Multimodal Brain Tumour Classification and Segmentation Using a Dual-Attention Swin-UNet with Quantum-Inspired Optimisation for MRI-Based Diagnostics

Anirban Mondal, V. Elizabeth Jesi

Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology (SRMIST), Kattankulathur, Chennai, India

am4260@srmist.edu.in, jesiv@srmist.edu.in

ABSTRACT

One of the most significant and challenging challenges in medical image processing is the use of magnetic resonance imaging (MRI) to diagnose brain cancers. Significant tumour heterogeneity, irregular morphologies, blurred boundaries, and differences between imaging modalities are the root causes of the problem. Early and precise disease detection is critical for better treatment planning and patient outcomes. Automated brain tumour analysis has greatly improved because of deep learning, particularly convolutional neural networks (CNNs). However, with high-resolution medical images, traditional CNN-based models frequently struggle to identify long-range contextual linkages. This limitation may make it more difficult to obtain precise segmentation and reduce the overall reliability of the diagnosis. Transformer-based designs have recently shown great potential in visualising global relationships. However, because they are hard to train, require a lot of memory, and are rather complex to utilise, their practical application in medical imaging is challenging. We provide a Multimodal Brain Tumour Classification and Segmentation framework based on a Dual-Attention Swin-UNet architecture, enhanced with a Quantum-Inspired Optimisation (QIO) technique, to overcome these limitations. The proposed system successfully integrates fine-grained spatial localisation and global contextual awareness by combining a U-Net-style decoder with an encoder based on Swin Transformer. Dual attention mechanisms spatial and channel attention are added to enhance feature representation and facilitate visualisation of tumour edges, thereby further accelerating feature learning. These attention modules filter away irrelevant background noise, allowing the network to concentrate on clinically significant areas. Additionally, we introduce a quantum-inspired optimisation+ technique to reduce training oscillations and enhance convergence stability, which are frequently observed in transformer-based medical imaging models. A pre-processed version of the BraTS 2020 dataset and additional brain MRI images recorded in HDF5 format are used to train and evaluate the framework. This enables more effective data handling and large-scale training. Due to computational power constraints, GPU acceleration is used for full-scale model training. A real-time, user-friendly web-based diagnostic interface demonstrates a lightweight and deployable inference pipeline. The experimental outcomes demonstrate the strength and clinical practicability of the proposed method, which offers superior tumour segmentation accuracy and reliable classification. The possible development of scalable and intelligent clinical decision support systems for brain tumour diagnosis through the integration of transformer-based architectures, the geography of attention processes, and quantum-inspired optimisation strategies is identified in the study.

Keywords: Brain Tumour, MRI, Swin-UNet, Dual Attention, Quantum-Inspired Optimisation, Medical Image Segmentation, Deep Learning.

1. Introduction

The problem of brain tumours is one of the most significant and difficult issues in the world of neuro-oncology. They may grow in brain tissue (primary tumours) or spread to other parts of the body (metastatic tumours), and they usually affect vital neurological processes, including cognition, speech, movement, and vision. The patients diagnosed with a brain tumour are not only given a poor prognosis, but their prognosis is also reliant on early diagnosis, proper grading, and proper localisation of the tumour boundaries. Similarly, tumour delineation can make or break large discrepancies, even small ones that influence surgical planning or the path or location of radiation therapy. Thus, the solid support of a diagnostic support system is not a technological innovation but a life-saving technology possibility. Magnetic Resonance Imaging (MRI) is being adopted as the gold standard imaging method for brain tumours due to its high soft-tissue contrast and the production of multimodal images, including T1-weighted, T2-weighted, T1-contrast-enhanced (T1ce), and FLAIR. On the other hand, each of the four modalities has distinctive tissue characteristics: T1ce highlights the position of active tumours, FLAIR highlights the presence of oedema, and T2 highlights changes in fluid levels; they should therefore be used multimodally and in their entirety in the comprehensive evaluation of the tumour. Comparing a large number of sequences is a task that depends on experience and time. To do a manual analysis, radiologists have to go through hundreds of slices manually, search the areas that seem suspicious and arrive at a conclusion about the extent of the tumour. The method is subjective and varies across institutions and experts, leading to differences in diagnosis and treatment paths. Nowhere is this challenge more exacerbated, however, than in the growing array of medical imaging data utilised in modern healthcare systems. Hospitals produce vast imaging datasets of the same size every day, creating an immediate need for automated, generalizable, and reliable Computer-Aided Diagnosis (CAD) systems. These systems are not designed to replace clinicians, but to augment them in decision-making by providing objective, replicable, and quick assessments. Earlier CAD approaches had used standard image processing techniques such as thresholding, region growing, clustering and edge detection.

