Open Access

OBM Neurobiology

LIDSEN Publishing Inc.

Original Research

Precision Brain Tumour Diagnosis with Convolutional Recurrent Attention Networks

S.L. Jany Shabu [1], L. Lakshmanan [1], J. Refonaa [1], S. Dhamodaran [1], Aimin Li [2], Saurav Mallik [3, 4, *]

1. Department of Computer Science and Engineering, Sathyabama Institute of Science & Technology, Chennai, India; E-Mails: Janyshabu.cse@sathyabama.ac.in; janyshabu@gmail.com; lakshmanan.cse@sathyabama.ac.in; refonna.cse@sathyabama.ac.in; dhamodaran.cse@sathyabama.ac.in
2. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, Shaanxi, China; E-Mail: liaiminmail@gmail.com
3. Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA; E-Mails: sauravmtech2@gmail.com; smallik@hsph.harvard.edu; smallik@arizona.edu
4. Department of Pharmacology and Toxicology, University of Arizona, Tucson, AZ, USA

* **Correspondence:** Saurav Mallik; E-Mails: sauravmtech2@gmail.com; smallik@hsph.harvard.edu; smallik@arizona.edu

Academic Editor: Antonio Meola

Collection: Tumors of the Central Nervous System

OBM Neurobiology
2026, volume 10, issue 3
doi:10.21926/obm.neurobiol.2603351

Received: May 25, 2026
Accepted: August 10, 2026
Published: August 28, 2026

Abstract

Histological heterogeneity and complex imaging patterns make brain tumour diagnosis difficult. In order to make timely clinical decisions and plan personalised treatment, MRI multi-class brain tumour classification must be accurate. Expert evaluation in conventional image interpretation increases diagnostic burden and inter-observer variability. While current AI-based tumour classification methods have improved, learning discriminative spatial-contextual representations and giving clinically relevant model interpretability remain challenges. This article offers Convolutional Recurrent Attention Network (CRANet) for multi-class brain tumour classification using T1-weighted MRI data to overcome these issues.

Convolutional Neural Networks (CNNs) extract hierarchical spatial features, Recurrent Neural Networks (RNNs) model contextual dependencies among deep feature representations, and a Channel-Spatial Attention Module emphasises diagnostically informative features during classification in the proposed architecture. Cross-validation evaluates generalisation performance, whereas noise reduction, intensity normalisation, and data augmentation improve data quality and model resilience. Experiments show that CRANet outperforms state-of-the-art deep learning methods in conventional diagnostic measures. Attention activation maps improve model transparency and clinician interpretation by visualising picture areas contributing to classification decisions, without localising tumours.

Keywords
Brain tumour diagnosis; convolutional neural networks; recurrent neural network; attention networks; classification accuracy

1. Introduction

Accurate and prompt identification is typically necessary to allow successful treatment of brain tumours, which rank among the most urgent health concerns. Diagnosis is more difficult since these tumours may vary greatly in kind, grade, and location [1]. It controls the neurological system and ensures that billions of cells work together. Malignant brain tumours are a major health risk because they disrupt neurologically-regulated organ systems and may even be fatal [2]. Seizures, headaches (both sudden and chronic), impaired vision, personality changes, neurological impairments, musculoskeletal difficulties, and vertigo are all possible symptoms of a brain tumour [3]. When they start in brain tissue, these tumours are called primary. When they get to the brain via the circulatory system from some other body area, physicians say that they are secondary. However, the speed with which brain tumours may be detected highly depends on the radiologist's level of expertise [4]. Establishing a diagnostic mechanism is crucial for utilizing MR imaging to diagnose malignancies. Using this approach, the diagnostic process may remain objective, and the likelihood of biased procedures can be significantly reduced. Medical science has changed dramatically due to ML and AI [5]. With new technologies, radiologists can better address health issues by classifying MRI images. Medical imaging is frequently used and shown to identify cancer [6]. Its non-invasiveness makes it notable. Medical imaging helps categorise brain tumours by providing a complete image of brain tissue. Different tumours vary in size, form, and density [7]. Tumours may present differently despite their superficial similarities. Medical databases include many images, making neural network MRI scan classification difficult [8]. The data quantities may increase as MRI image production from diverse perspectives improves [9]. The data should be pre-processed before being fed into various networks for improved classification accuracy [10]. CNNs' robust advantages are enhanced feature extraction skills and decreased pre-processing needs. Unfortunately, the diagnostic performances of conventional approaches are still insufficient for incorporation into clinical procedures. Convolutional neural networks (CNNs), a DL technique, have dominated the fast breakthroughs in computer vision during the last decade [11]. In contrast to more conventional forms of machine learning, DL can do tasks like classification, object identification, and

segmentation all at once after determining which characteristics are most relevant to the task [12]. In addition, its performance is enhanced by massive datasets, which is a huge advantage. Previous research has shown that DL effectively detects ICH on non-contrast head CT scans, which supports its use in clinical practice [13]. However, it is extensively recognized that deep learning models' performance is best investigated using unknown test data, ideally from an outside source, to determine the models' generalizability with precision. Nevertheless, there has been a dearth of research into the clinical workflow integration of DL models or the generalizability of DL on multi-centre data sets [14].

1.1 Problem Statement

Accurately identifying brain tumours using magnetic resonance imaging (MRI) scans continues to be a serious problem in medicine. Even though manual evaluation of magnetic resonance imaging (MRI) pictures by radiologists may provide useful insights, this technique is time-consuming and prone to human errors. The automated approaches that are now available often struggle to identify and categorize the many types of brain tumours effectively. This study builds a deep learning model to overcome current limitations in brain tumour detection.

1.2 Motivation

Brain tumours are growing and the healthcare sector is becoming more reliant on imaging technology, thus reliable and effective diagnostic tools are needed. Healthcare providers' manual diagnosis is limited by time and subjectivity. There is also the possibility that patients may suffer serious effects if they get an inaccurate or delayed diagnosis. As a result, using sophisticated machine learning models to automate the diagnostic process can greatly enhance healthcare results. Using deep learning, the aim is to develop a system capable of processing MRI images in a manner that is more effective, accurate, and consistent than the approaches that are now in use.

