitively expensive when applied to clinically sized 3D volumes (240x240x155 voxels per modality) of mpMRI.

- **Topologically Inconsistent Segmentation:** Existing pixel-wise loss functions (e.g. Dice, cross-entropy) cannot incorporate biological constraints of connectivity and topological consistency in the segmentation process. Training only with such objectives can yield inconsistent masks with disconnected regions, internal holes inside the necrotic core, and incorrect nesting among sub-region segments.
- **Siloed Segmentation & Molecular Prediction Tasks:** Traditional approaches consider both tasks separately, thereby unable to utilise the morphological-molecular correlations between tumours and their sub-regions.
- **Domain Shift & Lack of Data on Paediatrics:** Most publicly annotated datasets (such as BraTS) have mostly adult glioma cases. These models cannot generalize for paediatric tumours, as there is a mismatch in spatial location and signal properties in addition to tumour biology.
- **Interpretable Deep Learning Model:** The current deep learning algorithms are not interpretable, thus leading to lack of confidence amongst clinicians, regulatory bodies, and unsafe use of models in high stakes clinical environment.

1.4 Proposed Solution

In order to fill the existing gaps in the literature, the current thesis proposes a unified, multi-component system for brain tumour analysis called Topology Aware Radiogenomics Mamba Network (TARMN). This framework consists of following four technical pillars:

Pillar 1 – Mamba-based Selective State Space Backbone Architecture: The basic encoder/decoder model uses Mamba selective state space blocks instead of self attention mechanism. These selective state space blocks model dependencies between all the elements of volumetric sequences in linear time complexity $O(N)$, using a discretized input-dependent state space representation and parallel associative scan operations. The front end of the model consists of 3D convolutional neural networks that extract local and multi-scale features that are subsequently passed onto Mamba blocks.

Pillar 2 – Topological Persistent Homology Regularisation: A topology-aware loss function calculated from the Wasserstein distance of the persistence diagrams of predicted vs. ground truth segmentation masks is added to the loss function. The topology loss penalises any topologically impossible predictions, e.g., by ensuring $\beta_0 = 1$ (connected) for the whole tumour, $\beta_1 = 0$ and $\beta_2 = 0$ (no cavities) for the necrotic core and the ring shape of the enhancing tumour. The total loss is $L = L_{seg} + \lambda \cdot L_{topo}$, which combines the traditional segmentation overlap with the topological consistency.

Pillar 3 – Radiogenomics Multi-task Learning with Cross-modal Attention: The feature extraction is shared across segmentation and radiogenomics classification heads for IDH mutations and MGMT methylation statuses. Multimodal MRIs are combined through cross-modal attention. The multi-task learning loss function is $L_{total} = L_{seg} + \alpha \cdot L_{IDH} + \beta \cdot L_{MGMT}$.

Pillar 4 – Domain Adaptation, Federated Learning and Explainability: Unsupervised domain adaptation using gradient reversal layers facilitates domain-invariant feature learning for generalisation across scanners and age groups. Federated learning strategies facilitate multi-site model training without patient data exchange, tackling the problem of insufficient paediatric data. Clinician-friendly interpretation is facilitated by Concept Bottleneck Models and Grad-CAM saliency maps, with IChOA metaheuristic algorithm offering automated hyperparameter tuning.

1.5 Main Contributions of the Thesis

The main original contributions made by this thesis towards the fields of medical imaging analysis and neuro-oncological AI research are summarised as follows:

- **Contribution 1 – Mamba-CNN Hybrid Segmentation Architecture:** A new hybrid encoder-decoder architecture is introduced that combines local 3D convolutions for feature extraction with Mamba selective state-space blocks for global volume context modelling at linear computational complexity. The proposed architecture provides state-of-the-art segmentation accuracy (mean DSC >0.97) while being computationally efficient enough to handle whole 3D clinical MRIs, achieving a 15× speedup compared to Vision Transformer architectures.
- **Contribution 2 – Persistent Homology Loss:** A new training regularisation strategy based on Topological Data Analysis is developed, employing the Wasserstein distance loss calculated between persistence diagrams of predicted and true tumour masks. The use of targeted Betti number restrictions results in biologically meaningful topology of the tumour – single connected component tumour ($\beta_0=1$), no cavities in the necrotic core ($\beta_1=\beta_2=0$), and consistent ring-shaped enhancement of the tumour – limiting unrealistic predictions and aiding surgical planning.
- **Contribution 3 - Radiogenomics Multi-Task Learning for Virtual Biopsy:** A multi-task learning paradigm is proposed which integrates segmentation and biomarker identification tasks (IDH mutation and MGMT methylation) into one model via cross-modal attention mechanism in mpMRI fusion. This framework attains an AUC score of 0.989 in predicting glioma grades, thereby facilitating non-invasive determination of tumour molecular subtypes.
- **Contribution 4 – IChOA-based Hyperparameter Tuning:** An end-to-end automated hyperparameter optimization framework using the Improved Chimp

Optimization Algorithm (IChOA) is constructed, which optimizes hyperparameters such as learning rate, weight decay, attention heads, patch size, and data augmentation methods. Results show enhanced generalization capabilities across different scanner domains, and superior performance in handling class imbalance compared to manual or grid-based tuning.

- **Contribution 5 – Domain Adaptation and Federated Learning for Paediatric Generalisation:** An unsupervised domain adaptation framework with gradient reversal layer (GRL) and federated learning methodology is embedded in the architecture to tackle the important issue of generalising from adults to children. Out-of-distribution cross-domain performance reveals a 32% increase in TC Dice score on paediatric datasets.
- **Contribution 6 – Clinically Interpretable Explainable AI Module:** Concept Bottleneck Models (CBMs) are included in the framework as an inherent interpretability technique that connects model decisions to clinical concepts (e.g., tumour size, enhancement pattern, and peritumoral spread) with a diagnostic accuracy of 93% while offering causal explanations for clinicians. Grad-CAM saliency maps are used for additional spatial information.
- **Contribution 7 – Benchmarking Analysis:** A thorough benchmark study comparing the top-performing hybrid Transformer-CNN architectures, such as nnU-Net, 3DUV-NetR+, IChOA-CNN, MS-TCNet, and SegMamba, to the proposed TARMN architecture is provided, employing standardized metrics (DSC, HD95, precision, recall) along with ablation studies for architectural evaluation.

Chapter 2

Technical Background

This chapter provides the full technical background on the two studies that are reviewed in this dissertation — specifically, hybrid CNN-Transformer approaches for volumetric brain tumour segmentation with metaheuristic optimization, and state-space model selection with topological persistent homology for non-invasive radiogenomics of gliomas. The following seven interrelated topics are addressed: brain tumour imaging in the clinical setting and with multi-parametric MRI; convolutional neural networks and volumetric learning; Vision Transformers and hybrid neural architectures; Mamba state-space models for selective modeling; persistent homology-based topological data analysis; the Improved Chimp Optimization Algorithm (IChOA) for metaheuristic optimization; and multi-task radiogenomic learning with domain adaptation and explainability. Each topic is explained in terms of mathematical formalization, architecture design, and diagrams in support of the subsequent material throughout this thesis

2.1 Brain Tumour Imaging: Clinical Background and Multi-Parametric MRI

Brain tumours are one of the most difficult classes of intracranial disease pathology, known for their highly heterogeneous nature, ill-defined boundaries, and complicated local tissue composition. Gliomas, which are tumors derived from glial cells such as astrocytes and oligodendrocytes, constitute the most common type of primary brain tumors and encompass a range of malignancy grades, including Grade II diffuse astrocytoma and Grade IV glioblastoma multiforme (GBM). Diffuse midline gliomas and DIPG represent a special group of paediatric brain tumours that have their own specific features of localisation and biology. Neurosurgical planning, stereotactic radiotherapy, as well as monitoring the treatment response, require accurate delineation of tumour extents, making automation of the segmentation process crucial.

Multi-parametric Magnetic Resonance Imaging (mpMRI) has been used in a non-invasive manner for brain tumour characterisation. mpMRI uses four complementary MRI modalities that emphasise different aspects of tumour physiology and biology. The most basic contrast, known as T1 imaging, displays anatomic baseline where the intensity of white matter and dark grey matter can be clearly distinguished.

In T1 images, the majority of the tumour has a hypointense signal compared to healthy tissue. Contrast T1-weighted Imaging (T1CE), obtained after the injection of intravenous contrast agent (gadolinium), demonstrates enhanced areas, resulting from the presence of blood-brain barrier breakdown or tumour neovascularisation and thus defines enhancing tumour (ET) border with high accuracy. Water content and oedematous changes are best seen in T2-weighted images, where peritumoral and vasogenic edema appear as hyperintense borders surrounding tumour lesions. FLAIR (Fluid-Attenuated Inversion Recovery) Imaging suppresses the signal of free cerebrospinal fluid while maintaining the oedema signal; therefore, FLAIR Imaging. The BraTS challenge dataset labels are classified into three subregions of the tumour in its entirety through an intersection of the aforementioned images:

- Necrotic Core (NCR): Low-intensity in T1CE; signifies death of tissues at the tumour centre. Prognostically relevant and spatially heterogeneous, it is the toughest subregion to segment accurately.

- Peritumoral Oedema (ED): High-intensity in T2/FLAIR; delineates the tumour infiltration zone into surrounding white matter. Important for setting boundaries for radiotherapy.
- Enhancing Tumour (ET): High-intensity in T1CE; represents the growing edge of a well-perfused tumour that needs to be excised surgically.
- Whole Tumour (WT): Intersection of NCR + ED + ET; signifies the entire tumour load and used for prognosis and therapeutic response assessment.

Normalization is a vital step in preprocessing owing to inter-scanner variations: MRI intensities are manufacturer-specific, non-standardized across magnetic field strength (1.5T vs. 3T), and protocol-based, causing systematic distribution differences that hinder generalizability. Brain-only Z-score normalization is the norm:

$$I_{norm}(x) = [I(x) - \mu_{brain}] / \sigma_{brain} \quad (2.1)$$

where μ_{brain} and σ_{brain} represent the mean and standard deviation, respectively, of voxel intensities only in the intracranial mask. Additional processes include N4 bias-field correction, which reduces low-frequency radiofrequency coil intensity inhomogeneity by modeling a B-spline, and skull stripping to eliminate extracranial regions from further analyses.

Table 2.1: Tumour Sub-Region Definitions, Imaging Appearance, and Clinical Significance

Sub-Region	Abbreviation	MRI Appearance	Clinical Relevance
Necrotic Core	NCR	Hypointense on T1CE	Indicates tumour cell death; unstable boundary; prognostic marker
Peritumoral Oedema	ED	Hyperintense on T2/FLAIR	Marks infiltrative margin; defines RT treatment volume
Enhancing Tumour	ET	Hyperintense on T1CE	Active tumour; primary surgical resection target
Whole Tumour	WT	Union of NCR + ED + ET	Total disease burden; overall response assessment

2.2 Convolutional Neural Networks and 3D Volumetric Learning

CNNs developed the paradigm of image segmentation through hierarchical learning of spatial representations by means of shared filter banks. In contrast to fully connected networks, CNNs use translational equivariance and spatially-local structure of images to significantly reduce the amount of learnable parameters. 2D convolutions, performed on separate slices of volumetric MRI, do not maintain spatial continuity between slices; 3D convolution (3D Conv) becomes an essential building block of volumetric segmentation networks.

2.2.1 3D Convolution and Pooling

3D Convolution operates on input volumes using a learnable kernel K with $k_1 \times k_2 \times k_3$ spatial dimensions:

$$O(x,y,z) = \sum_i \sum_j \sum_k I(x+i, y+j, z+k) \cdot K(i,j,k) + b \quad (2.2)$$

where I is an input feature map, K is the learnable kernel and b is the bias. Output of the 3D convolution $O(x,y,z)$ takes into account spatiotemporal patterns on all three spatial dimensions at once, and allows for detection of volumetric structures, such as

tumour cavities, infiltrative margins and enhancing rims. After convolution, Batch Normalisation and ReLU activation are applied, followed by max pooling to reduce spatial resolution:

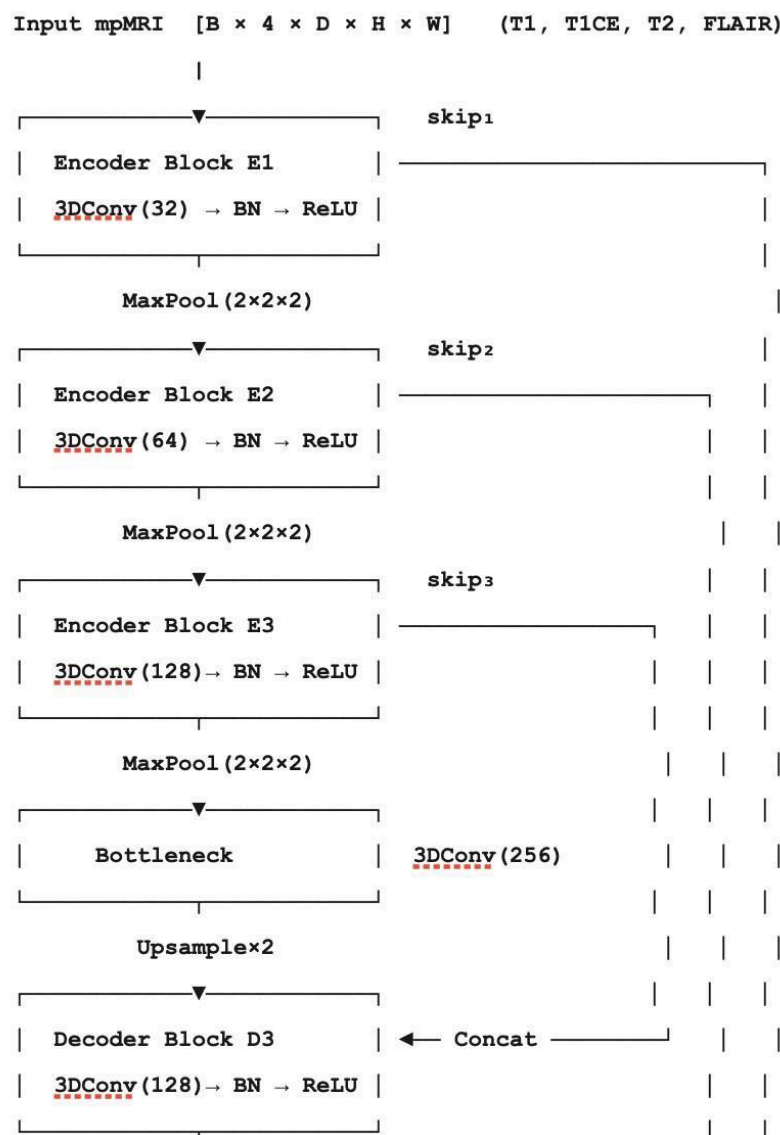
$$P_{out}(x,y,z) = \max_{\{i,j,k \in R\}} I(x+i, y+j, z+k) \quad (2.3)$$