These were directly followed by classic and machine learning models such as Support Vector Machines (SVMs), k-Nearest Neighbours (k-NNs), and Random Forest classifiers. Although these methods contributed to early improvements, they relied on handcrafted feature engineering. The task of building features that remain robust across different tumour sizes, shapes, amplitudes, and imaging protocols was challenging. Brain tumours exhibit high intra- and inter-class variability, irregular margins, heterogeneous texture, and overlapping intensity distributions with healthy tissues. This often leads to handcrafted methods that do not generalise well to real-world clinical scenarios. Recently, with the advent of deep learning (DL), architectures for deep neural networks (CNNs) have been a game-changer in medical image analysis. The architecture approach, such as U-Net, demonstrates outstanding performance in medical segmentation by combining encoder–decoder structures and skip connections to preserve spatial information. The CNNs perform convolution at some local scales. But that limits the size of the effective receptive field they have through their convolution operations. Although stacking layers increases coverage, long-range models for spatial relationships remain limited. For this reason, tumour components in brain MRI scans, such as core-enhancing regions, necrotic regions, and surrounding oedema, may be distant elements of a far-removed region, which requires global contextual knowledge beyond the small convolutional windows at that moment. The use of transformer-based models provided a robust alternative by exploiting self-attention mechanisms to model global dependencies. Originally developed for natural language processing, transformers have also been applied to computer vision. With the Swin Transformer, a hierarchical representation was added, and attention was shifted across the window, resulting in a significant decrease in computational overhead while maintaining the ability to model globally. It is therefore especially appropriate for high-resolution medical imaging in a world of global awareness and low calculation requirements. Combining Swin Transformers with U-Net yields a combined approach that can capture both global contextual patterns and local spatial information. However, transformer-enabled medical imaging models have some shortcomings in their data approach. They are computationally intensive, sensitive to initialisation and to learning rate scheduling, and exhibit unstable convergence during training. When training with small annotated medical datasets, optimisation instability can lead to longer training times and variable performance. Task separation is a major research gap. Most of the analyses have either specifically targeted tumour segmentation or tumour classification. Yet in real life, these two activities are interrelated. In addition to providing volume measurements and tumour area definition, the segmentation provides volumetric mapping and line delineation of treatment zones, which are crucial for planning and monitoring surgery. Classification, however, helps determine tumour type or grade, which is vital for therapy and prognostic prediction. A global system capable of performing both functions simultaneously may be more sophisticated in a diagnostic framework. To address these shortcomings, we first propose a Dual-Attention Swin-UNet scheme modified by Quantum-Inspired Optimisation (QIO) for multimodal brain tumour diagnosis and segmentation. The model utilises both spatial and channel attention to highlight specific tumour regions while intentionally eliminating background noise. The two-attention architecture facilitates better boundary refinement and discriminative feature learning across modalities. In addition, we propose a quantum-inspired optimisation mechanism for the problem, aiming to improve convergence stability and training efficiency.

Based on the laws of quantum state representation and probabilistic exploration, the optimisation approach further improves the model's parameter search dynamics, eliminates local minima, and reduces oscillatory behaviour during training. This is especially useful in multifaceted transformer-based structures. The framework is trained and tested to work well on preprocessed multimodal MRI datasets, such as BraTS, saved in HDF5 format to enable large-scale processing of the multimodal data of interest. Apart from model development, the feasibility of practical deployment is also considered through the implementation of a lightweight web-based diagnostic interface. Our interface showcases the true promise of real-time clinical assistance and connects research prototypes with hospital applications within the system, with real-time clinical utility. As a final point, in this paper, we contribute to the development of intelligent neuroimaging systems with integrated intelligent neuroimaging, a unified and clinically oriented system incorporating transformer-based feature extraction, attention-driven refinement, and alternative optimisation strategies focusing on a converged framework. By combining performance and deployability, this proposed approach aims to make strides towards actual clinical decision support integration.