1.3 Novelty and Contributions

- The paper introduces a hybrid model combining CNNs and RNNs to enhance feature extraction from MRI scans and better capture temporal needs in brain tumour data.
- Attention mechanisms such as the Channel-Spatial Attention Module model are incorporated into the system to selectively focus on the most appropriate regions of the MRI images, improving the network's ability to detect tumours with higher accuracy.
- The model outperforms classical deep learning in classification accuracy and diagnostic precision.

1.4 Organization of the Paper

The introduction discusses brain tumour diagnosis's origins and relevance, and Section 2 discusses approaches and their limitations. Section 3 describes the Convolutional Recurrent Attention Network framework. Section 4 evaluates the proposed model on benchmark datasets against existing models. In Section 5, the article discusses the model's therapeutic implications and suggests additional research.

2. Related Works

This section delineates the method for classifying MRI brain images into tumour or non-tumour categories. The vast array of medical image analysis applications in healthcare, especially inpatient investigation and diagnosis, has piqued the interest of researchers and observers during the last two decades. Strategies for classifying brain pictures and analyzing brain architecture based on machine learning have been proposed in studies.

2.1 Deep Learning Architectures for Brain Tumour Recognition

Takowa Rahman and Md Saiful Islam [15] suggested parallel deep CNN (PDCNN) for MRI brain tumour recognition and classification. The model can learn local and global information thanks to two deep CNNs combined simultaneously with varying window sizes. This allows for the advantages of parallel pathways. Three different kinds of MRI datasets are employed to test how well the approach works. With accuracies of 97.33%, 97.60%, and 98.12%, respectively, the binary tumour identification Figshare dataset-II, dataset-I, and Multi-class Kaggle dataset-III are available.

Muhammad Attique Khan et al. [16] proposed the deep saliency map and enhanced dragonfly optimization algorithms for Multi-modal brain tumour recognition and classification. The first phase of the suggested model includes fusion-based contrast augmentation. In the next step, saliency map-based tumour segmentation is suggested and mapped on active contour pictures. Next, pre-trained CNN models dubbed EfficientNetB0 were fine-tuned and trained using improved tumour localization pictures. Finally, an extreme learning machine classifies the best attributes. The simulation procedure enhanced accuracy to 95.14%, 94.89%, and 95.94% on 3 public databases.

Abdullah A. Asiri et al. [17] recommended the g Fine-Tuned CNN with ResNet50 and U-Net Model for Brain Tumor Recognition and Classification. To identify and categorize photos as having tumours or not, the suggested CNN and ResNet50 models are used. The U-Net model is used for accurate tumour area segmentation. These metrics evaluate the model's performance: similarity index (SI), intersection over unions (IoU), accuracy, and dice similarity coefficient (DSC). Here are the findings from the ResNet50 model: DSC: 0.95, IoU: 0.91, SI: 0.95. ResNet50 with U-Net had the best results compared to all other models regarding tumour region classification and segmentation accuracy.

Ayesha Jabbar et al. [18] discussed the CapsNet with the VGGNet models (Caps-VGGNet) for Brain Tumour Recognition and Multi-Grade Segmentations. The given model automatically extracts and classifies features, which solves the issue of big datasets being needed. The efficacy of the recommended technique was measured utilizing the Brats-2020 and Brats2019 datasets, which include high-quality pictures of brain tumours. On the Brats20 database, the shown hybrid model attained precise results with an accurateness of 0.99, a specificity of 0.98, and a recall of 0.99.

Muneeb A. Khan and Heemin Park [19] proposed the convolution block base architecture (CBBA) for Multi-class Brain Tumour Recognition utilizing MRI. With a precision of 97.63%, the accuracy rate of 97.52%, specificity of 98.32%, recall of 97.18%, and F1-score of 97.36%, the model's outstanding diagnostic accuracy is highlighted by extensive assessments on three varied datasets. These findings have surpassed contemporary approaches, comprising existing models like VGG19, VGG16, EfficientNet, MobileNet, Xception, ResNet50, and DenseNet121. Radiology diagnostics and brain tumour detection stand to benefit greatly from its versatility, which might lead to widespread clinical use.

2.2 Data Augmentation and Synthetic Image Generation

Saswati Sahoo et al. [20] deliberated the generative adversarial networks and convolution neural network (GAN-CNN) for Brain Tumour Detection. A GAN ensemble was utilized to increase the brain tumour imaging data, which is input into a hybrid modulated CNN methodology for classification. With 98.85% for accuracy, 98.45% for precision, 97.2% for recall, 98.11% for F1-score, and 98.09% for negative predictive value (NPV), PGGAN was the best performer in the study. Furthermore, a progressive-growing GAN (PGGAN) determines an extremely low latency of 3.4 s.

Saurabh Mandloi et al. [21] investigated DL and layer-wise relevance propagation (LRP) for an explainable brain tumour recognition and classification model. The first step is to create synthetic MRI pictures of different classes utilizing conditional generative adversarial networks (cGANs) to address data imbalance and overfitting. LRP were used to understand the model's output in the final stage. The tumour detection model has an accuracy of 99.6% during training, 99.2% during validation, and 99.0% during testing. The experiment outcomes validate that the pre-trained model from InceptionResNetV2 outperforms the other pre-trained model. Nevertheless, EfficientNet-B0 outperformed the other models in tumour categorization. Achieving precision levels of 99.3% in training, 99.2% in validation, and 100% in testing, respectively.

2.3 Segmentation Techniques for Tumour Localization

Usharani Bhimavarapu et al. [22] presented the Fuzzy C-Means Segmentation Algorithm (FCMSA) for Brain Tumour Detection and Categorization. The most important form, texture, and hue features are chosen to simplify the process. The tumours are classified with a recall of 99.25%, precision of 99.14%, and accuracy of 98.56% using the enhanced Extreme Learning system. Compared to previous models, the suggested classifier always outperforms them in accuracy for all types of tumours. The suggested model outperforms competing models in accuracy by a margin of 1.21% to 6.23%.