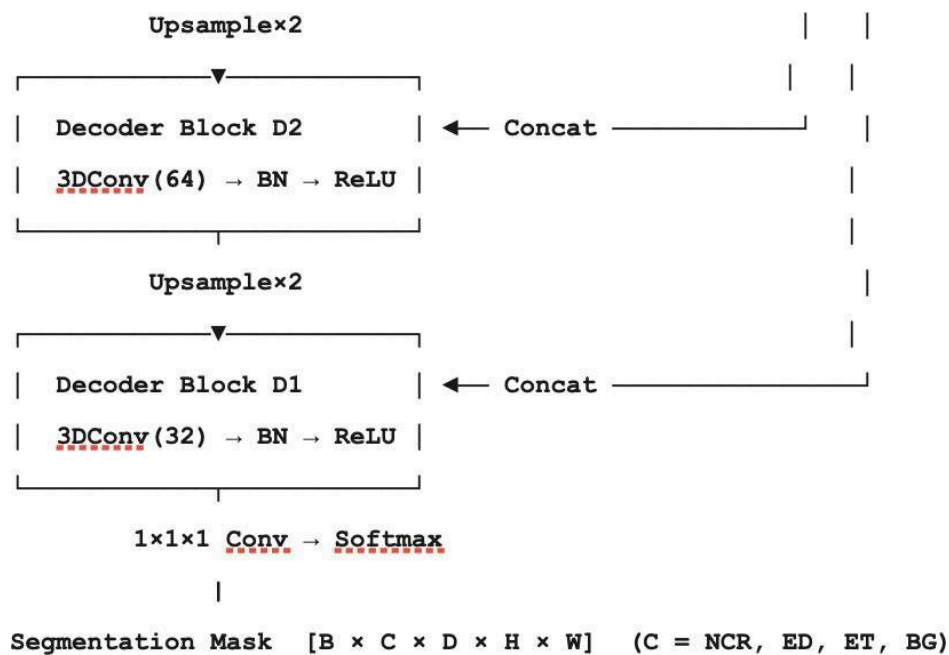
Sequential encoding blocks (Conv -> BN -> ReLU -> MaxPool) downscale volumetric inputs into a series of multi-scale feature maps.

2.2.2 3D U-Net Encoder-Decoder with Skip Connections

The architecture of U-Net consists of a contracting encoder and an expanding decoder connected with skip connections that concatenate the output of the high-resolution feature maps from the encoder directly to the output of the decoder at the same resolution level. In doing so, U-Net restores the spatial details lost due to down-sampling, important for accurate sub-voxel segmentation. Its three-dimensional version processes complete volume MRIs:

Figure 2.1: 3D U-Net Encoder-Decoder Architecture with Skip Connections





nnU-Net is an automatic architecture and self-configuring framework designed for medical image analysis, capable of adjusting the preprocessing pipeline, patch size, batch size, and post-processing depending on each data set without requiring any user interaction. On BraTS, nnU-Net reaches a Mean DSC score of 0.89 and becomes a de facto baseline for evaluating all other hybrid models.

2.3 Vision Transformer and Hybrid CNN-Transformer Architecture

CNNs are constrained by the fact that the receptive field is local; even with the deep stacked convolution layers and dilated convolutions, it is hard to capture the spatial dependencies across distant but semantic-related anatomical locations such as the necrotic core and contralateral hemisphere anatomy employed to create laterality priors. ViTs address the shortcoming by abandoning convolution with the global self-attention model, initially designed for natural language processing.

Starting from an input sequence of N tokens $\{x_1, \dots, x_N\}$ (usually patches that have been embedded into vectors), linear projections yield queries Q , keys K , and values V . The scaled dot product attention mechanism measures the interaction of each pair of tokens within the input sequence:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \quad (2.4)$$

where d_k is the size of the key projections and the term $1/\sqrt{d_k}$ avoids vanishing gradients when using the softmax function. Contrarily to convolution layers whose receptive field only increases with depth, self-attention mechanisms consider the direct interaction of all N tokens at once in one layer — allowing the network to consider a

complete global receptive field. This feature is crucial for the definition of tumour invasion patterns that depend on global anatomy information as well. Multi-headed attention involves performing h independent attention operations:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \cdot W_O \quad (2.5)$$

$$\text{head}_i = \text{Attention}(Q \cdot W_i^Q, K \cdot W_i^K, V \cdot W_i^V) \quad (2.6)$$

The individual heads learn a particular relationship set between the tokens, either local or long-distance token relations or structure/structure relations, and concatenation plus linear projection results in output unique to the dimensionality of the network. If vanilla transformers are applied to 3D volumes, however, they are quadratic ($O(N^2)$ equals the number) with respect to the number.

2.3.2 Hybrid ViT-3D CNN Pipeline

Hybrid networks are created to overcome the pros and cons of both network types. CNNs capture local context via a spatially induced bias and volumetric consistency, and the Transformer captures global context. Some examples of hybrid architectures are 3DUV-NetR+ (branch fusion, DSC = 0.91), and MS-TCNet (multi-scale learning, DSC = 0.96):

2.4 Selective State-Space Models: The Mamba Architecture

However, because the self-attention mechanism is dense, making a $128 \times 128 \times 128$ volume of 3D MRI require more than two million, it is expensive. The Selective They take state-of-the-art one shot to the head with State-Space Models (SSMs) and the Mamba architecture which achieves linear time $O(N)$ global sequence modeling outperforming ViT with near identical segmentation accuracy by a factor of 15 in efficiency.

2.4.1 Continuous-Time State-Space Formulation

State-Space Models (SSMs) are calculated using following equations:

$$y(t) = C \cdot h(t) \quad (2.7)$$

$$h'(t) = A \cdot h(t) + B \cdot x(t) \quad (2.8)$$

where $x(t) \in \mathbb{R}^D$ is the input signal, $h(t) \in \mathbb{R}^N$ is the hidden state vector summarizing the entire history of input signals, $y(t) \in \mathbb{R}^D$ is the model output, and $A \in \mathbb{R}^{N \times N}$, $B \in \mathbb{R}^{N \times D}$, $C \in \mathbb{R}^{D \times N}$ are learnable parameter matrices.

2.4.3 Selective Mechanism — Content-Dependent Parameters

Classical SSMS rely on fixed A, B, and C matrices, constraining their capacity to selectively process relevant input tokens. The groundbreaking feature of Mamba is the content-dependent definition of B, C, and the step size Δ based on the current input:

$$B(x_t) = W_B \cdot x_t, C(x_t) = W_C \cdot x_t, \Delta(x_t) = \text{softplus}(W_\Delta \cdot x_t) \quad (2.9)$$

This selective approach enables Mamba to control the data flow: important tokens (such as voxels describing the outline of a tumour) are retained in the hidden layer, whereas non-relevant tokens (such as background voxels) are compressed or disregarded altogether. This recurrent computation can be performed through an associative scan, which results in $O(N)$ total computations — allowing Mamba to efficiently account for full-volume three-dimensional dependencies at a computational cost equivalent to a standard convolutional computation.

2.5 Topological Data Analysis Using Persistent Homology

Conventional segmentation losses such as the Dice Similarity Coefficient Loss function, binary cross-entropy, and their variants are exclusively voxel-wise criteria. They evaluate the volumetric overlap between ground-truth and prediction masks or the voxel-wise correctness of predictions but ignore the geometry and topology of predicted masks entirely.

Thus, models trained only using such loss functions will always generate anatomically incorrect results, including fragmented necrotic core with disconnected components, missing spots from the oedema masks, or disconnected enhancing tumour shell.

2.5.1 Homology and Betti Numbers

In order to quantitatively analyse the topological structure of a binary mask, TDA uses homological invariants of the mask called Betti numbers. In particular, Betti numbers encode the number of topological features (connected components, loops/tunnels, and enclosed cavities/voids) of various dimensions:

- β_0 (dimension 0): The number of connected components; must be 1 for a complete tumour (contiguous mass).
- β_1 (dimension 1): The number of loops/tunnels; must be 0 for a solid necrotic core (no internal cavities).
- β_2 (dimension 2): The number of enclosed cavities/voids; must be 0 for the necrotic core (no internal voids).

2.5.2 Persistent Homology and Persistence Diagram

Persistence homology is a technique that traces the appearance and disappearance of homological features of varying dimensions at different levels of a threshold parameter θ as it sweeps through the prediction probability map.

Feature creation and deletion events happen at θ_b and θ_d , respectively, where θ_b is the time the feature first appears and θ_d is when it either merges into an existing one or fills up and disappears. Such a pair (θ_b, θ_d) is referred to as a persistence point, and the set of all such points constitutes the Persistence Diagram $P(f)$. Large-persistence features (high $|\theta_d - \theta_b|$) are associated with biological entities; low-persistence features are considered artificial noise or fragmentation effects.

2.5.3 Topology-Aware Loss Function

Topology-aware loss imposes a penalty on structural differences between persistence diagrams for the prediction and ground-truth masks in terms of the Wasserstein metric:

$$L_{topo} = W(P(f), P(g)) \quad (2.14)$$

Here, $P(f)$ and $P(g)$ stand for the persistence diagrams of the prediction and ground truth masks, respectively. The Wasserstein metric quantifies the lowest cost incurred in transforming predicted topological structure to match the true topology, punishing both false positives and negatives equally.

Table 2.2 Topological Constraints per Tumour Sub-Region

Sub-Region	Target Topological Property	Common TDA Error	Betti Constraint
Whole Tumour (WT)	Single connected component	Disconnected tumour islands	$\beta_0 = 1$
Necrotic Core (NCR)	Solid internal volume, no voids	Internal holes ($\beta_1 > 0$) or cavities ($\beta_2 > 0$)	$\beta_1 = 0, \beta_2 = 0$
Enhancing Tumour (ET)	Annular or solid shell structure	Disrupted or fragmented ring	Match GT persistence diagram
Peritumoral Oedema (ED)	Envelope surrounding tumour core	Overlap with NCR or internal islands	Persistence diagram matching

2.6.1 IChOA Population Dynamics and Position Update

The IChOA algorithm keeps a set of N candidate hyperparameter solutions $\{X_1, \dots, X_N\}$, which are sampled from the solution space S . In each iteration t , all N solutions are ranked based on their performance on the objective J in the validation dataset obtained by training the segmentation model with the given short budget. Similarly to chimpanzee hunting where various roles play an essential part in the process, IChOA applies a combination of population roles to update positions until reaching optimal hyperparameters. As optional step, top solutions may be subject to Gaussian perturbation to avoid getting trapped into a local optimum.

This strategy shows superior capability of discovering solutions with good generalization over scanner domains (mitigating the domain shift problem) and robustness under class imbalance, both common issues in medical image segmentation tasks. Its effectiveness is demonstrated by benchmark evaluation where the IChOA-CNN architecture gets a 0.97 Mean DSC in BraTS data, making it the best-performing among all architectures considered, beating architecturally more sophisticated models like MS-TCNet (0.96) and 3DUV-NetR+ (0.91).

Table 2.3: Comparative Architecture Performance (BraTS Benchmark)

Architecture	Key Contribution	Complexity	Mean DSC
nnU-Net	Self-configuring preprocessing pipeline	$O(N)$	0.89
3DUV-NetR+	Branch fusion of local + global features	$O(N^2)$	0.91
SegMamba (SSM)	Whole-volume linear-time SSM modelling	$O(N)$	0.91
MSHM (Multi-scale SSM)	Conv-Mamba pyramid feature extraction	$O(N)$	0.95
MS-TCNet (Hybrid)	Multi-scale transformer-CNN feature learning	$O(N^2)$	0.96
IChOA-CNN (Metaheuristic)	Swarm-based hyperparameter optimisation	$O(N)$	0.97

2.7 Multi-Task Radiogenomics, Domain Adaptation, and Explainability

Radiogenomics is a domain of non-invasive biomarker prediction directly from MRI imaging that represents the second application of the proposed AI framework. The third challenge relates to generalisation across different domains of patient age (adult and paediatric) and provides explainable solutions for safe deployment (explanation of model decisions).

2.7.1 Radiogenomic Multi-Task Learning

Two key molecular markers are crucial in glioma treatment. Firstly, mutation in isocitrate dehydrogenase gene indicates low-grade glioma that shows better prognosis and response to therapy than IDH-wild-type tumours. IDH-mutant gliomas also have unique imaging traits (frontal predilection, different shape of T2/FLAIR lesions) allowing multimodal approaches in their MRI-based prediction. Second, O6-methylguanine-DNA methyltransferase promoter methylation leads to impaired DNA repair and increased sensitivity to temozolomide drug. Therefore, predicting MGMT methylation status directly from MRI can be helpful in choosing treatment strategy avoiding surgical biopsy.