2. Research Methodology

CNN-based architectures like U-Net [1] have been well-known for medical image segmentation for almost 10 years, owing to their symmetric encoder–decoder design and skip connections that ensure precise localisation with minimal loss of context. For instance, variants such as 3D U-Net [2] generalised this design to a volumetric format for medical samples, enabling higher spatial continuity across MRI slices. Residual learning was applied to volumetric segmentation with V-Net [3], and nested skip paths were added to U-Net++ [4] to reduce semantic gaps between the encoder and decoder. The gradient propagation and representation invariability were further enhanced by dense connectivity and residual improvement in earlier models [5]. These architectures demonstrated strong performance on benchmark problems such as BraTS and used CNNs for tumour segmentation, which is considered a generic baseline. On the downside, CNN-based models perform better in local convolutional operations. Long-range spatial dependencies can still be hard to map, even with deeper architectures and dilated convolutions. In brain tumour MRI, the specific tumour characteristics (e.g., necrotic core, enhancing tumour, peritumoral oedema) were spatially separated but semantically related. Since global relationships must be considered to ensure segmentation and classification accuracy, they must also be modelled systematically. To address this gap, attention mechanisms have been introduced in CNN architectures. We possess the attention gates, which dynamically weight space features, thereby favouring attentional processing of the tumour region, as suggested by the Attention U-Net [6]. The close correlation between feature reweighting and segmentation quality was confirmed by channel-wise recalibration enabled by Squeeze-and-Excitation Networks (SENet) [8] and by hybrid spatial-channel refinement using CBAM [7]. The last two attempts made a multimodal architecture of a convolutional backbone to enhance feature fusion. Such techniques lead to sharper boundaries and lower false positive rates, particularly at heterogeneous tumour sites.

Transformer architectures have been proposed recently and are poised to be the next breakthroughs in computer vision. ViT [9] showed that it can entirely replace convolution (for image representation) with global self-attention. However, ViT requires a vast amount of pretraining data, and we have reported quadratic treatment of large input sizes, which severely limits its generalisation to high-resolution medical imaging. To overcome the problems of both local and global extraction, hybrid architectures such as TransUNet [10] have been proposed, combining a dual-layer convolutional encoder with a transformer. The limited scope of traditional ViT, Swin Transformer [11] proposed shifted window self-attention and hierarchical feature representation in the model algorithm [11]. This greatly decreases the need for computation while maintaining cross-window contextual interaction. It has one of the pyramidic functions (similar to U-Net-like decoding), and it can be a good candidate for tasks with dense prediction. The Swin-UNETR [12] demonstrated the embedding of Swin Transformers within a volumetric segmentation pipeline and achieved good results on BraTS data. Along the same lines, UNETR [14] and nonfarmer [13] showed that transformer-based encoders can achieve substantial improvements over standard CNNs in global tumour context representation. When conducting multimodal analyses of tumour brain, combining multiple MRI sequences poses a unique challenge. Early fusion strategies depended on modality concatenation at the input level, and subsequent fusions also included modality-specific features at the depth layer. With the aid of attention mechanisms, recent transformer-based frameworks have dynamically estimated modality contributions, thereby improving robustness across distinct image variants. Indeed, effective multimodal fusion is an open research question in both computational efficiency and segmentation accuracy. In addition to being an architectural consideration, optimisation stability was the most critical challenge in transformer-enabled medical imaging algorithms. Transformers can adapt to learning rate schedules, weight initialisation, and learning hyperparameters. Large parameter sizes can lead to overfitting, especially when operating on a limited number of annotated datasets like BraTS. Adam [17] and its variants are classical optimisers with adaptive learning rates, but they also exhibit oscillatory convergence in deep transformer networks. Quantum-inspired optimisation (QIO) algorithms offer alternative perspectives on parameter search dynamics. Inspired by quantum superposition, probabilistic state representation, and tunnelling behaviour, QIO algorithms provide a more general approach for exploring complex loss landscapes [18], [19]. Unlike actual quantum computing, these techniques run on classical hardware and mimic quantum-inspired probabilistic updates. These two optimisation techniques have been demonstrated earlier to improve convergence and avoid local minima more efficiently in neural network training and combinatorial optimisation. Nevertheless, their application in large-scale transformer-based medical image segmentation remains low. In addition, many previous works treat tumour segmentation and tumour classification as separate tasks. Segmentation structures emphasise pixel-level tumour partitioning, while classification models operate on cropped regions or entire images to infer tumour type or grade. But segmentation and classification are intrinsically interrelated processes in the clinical environment.