Homa and Rajini [23] introduced the reduced complexity spatial fusion CNN (RCSF-CNN) for Enhancing Brain Tumour Detection and Classification. Image features, including energy, contrast, entropy, variance, smoothness, and standard deviation, are extracted to capture critical detecting qualities. The proposed approach demonstrates the efficacy and superiority of the enhanced DL technique for brain cancer detection when coupled with the discrete orthogonal Stockwell transform (DOST) as a transitional step. The entropy, energy, and contrast values for brain tumour detection are 0.008, 0.8155, and 0.354, respectively.

2.4 Optimization Techniques for Model Performance

Asif Raza et al. [24] examined Inceptionv3 and DenseNet121 for brain tumour classification. The experiments showed that the DenseNet-121 model, which used TL, more accurately identified and classified brain tumours. DenseNet-121 had 92.11% precision, 99.95% accuracy, 96.7% recall, and 94.8% F1-measure score in the classification test. With a validation accuracy of 92.42% and training accuracy of 100%, DenseNet-121 showed promise as a reliable approach for diagnosing brain cancers.

Muthuvel Arumugam et al. [25] recommended the Crossover Smell Agent Optimized Multi-layer Perception (CSA-MLP) for accurate brain tumour classification on MRI images. The images are

selected from MR, Brain MRI, and Brain tumour databases after noise removal. CNN classifier was also used to distinguish healthy and sick brain cells. The Multi-Layer Perceptron increased the classification method's accuracy and performance. MLP and CSA optimisation improve classification accuracy. The recommended method outperformed prior methods with an accuracy of 98.56%.

Ganie and Pacal [26] proposed DiSCNet, a lightweight framework incorporating Directional Split Convolution (DiSC), Global Response Normalization (GRN), and Efficient Channel Attention (ECA) to improve robustness against scanner and protocol variability. Evaluated on a unified benchmark of 17,888 MRI images collected from five public datasets, DiSCNet achieved 99.22% accuracy, 99.16% precision, 99.30% recall, and 99.23% F1-score while requiring only 2.78 million trainable parameters and 0.85 GFLOPs, outperforming 71 contemporary CNN-, Vision Transformer-, and hybrid-based models. Ganie and Pacal [27] introduced GBA-NET, a gated bottleneck and attention-driven architecture for ischemic stroke segmentation across CT and diffusion-weighted MRI, where gated bottleneck blocks and attention mechanisms effectively enhanced contextual feature representation and lesion boundary delineation, yielding superior Dice and IoU performance across the evaluated benchmark datasets. Cakmak and Pacal [28] compared 11 Transformer architectures for four-class brain tumour classification, including ViT, DeiT, and Swin Transformer variants. All evaluated models had classification accuracy above 98.8%, with Swin-Small and Swin-Large reaching 99.37%. Notably, Swin-Small had identical diagnostic performance with 48.84 million parameters, 17.09 GFLOPs, and 0.54 ms inference time, a much better accuracy-efficiency trade-off than Swin-Large. Pacal [29] proposed a CNN-based breast ultrasound classification strategy using Chaotic Learning Rate Scheduling (CLRS), showing that adaptive learning-rate optimisation improves convergence stability and classification performance without increasing architectural complexity, emphasising the importance of optimisation strategies for robust medical image analysis.

The augmented reality decision support system AR-EpiAid by Anari et al. [30] integrates multimodal epilepsy data for real-time clinical interpretation, showing that multimodal information fusion and interactive visualisation can improve diagnostic decision-making. The Explainable Attention-Guided Swin Transformer Network by Ranjbarzadeh et al. [31] used hierarchical self-attention and explainable attention maps to accurately delineate brain tumour volumes from 3D MRI and improve model interpretability for clinical applications. Kasgari et al. [32] developed a Spiking Convolutional Neural Network (SCNN) based on a Spike-Timing-Dependent Plasticity (STDP) learning mechanism for glioma brain tumour segmentation, showing that biologically inspired spiking neural architectures can capture complex tumour characteristics while reducing computational redundancy and improving segmentation A Global-Local 3D Brain Tumour Segmentation Model by Ranjbarzadeh et al. [33] uses Vision Transformers and Axial State-Space Modelling to jointly learn global contextual representations and local anatomical features from volumetric MRI data, improving segmentation accuracy and long-range dependency modelling.

2.5 Summary and Novelty

Table 1 shows the summary of the related study. Deep learning methods for brain tumour classification mostly use standalone Convolutional Neural Networks (CNNs) to learn spatial representations or sequentially combine CNNs with recurrent architectures and attention mechanisms to optimise contextual dependency modelling and feature selection. Several attention-based methods improve feature saliency, while recurrent models capture contextual information

without adaptive feature weighting during representation learning. The proposed Convolutional Recurrent Attention Network (CRANet) uses a unified architecture to convert hierarchical convolutional feature maps into contextual feature sequences and refine recurrent hidden representations before classification. This integrated representation learning strategy enables simultaneous optimization of spatial feature extraction, contextual dependency modeling, and discriminative feature weighting within a single end-to-end framework. Consequently, CRANet improves feature discrimination and classification robustness while providing interpretable attention activation maps that support diagnostic decision-making.

Table 1 Summary of the related works.

Author	Method	Advantages	Limitations
Takowa Rahman and Md Saiful Islam [15]	Parallel Deep CNN (PDCNN)	Learns both local and global features; Parallel pathways for better accuracy	Computationally intensive; Requires large datasets for effective training
Muhammad Attique Khan et al. [16]	Deep saliency map, improved dragonfly optimization, and EfficientNetB0 fine-tuning	High accuracy with improved tumour localization and contrast enhancement	High complexity; Requires extensive fine-tuning of models
Abdullah A. Asiri et al. [17]	Fine-tuned CNN with ResNet50 and U-Net	High accuracy for classification and segmentation (IoU, DSC, SI metrics)	Complex architecture; High computational requirements
Ayesha Jabbar et al. [18]	CapsNet with VGGNet for brain tumour detection and multi-grade segmentation	High specificity, recall, and accuracy; Handles limited datasets effectively	Needs large datasets for full potential; May not generalize well for other datasets
Saswati Sahoo et al. [20]	Generative Adversarial Networks and Convolutional Neural Networks (GAN-CNN)	High accuracy; Low latency; Effective augmentation of MRI data	GAN training can be unstable, Limited to specific types of tumours
Usharani Bhimavarapu et al. [22]	Fuzzy C-Means Segmentation Algorithm (FCMSA)	High recall and precision: Effective in classifying different tumour types	It may not be as robust for more complex cases; it Requires careful feature selection
Homa and Rajini [23]	Reduced Complexity Spatial Fusion CNN (RCSF-CNN) with Discrete Orthogonal Stockwell Transform (DOST)	High efficacy and superior performance for tumour detection	Requires significant computational resources; Needs precise feature extraction