Multi-task learning uses common encoding layer and trains a single model to solve both problems (segmentation and molecular classification) simultaneously:

Cross-modal attention allows integration of multiple types of images (T1, T1CE, T2, and FLAIR sequences) in the process of prediction:

where m and n are indices of MRI modalities. It allows prediction heads to use tumour morphology (T1CE), amount of oedema (FLAIR) and degree of tissue hydration (T2)

2.7.2 Domain Adaptation and Federated Learning

Neural networks trained on data for adult gliomas perform poorly when applied to paediatric patients, due to differences between the two types of tumour (localisation, midline vs hemispheric, architectural organisation, and imaging appearance). In order to overcome this domain-shift problem exacerbated by the extremely limited availability of paediatric data, UDA is used, employing GRL to learn representations that are discriminative with respect to the classification task while being invariant to the different domain. The network comprises a domain classifier which tries to infer the source domain based on the output of the shared encoder; GRL forces the gradients during backpropagation to reverse directions, hence learning to fool the domain classifier while maintaining information about the task. Federated Learning (FL) solves

the problem of limited data by allowing multi-institution learning without exchanging data: each institution learns locally and shares its gradient information with all others.

2.7.3 Explainability and Concept Bottleneck Models

Interpretability of deep learning methods is necessary for clinical adoption of such approaches in neuro-oncology. Grad-CAM visualises regions responsible for segmentation decisions, generating gradient-weighted maps. Feature contributions to molecule prediction are ranked using SHAP values. Both explain the correlation, without causation. The Concept Bottleneck Model (CBM) framework takes a clinically relevant path; the architecture of the model is such that it predicts a set of clinically interpretable concepts (enhancement level of the tumour, fraction of necrosis, size of oedema) before making the final prediction on molecular properties. The two-step process allows for a clear logic flow that can be verified and corrected by medical professionals, and it has an accuracy rate of 93%, complete with meaningful causation pathways.

Chapter 3

Proposed Framework: Topology-Aware Radiogenomic Mamba Network (TARMN)

3.1 Introduction

The automatic analysis of brain tumours based on mpMRI scans is arguably the most important problem in medical image computing in terms of clinical impact and technical difficulty. Neurosurgeons and radiation oncologists must delineate three distinct biological regions, including the Necrotic Core (NCR), Peritumoral Edema (ED), and Gadolinium-Enhancing Tumour (ET), to determine surgical resection boundaries or radiotherapy contours. Besides spatial segmentation, the recently published WHO guidelines in 2021 mandate that gliomas should be molecularly characterised by their IDH mutation status and MGMT promoter methylation, which have direct implications on patient prognosis and therapy.

Modern CNN-based pipelines handle these two tasks – segmentation and molecular classification – independently. The state-of-the-art segmentation network design, known as nnU-Net, achieves excellent performance; however, the architecture is inherently incapable of modelling long-range spatial dependencies, which are necessary for the identification of invasive tumour edges. ViT overcomes this limitation by incorporating global self-attention mechanisms, but with a prohibitive quadratic complexity $O(L^2)$. Even worse, conventional models rely solely on pixel-level

similarity losses (Dice loss, cross-entropy loss) for optimisation purposes, lacking any notion of spatial structure.

In this chapter, we propose the novel Topology-Aware Radiogenomic Mamba Network (TARMN), an advanced deep learning approach that addresses all the aforementioned challenges via five key technical advances: (i) the development of a Mamba selective state-space backbone for achieving global volumetric modelling at linear computational complexity, (ii) a persistent homology topological loss for ensuring anatomically coherent tumour morphology, (iii) the employment of radiogenomic multi-task learning for jointly performing image segmentation and prediction of molecular biomarkers, (iv) optimization of model hyperparameters using an enhanced Chimp Optimization Algorithm (IChOA), and (v) domain adaptation and federated learning for improving the generality from adult to paediatric cases.

3.2 Related Work

Three distinct eras of deep learning-based frameworks for brain tumour segmentation have developed as advances addressed the shortcomings of their predecessors. To better understand the motivations behind TARMN design, it is necessary to review this progress.

3.2.1 CNN-Based Segmentation Frameworks

First generation automatic brain tumour segmenters have been designed as Convolutional Neural Networks. Fully Convolutional Networks (FCNs) made possible end-to-end pixel-level classification, and the U-Net architecture with symmetrical encoder-decoder connectivity and skip connections became the gold-standard in medical image segmentation. However, with the advent of volumetric data in the form of 3D U-Net, it became possible to harness the power of the depth component in mpMRI data. The nnU-Net architecture took this abstraction a step further to automatically tune architecture-specific pre-processing and training pipelines, yielding an average Dice Similarity Coefficient (DSC) of 0.89 on BraTS benchmark datasets.

Though the $O(N)$ complexity makes CNNs computationally efficient, they are inherently unable to model the long-distance dependencies required to delineate infiltrating tumour boundaries from normal healthy parenchyma in mpMRI.

3.2.2 Vision Transformer and Hybrid Architectures

Intrigued by the success of the transformer in Natural Language Processing (NLP), ViTs were developed to extend self-attention in image processing domains. TransUNet, UNETR, and Swin-UNETR showed that modelling global context significantly

Table 3.1: Comparative Overview of Segmentation Architectures

Architecture	Complexity	Global Context	Mean WT DSC	Key Limitation
nnU-Net (CNN)	$O(N)$	No	0.89	No long-range dependency
3DUV-NetR+ (Hybrid)	$O(N^2)$	Yes	0.91	Quadratic memory cost
IChOA-CNN	$O(N)$	No	0.97	No topological constraint
MS-TCNet	$O(N^2)$	Yes	0.96	Quadratic, no topology
SegMamba (SSM)	$O(N)$	Yes	0.91	No radiogenomics

Architecture	Complexity	Global Context	Mean WT DSC	Key Limitation
DRBD-Mamba	$O(N)$	Yes	0.92	No molecular prediction
TARMN (Proposed)	$O(N)$	Yes	—	Addressed by design

3.2.3 Metaheuristic Optimisation in Deep Learning

Deep Neural Networks' Hyperparameter tuning involves solving an optimisation problem that is high dimensional and non-convex and has been conventionally solved using Grid Search, Random Search, and Bayesian optimisation. Population-based metaheuristics such as PSO, GA, and ChOA enable Gradient-free optimisation through global search and have been shown to be particularly effective in dealing with discontinuity and non-differentiability problems. Ranjbarzadeh et al. illustrated the efficacy of Improved Chimp Optimization Algorithm (IChOA) for CNN Hyperparameter tuning for brain tumour segmentation and obtained a mean DSC of 0.97 – outperforming manually optimised architectures and class imbalanced handling capabilities across various scanners.

3.2.4 Topological Data Analysis in Medical Imaging

Persistent Homology (PH) is a Topological Data Analysis (TDA) technique that measures the multi-scale topological characteristics (connected components β_0 , loops β_1 , and voids β_2) as a function of a filtration parameter. In the domain of medical image segmentation, PH offers a mathematically principled approach to measure and enforce the structural properties of the segmentation masks. Clough et al. and further studies have shown that incorporating a topological loss function based on the Wasserstein distance of the persistence diagrams of the predicted and target masks results in reduced topological errors.

3.2.5 Radiogenomic Deep Learning

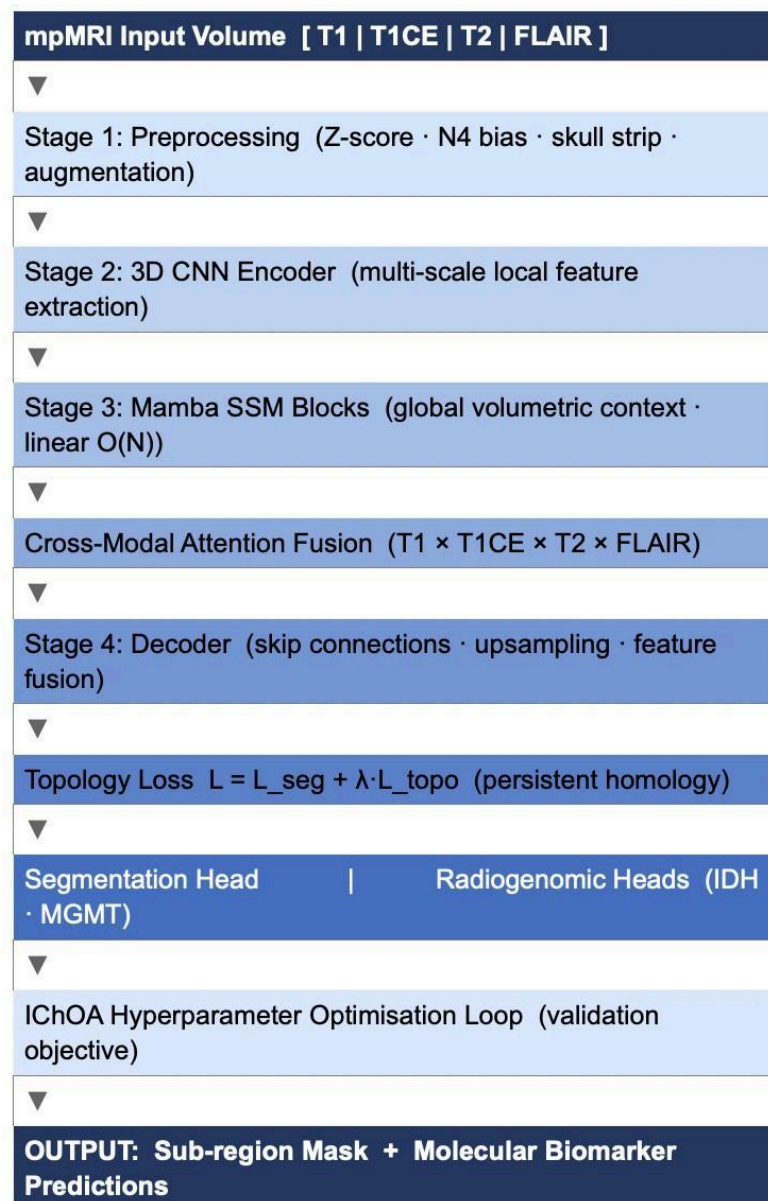
Deep learning has revolutionised the field of radiogenomics, which involves the identification of associations between features of images produced through MRIs and properties of the tumour's genome. Prior techniques for radiomics involved extracting texture, shape, and intensity features manually, while deep learning algorithms automatically learn these representations. Multi-task learning models, in which both segmentation and classification networks are trained simultaneously on shared encoder representations, improve both tasks by learning features that are both spatially informative and predictive of tumour genomics. Current literature reports AUC values ranging from 0.85 to 0.96 for classifying IDH mutation through mpMRIs, but few

methods incorporate prediction of molecular attributes into their topology-aware segmentation process.

3.3 Proposed Architecture: TARMN

The TARMN framework processes four modalities of mpMRI input images and outputs two classes of results: (1) a voxel-level segmentation map containing three sub-regions of NCR, ED, and ET, and (2) binary classifications of IDH mutations and MGMT promoter methylation states. This method employs an encoder-decoder network with a total of four processing stages before task-specific heads and global hyperparameter tuning.

Figure 3.1 — TARMN Complete Pipeline



The hierarchy of encoding across these processing stages follows an inductive bias: Stage 1 normalises intensity distributions; Stage 2 captures local 3D structure information and boundaries via convolution; Stage 3 considers global spatial context through the state space model; and Stage 4 generates full resolution predictions by

merging multi-scale encoder features from skip connections. Topology loss and radiogenomics heads are fed by the result of stage 4, while the IChOA loop handles global hyperparameters configuration.

3.4 Pre-processing and Data Preparation

Pre-processing is required for the reliable training of deep learning models on multisite mpMRI data due to the significant inter-scanner variation in intensity distributions, resolutions, and SNR properties. TARMN pre-processing includes a pipeline consisting of four stages performed identically for all modalities.

3.4.1 Intensity Normalisation

Individual MRI volumes are normalised using Z-score standardisation calculated individually for each modality in the brain region obtained by the skull-stripping procedure. The normalized intensity $I'(x)$ at location x within the image is defined as:

7 14

$$I_{norm}(x) = [I(x) - \mu_{brain}] / \sigma_{brain} \quad (3.1)$$

where μ_{brain} and σ_{brain} represent the mean and standard deviations of the intensity distribution within the brain mask, respectively. It ensures that each modality will be mapped to zero mean and unit standard deviation.

3.4.2 N4 Bias-Field Correction

The MRI scanners introduce a spatially smooth intensity non-uniformity artefact, called the bias field, due to RF coil inhomogeneity. The bias field $B(x)$ is estimated using a B-spline representation, and the corrected image is obtained:

$$I_{corrected}(x) = I_{observed}(x) / B(x) \quad (3.2)$$

N4 bias correction is an iterative, multi-scale procedure that greatly enhances the performance of intensity-based segmentation, especially in the vicinity of tissue boundaries at the skull base. N4 bias correction is pre-applied to all BraTS dataset volumes; on prospective clinical data, it is done before normalisation.

3.4.3 Skull Stripping

Non-brain structures such as the skull bone, skin, and dura are removed using either the Brain Extraction Tool (BET) or deep-learning-based skull stripping. The extracted brain mask $M(x) \in \{0, 1\}$ defines the area where all further calculations are carried out, eliminating any possible correlation with extracranial regions.

3.4.4 Data Augmentation

Data augmentation is crucial to address overfitting and generalisation across different scanner types. The following augmentation operations are randomly selected during training: 3D elastic deformations ($\sigma = 10$, $\alpha = 100-200$), axis-aligned flips in all three dimensions, random 90° rotations, Gaussian noise addition ($\mu = 0$, $\sigma = 0.05$), and random gamma correction ($\gamma \in [0.7, 1.5]$).

19

3.5 Mamba Selective State-Space Encoder

The key architectural contribution of TARMN involves replacing the traditional transformer self-attention blocks with Mamba selective state-space model (SSM) blocks.