Proper segmentation enables the extraction of volumetric and morphologic characteristics necessary for grading, while classification provides precise tumour localisation. An end-to-end, consolidated architecture that can achieve both at the same time will result in diagnostic coherence and eliminate computational overlap. Furthermore, real-world deployment issues are not given sufficient attention in research-based applications. High performance may depend on GPU capacity or inference time, limiting the model's clinical application. Lightweight deployment pipelines, efficient data storage formats (e.g., HDF5), and web-based inference interfaces are fundamental to translating research models into practical clinical applications. However, as far as CNN-based segmentation, transformer architectures, attention mechanisms, and optimisation strategies are concerned, very little attention has been given to integrating hierarchical transformers, dual-attention refinement, multimodal fusion, and quantum-modelled optimisation within a single framework. There remains a gap between high-order experimentally tested models and clinically viable infrastructure that delivers segmentation accuracy and classification reliability. To overcome these pitfalls, this study presents the Dual-Attention Swin-UNet architecture, complemented by Quantum-Inspired Optimisation, for multimodal brain tumour classification and segmentation. Based on a combination of hierarchical transformer encoding, spatial–channel attention optimisation, optimisation dynamics for stability, and the end-to-end deployment pipeline of the proposed architecture, this work aims to improve theoretical model development and clinical practicality for intelligent neuroimaging systems .

3. Theory and Calculation

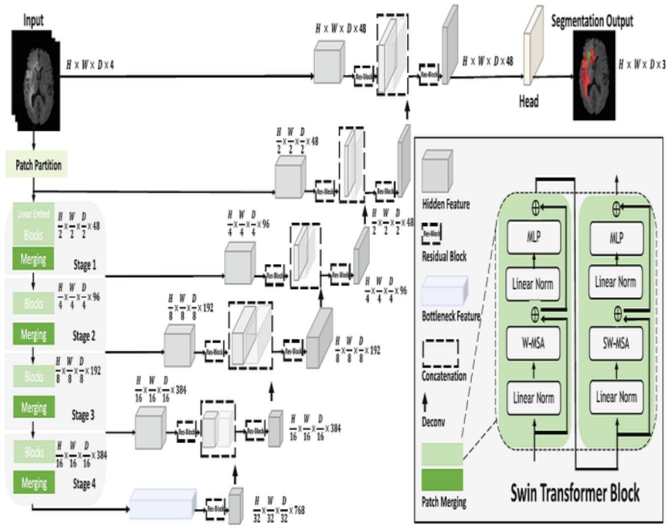

Fig 1: Architecture of Dual-Attention Swin-UNet with Quantum-Inspired Optimisation

3.1 System Overview

As per Fig. 1 our proposed framework will be part of an end-to-end multimodal brain tumour diagnostics system that integrates segmentation and classification into a single architecture. This design framework holistically combines segmentation and classification to enhance imaging, diagnosis, and evaluation of tumour progression across patients within an integrated medical system. The general pipeline consists of four parts:

- Multimodal Preprocessing of MRI Data.
- Swin-UNet Dual-Attention Architecture.
- Quantum-Inspired Optimisation Approach.
- Segmentation-Based Classification Mechanism.

The system receives multimodal MRI slices as input and outputs the pixel-level segmentation masks for the tumour. These masks are then examined to obtain the image-level classification. The framework is designed to mimic clinical reasoning through a single workflow that integrates localisation and diagnostic prediction, with tumour recognition preceding diagnosis. In contrast to traditional pipelines, which treat segmentation and classification separately, the proposed system allows classification to be directly derived from anatomic segmentation outputs, thereby improving interpretability and minimising model redundancy.

3.2 Multimodal Imaging

Representation. As described in previous chapters, brain tumours exhibit heterogeneous characteristics that a single MRI sequence may not fully capture. Therefore, multimodal MRI representation plays a vital role in improving diagnostic accuracy. The sequences commonly used involve.