Saurabh Mandloi et al. [21]	Deep learning with conditional GAN and Layer-wise Relevance Propagation (LRP)	High classification accuracy and model interpretability with LRP	Sensitive to data imbalance, the GAN model can be prone to overfitting
Asif Raza et al. [24]	Inceptionv3 and DenseNet121 for brain tumour classification	High accuracy and precision; Excellent for reliable diagnostic purposes	DenseNet-121 is computationally expensive and may require fine-tuning for optimal performance
Muthuvel Arumugam et al. [25]	Crossover Smell Agent Optimized Multi-layer Perception (CSA-MLP)	High accuracy; Improved classification results with Multi-Layer Perceptron	Optimization can be slow; Multi-layer perceptron can be sensitive to noise
Muneeb A. Khan and Heemin Park [19]	Convolution Block Base Architecture (CBBA) for multi-class brain tumour detection	High diagnostic accuracy across multiple datasets; Versatile and adaptable to various datasets	Requires extensive evaluation across diverse datasets; Limited to MRI data

2.6 Ethics Approval Statement

This research paper based on computational study involves no human participants, patient contact, or clinical data, and thus requires no ethical approval.

3. Proposed Convolutional Recurrent Attention Network (CRANet)

Abnormal cell growth in the brain is the characteristic of brain tumours, one of the most lethal types of cancer. Although diagnosing brain tumours early is crucial for effective treatment, skilled physicians' slow, laborious, and error-prone manual segmentation might slow the diagnostic process. For this reason, it is important to create automated algorithms that can segment brain tumours. High levels of inter- and intra-tumour variation in morphology, texture, and form make precise segmentation of the core and increased tumour areas challenging. The proposed Convolutional Recurrent Attention Network (CRANet) combines CNNs and RNNs to extract spatial features from T1-weighted MRI imaging data, namely MRI images. The dataset is taken from the Brain Tumour MRI Dataset [34]. This research uses 2D contextual data and 3D sequential spatial data to segment brain tumours using two distinct inputs, i.e., 2D and 3D patches of MRI images. This research first pre-processed the scan to improve the tumour contrast and standardize the dimensions. Using the suggested CRANet architecture, this research subsequently conducted slice-by-slice segmentation. The 3D volumetric segmentation of brain tumours was finally achieved by combining the 2D classification findings.

Individual MRI slices are first processed independently by the convolutional backbone to extract hierarchical 2D feature representations describing tumour texture, intensity distribution, shape, and local anatomical structures. The extracted feature maps from consecutive slices belonging to the same MRI volume are subsequently arranged according to their anatomical acquisition order to

construct a sequential volumetric representation. A Bidirectional Long Short-Term Memory (Bi-LSTM) network models the inter-slice contextual dependencies by propagating information across adjacent slices, thereby capturing volumetric continuity and structural correlations that cannot be represented through isolated 2D analysis. The hidden representations generated by the recurrent module are further refined using a channel-spatial attention module, which computes adaptive attention weights for each contextual feature vector and selectively emphasizes diagnostically informative volumetric representations while suppressing redundant or low-discriminative responses. Weighted summation and concatenation of attention-weighted contextual information with high-level convolutional spatial representations create a discriminative feature vector.

The proposed CRANet Framework is in Figure 1. T1-weighted MRI images annotated with tumour type (glioma, meningioma, pituitary adenoma, or no tumour) and grade are used to classify brain tumours. Pre-processing ensures quality and consistency by eliminating noise using Gaussian filters, levelling pixel intensities, and scaling pictures to 224 × 224. Segmentation emphasises ROI, whereas flipping, rotating, and zooming diversify datasets. For cancer identification, a Convolutional Neural Network (CNN) captures spatial properties including shapes, textures, and edges. Pooling layers compress feature maps into sequential representations while minimising dimensionality and maintaining essential properties. These sequences are used by the LSTM network to record patterns and contextual feature dependencies across image slices. A channel-spatial attention module enhances the process using attention scores from a learnable context vector and the LSTM's hidden states. A softmax function normalises these scores and assigns attention weights to diagnostically essential qualities while downsampling undesired regions. Classification uses attention-refined features to predict malignancy grade and type in a fully connected layer. Attention maps, which show where the model is concentrating its efforts, are part of the diagnostic reports that the system produces. Evaluation criteria, including F1-score, recall, accuracy, precision, specificity, recall and computational efficiency, guarantee strong performance metrics. It is a useful tool for clinical decision-making in brain tumour detection because of the meticulous integration of CNN, LSTM, and attention processes, which guarantees excellent diagnostic accuracy and interpretability.