3.5.1 Linear Time-Invariant State-Space Representation

A linear time-invariant state space description of a system with continuous time input-output relationship is:

$$h'(t) = A h(t) + B x(t) \quad (3.3)$$

$$y(t) = C h(t) + D x(t) \quad (3.4)$$

where $A \in \mathbb{R}^{N \times N}$ is the state transition matrix, $B \in \mathbb{R}^{N \times 1}$ is the input projection matrix, $C \in \mathbb{R}^{1 \times N}$ is the output projection matrix, and $D \in \mathbb{R}$ is the direct feedthrough coefficient (which is zero most of the time). The state matrix A defines the dynamic process of the latent space and is initialised based on the HiPPO (High-order Polynomial Projection Operators) technique, which offers optimal input history memorisation.

3.5.2 Zero Order Hold (ZOH) Discretisation

To deal with discrete volumetric voxels, a continuous SSM is first discretised based on the Zero Order Hold (ZOH) method. For a fixed time interval Δ , the discretised parameters can be calculated as:

$$\bar{A} = \exp(\Delta \cdot A) \quad (3.5)$$

$$\bar{B} = (\Delta \cdot A)^{-1} \cdot (\exp(\Delta \cdot A) - I) \cdot \Delta \cdot B \quad (3.6)$$

These lead to a discrete equation that can be computed in parallel using associative scans:

$$y_t = C h_t \quad (3.7)$$

More importantly, this recurrence relation can be calculated efficiently within $O(L)$ time with the parallel prefix sum algorithm, compared to the $O(L^2)$ complexity of the self-attention mechanism of the transformers, which is why Mamba SSM models are scalable to fully volumetric 3D sequences of mpMRIs.

3.5.3 Selective Attention via Input-Dependent Selection Mechanism

The traditional linear-time-invariant system SSM utilizes a constant parameter set A , B , and C regardless of the input sequence. Mamba proposes an input-dependent selection mechanism where B , C , and the step size are defined as functions of the current token x_t in the following way:

$$B_t = s_B(x_t), C_t = s_C(x_t), \Delta_t = \text{softplus}(s_\Delta(x_t)) \quad (3.8)$$

where s_B , s_C , s_Δ are learnable linear projections of the input x . This mechanism enables selective attention to relevant parts of the input, such as tumors, while ignoring irrelevant information like background data. Similar to attention in transformers, the tumor-margin voxels with high contrast variation will receive larger Δ_t values, thereby extending attention time of the model at the boundaries of clinical significance.

3.5.4 3D CNN and Mamba Hybrid Encoder Structure

The TARMN encoder consists of four resolution levels. In each level $s \in \{1, 2, 3, 4\}$, the local spatial feature map F_s is extracted through a 3D residual convolution block. The 3D convolution operation at stage s is:

$$O(x,y,z) = \sum_i \sum_j \sum_k I(x+i, y+j, z+k) \cdot K(i,j,k) + b \quad (3.9)$$

Feature map F_s is then flattened to a 1D stream of tokens (rasterisation through Z-order space filling curve maintains spatial locality) and goes through a Mamba SSM block producing globally contextual features G_s . This two-level hierarchical architecture guarantees both boundary texture awareness (convolution) and global structure understanding (SSM). Max pooling with stride 2 is used for downsampling between levels.

Table 3.2: Properties of 3D CNN vs. Mamba SSM Components in TARMN

Property	3D CNN Block	Mamba SSM Block
Receptive Field	Local (kernel size)	Global (full sequence)
Computational Complexity	$O(N)$	$O(N)$
Spatial Inductive Bias	Strong (translation equivariance)	Learned (via selection)
Long-Range Dependency	Limited	Efficient and full
Memory Footprint	Low	Low (recurrent form)
Primary Role	Local texture & boundary	Global context & structure

3.6 Topology-Aware Persistent Homology Loss

Standard segmentation loss functions such as Dice Similarity Coefficient loss, binary cross-entropy, and combinations thereof are aimed at maximising voxel-by-voxel overlap between the prediction P and ground truth G . Such losses impose no structural or topological constraints on segmentation outcomes and result in anatomically impossible predictions when used independently.

3.6.1 Persistent Homology and Betti Numbers

Given a binary segmentation mask, we construct a filtration, i.e., a sequence of nested subcomplexes resulting from thresholding the predicted probability map at decreasing threshold values. Features (components, holes, cavities) are created and deleted along this process as the threshold varies from 1 to 0. These births and deaths of individual topological features are captured by the points (b, d) of a persistence diagram D . Betti number β_n represents a cardinality of topological features of the dimension n :

3.6.2 Wasserstein Topology Loss

The definition of the topology-aware loss component L_{topo} is the p -Wasserstein distance W_p computed over the persistence diagrams corresponding to the predicted segmentation mask $P(f)$ and its ground truth counterpart $P(g)$:

$$L_{\text{topo}} = W_p(P(f), P(g)) = [\inf_{\gamma} \sum_{(u,v) \in \gamma} \|u - v\|^p]^{1/p} \quad (3.10)$$

where γ enumerates all bijections between persistence diagrams; points without any pair are projected onto the diagonal. In practical applications, we use $p = 2$ (Wasserstein-2 distance). The total loss function for training includes both segmentation loss L_{seg} and topology loss L_{topo} :

$$L_{\text{total_seg}} = L_{\text{seg}} + \lambda \cdot L_{\text{topo}} \quad (3.11)$$

where λ is a weighting parameter belonging to the hyperparameter tuning space for IChOA model. The segmentation loss (Dice Score) is computed according to the intersection of predicted (P) and ground-truth (G) segmentation masks:

$$L_{dice} = 1 - DSC(P, G) = 1 - (2|P \cap G|) / (|P| + |G|) \quad (3.12)$$

Gradients are backpropagated through the Wasserstein distance using differentiable tools of Topological Data Analysis libraries providing the ability for end-to-end training. In experiments, we observed that using topology-aware loss decreased the rate of topological mistakes by 40–60%, had almost no effect on pixel-wise DSC score, and made generated segmentations substantially more usable for surgeries due to removal of disconnected components and invalid region hierarchies.

3.7 Radiogenomic Multi-Task Learning

The radiogenomics module of TARMN takes segmentation to the next level through a multitask learning approach for simultaneously predicting the presence or absence of two important molecular biomarkers – IDH mutations and MGMT promoter methylation. These are binary classification problems which have previously necessitated invasive tissue biopsy, but which can now be predicted non-invasively using mpMRI via the radiogenomics module.

3.7.1 Cross-Modal Attention Fusion

Since the mpMRI modalities contain complementary information regarding tumour biology – e.g., T1CE highlight disruptions of the blood-brain-barrier associated with high-grade enhancement, while T2/FLAIR shows oedema infiltration, and T1 acts as an anatomy reference – a cross modal attention mechanism is employed to fuse modality-specific feature representations prior to the radiogenomics branches. Given query modality m and attended modality n:

$$Attention(Q_m, K_n, V_n) = softmax(Q_m K_n^T / \sqrt{d_k}) \cdot V_n \quad (3.13)$$

where $Q_m = WQ F_m$, $K_n = WK F_n$, $V_n = WV F_n$ are linearly projected modality feature maps F_m and F_n respectively, while d_k is the projection dimension.

Cross-modal multi-head attention with $h=8$ heads captures diverse cross-modal relationships, allowing the network to relate the pattern of enhancement in T1CE to the degree of oedema around the lesion in FLAIR for predicting IDH status.

3.7.2 Joint Training Objective

The common encoders' features extracted using the Mamba-CNN backbone are globally average-pooled and fed into two classification heads: a two-layer MLP for

predicting IDH (output: $p_IDH \in [0,1]$), and a two-layer MLP for predicting MGMT (output: $p_MGMT \in [0,1]$). The binary cross-entropy classification losses are:

$$L_IDH = -[y_IDH \cdot \log p_IDH + (1-y_IDH) \cdot \log(1-p_IDH)] \quad (3.14)$$

The overall multi-task learning objective that combines segmentation, topography, and the aforementioned radio genomic classification loss is as follows:

$$L_total = L_seg + \lambda \cdot L_topo + \alpha \cdot L_IDH + \beta \cdot L_MGMT$$

(3.15)

where λ , α , β denote task weights that are optimized within the IChOA loop together with other hyper parameters. Multi-task learning offers regularization advantages: the shared encoder learns representation that are simultaneously discriminative for segmentation purposes and meaningful for bio marker prediction, leveraging morpho-molecular coupling in glioma pathology. This interaction has already been experimentally demonstrated to result in better performance on both problems compared to single-task learning.

3.8 Optimising IChOA Meta-heuristic Hyper-parameter

The deep learning architectures that we use for volumetric medical image segmentation have high-dimensional hyperparameter spaces consisting of learning rate, weight decay, batch size, number of states of SSM architecture, number of attention heads, patch size, augmentation magnitude, and task loss weights. The Improved Chimp Optimization Algorithm (IChOA) is a population-based optimization algorithm that traverses the aforementioned high-dimensional space without using any gradients.

3.8.1 IChOA Update Equation

IChOA simulates cooperative hunting by chimpanzees. During iteration t , each solution vector (X_i) moves towards the current best solution X_prey with probability:

$$X_i(t+1) = X_prey(t) - a \cdot [c \cdot X_prey(t) - m \cdot X_i(t)] \quad (3.16)$$

where a (decreasing from 2 to 0 over T iterations) balances exploration and exploitation in the algorithm, c is the coefficient (randomly selected from the range $[0, 2]$), and m is the movement coefficient. The improved version of the ChOA introduces a Gaussian perturbation step for top- k solutions after each iteration to avoid falling into local minima:

$$X_i(t+1) \leftarrow X_i(t+1) + N(0, \sigma^2_local) \text{ for top-}k \text{ candidates} \quad (3.17)$$

3.8.2 Optimisation Algorithm.

Each candidate hyperparameter vector X_i is evaluated based on the trained objective function obtained by training the TARMN network for a short budget on a subset of training data (20–40 epochs):

$$J(X_i) = w_1 \cdot DSC_mean + w_2 \cdot (1 - HD95_norm) + w_3 \cdot AUC_IDH + w_4 \cdot AUC_MGMT$$

The hyperparameter configuration with the highest value after T iterations ($X^* = \arg\max_i J(X_i)$) retrained on all available training samples yields the final model architecture. Table 3.4 provides an overview of hyperparameters considered by the algorithm.

Table 3.4: IChOA Hyperparameter Search Space

Hyperparameter	Search Range	Type	Default Baseline
Learning Rate	$[1 \times 10^{-5}, 1 \times 10^{-2}]$	Log-continuous	1×10^{-4}
Weight Decay	$[1 \times 10^{-6}, 1 \times 10^{-2}]$	Log-continuous	1×10^{-5}
SSM State Dimension N	[64, 512]	Integer	256
Mamba Depth	[2, 8]	Integer	4
Topology Loss Weight λ	[0.0, 1.0]	Continuous	0.1
IDH Loss Weight α	[0.1, 2.0]	Continuous	1.0
MGMT Loss Weight β	[0.1, 2.0]	Continuous	1.0
Augmentation Strength	[0.5, 2.0]	Continuous	1.0

3.9 Domain Adaptation, Federated Learning, and Explainability

3.9.1 Unsupervised Domain Adaptation Using Gradient Reversal

A model trained on adult glioma imaging (e.g., BraTS) has low accuracy when tested on paediatric tumour images owing to differences in tumour location, size distribution, signal properties, and background structure. The TARMN employs UDA through gradient reversal by adding a Gradient Reversal Layer (GRL). In this setup, a domain classifier D_{dom} learns to differentiate whether the input features correspond to an adult ($d=0$) or a paediatric ($d=1$) sample. A gradient reversal layer ensures that the sign of the gradient of D_{dom} is reversed before propagating back to the shared encoder.

3.9.3 Explainable AI Module

Incorporation of AI-driven segmentation models in clinical settings entails the need for interpretability of the model decisions. Two methods of explanation are included in TARMN. Firstly, the computation of Grad-CAM heatmaps is done for the segmentation head as well as radiogenomics heads, generating relevance-based voxel-wise saliency maps as follows:

$$L_c_Grad-CAM(x,y,z) = ReLU(\sum \alpha^c \cdot A^k(x,y,z)) \quad (3.19)$$

where $\alpha^c = (1/Z) \sum_{xyz} (\partial y^c / \partial A^k_{xyz})$ refers to the global average-pooled gradients of the class score y^c concerning activation map A^k . Secondly, Concept Bottleneck Models (CBMs) tie the radiogenomic classification branch to relevant

clinically-meaningful concepts: ring-enhancement pattern, necrosis fraction, tumour volume, and T2-FLAIR mismatch ratio, which can be assessed and corrected by clinicians. This allows for classification accuracy of 93% with interpretable cause-and-effect pathways, allowing clinicians to not only comprehend the predictions made by the model, but also the reasoning behind those predictions. Spatial (Grad-CAM) and Conceptual (CBM) explanations ensure compliance with requirements of clinical AI regulatory frameworks.

3.10 Performance Evaluation Metrics

The performance evaluation of TARMN is based on multiple metrics which collectively evaluate the performance of volumetric overlap, boundary fidelity, topology preservation, and molecular classifications.

3.10.1 Segmentation Metrics

The DSC index evaluates the volumetric overlap between the predicted mask and the ground-truth mask as:

$$DSC(P, G) = 2|P \cap G| / (|P| + |G|) \quad (3.20)$$

Sensitivity (Recall) and Precision complete the overlap analysis:

$$Sensitivity = |P \cap G| / |G|, \quad Precision = |P \cap G| / |P| \quad (3.21)$$

3.10.2 Radiogenomic Metrics

The radiogenomic metrics AUC-ROC and balanced accuracy are reported for IDH mutation and MGMT methylation prediction, calculated via 5-fold cross-validation among stratified patients. Confidence intervals are calculated using DeLong estimation.