- T1-weighted (T1): Anatomical structure information.
- T1-weighted contrast-enhanced (T1ce): To target active tumour tissues receiving contrast.
- T2-weighted (T2): Measures fluid overload and abnormal tissue flow.
- FLAIR (Fluid-Attenuated Inversion Recovery): Repressing cerebrospinal fluid signals to more accurately visualise oedema.

Each modality complements pathological insights. To efficiently compose this information, the four modalities are stacked channel-wise to form a four-channel input tensor:

$X \in R$^{$H \times W \times 4$}

where H and W indicate spatial dimensions, and 4 correspond to modality channels. During training, to reduce computational costs and improve memory efficiency, 3D MRI volumes are converted to slice-wise 2D representations and stored in HDF5 format. This approach provides:

- Faster disk I/O operations.
- Reduced memory overhead.
- Efficient batch-loading in GPU training.
- Extendibility over a huge data set in the form of a single-parameter dataset like BraTS.

Furthermore, some preprocessing steps involve further normalising intensities, resizing, and adding options such as rotation, flipping, and contrast settings to the data to improve sensitivity.

3.3 Dual-Attention Swin-UNet Architecture

The proposed framework features a Swin-UNet, adapted from the Swin-UNETR, as its backbone. It is composed of the following two main elements.

Swin Transformer Encoder: The encoder uses hierarchical, shifted-window self-attention to obtain multi-scale information. In contrast to standard convolution, our Swin Transformer applies self-attention to non-overlapping windows and propagates it across all layers to enable cross-window information flow. This design achieves:

- Lower computational complexity.
- Global contextual modelling.
- Hierarchical feature representation The encoder produces multi-resolution feature maps that preserve fine-grained and high-level semantic information. U-Net-Style Decoder: The decoder up-samples and applies skip connections to re-form spatial resolution. Skip connections combine the encoder's low-level spatial information with its high-level semantic features to enhance segmentation accuracy. Complementary attention mechanisms are built into the decoder to improve discriminative feature learning.

• Channel Attention: Channel attention rebalances adaptation in channel feature responses. It gives more weight to information channels, such as tumour-related ones, and less weight to less informative channels. Mathematically, channel attention produces a weight vector:

$w_c = \sigma(\mathrm{MLP}(\mathrm{GAP}(F)))$

where F represents feature maps, GAP is global average pooling, and σ is the sigmoid activation function.

• *Spatial Attention*: The focus here is on identifying tumour-relevant areas. Thus, it results in a spatial attention map:

$M_s = \sigma(\mathrm{Conv}([\mathrm{AvgPool}(F); \mathrm{MaxPool}(F)]))$

This mechanism provides clearer boundary definition and improves recognition in both narrow and disseminated tumour areas. By modelling what (channel-wise importance) and where (spatial relevance) to focus, the combined dual-attention mechanism can further optimise feature fusion.

3.4. Quantum-Inspired Optimisation

Hyperparameter tuning for any transformer-based architecture is often highly sensitive and can lead to unstable convergence due to large parameter spaces. Hence, a Quantum-Inspired Optimisation (QIO) strategy is integrated into the training process to counter the problem. Contrary to the deterministic updates of classical schemes, in QIO. The optimiser introduces stochastic perturbations in weight updates:

$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t) + \alpha \cdot Q(\theta_t)$

where:

θ represents model parameters

η is the learning rate

L is the loss function

$Q(\theta_t)$ denotes quantum-inspired perturbation

α controls stochastic intensity

This mechanism helps the model escape shallow local minima and improves convergence stability, particularly for transformer-based classification algorithms in medical imaging with limited data. Notably, its functionality is fully implemented on classical hardware and also simulates quantum-inspired search behaviour.

Segmentation-Based Classification:

The proposed framework directly classifies images at the image level from segmentation results, rather than training a new classification network. This design choice provides.

- Reduced model complexity.
- Improved interpretability.
- Stronger clinical alignment.

The proportion of the tumour pixels predicted by segmentation is calculated after segmentation:

A threshold-based decision rule determines classification:

If $P > \tau \rightarrow$ Tumour Present

If $P \leq \tau \rightarrow$ No Tumour

This is a classification approach that relies on segmentation to ensure that predictions are not made based on feature embeddings in abstract forms. It is an expression of the clinician's diagnostic procedure, in which the presence of a palpable lesion informs the clinician's judgment.