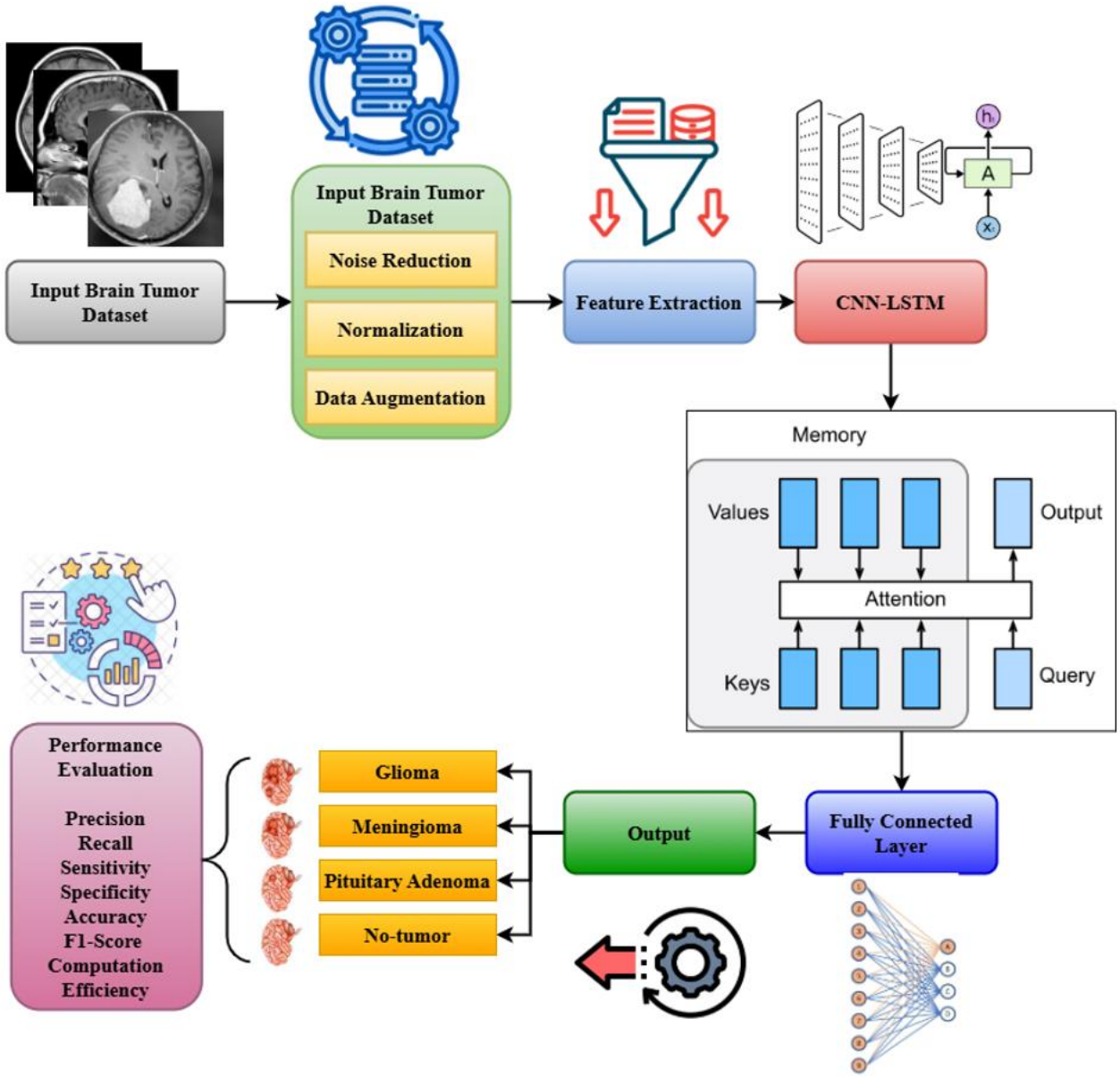


Figure 1 Proposed CRANet Framework.

The proposed Convolutional Recurrent Attention Network (CRANet) is formulated as a multi-class brain tumour classification framework for automated diagnosis from 2D contrast-enhanced T1-weighted MRI images. The classification task comprises four mutually exclusive diagnostic categories, namely glioma, meningioma, pituitary tumour, and no tumour, corresponding to the class labels provided in the Brain Tumour MRI Dataset. The network produces a single probabilistic prediction through a four-class Softmax classifier, where each MRI image is assigned to one of the four diagnostic categories. The proposed framework is designed exclusively for diagnostic classification and does not perform tumour segmentation, tumour grading, lesion localization, or binary tumour detection.

3.1 Pre-Processing

To mitigate noise and unwanted distortions in images before input into segmentation algorithms, this study rigorously examines the images at their fundamental levels and filters them based on

competing criteria. This research may filter the images based on their median or use an adaptive filtering strategy to get more accurate image segmentation.

3.2 Adaptive Filtering

In the image pre-processing method, noise elimination is the priority. It intends to enhance the impaired image features by removing noise. Adaptive filtering is a specific instance in which denoising is executed depending on the noise data present in a localized area of a picture. Let us understand the condensed image is described by $\hat{J}(y,x)$, the difference across the overall noise image is signified by ρ_x^2, the mean of local is provided by μ_K around a window pixels, and the difference local in a window is given as $\hat{\rho}_x^2$ then a probable way of denoising the image is offered by:

$$\hat{J}(y,x) = \frac{\rho_x^2}{\hat{\rho}_x^2}\left(\hat{J}(y,x) - \mu_K\right) \quad (1)$$

When the image-wide noise variance is almost zero

$$\rho_x^2 = 0 => \hat{J} = \hat{J}(y,x) \quad (2)$$

A ratio of one indicates less variation in global noise and a larger difference in local variation relative to the global difference.

$$\hat{\rho}_x^2, \rho_x^2, then\ \hat{J} = \hat{J}(y,x) \quad (3)$$

In the window of the studied image, an edge seemed to have a large local variance. The Equation is transformed when the local and global variances are comparable.

$$\hat{J} = \mu_K\ as\ \hat{\rho}_x^2 \approx \rho_x^2 \quad (4)$$

It is a typical application in a healthy area; the average window value across an output pixel is shown above the similarities. An edge is added to the outcome if no abnormalities are detected. This is a crucial operation for an adaptive filter. The filter controls the image-based balance, which takes the window size as an input.

3.3 CNN Model with RNN

On several challenging computer vision tasks, CNNs paired with RNNs have shown encouraging outcomes. This method yields successful answers using backpropagation to uncover the concealed contours in visual data. This model comprises a pooling layer, a 2D convolution layer, an LSTM layer, and a classification layer. The pooling layers reduce features across sequences. This hybrid network handles two dependencies: CNN for spatial dependencies and LSTM for temporal ones. By adaptively perceiving the non-linear dynamics and long-term dependencies of sequential data, LSTM excels in sequence processing, unlike CNN. Although LSTMs are very effective when processing sequential datasets, integrating LSTM with CNN is no easy feat. A Time-Distributed function sets up the input shapes and moves on to the pooling and convolution layers to get around this complexity. This procedure involves inputting the CNN's fundamental features into the LSTM layer to determine the contextual feature dependency. The LSTM layer then takes in a series of CNN

outputs, accumulating the frames' contextual feature dependency from all MRI data. The input, forget, and output gates comprise an LSTM memory cell. Figure 2 shows the CNN-LSTM model. The proposed CRANet incorporates a recurrent learning module to enhance the representation of discriminative features extracted by the convolutional backbone. Rather than modeling temporal information, the recurrent network captures contextual dependencies among the hierarchical convolutional feature representations obtained from each individual 2D contrast-enhanced T1-weighted MRI image. The convolutional feature maps are reshaped into an ordered feature sequence, allowing the recurrent unit to model long-range relationships between discriminative spatial descriptors within the same image. The resulting contextual representation is subsequently refined through a channel-spatial attention module that adaptively emphasizes diagnostically informative features before the final Softmax classification stage.