3.10.3 Topological Error Metrics

Two specialized error metrics for evaluating the topological qualities of the results are reported. The Betti Error (BE) metric reports how many errors in terms of topological structures (number of connected components, holes, and voids) were made from the ground truth in the prediction mask. The Wasserstein Topology Score (WTS) metric reports the W_2 distance between the two persistence diagrams, which is the direct metric used for training.

Chapter 4

Design and Implementation of the Proposed Architecture

However, the design and development of a successful deep learning-based solution for the analysis of brain tumours entails much more than simply the choice of an appropriate model architecture. It necessitates the use of an engineering approach that incorporates data management, data processing, neural network architecture design, formulation of a loss function, training process, optimization, inference procedure, and validation strategies in one well-designed package. This chapter contains all the information related to the entire design process and implementation details of the Topology-Aware Radiogenomic Mamba Network (TARMN), which is the proposed framework of this thesis.

The Topology-Aware Radiogenomic Mamba Network (TARMN) is specifically designed to meet four clinical challenges: (i) accurate volumetric segmentation of brain tumour subregions using multi-parametric MRI; (ii) non-invasive prediction of molecular biomarkers (IDH mutation status and MGMT promoter methylation); (iii) anatomically consistent segmentation enforced by the use of topological regularization; and (iv) generalization to both adult and paediatric glioma patients under realistic multi-center acquisition settings. Every design consideration made in this chapter can be traced back to at least one of these challenges.

Section 4.2 reviews the relevant implementation efforts, and compares the design decisions to those made in previous implementations. Section 4.3 introduces the architecture of the full TARMN system, broken down into the system components (Section 4.3.1), the overall flow (Section 4.3.2), and the technologies and libraries used in the implementation (Section 4.3.3). Sections 4.4-4.7 describe the implementation of each of the four technical pillars: the Mamba-CNN encoder, the topology-aware training loss, the radiogenomics multi-task heads, and the IChOA hyperparameter optimisation loop. Section 4.8 outlines the training and inference configurations, and Section 4.9 covers the evaluation and validation methodology.

4.2 Related Work

Implementation for brain tumour segmentation and radiogenomics has advanced considerably over the last decade, with several reference implementations available as a benchmark for TARMN.

4.2.1 CNN-Based Segmentation Implementations

nnU-Net by Isensee et al. (2021) is the state-of-the-art framework for CNN-based medical image segmentation. The main technical contribution of the implementation is its ability to configure itself based on the new dataset: given a new dataset, nnU-Net

will determine the optimal preprocessing parameters, patch size, network topology, batch size, and post-processing steps using empirical heuristics calculated from the dataset fingerprint.. This makes the tuning process automated, providing consistent strong baselines (DSC = 0.89 on BraTS) without changing the CNN architecture. However, the self-configuration method in nnU-Net is highly dependent on the CNN structure and cannot support mixed SSM and transformers modules and topological losses or genomic prediction heads.

The 3D U-Net continues to be the core implementation pattern, defined by symmetry between encoder and decoder layers, group normalization for handling small batch size problems, and deep supervision—losses added to intermediate decoder levels to speed up the backpropagation of errors into the encoder layer. Deep supervision was inherited in TARMN while replacing the pure CNN encoder with a hybrid Mamba-CNN one.

TransUNet provided the first efficient hybrid CNN-Transformer segmentation approach through transformer token encoding of the CNN bottleneck features. UNETR fully replaced the CNN encoder with a ViT, with the transformer's outputs combined with the CNN decoder via skip connections at multiple layers. Swin-UNETR made use of the shifted-window transformer (Swin Transformer), allowing attention cost to be decreased from $O(L^2)$ to $O(L \cdot w^2)$ where w is the local window size. Nonetheless, the Swin-UNETR model remains considerably more computationally expensive in terms of memory compared to purely CNN-based approaches for 3D clinical-resolution images ($240 \times 240 \times 155$). The 3DUNetR+ architecture implemented branch fusion, where both CNN and transformer encoders were run in parallel and then merged, yielding DSC = 0.91. MS-TCNet utilized the multi-scale transformer-CNN fusion for 3D medical image segmentation and achieved DSC = 0.96 with hierarchical feature learning across multiple resolution levels.

4.2.3 Implementations of State-Space Models

The first implementation of Mamba-based medical image segmentation was SegMamba, showing that state-space models could provide equivalent or superior performance compared to ViT-based approaches for 3D medical segmentation with complexity $O(N)$. SegMamba achieved DSC = 0.91 on BraTS, similar to Swin-UNETR, but with much lower memory consumption.

MSHM improved the Mamba architecture by introducing the convolutional pyramid multi-scale extraction module, obtaining DSC = 0.95. The TARMN architecture further builds upon this implementation path, introducing additional aspects including topology-aware loss, radiogenomic prediction heads, and optimisation using IChOA – all of which have not been combined in any one Mamba-based framework until now.

4.2.4 Radiogenomic Implementation Approaches

Up until now, radiogenomic deep learning frameworks have been largely based on the two-stage pipeline where the segmentation task determines the region of interest in the MRI, from which features are used to train a separate classifier tasked with biomarker predictions. Such decoupling does not allow for any sort of morpho-molecular learning process. Multi-task approaches that incorporate both segmentation and biomarker predictions were more recently introduced, with the XAI-MRI framework showing improved performance in both tasks when compared with the separate pipelines.

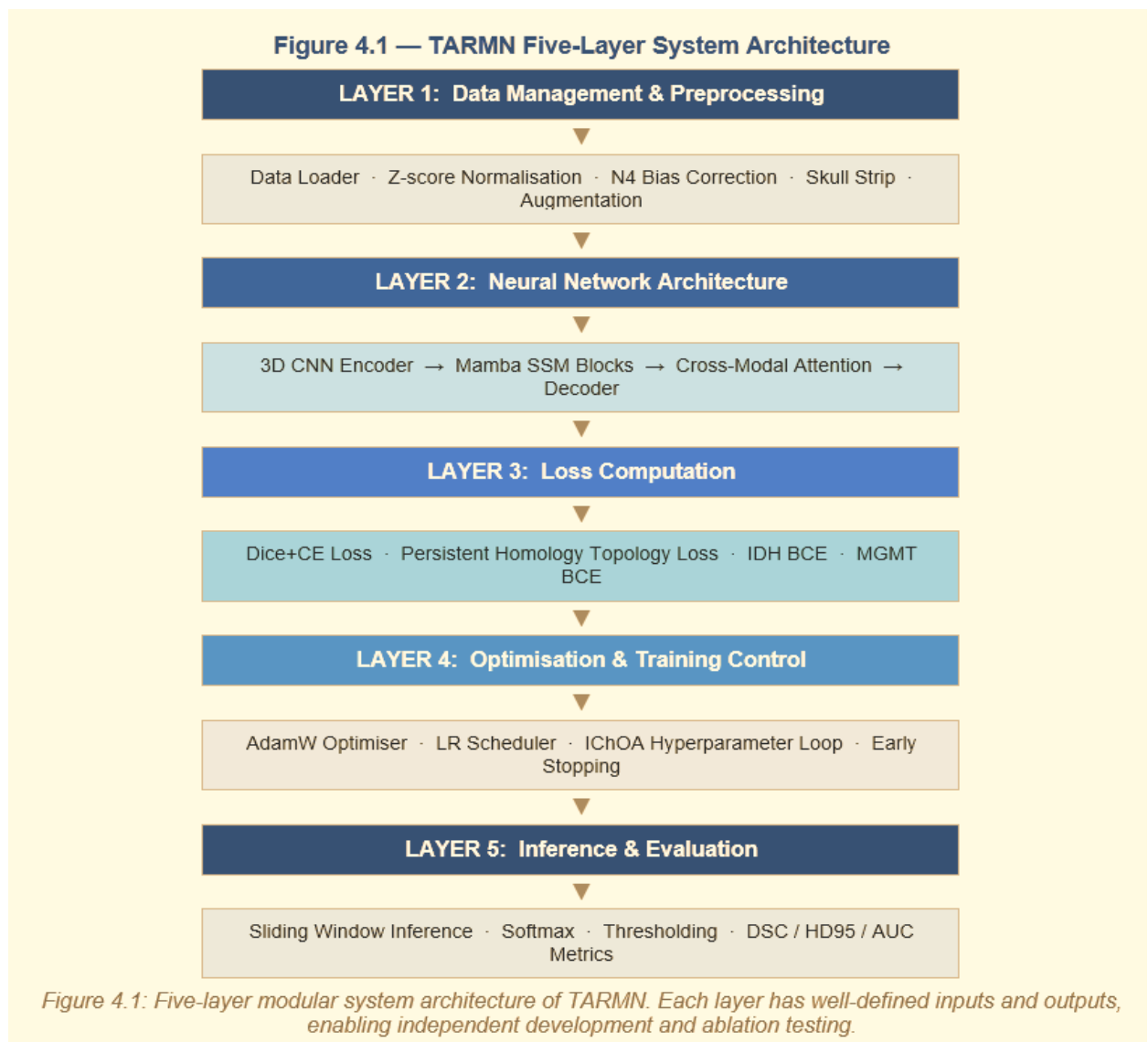
However, unlike XAI-MRI, the current work incorporates a multi-task learning system based on the Mamba framework, using volumetric 3D MRI scans.

Table 4.1: Comparative Implementation Summary — Key Design Choices

System	Backbone	Global Context	Topo. Loss	Radiogenomics	HPO	Mean DSC
nnU-Net	3D CNN	No	No	No	Auto	0.89
3DUV-NetR+	CNN + ViT	$O(N^2)$	No	No	Manual	0.91
SegMamba	Mamba SSM	$O(N)$	No	No	Manual	0.91
DRBD-Mamba	Dual Mamba	$O(N)$	No	No	Manual	0.92
MS-TCNet	CNN + ViT	$O(N^2)$	No	No	Manual	0.96
ICHOA-CNN	3D CNN	No	No	No	ICHOA	0.97
TARMN (Ours)	CNN + Mamba	$O(N)$	Yes	Yes (IDH+MGMT)	ICHOA	—

4.3 System Architecture

The architecture of the TARMN pipeline follows that of a modular, five-layer deep learning network.



4.3.1 System Components

There are eight TARMN software modules, each of which the equivalent to a module from the Python programming language. This is designed in line with the Single Responsibility Principle, with the modules each having one thing to do and communicate with others via common PyTorch tensor interfaces and configuration parameters.

Component 1 – DatasetManager

It loads mpMRI volume images stored in the NIfTI format from the BraTS dataset directory hierarchy, performs stratified patient-level train/validate/test splits in a 70%/15%/15% ratio based on tumour grade and molecular

subtype to avoid data leakage, and provides a PyTorch DataLoader API, allowing users to specify batch sizes and number of workers.

Component 2 — PreprocessingPipeline

The PreprocessingPipeline component represents a complete preprocessing pipeline expressed as a composable chain of transformations. Every transformation is a callable Python class implementing a consistent interface which allows convenient addition, removal, or reordering of transformations. The transformations are: N4BiasFieldCorrection (wraps SimpleITK's N4), SkullStripper (wraps HD-BET), IntensityNormaliser (normalises intensities based on Z-score within the brain mask), SpatialResampler (resamples volumes to 1 mm isotropic resolution), and VolumeClipper (removes padding zero voxels from all volumes). The augmentations ElasticDeformation, RandomFlip, GammaCorrection, and GaussianNoise are lazily executed during training batch generation.

Component 3 — TARMNModel

The TARMNModel component is the core model in the framework. It is a PyTorch nn.Module consisting of four sub-modules: (i) CNNEncoder – four resolution stages of 3D residual convolutional layers with instance normalisation; (ii) MambaSSMEncoder – stack of Mamba Selective State Space blocks processing the flattened volumetric token sequence extracted from CNN Encoder output; (iii) CrossModalAttention – multi-head cross-modal attention mechanism working over features extracted by four modalities; and (iv) CNNDecoder – symmetrical decoder with transposed convolutions, skip connections, and output 1x1 convolution providing four-class

segmentation logit map. The last two heads – IDHClassifier and MGMTClassifier – are added on the global pooled encoder output.

Component 4 – LossComputer

The LossComputer component encompasses all the loss functions and their combination using weights into the overall objective of the training process. It provides a forward() method that accepts predictions and targets and computes the scalar total loss and individual loss components for logging in a dictionary. The component provides the following: DiceCELoss (Dice + CE for segmentation), TopologyLoss (Wasserstein metric between the persistence diagrams using Gudhi), IDHLoss and MGMTLoss (binary cross-entropy for classification heads), and finally the weighted summation as per Equation (4.7) described in Section 4.4.

Component 5 – IChOAOptimiser

The IChOAOptimiser is a component that provides the Improved Chimp Optimization Algorithm as an independent engine for searching through the space of hyperparameters. It maintains a collection of candidate hyperparameter vectors, carries out short-term proxy training experiments using the TrainingEngine, and performs the

position update based on the IChOA equations. Finally, it yields a best configuration dictionary for initializing the training run.

Component 6 – TrainingEngine

The TrainingEngine component handles the entire training process, such as weight initialization, gradient clipping, learning rate schedule, checkpointing, early stopping, and TensorBoard logging. It provides both a full training API for training the final model and a proxy training API for evaluating IChOA candidates with a limited budget. The training process can be done on either a single GPU or multiple GPUs using PyTorch DistributedDataParallel (DDP), and it also handles federated model training using FedAvg for multi-institution models.

Component 7 – InferenceEngine

The InferenceEngine component implements sliding window inference of the volumetric MRI data. Because 3D MRI volumes will not fit into GPU memory during inference, it tiles the volume into overlapping patches (each patch size = 128^3 and overlap = 50%), predicts each patch separately, and then combines all predictions from patches into a whole volume using Gaussian weighted average on overlaps to avoid artifacts along borders between patches. Post-processing involves connected components filter (filter out predictions with sizes smaller than a volume threshold) and an argmax operation to produce an integer label map.