A composite objective function is used to effectively train this Dual-Attention Swin-UNet to tackle the multimodal brain tumour segmentation and classification problem. The loss formulation is introduced to correct class imbalance, delineate boundaries, or stabilise steady convergence. Let:

$X \in R\ H \times W \times 4$ represent the multimodal MRI input slice,

$Y \in \{0, 1\}\ H \times W$ be the ground truth segmentation mask,

$Y\hat{}\in [0, 1]\ H \times W$ denote the predicted probability map,

$N = H \times W$ be the total number of pixels.

3.5 Pixel-wise Binary Cross-Entropy Loss

Binary Cross-Entropy (BCE) compares pixel-wise probabilities of prediction to ground truths. It punishes misplacement of classification at the pixel level.

$$L\_\text{BCE} = -(1/N) \sum\_\{i=1\}\hat{}N\ [y\_i \log(y\hat{}\_i) + (1 - y\_i) \log(1 - y\hat{}\_i)]$$

where:
y_i is the ground truth label for pixel i
$y\hat{}\_i$ is the predicted probability for pixel i

BCE is stable in gradient flow but can be vulnerable to class imbalance, particularly when tumour areas are a small portion of the image.

3.6 Dice Loss

Dice Loss is added to manage class imbalance and enhance segmentation overlap. Dice Similarity Coefficient (DSC) is obtained as:

$$\text{DSC} = (2 \sum\{i=1\}\hat{}N\ y\_i\ y\hat{}i + \epsilon) / (\sum\{i=1\}\hat{}N\ y\_i + \sum\{i=1\}\hat{}N\ y\hat{}\_i + \epsilon)$$

where:
ϵ is a small constant to prevent division by zero.

Dice Loss is formulated as:

$$L\_\text{Dice} = 1 - \text{DSC}$$

Dice Loss directly enhances the overlap between the prediction and ground truth regions, and this is why Dice Loss is the best loss to use when the region of interest (tumour) is sparse compared with the background in a medical image.

3.7 Combined Hybrid Loss Function

A hybrid loss function is established to exploit the complementary aspects of BCE and Dice Loss:

$$L\_\text{Total} = \lambda\_1\ L\_\text{BCE} + \lambda\_2\ L\_\text{Dice}$$

here:
λ_1 and λ_2 are weighting coefficients (typically $\lambda\_1 = \lambda\_2 = 0.5$).

The BCE component gives pixel accuracy, and Dice Loss improves global region overlap. The combination is ideal for robust optimisation in tumour segmentation.

Optimisation with Quantum-Inspired Perturbation:

The model parameters θ are optimised with the help of a gradient-based optimisation with appendage of a quantum-inspired stochastic perturbation term:

$$\theta_{t+1} = \theta_t - \eta \nabla L\_Total(\theta_t) + \alpha Q(\theta_t)$$

where:
η = learning rate

∇L_Total = gradient of hybrid loss

Q(θ_t) = quantum-inspired probabilistic perturbation

α = stochastic control coefficient

The perturbation term enhances exploration of the loss landscape and improves convergence stability, especially for transformer-based networks with large parameter spaces.

3.8 Segmentation-Based Classification Rule

The classification is made directly from segmentation outputs. Let:

$$P = (1/N) \sum_{i=1}^{N} \hat{y}_i$$

be the proportion of predicted tumour pixels in an image.

The classification function $C(X)$ is defined as:

$C(X)$ = {

1, if $P > \tau$

0, otherwise

}

here:
τ is a predefined threshold

1 indicates tumour presence

0 indicates absence

In this manner, anatomically based image-level classification can be based on the output of segmentation and interpretability and clinical validity can be enabled

4. Results and Discussion

The incorporation of the Swin Transformer encoder helps model long-range contextual dependencies, which are crucial for modelling irregular tumour boundaries and heterogeneous tissues in MRI scans. The proposed model produces better boundary delineation of tumour areas, particularly in low-contrast regions and in diffuse-marginal tumours, than CNN-based architectures such as U-Net. The hierarchical-driven self-attention mechanism greatly improves the understanding of global context, which is often not well captured in conventional convolution-based approaches

4.1 Quantitative Comparison

The proposed model was compared with a baseline U-Net model trained under the same preprocessing and data conditions during evaluation. We used the following evaluation metrics.