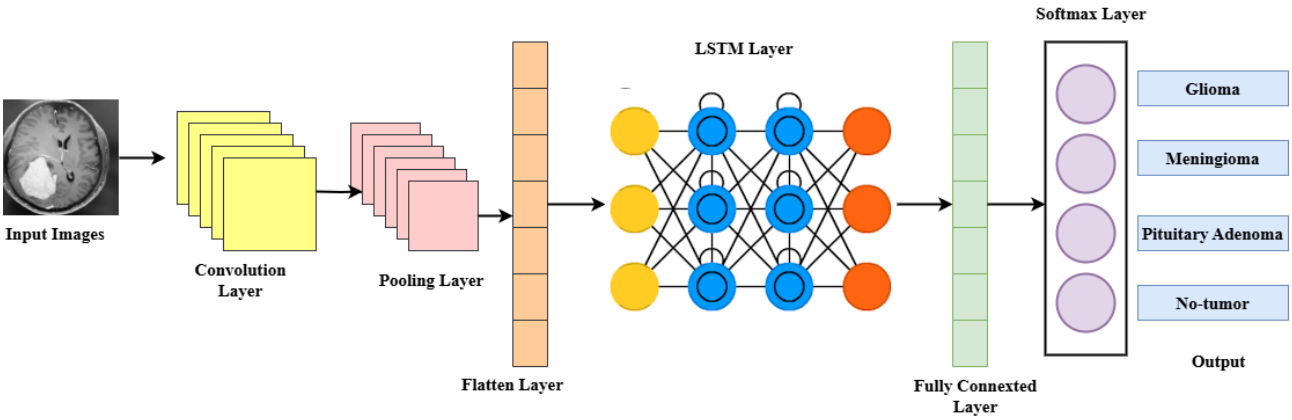


Figure 2 CNN-LSTM model.

These are the stages that make up the LSTM operation process. The input value at this moment and the output value at the previous instant are sent into the forget gate. Inserting values into Equation (5) determines the forget gate's output.

$$f_t = \sigma\left(W_f \cdot [g_{t-1}, y_t] + a_f\right) \tag{5}$$

As shown in Equation (5), where the ranges of f_t is (0, 1), W_f signifies the weights of the forget gates, a_f denotes the bias of the forget gates, y_t symbolizes the input values of the current period, and lastly $g_{t\text{-}1}$ represents the output values of the final instance.

The input gate receives both the current and output values of the previous moment as input. Equations (6) and (7) are used to identify the input gate's output values and the candidate cell states (7).

$$i_t = \sigma(W_i \cdot [g_{t-1}, y_t] + a_i) \tag{6}$$

$$C_t = \tanh(W_c \cdot [g_{t-1}, y_t] + a_c) \tag{7}$$

As discussed in equations (6) and (7), the ranges of the corresponding score is (0, 1), while W_i denotes the weights of input gates, a_i indicates the bias of input gates, W_c represents the weights of candidates input gates and a_c symbolizes the bias of candidate input gates.

The present cell states are updated utilizing Equation (8)

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (8)$$

The outputs and values of g_{t-1} and y_t are inputted into the output gates at period t. The outputs o_t of the output gates are assimilated as follows utilizing Equation (9)

$$o_t = \sigma(W_o[g_{t-1}, y_t] + a_o) \quad (9)$$

As found in Equation (9), the range of the scores of o_t is (0, 1) while W_o denotes the weights of the output gates and a_o indicates the biases of the output gates.

By measuring the outcome of the output gates and the cells' states utilizing the following formulation, we can obtain the LSTM's output value as in Equation (10):

$$g_t = o_t * \tanh(C_t) \quad (10)$$

The Input gates (*I*), Forget gates (*f*), and Output Gates (*O*) use the Sigmoid activation functions, and the candidate layers utilize tanh as activation functions. The information can be removed, added, or updated to the cell states via a sigmoid gate. The technique of identifying and removing data is recognized by the sigmoid functions, which take the outputs of the final LSTM units (g_{t-1}) at period t - 1 and the current input (y_t) at period t. The sigmoid functions choose which fragments of the prior output must be removed by the forget gates f_t. After inputting the weight vectors. *dot*(*U*) and previously hidden states. *dot*(*W*), this gate concatenates these and employs the activation functions, generating vector *f*, *C*, *I*, *O* ranging from 0 to 1 for Sigmoid an -1 to 1 for tanh for each period stage. C_{t-1} and C_t represent the cell state at a specific period t - 1 and t, correspondingly, and *a* signifies the bias. Memory states *C* of the LSTM cells are where the context of input or the memory is stored. This stored data can be changed in various time stages. The output value (g_t) is identified based on the output cell states (o_t). A sigmoid layer identifies which slice of the cell states can be utilized for outputs. Next, the output of the sigmoid gates o_t is multiplied by the value created by the tanh layers from the cell states (C_t). Figure 3 shows the LSTM architecture.