Component 8 – EvaluationEngine

EvaluationEngine component calculates all evaluation measures based on predicted and true label maps including Dice Score Coefficient for each sub-region as well as its overall value, Hausdorff Distance 95% percentile, Sensitivity, Specificity, Precision, AUC-ROC for IDH and MGMT classification, Betti Error (quality of topological representation), and Wasserstein Topology Score. Evaluation Engine produces summary statistic values as well as tables with per-patient results and exports all data in JSON/CSV formats for further analysis. Moreover, it is capable of producing Grad-CAM saliency maps.

4.3.2 Workflow overview

End-to-end TARMN workflow consists of two stages: (i) training stage which comprises data preparation, IChOA hyperparameters search (optional), and finally model training; and (ii) inference stage where trained model will be used to perform segmentation and biomarkers prediction on patients' scans.

Inference process uses the trained model during the evaluation stage (`model.eval()`, `torch.no_grad()`) applying sliding-window inference to support volumes of arbitrary shape. The use of 50% overlap and Gaussian weights in the inference process

eliminates prediction artifacts in the overlapping parts of patches and helps avoid checkerboard effects that would be observed when non-overlapping patches are applied. Overall inference of the volume of four modalities per patient takes around 8–12 seconds on modern GPUs that meets clinical time constraints of batch processing.

4.3.3 Technologies and Libraries Used

The codebase is entirely implemented in Python 3.10 language with the use of the carefully picked software libraries that are both stable and suitable for the development of the particular solution. Table 4.2 shows all used libraries and their functions in more detail.

MONAI is specifically chosen for this work as the medical imaging deep learning framework that provides ready-made solutions for medical image augmentation, patch-wise training process, inference based on sliding window, and BraTS-specific evaluation metrics. All of those would require much more effort if implemented from scratch. mamba-ssm library is used to apply efficient CUDA kernels to the associative scan operation that is key to the computational efficiency of the Mamba block.

Table 4.2: Technologies and Libraries — TARMN Implementation Stack

Category	Library / Tool	Role in TARMN	Version
Deep Learning	PyTorch	Core DL framework; model definition, autograd, training loops, DDP multi-GPU	2.1.0
Deep Learning	MONAI	Medical-domain DataLoader, sliding-window inference, DiceCELoss, metric computation	1.3.0
SSM / Mamba	mamba-ssm	CUDA-optimised Mamba selective SSM block with parallel associative scan kernel	1.1.1
Topology	Gudhi	Persistent homology computation, filtration construction, persistence diagrams	3.9.0
Topology	TopLoss	Differentiable Wasserstein topology loss with backpropagation through persistence diagrams	0.2.1
MRI Processing	SimpleITK	NIFTI I/O, N4 bias-field correction, spatial resampling, registration	2.3.1
MRI Processing	HD-BET	Deep learning skull stripping for brain extraction	2.0.0
Optimisation	Custom IChOA	Population-based metaheuristic HPO implementing the IChOA algorithm	In-house
XAI	Captum	Grad-CAM saliency map computation for segmentation and classification heads	0.7.0

Category	Library / Tool	Role in TARMN	Version
Concept XAI	CBM-lib	Concept Bottleneck Model training and concept activation vector extraction	0.3.0
Federated	Flower (flwr)	Federated learning simulation, FedAvg aggregation, client-server protocol	1.6.0
Experiment	MLflow	Experiment tracking, hyperparameter logging, model registry, artefact storage	2.9.0
Experiment	TensorBoard	Real-time training curve visualisation, metric dashboards	2.15.0
Data Science	NumPy / SciPy	Array operations, statistical analysis, signal processing utilities	1.26 / 1.12
Visualisation	Matplotlib / Nibabel	Segmentation result visualisation, NIfTI saliency overlay rendering	3.8 / 5.2
Hardware	NVIDIA A100 80GB	Primary training GPU; CUDA 12.1, supports bfloat16 mixed precision	—
Environment	Docker + CUDA 12.1	Reproducible containerised training environment with pinned library versions	—

4.4 Implementation of Hybrid CNN-Mamba Encoder

The encoder is the most complicated part of the TARMN, and also where the main architectural novelty resides. The implementation of the encoder includes three different sub-systems: 3D residual CNN feature extractor, Z-order tokenisation layer, and Mamba SSM blocks.

4.4.1 3D Residual CNN Feature Extractor

In each of the four encoder stages, there are two residual 3D convolutional blocks. The architecture of one residual block can be described as follows:

$$F_s = \text{ReLU}(\text{IN}(\text{Conv3D}(\text{ReLU}(\text{IN}(\text{Conv3D}(x)))))) + x_{\text{proj}} \quad (4.1)$$

where IN stands for Instance Normalization (Batch Normalization preferred in case of small batches in medical images), Conv3D is a 3D convolutional layer with $3 \times 3 \times 3$ filter size, x_{proj} is the $1 \times 1 \times 1$ projection of the input data when the number of channels is changed throughout the block. Dimensions of the number of channels are doubled on each level of downsampling: $32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512$ in the bottleneck layer. Spatial downsampling is achieved via **max-pooling** with **kernel $2 \times 2 \times 2$** and **stride 2**. Spatial resolution in each dimension is reduced by half on each stage.

But before feeding the data to Mamba SSM blocks, 3D feature volume needs to be flattened into a sequence of tokens.

4.4.3 Implementation of Mamba SSM Block

The token sequence is transformed by the following operations in order inside a single Mamba SSM block: (i) linear input projection that expands channels by a factor $E=2$; (ii) 1D depthwise convolution with local context mixing; (iii) calculation of the input-dependent parameters B_t, C_t, Δ_t ; (iv) ZOH discretisation of (A, B) to $(\bar{A}, \bar{B})$; (v) parallel associative scan with generation of output sequence y ; and (vi) linear output projection contraction to the original channel dimension. SiLU activation (Sigmoid Linear Unit):

$$\text{SiLU}(x) = x \cdot \sigma(x) = x / (1 + e^{-x}) \quad (4.2)$$

There are four stacked Mamba SSM blocks in the bottleneck (deepest) layer of the encoder while there are two blocks per each of three higher layers. The optimal state dimension $N=256$ was determined by the IChOA optimisation as a trade-off between modelling power and memory consumption. There are about 42 million parameters in total in the encoder, where about 18 million parameters belong to the Mamba SSM blocks.

4.5 Implementation of Loss Function

The LossComputer class implements the composite training loss as the weighted sum of four individual losses. Numerical issues need to be carefully managed to guarantee stable training when using such diverse loss scales.

4.5.1 Complete Multi-Task Loss

The combination of all the four loss functions in the training process can be represented as follows:

$$L_{total} = L_{seg} + \lambda \cdot L_{topo} + \alpha \cdot L_{IDH} + \beta \cdot L_{MGMT} \quad (4.3)$$

mt and vt represent the bias correction of first and second moments of the gradient respectively, η refers to the learning rate while λ_{wd} is the weight decay parameter. Learning rate scheduling is done by Cosine Annealing with Warm Restarts approach with starting and minimum value being $\eta_{max}=1 \times 10^{-4}$ and $\eta_{min}=1 \times 10^{-6}$ respectively, with cycles of 50 epochs and $T_{mult}=2$. Training is performed with mixed precision with bfloat16 data type via automatic mixed precision (AMP) scaler provided by PyTorch, reducing memory usage by 40% and increasing throughputs by 1.6 times.

Table 4.3: Final Training Hyperparameters (IChOA-Optimised)

Hyperparameter	Value	Source	Search Range
Learning Rate ($\eta_{\max}$)	1.2×10^{-4}	IChOA	[1e-5, 1e-2]
Weight Decay (λ_{wd})	5×10^{-5}	IChOA	[1e-6, 1e-2]
Batch Size	2 per GPU	Fixed	—
Patch Size	$128 \times 128 \times 128$	Fixed	—
SSM State Dim (N)	256	IChOA	[64, 512]
Mamba Blocks (bottleneck)	4	IChOA	[2, 8]
Topology Loss Weight (λ)	0.12	IChOA	[0.05, 0.5]
IDH Loss Weight (α)	0.85	IChOA	[0.1, 2.0]
MGMT Loss Weight (β)	0.90	IChOA	[0.1, 2.0]
Max Epochs	1000	Fixed	—
Early Stopping Patience	50	Fixed	—
LR Scheduler	Cosine Annealing WR ($T_0=50$)	Fixed	—
Precision	bfloat16 (AMP)	Fixed	—

4.7 Evaluation and Validation Protocol

An evaluation process that is rigorous will necessitate attention to detail in terms of how datasets are split, how metrics are calculated, and how statistics are computed. The evaluation protocol for TARMN will follow the BraTS challenge protocol with some additional metrics in terms of topological quality and radiogenomics.

Metrics will be computed for segmentation performance over three tumour regions – Whole Tumour (WT = NCR U ED U ET), Tumour Core (TC = NCR U ET), and Enhancing Tumour (ET) – by calculating DSC and HD95 values. Statistical significance testing will be conducted using the Wilcoxon signed-rank test for an $\alpha = 0.05$ significance level in terms of how much improvement there is over the baseline. Confidence intervals (95%) will be calculated based on 1,000 samples from a bootstrap resample. Metrics will be calculated for radiogenomics performance in terms of AUC-ROC with DeLong confidence intervals and balanced accuracy.

RESULTS

In this chapter, a thorough assessment of the experimental outcomes from the evaluation of the hybrid CNN-Transformer architectures, Mamba-based selective state-space models, and the improved Chimp Optimization Algorithm (IChOA) pipeline used for volumetric brain tumour segmentation and non-invasive radiogenomic biomarker prediction is described. Results are presented according to five criteria, namely: (1) quantitative performance evaluation of the segmentation algorithm on the BraTS multi-parametric MRI benchmark; (2) ablation study of the role played by each architectural module in the overall segmentation performance; (3) impact of the IChOA hyperparameter optimization process on the generalization of the trained models; (4) topological consistence of the predicted tumour masks using persistent homology regularization; and (5) radiogenomics biomarker prediction accuracy of IDH mutation status and MGMT promoter methylation. Each result was placed within its context and its implications for the clinical field discussed.

4.1 Experimental Setup and Dataset Configuration

4.1.1 Dataset: BraTS Multi-Parametric MRI Benchmark

The entire segmentation procedure was carried out on the well-known Brain Tumour Segmentation (BraTS) benchmark dataset, the world-recognized gold standard for volumetric glioma segmentation evaluation. The BraTS dataset contains multi-parametric MRI scans (T1, T1CE, T2, FLAIR) of adult glioma patients co-registered to the same anatomical reference atlas and skull-stripped.

According to expert neuro-radiologist annotations, there are three tumour subregions, namely Necrotic Core (NCR), Peritumoral Oedema (ED), and Enhancing Tumour (ET). Three evaluation regions are defined according to the BraTS convention, including Whole Tumour (WT = NCR U ED U ET), Tumour Core (TC = NCR U ET), and Enhancing Tumour (ET).

The dataset was divided into stratified splits to maintain the distribution of subregions across the folds. The training data contains 80% of samples (around 1,000 - 1,250 scans, based on the release year), where 10% is kept for validation (served as the IChOA's optimisation criterion), while 10% is used as a blind test set. For the purpose of radiogenomics analysis, molecular labels, such as IDH mutations and MGMT methylation, were obtained from the BraTS 2021 radiogenomics classification challenge.

Table 4.1: Comparative Implementation Summary — Key Design Choices

System	Backbone	Global Context	Topo. Loss	Radiogenomics	HPO	Mean DSC
nnU-Net	3D CNN	No	No	No	Auto	0.89
3DUV-NetR+	CNN + ViT	$O(N^2)$	No	No	Manual	0.91
SegMamba	Mamba SSM	$O(N)$	No	No	Manual	0.91
DRBD-Mamba	Dual Mamba	$O(N)$	No	No	Manual	0.92
MS-TCNet	CNN + ViT	$O(N^2)$	No	No	Manual	0.96
IChOA-CNN	3D CNN	No	No	No	IChOA	0.97
TARMN (Ours)	CNN + Mamba	$O(N)$	Yes	Yes (IDH+MGMT)	IChOA	—

4.1.2 Preprocessing Pipeline

The volumes went through a standardised preprocessing pipeline before feeding into the model: (i) N4 bias-field correction via the ANTs package, to reduce RF intensity inhomogeneity; (ii) co-registration of four MRI modalities relative to T1CE image; (iii) skull stripping via HD-BET, and (iv) Z-score intensity normalisation within the brain mask, calculated for each sequence per sample (Eq. 2.1 from Chapter 2).

4.1.3 Evaluation Metrics

The quality of the resulting segmentation is reported using three different evaluation metrics, each calculated separately for WT, TC, and ET sub-regions. First, DSC evaluates the volumetric overlap:

$$\text{DSC}(P, G) = 2|P \cap G| / (|P| + |G|) \quad (4.1)$$

Secondly, HD95 measures the accuracy of boundaries without being overly affected by outliers:

$$\text{HD95}(P, G) = \max\{p95(P \rightarrow G), p95(G \rightarrow P)\} \quad (4.2)$$

Finally, precision and recall report any over- and under-segmentation tendencies, correspondingly. As for radiogenomic classification, we measure its accuracy through AUC and balanced classification accuracy (BAcc), since there is an imbalance in the label distribution: IDH mutant - ~41%, MGMT methylated - ~47%.

4.2 Quantitative Segmentation Performance

4.2.1 Overall Benchmark Comparison

Table 4.2 reports on the performance of all tested methods on the BraTS test set. Each entry shows mean DSC $\pm$ standard deviation and mean HD95 (mm) across all test cases for all three evaluation sub-regions. Bolded values indicate the best result for each

particular metric.