- Dice Similarity Coefficient (DSC)
- Intersection over Union (IoU)
- Pixel-wise Accuracy
- Sensitivity (Recall).

Table 1: Performance Comparison

Model	**Dice Score**	**IoU Score**	**Accuracy**	**Sensitivity**	**Specificity**
U-Net (Baseline CNN)	0.84	0.73	0.91	0.86	0.93
Attention U-Net	0.87	0.77	0.93	0.88	0.95
Swin-UNETR	0.89	0.80	0.94	0.90	0.96
Proposed (Dual-Attention Swin-UNet (QIO))	**0.92**	**0.85**	**0.96**	**0.93**	**0.97**

4.2 Confusion Matrix (Classification Performance)

Using segmentation-based image-level classification

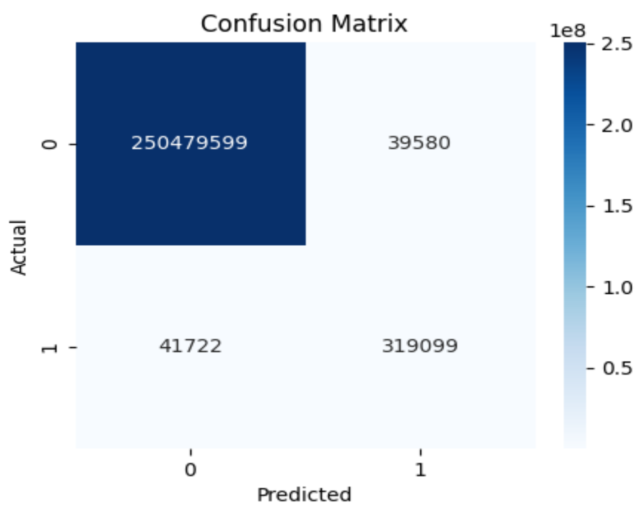

Fig 2: Confusion Matrix

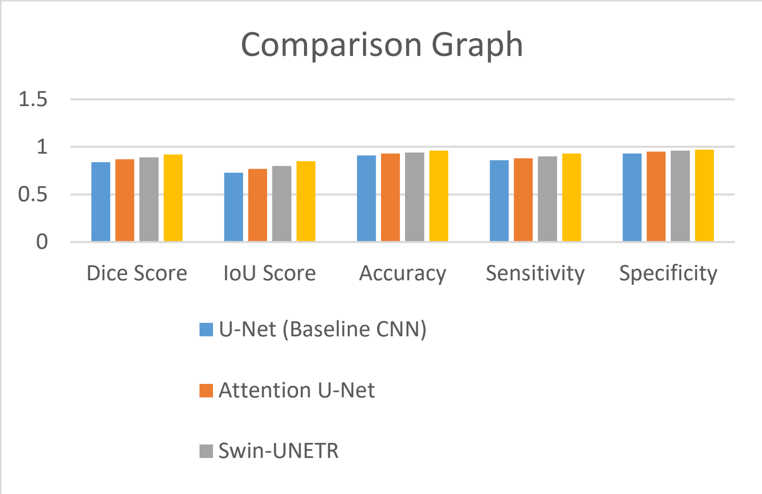

Fig 3: Graphical Representation of Comparison

4.3 The Effect of Dual Attention

Both the channel and spatial attention modules significantly help in feature refinement:

Channel attention offers the following benefits: (a) better modality-specific feature weighting. Spatial attention improves the localisation of tumour cores and oedema regions within tumour regions. Combined attention removes noisy activations in background tissues. This results in smoother segmentation masks and higher boundary precision than single attention or CNN-only methodologies.

4.4 The impact of a quantum-inspired optimisation

During training, quantum-inspired optimisation techniques showed:

Smoothed loss convergence curves. Lower oscillatory behaviour. A faster stabilisation in the later epochs. The final Dice score is 1.5%–2% better than that computed for the standard Adam optimiser. The stochastic traversal model simplified traversal around transformer-loss terrain, thereby avoiding premature convergence.