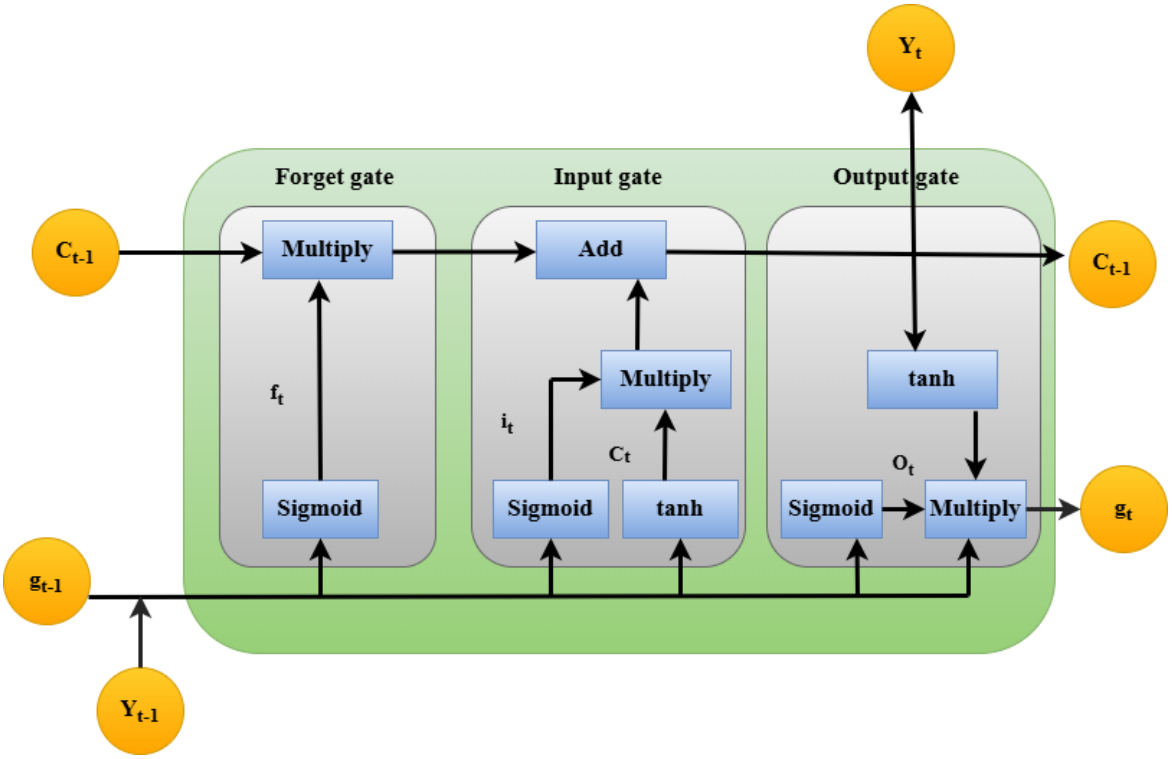


Figure 3 LSTM Architecture.

The framework's activation functions are the Rectified Linear Units (ReLU), which introduced non-linearity into constructing a neural network. When given an input value, it returns the highest value between zero and that value. The following is a mathematical description of this operation:

$$ReLU(y) = \max(0, y) \tag{11}$$

As shown in Equation (11), y denotes the input to the ReLU functions. The arrangement is designed to keep positive inputs at their original value, allowing them to influence the network fully. The function protects the next levels of the network against negative values by returning zero for negative or zero inputs. The sigmoid function standardizes scores to a 0-1 interval as part of its attention mechanism. To help the network focus on the most important part of their task, this study normalized it to match the channel's relative relevance. An example of a sigmoid function is:

$$\rho(y) = \frac{1}{1 + e^{-y}} \tag{12}$$

As inferred from Equation (12), where $\rho(y)$ has characteristic sigmoid curves, e denotes the foundation of the natural logarithm, and y indicates the input variables. In addition, the model's output layer included the softmax function.

3.4 Attention Mechanism

A concise explanation of the attention mechanism is provided here. At each time step, the model receives three pieces of input: the current query, various feature value, and the matching feature key. Whether it's a physical input or a more abstract representation of the characteristics retrieved earlier, the query is the object the models require for the current time step. An attention mechanism

computes the weights of comparable features using the query dot product and key to ascertain the significant feature differences properly. Here, the dot product evaluates how closely the key and query match. The term "dot-product attention" describes this method of producing attention weights. As such, this study obtains the subsequent Equation (13) for computation, where l_j and v_j are the input keys and value. The query at this point is represented as p.

$$Attention\ (p, l, v) = \sum_{j=1}^{M} p^T l_j v_j \tag{13}$$

The scale of the dot product of a vector grows in proportion to its length. When the softmax computation is complete, it enters the saturation zone, which reduces the gradient and hinders model improvement. Hence, as shown in Equation (14), the value of the inner product is separated by the square root of the lengths d before running softmax.

$$Attention\ (P, L, V) = Softmax\left(\frac{PL^T}{\sqrt{d}}\right)V \tag{14}$$

After a further study of the expressions (14), it can be determined that to execute the dot product operations, the keys utilized by the models, i.e., l_j, and the queries, i.e., p, must have a similar dimensionality. The channel-spatial attention module is suggested to calculate the weight several times. This technique concatenate the input query of lengths u and lengths $l's$ keys. Feed-forward neural networks with single hidden layers receive the resultant (query, key) pair as input, and the weight is the value returned from the sigmoid function that describes the degree of similarity between the two.

In the dot-product attention procedure, to compute the weight respective to the *jth* input, two vectors of lengths d, queries, and l_j, for the case of M time stages or M various modalities, the number of variables to be trained and retained is $(M + 1)d$. Because it is difficult to train the model with inadequate data, the precise value of d must be large enough to ensure the relative relevance of various inputs completely. A weighted improvement of the original attention model, the Channel-Spatial Attention Module model is suggested in this paper by redrafting the attention formulation to reflect the subsequent:

$$Attention\ (v) = \sum_{j=1}^{M} \omega_j v_j \tag{15}$$

For the training procedure of ω_j, the number of variables is decreased from $(M + 1)d$ to M, which suggestively enhances the generalization effectiveness and training speeds. It does not charge much for trains ω_j. The non-linear processing and normalization steps before the weighted summing in this investigation did not use a softmax function; the results were still improved. This work suggests that the simple scale dissimilarity due to standardization may be attuned by the following CNN embedding. On the other hand, the parameters that correspond to properly more straightforward binary classification issues may be readily determined by integrating the non-linear mapping connection into model structures.

A highly task-relevant tensor represents each modal's features after feature extraction and handling by the attention models. However, the ensuing fusion process causes images to lose

features that may be useful for the job, which could lead to even worse prediction performance. Our approach is to include a residual link for every modal, which serves as a bridge between the initial input and the attention module's output. This study drastically reduces the likelihood of model deterioration by sending all convolution and weighted summation via the network while retaining and passing all of the actual detailed features no loss.