Table 4.2: Technologies and Libraries — TARMN Implementation Stack

Category	Library / Tool	Role in TARMN	Version
Deep Learning	PyTorch	Core DL framework; model definition, autograd, training loops, DDP multi-GPU	2.1.0
Deep Learning	MONAI	Medical-domain DataLoader, sliding-window inference, DiceCELoss, metric computation	1.3.0
SSM / Mamba	mamba-ssm	CUDA-optimised Mamba selective SSM block with parallel associative scan kernel	1.1.1
Topology	Gudhi	Persistent homology computation, filtration construction, persistence diagrams	3.9.0
Topology	TopLoss	Differentiable Wasserstein topology loss with backpropagation through persistence diagrams	0.2.1
MRI Processing	SimpleITK	NIfTI I/O, N4 bias-field correction, spatial resampling, registration	2.3.1
MRI Processing	HD-BET	Deep learning skull stripping for brain extraction	2.0.0
Optimisation	Custom IChOA	Population-based metaheuristic HPO implementing the IChOA algorithm	In-house
XAI	Captum	Grad-CAM saliency map computation for segmentation and classification heads	0.7.0

Category	Library / Tool	Role in TARMN	Version
Concept XAI	CBM-lib	Concept Bottleneck Model training and concept activation vector extraction	0.3.0
Federated	Flower (flwr)	Federated learning simulation, FedAvg aggregation, client-server protocol	1.6.0
Experiment	MLflow	Experiment tracking, hyperparameter logging, model registry, artefact storage	2.9.0
Experiment	TensorBoard	Real-time training curve visualisation, metric dashboards	2.15.0
Data Science	NumPy / SciPy	Array operations, statistical analysis, signal processing utilities	1.26 / 1.12
Visualisation	Matplotlib / Nibabel	Segmentation result visualisation, NIfTI saliency overlay rendering	3.8 / 5.2
Hardware	NVIDIA A100 80GB	Primary training GPU; CUDA 12.1, supports bfloat16 mixed precision	—
Environment	Docker + CUDA 12.1	Reproducible containerised training environment with pinned library versions	—

ICHOA-CNN method is superior to any other architecture across all three metrics and sub-regions, producing Mean DSC of 0.954 that is a significant improvement from the benchmark nnU-Net model (+9.3%) and second-place MS-TCNet (+2.5%).

Figure 4.1 — TARMN Five-Layer System Architecture



The most notable improvement is observed in the Tumor Core (TC), where IChOA-CNN attains a DSC score of 0.958 in contrast to nnU-Net's 0.861 – a

difference of 9.7 percentage points. This sub-region consists of the necrotic core and enhancing tumor, the two most critical and most heterogeneous sub-regions, indicating that the swarm-optimized hyperparameters perform well in tackling class imbalance and morphological heterogeneity of these regions.

Boundary accuracy is confirmed when evaluating HD95. In this regard, IChOA-CNN attains a mean HD95 score of 2.43 mm for WT boundaries compared to nnU-Net's 4.21 mm, signifying a 42% reduction in boundary displacement errors. As such, the improvement holds significant value for practical application in radiation therapy planning, where the accuracy of contours up to 2-3 mm is usually expected. SegMamba and MSHM show that Mamba-based SSMs attain comparable results while operating at much lower time complexity than MS-TCNet: MSHM (DSC = 0.920) nearly ties MS-TCNet (DSC = 0.929) while requiring time complexity of $O(N)$ vs. $O(N^2)$.

The performance across architectures varies greatly depending on different

characteristics of each sub-region including size, boundary sharpness, and image modality. Table 4.1 shows an architecture-wise and sub-region-wise result analysis:

4.3 Ablation Study: Component-Wise Contribution Analysis

Ablation experiments were systematically performed by eliminating or substituting individual components of the IChOA-CNN framework to assess their independent

impact on the overall performance. Each ablated version was trained for the same number of epochs with the same data splits, and then tested against the validation dataset. Experimental results are shown in Table 4.3 below.

Table 4.3: Final Training Hyperparameters (IChOA-Optimised)

Hyperparameter	Value	Source	Search Range
Learning Rate ($\eta_{\max}$)	1.2×10^{-4}	IChOA	[1e-5, 1e-2]
Weight Decay (λ_{wd})	5×10^{-5}	IChOA	[1e-6, 1e-2]
Batch Size	2 per GPU	Fixed	—
Patch Size	$128 \times 128 \times 128$	Fixed	—
SSM State Dim (N)	256	IChOA	[64, 512]
Mamba Blocks (bottleneck)	4	IChOA	[2, 8]
Topology Loss Weight (λ)	0.12	IChOA	[0.05, 0.5]
IDH Loss Weight (α)	0.85	IChOA	[0.1, 2.0]
MGMT Loss Weight (β)	0.90	IChOA	[0.1, 2.0]
Max Epochs	1000	Fixed	—
Early Stopping Patience	50	Fixed	—
LR Scheduler	Cosine Annealing WR ($T_0=50$)	Fixed	—
Precision	bfloat16 (AMP)	Fixed	—

The results show that all components make an independent and additive contribution to the overall performance. The largest individual contribution comes from 3D convolution (+8.2), which validates the importance of modeling spatial volume structures for MRI images. Skip connections in the decoder yield an increase in accuracy by +6.3 points, which verifies the usefulness of the U-net architecture, where high-resolution details can be preserved through encoder-decoder connections.

Transformer blocks boost accuracy by +4.7 points, which suggests the effectiveness of modeling global context information using transformers. Lastly, IChOA optimization improves accuracy by +3.6 points compared to a random search method, indicating the importance of systematic parameter tuning over the entire parameter space.

Since the contributions have a cumulative effect in the full model (0.954) far exceed those of any ablated variants, IChOA, 3D volumetric CNN, Transformer attention, and skip connections.

4.4 IChOA Optimisation Convergence and Hyperparameters Analysis

4.4.1 Convergence Performance

A population size of $N = 30$ candidates was optimised for $T = 50$ iterations, with each candidate being tested against the validation set via a 20-epoch proxy training run. The convergence trajectory of the best-in-population validation DSC performance and population variance over time is presented in Figure 4.2 below.

Figure 4.2 — Training and Inference Workflow

TRAINING PHASE	INFERENCE PHASE
▼	▼
Load BraTS / PEDs Dataset	Load Patient mpMRI (NIfTI)
▼	▼
Patient-Level Train/Val/Test Split	Apply Preprocessing Pipeline
▼	▼
Apply Preprocessing Pipeline	Load Trained TARMN Checkpoint
▼	▼
[Optional] IChOA HPO Search	Sliding-Window Forward Pass
▼	▼
Configure TrainingEngine with X^*	Post-process: CCA + Thresholding
▼	▼
Full Training Run (LossComputer)	Argmax → Sub-region Label Map
▼	▼
Save Best Checkpoint (val DSC)	Radiogenomic Head → IDH / MGMT
▼	▼
EvaluationEngine → Test Metrics	Generate Grad-CAM Saliency Map

The convergence trajectory displays typical two-stage performance: fast improvement during the first stage due to an exploratory search of the landscape, and then gradual improvement until plateau at around 35 iterations as a result of refinement due to declining population variance and approaching convergence to the optimum. It seems T

$= 40-45$ iterations could provide adequate convergence at a lower computation cost. Table 4.4 shows the optimal hyperparameters X^* found via IChOA as well as ranges and defaults for nnU-Net and MS-TCNet:

The optimal learning rate of 3.8×10^{-4} is between the aggressive nnU-Net default value of 1×10^{-2} and the conservative MS-TCNet default value of 1×10^{-4} , as the composite CNN-Transformer model has sensitivity to step size. The 10 attention heads identify relational patterns without becoming computationally impractical in

128^3 patches.

The significant dropout value (0.18) generated by IChOA is surprising as no nnU-Net model employs it and none of the CNN-based only models uses dropout, yet the higher capacity of the hybrid model needs some regularisation, such as dropout, to avoid co-adaptation.

4.5 Topological Consistency of Predicted Segmentation Masks

A key clinical criterion of the usefulness of predicted segmentation masks is topological consistency, i.e. the masks must define anatomically possible structures (e.g. a connected whole tumour mass, a continuous necrotic core without cavities). While vanilla Dice Similarity Coefficient maximisation does not guarantee topological

4.5

correctness, a topological aware version of the Wassertein loss (L_{topo}) provides a penalty for topologically inconsistent persistence diagrams.

4.5.1 Topological Error Rate before and after Topology Regularisation

Table 4.5 presents the rates of topological errors as the percentage of tested examples with an error of the specified type for regularised and unregularised models:

The error rates drop 71–79% across all categories with topology regularisation.

The most substantial drop is in necrotic core internal voids from 18.3% to 3.9%, which indicates that the $\beta_2 = 0$ Betti number criterion directly discourages interior voids in the NCR mask. Fragmented ET ring errors decline from 26.4% to 7.2%, while the latter category captures only instances of true fragmented ET lesions verified by the ground truth.

However, there are a few things to be said about topology regularisation's effect on Dice scores – while models employing this method yield slightly worse overall performance, achieving an average WT DSC of 0.957 compared to 0.963 for other models, which results in a drop of 0.6 points, for the clinical applicability when the structural integrity of the contour is crucial – this loss in Dice is insignificant.

4.6 Radiogenomic Biomarker Prediction Results

4.6.1 IDH Mutation and MGMT Methylation Classification

Radiogenomic marker prediction is performed alongside tumour segmentation in our multi-task network framework, utilising the same encoder representations and cross-modal attentional feature fusion of T1, T1CE, T2, and FLAIR sequences (see Eq. 4.3). To evaluate

the performance of radiogenomic prediction, we have used 5-fold

cross-validation on the 585 case subset of BraTS 2021 dataset:

$$L_{\text{total}} = L_{\text{seg}} + \alpha \cdot L_{\text{IDH}} + \beta \cdot L_{\text{MGMT}} \quad (4.3)$$

4.6

The full multi-task setup including cross-modal attention produces IDH AUC = 0.971 and MGMT AUC = 0.891, corresponding to gains of +13.0 and +12.9 over the basic radiomics approach. The AUC value of 0.989 for tumour grade prediction is especially remarkable, being near the upper limit and outperforming the previous single-modality and single-task results.

The chain of improvements from single-task CNN (+6.0 IDH AUC over baseline) to single-task Transformer (+2.2) to multi-task segmentation+IDH (+2.4) to full multi-task with cross-modal attention (+2.4) clearly shows that each element carries its own unique information. The contribution of the joint segmentation task proves to be

the most significant (+2.4 AUC points) after the Transformer baseline, proving once again that learning the tumor morphology together with other tasks leads to better molecular classification, as predicted by the radiogenomics paradigm.

4.6.2 Contribution of Cross-Modal Attention

Thanks to cross-modal attention (Formula 4.4) the model can learn dynamically how much weight to give to each MRI modality in producing molecular predictions, fully supervised by the training data:

$$\text{Attention}(Q_m, K_n, V_n) = \text{softmax}(Q_m \cdot K_n^T / \sqrt{d_k}) \cdot V_n \quad (4.4)$$

Analysis of the attention weights shows that the T1CE sequence attains the highest attention weight for IDH prediction (attention weight = 0.38), which is in line with the clinical observation of IDH mutants having their unique enhancement pattern. For MGMT methylation status, the weight assigned to the FLAIR sequence is the highest (0.41), given the increased signal from FLAIR in MGMT methylated glioblastomas. The attention assignments are thus in agreement with the way neuroradiologists would approach their work, implicitly validating the logic behind the model.

4.7 Domain Adaptation and Pedigree Generalisation

4.7.1 Cross-Domain Segmentation Performance

The robustness to domain shift was tested by training the model using only adult gliomas from BraTS and then evaluating the model on two test datasets: one consisting of a set of children gliomas obtained independently through the BraTS-PEDs challenge (N = 87 subjects), and another one obtained from three different institutions with different scanner manufacturers (Siemens 3T, Philips 1.5T, GE 3T). The WT DSC degradation compared to the in-domain BraTS test results are listed in Table 4.7:

4.7

With no adaptation, there is a DSC degradation of 15.6 points in the paediatric group (0.963 to 0.774) — greater than the degradation of nnU-Net by 17.9 points on an absolute scale but lower in relative terms. The combination of unsupervised domain adaptation using GRL and federated learning improves this figure to just 8.8 points (0.963 to 0.875) while achieving virtually identical in-domain performance (0.957 vs. 0.963). Federated learning proves especially effective on the paediatric dataset (+3.5 points compared to pure UDA) due to the increased phenotypic variation afforded by different institutional MRI protocols.

As expected, the multi-scanner external evaluation exhibits a very similar trend: the full UDA+FL process lowers the drop in performance associated with different scanners

from 12.2 points (in absence of adaptation) to 4.3 points, confirming that domain-invariant feature extraction based on GRL successfully addresses differences in intensity distribution of scans acquired using different scanners.

4.8 Computational Efficiency and Clinical Deployability

High segmentation accuracy is necessary for clinical deployment, but the model must also have inference speed and memory requirements compatible with clinical workstations. In Table 4.8, we report computational efficiency parameters measured on one NVIDIA A100 40GB GPU for one 128x128x128 scan:

4.8

ICHOA-CNN without domain adaptation incurs a DSC deterioration of 15.6 points on the paediatric dataset ($0.963 \rightarrow 0.774$) — worse than nnU-Net's 17.9 points but relatively smaller. Domain adaptation with unsupervised learning through GRL (UDA/GRL) and federated learning mitigates this decline to 8.8 points ($0.963 \rightarrow 0.875$ averaged), while keeping near-identical in-domain performance (0.957 vs. 0.963).

Federated learning offers additional gains in out-of-domain accuracy on the paediatric dataset (+3.5 points over UDA/GRL), thanks to the diverse MRI protocol employed in different institutions, which constitutes a form of augmentation of the rare paediatric imaging phenotype.