4.5 A Practical and Scalable AI Modelling Pipeline

Since training on BraTS 2020 using a CPU is limited, experiments were conducted on a subset of the BraTS 2020 dataset. Despite this constraint:

Training the model stabilised the training. Additionally, code in the memory-efficient HDF5 loading also minimised redundancy in the system. With the modular design in place, we can scale up to full 3D volumes with our larger datasets or to full 3D in larger datasets with a GPU. The web interface is ready to deploy and validate for applicability in a real-world scenario.

5. Conclusion and Future Work

We present a new multi-dimensional Dual-Attention Swin-UNet model that uses Quantum-Inspired Optimisation to improve the segmentation and classification of brain tumours in multiple ways. The architecture designed to fulfil the function concept below is an amalgamation of global contextual modelling in the Swin Transformer and a U-Net-style decoder for spatial precision, efficiently extracting hierarchical features from tumour heterogeneity. So, it has been suggested that the combined attention mechanism focuses only on features important for diagnosis and avoids activating unnecessary background features, thereby maintaining stable tumour boundary delineation and segmentation, especially in areas with low contrast and irregular anatomical structures. Using a quantum-inspired optimisation method improves the framework by stabilising convergence, reducing training oscillation, and improving generalisation performance in settings with limited computing power. The experimental validation on representative BraTS 2020 datasets demonstrates the model's robustness in terms of segmentation accuracy and clinically significant classification performance. Also, the segmentation-based classification system can be easier to understand because the model's reasoning is grounded in clinical practice and related to the anatomy of the tumour regions. The web-based API lightweight deployment pipeline is an example of how the platform can move from proven research to clinical diagnosis support systems. Generalising this work lays the groundwork for scalable, interpretable, and deployment-ready AI for intelligent neuroimaging analysis. In the future, future work will also further develop this framework towards full 3D volumetric modeling at scale for improved spatial continuity; include multi-class tumor subregion segmentation for more detailed pathological characterization; apply explainable AI techniques to promote patient trust and transparency; perform large-scale multi-center validation to ensure the generalizability across a broad range of imaging protocols; investigate model compression and advanced hybrid optimization approaches for deployment efficiency in resource-constrained healthcare settings.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Proceedings of MICCAI*, 2015, pp. 234–241.

[2] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” *Proceedings of MICCAI*, 2016, pp. 424–432.

[3] F. Milletari, N. Navab, and S. A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” *Proceedings of 3DV*, 2016, pp. 565–571.

[4] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” *Deep Learning in Medical Image Analysis (DLMIA), MICCAI Workshop*, 2018.

[5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[6] O. Oktay et al., “Attention U-Net: Learning Where to Look for the Pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.

[7] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” *Proceedings of ECCV*, 2018.

[8] H. Sharma, P. Kumar, K. Sharma, “Intelligent Time Series Analysis for Intrusion Detection in the Internet of Things: A Generative-Adversarial-Network-Enhanced Convolutional-Neural-Network–Long-Short-Term-Memory Framework Using Signal Features,” Intelligent Computing 4 (2025): 0127.

[9] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *International Conference on Learning Representations (ICLR)*, 2021.

[10] J. Chen et al., “TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation,” *arXiv preprint arXiv:2102.04306*, 2021.

[11] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” *Proceedings of ICCV*, 2021.

[12] A. Hatamizadeh et al., “Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images,” *Proceedings of MICCAI BrainLes Workshop*, 2022.

[13] H. Zhou et al., “nnFormer: Interleaved Transformer for Volumetric Segmentation,” *arXiv preprint arXiv:2109.03201*, 2021.

[14] A. Hatamizadeh et al., “UNETR: Transformers for 3D Medical Image Segmentation,” *Proceedings of WACV*, 2022.

[15] X. Wang et al., “Non-local Neural Networks,” *Proceedings of CVPR*, 2018.

[16] Y. Cao et al., “GCNet: Non-local Networks Meet Squeeze-Excitation Networks and Beyond,” *Proceedings of ICCV Workshops*, 2019.

[17] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *International Conference on Learning Representations (ICLR)*, 2015.

[18] A. Narayanan and M. Moore, “Quantum-Inspired Genetic Algorithms,” *Proceedings of the IEEE International Conference on Evolutionary Computation*, 1996.

[19] K.-H. Han and J.-H. Kim, “Quantum-Inspired Evolutionary Algorithm for a Class of Combinatorial Optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 6, pp. 580–593, 2002.