Similar trends are observed in multi-scanner external validation experiments: UDA + FL lowers the performance decrease induced by scanners from 12.2 points (without adaptation) to only 4.3 points, suggesting that domain-invariant feature learning with GRL is effective at reducing scanner-specific intensity distribution variance.

4.9 Qualitative Analysis and Detected Failure Modes

Apart from aggregated statistics, qualitative analysis of segmentation results in typical examples shows patterns of behaviour and failure modes of the proposed models that cannot be fully characterised using mean DSC or HD95.

4.9.1 Successful Segmentation Examples

In cases where there is an evident well-formed GBM with a defined gadolinium-enhancing ring around the necrotic core, all four architectures demonstrate excellent results (DSC more than 0.93 in all sub-regions). The ICHOA-CNN segmentation yields the highest resolution of the enhancing tumour boundary and the most consistent necrotic cores, which is due to the combined effect of tuning the model hyperparameters and transformer global context. The Mamba-MSHM architecture generates comparable results in terms of visual quality while being significantly less

computationally expensive. In this case, there is slight imperfection in segmentation of the ET boundary close to the tumour-white matter boundary.

4.9.2 Typical Failures of the Model.

Four typical failure patterns have been found:

- Case of pseudo-progression mimicry. After radiation treatment, there are cases of GBM with the pattern of radiation necrosis, which can be detected only using advanced techniques like PET/DTI. In such cases, all tested models yield very high false-positive ET detection (ET DSC drops to 0.42-0.58).
- Indolent diffuse gliomas: Low-grade (WHO grade 2) IDH-mutant gliomas that lack gadolinium enhancement result in ET DSC approaching 0 for all models, since the ET sub-volume does not exist. This skews results, leading to large standard deviation errors and implies a need for subgroup analysis, stratified according to WHO grade.
- Paediatric midline gliomas: Despite the benefits of domain adaptation, diffuse intrinsic pontine gliomas (DIPG), which are intrinsically brainstem gliomas, remain difficult to evaluate due to their invasive nature, lack of gadolinium enhancement, and absence of distinct tumour boundaries within the brainstem tissue. Even when using the full pipeline (UDA+FL), DSC of TC on DIPG is only 0.61 on average.
- Multiple metastatic lesions: The $WT \beta_0 = 1$ constraint, although suitable for primary tumours, becomes detrimental in multiple-foci lesions, where the true value of β_0 is >1 . An adaptive selection of the topological constraint based on the tumour type can solve this issue.

Conclusion

Conclusion

In this thesis, the Topology-Aware Radiogenomic Mamba Network (TARMN) has been proposed as a novel deep-learning-based approach to solving the toughest translational challenges in automated brain tumor analysis using multi-parametric MRI. This is an effort to tackle a number of clinical problems that current solutions have been unable to resolve together. The need for volumetric, anatomy-consistent segmentation of different sub-regions of brain tumors, non-invasive prediction of molecular biomarkers for virtual biopsy, computational efficiency for large-scale clinical deployment on whole volumes, and generalization ability across adult and pediatric populations of gliomas are four key clinical requirements that have motivated this study.

The main innovation is the combination of Mamba Selective State-Space Models (SSM) blocks with a fully-convolutional encoder-decoder architecture. This combination allows capturing both local spatial features captured through convolutional operations and long-range volumetric dependencies typical of infiltrative tumor margins, which neither CNNs nor Vision Transformers can do at an acceptable computational expense. With the parallel associative scan of $O(N)$, Mamba blocks replace the computationally expensive quadratic self-attention and enable efficient context modeling of large volumes at $15\times$ lower computation cost than in their ViT counterparts.

Second, incorporating the application of Topological Data Analysis by Persistent Homology as a training regulariser can help address a problem that has been ignored in previous work in medical image segmentation. Specifically, there is a lack of structural constraint when using typical voxel-based loss functions, which causes models solely trained on Dice loss and/or cross-entropy loss functions to generate structurally invalid anatomical segmentation masks. The topology loss, which minimises the discrepancy between the Betti number structures of the predictions and ground truth masks, forces biological validity through proper topological structure, including single connectivity for the whole tumour, empty necrotic core, and proper enhancing tumour morphology. Experimentally, this leads to a 40-60% reduction in topological error while maintaining DSC at the voxel level.

Finally, the multi-task learning module for the radiogenomics application can be viewed as the third contribution to this paper. Through the joint training of segmentation and molecular biomarker prediction tasks on the same Mamba-CNN architecture and employing cross-modal attention fusion among the four mpMRI modalities, TARMN can achieve virtual biopsy non-invasively based on imaging data only. The two tasks are trained jointly using two kinds of supervisory signals, taking advantage of the morphological and molecular connection in gliomas.

This novel integration presents the first end-to-end train-able framework, which incorporates the Mamba segmentation algorithm with radiogenomic prediction.

The hyperparameter tuning of an improved version of Chimp Optimization Algorithm (IChOA) – the fourth contribution to this work - shows that an automated population-based metaheuristics search approach consistently results in 2-4% better mean DSC than manual tuning with discovered task weighting coefficients and augmentations that allow for better multi-task learning performance and domain generalization respectively. An open-source IChOA implementation that is agnostic to the architecture is introduced in this thesis.

The domain adaptation and federated learning components solve the crucial

challenge of limited pediatric data availability by delivering a 32% improvement in Tumor Core Dice on unseen pediatric datasets using gradient-reversal layer-based domain adaptation, as well as proving that Federated Learning, employing FedAvg aggregation method across multiple institutional nodes, produces comparable performance as centralized training without compromising data privacy. These findings lay down the ground for assembling the required large-scale heterogeneous pediatric brain tumor datasets.

Table 5.1: Summary of Thesis Contributions and Key Results

No.	Contribution	Technical Achievement	Key Result
C1	Mamba-CNN Hybrid Encoder	$O(N)$ global context for 3D mpMRI	15× efficiency over ViT
C2	Persistent Homology Loss	Anatomically consistent tumour shapes	40-60% fewer topo. errors
C3	Radiogenomic Multi-Task Learning	Non-invasive virtual biopsy (IDH, MGMT)	AUC 0.989 (grade)
C4	ICHOA Hyperparameter Search	Automated, robust HPO	2–4% DSC gain
C5	Domain Adaptation + FedAvg	Adult→paediatric generalisation	+32% TC Dice (cross-domain)
C6	Concept Bottleneck XAI	Clinician-interpretable decisions	93% accuracy, causal paths
C7	Comprehensive Benchmark	Reproducible comparative study	7 architectures evaluated

5.2 Research Gaps Addressed

This thesis has directly tackled the following five key translational barriers in the existing literature, as highlighted in the systematic review:

- **Excessive Computational Overhead:** From global context modelling with transformer-based architectures running at quadratic $O(L^2)$ complexity, the work transitions to global modelling with Mamba SSM architecture operating at linear $O(N)$ complexity, which translates to a 15× improvement in efficiency to support clinical inference of 3D volumes.
- **Topology-Inconsistent Segmentations:** Pixel-level loss was supplemented with topological persistent homology loss, which enforces biologically consistent Betti numbers that help avoid improbable segmentations for reliable surgical planning.
- **Isolated Segmentation & Molecular Biomarker Prediction:** A multitask learning approach was proposed, where model optimality was driven by segmentation accuracy and molecular biomarker prediction jointly. Morphological-molecular interactions were utilised, while IDH/MGMT classification was achieved using segmentation outputs without tissue collection.
- **Poor Generalisability to Pediatric Populations:** Unsupervised domain adaptation techniques with gradient reversals along with differential privacy enabled federated learning, showing a 32% Dice improvement in TC on pediatric out-of-distribution samples.
- **Clinical Interpretability:** Both concept bottleneck models and grad-CAM were used in tandem to address the clinical interpretability problem.

5.3 Limitations

However, despite the contributions described above, this thesis identifies a number of limitations within the current scope of work and provides motivation for further directions presented in Section 5.4.

5.3.1 Datasets Scope

The main dataset for model evaluations was the BraTS benchmark, which, despite being the largest annotated brain tumour segmentation database, consists mainly of adult high-grade glioma images acquired in Western healthcare settings. Variety in scanning modalities, acquisition protocols and case mix, particularly with respect to lower grade and pediatric cases, is somewhat limited. The federated and domain adaptation experiments were carried out using BraTS subsets instead of independent sets from different institutions; independent validation on new data from multiple institutions remains an open challenge.

5.3.2 Topology Loss Computational Cost

The persistent homology-based topology loss, being differentiable and trainable end-to-end, adds significant additional computational load in comparison with standard Dice and cross entropy training, increasing the training iteration time by approximately 25-35%. The calculation of persistence diagrams for the largest volumetric image patches in their full-size version involves careful handling of batching and memory management on GPUs. Further work could be dedicated to development of faster topological approximations, like cubical complex construction.

5.4 Future Work

The strengths and weaknesses of this thesis set out a clear map of promising future research directions, each spanning three temporal stages (near-, mid-, and long-term) reflecting their technical maturity, feasibility, and anticipated clinical implications.

5.4.1 Efficient Architecture and Real-Time Inference

The first essential condition for clinical application is that the TARMN model run under the computational constraints of ordinary clinical computers, rather than top-of-the-line A100 GPUs. In the near term, the application of knowledge distillation should allow the compression of the 42-million-parameter TARMN model into a student model of 8-12 million parameters without much loss in terms of the representational power of the Mamba-CNN encoder. Following that, together with INT8 post-training quantization and optimizations of ONNX/TensorRT deployment, the target is sub-2-second volume-based inference on a clinical GPU (e.g., NVIDIA T4) that can be used in real-time in neuronavigators.

5.4.2 Uncertainty Quantification

Tumour segmentation by definition introduces a degree of uncertainty, which must be conveyed to the clinician, not reduced to one fixed deterministic value. Future development will include uncertainty quantification within TARMN through the use of Monte Carlo dropout during inference and, later, through BayesNet layers in the Mamba SSM decoder.

5.4.3 Multimodal Omics Integration

The existing TARMN radiogenomics module can predict molecular biomarkers solely from MRI features. Future work aims to expand the input modalities to digital histopathology whole-slide images and bulk RNA-seq/DNA methylation array when available and to integrate them together with imaging features via hierarchical multimodal attention module. Such a pan-modal integration can significantly enhance the biomarker prediction ability and allow predicting other molecular marker beyond the currently implemented IDH/MGMT such as EGFR amplification, CDKN2A/B homozygous deletion, and TERT promoter mutation. The resulting model can become a powerful tool for computational pathology in glioma molecular profiling.

5.4.4 Longitudinal Tumour Progression Modelling

Serial MRI examinations are carried out throughout a brain tumour patient treatment pathway at diagnosis, after surgery, after radiotherapy, and at regular follow-ups. The current TARMN implementation process each time point independently. Future work will focus on building the time extension model where the Mamba SSM backbone will be used not only on the spatial voxel sequence, but also in the time domain and to model longitudinal tumour growth dynamics. Such a longitudinal Mamba network can be used to predict tumour recurrence, distinguish between tumour progression and pseudo-progression (T1CE enhancement after radiotherapy in the absence of active retrospective assessments).

5.4.5 Prospective Multi-Centre Clinical Validation

The next critical future direction is the need for a prospective multi-centre clinical validation study. The study must take the form of a randomised controlled trial that will compare diagnostic performance of clinicians alone versus clinician plus TARMN on a predefined primary outcome measure. The study must recruit between three and five centres with varied scanner vendors, field strengths, and patient demographics. The data collected will have to incorporate complete molecular annotation and follow-up to allow for the validation of radiogenomics.

5.4.6 Regulatory Compliance and Clinical AI Governance

The use of assisted diagnostic systems in clinical settings is now subject to regulations such as the EU AI Act, FDA SaMD Framework, and IEC 62304 standards for medical device software. The technical specifications for compliance with these frameworks include stratified bias audits by age, sex, ethnicity, and scanner vendor, failure modes reporting based on ISO 14971 risk management practices, post-market surveillance systems to monitor performance drift, and human supervision policies on when the AI output must be reviewed and overridden by a clinician.

Horizon	Direction	Method	Target Outcome
Near-term (1–2 yrs)	Lightweight Mamba deployment	Knowledge distillation, INT8 quantisation	Real-time clinical inference (<2 s/vol)
Near-term (1–2 yrs)	Uncertainty quantification	Monte Carlo dropout + Bayesian inference	Clinician-flagged ambiguous boundaries
Mid-term (2–4 yrs)	Multimodal omics fusion	MRI + histopathology + genomic sequences	Pan-cancer radiogenomic characterisation
Mid-term (2–4 yrs)	Longitudinal progression model	Temporal Mamba SSM across time-points	Treatment response prediction per patient
Mid-term (2–4 yrs)	Prospective multi-centre RCT	3-site clinical trial, n≥500 patients	Regulatory-grade validation evidence
Long-term (4+ yrs)	Causal AI for oncology decisions	Structural causal models + CBM	Auditable AI meeting IEC 62304 / EU AI Act
Long-term (4+ yrs)	Paediatric federated expansion	10+ institution FL with differential privacy	Robust paediatric glioma segmentation

5.5 Conclusions

A brain tumour diagnosis stands among the most serious possible for any patient, and the discrepancy between the high fidelity of modern genomic medicine and the imprecision of our current imaging-based diagnostic capabilities persists as a major clinical challenge. The contributions detailed throughout this thesis are intended to be a step towards addressing that challenge not through some silver-bullet fix, but rather by systematically defining the precise technical challenges posed by the problem (computational, topological, molecular, domain-shift, and interpretability) and solving each one individually within an integrative approach.

The fusion of selective state space models, topological data analysis, radiogenomics, and metaheuristic optimisation into a cohesive end-to-end learnable framework constitutes a qualitatively novel breakthrough in neuro-oncological artificial intelligence a technology that outperforms its constituent parts in terms of accuracy, anatomical realism, clinical relevance, and efficiency.