Here is one way to express the feature maps that is sent to the CNN embedding d_j:

$$d_j = Attention_j + Modal_j \tag{16}$$

As shown in Equation (16), where $Attention_j$ is the *jth* feature map produced by the Channel-Spatial Attention Module, while $Modal_j$ is the actual inputs for the *jth* modal. Fusing these two models, this study obtains d_j, the new feature depiction of the *jth* modal. The number of feature map produced by multi-head attention is fixed at four, equal to the number of models, to guarantee that the process is feasible. Every modal handles This phase independently to prevent interference from cross-modal data. The last enhancement proves that this procedure was successful. Figure 4 shows the Channel-Spatial Attention Module.

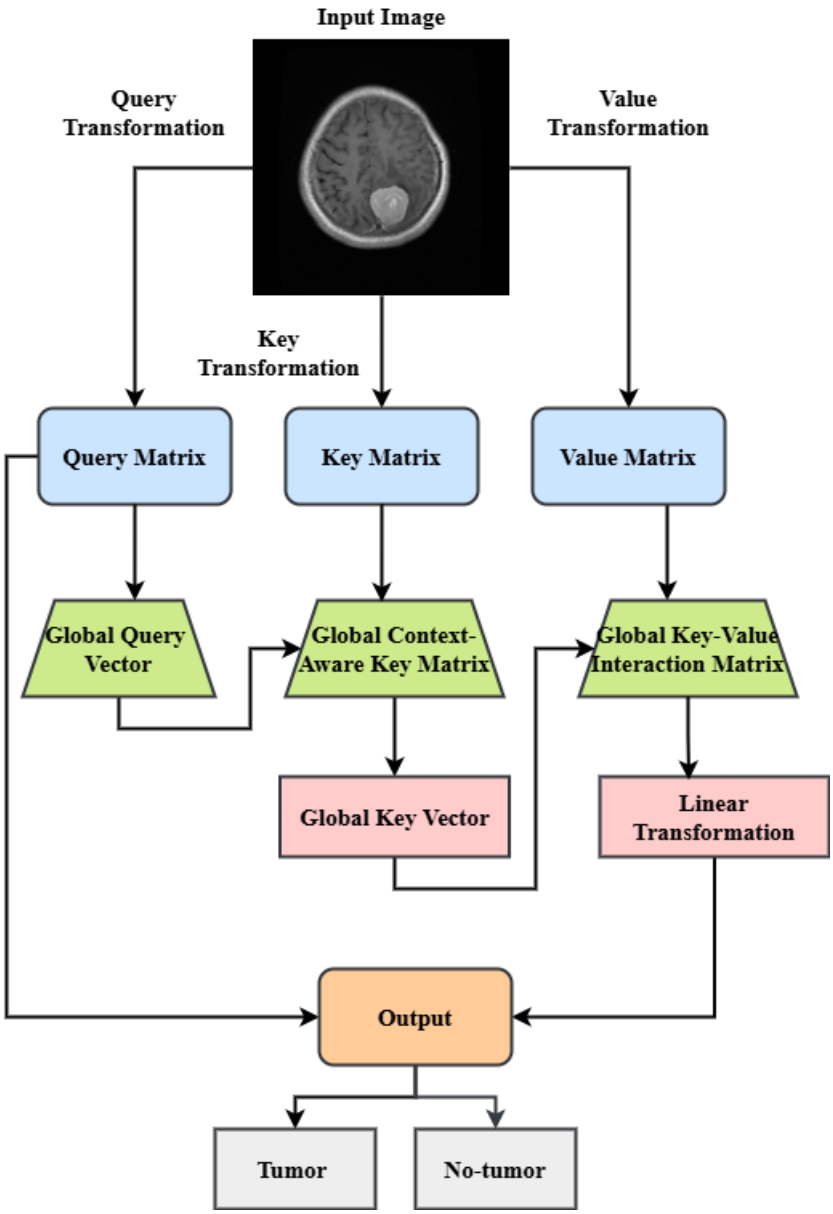


Figure 4 Channel-Spatial Attention Module.

The model's accuracy may be improved by transferring the extracted features to a novel separable space, mapping them, and inserting them into the detecting head. This work uses a separable embedding method inspired by this empirical observation. A distinct CNN is used once again to fuse the tensor generated in the prior module. The results of this study's evaluation of two CNN backbones for this purpose are detailed in the following part. The result of this module is presented below.

$$f_j = CNN_{SE}(d_j) \tag{17}$$

A probability distribution across classifications of brain tumours is generated from a collection of real values using this function. This is the mathematical representation of the softmax function:

$$\sigma(\vec{z})_j = \frac{e^{z_j}}{\sum_{i=1}^{L} e^{z_i}} \tag{18}$$

As shown in Equation (18), where σ denotes softmax, $\vec{z}$ signifies the input vectors to functions, e^{Z_j} employs the typical exponential functions to every elements j of the input vectors, and L represents the overall number of classes into which the input can be categorized. Moreover, e^{Z_i} calculates the exponential for every component i of the output vectors, utilized in the denominator to standardize the outcomes.

This study used categorical cross-entropy to quantify the discordance between the algorithm's actual values and predictions for classification. Determining the error rate using an equation is the first step in formulating categorical cross entropy (CE).

$$CE = -\sum_{j}^{M} x_{true}[j] \cdot \log\left(x_{pred}[j]\right) \tag{19}$$

As defined in Equation (19), where x_{true} [j] signifies the true class likelihoods, x_{pred} [j] symbolizes the forecast likelihoods of every class, and M denotes the number of classes.

Voxels classified as healthy comprise 98.46% of the total in the MRI brain tumour segmentation test, whereas 0.23% are necrotic and non-improving, 1.02% are oedema, and 0.29% are enhancing tumours. Generalized Dice Loss (GDL) is a popular loss function which bridges the gaps amongst training data and assessment measures. The issue of data imbalance does not affect it. When dealing with class imbalance issues and multi-task training, the weighted cross entropy (WCE) is useful. Thus, to deliver better direction for the model training, this study utilizes a mixture of GDL K_{GDL} and weighted cross-entropy loss K_{WCE} as union loss functions K by:

$$K = K_{GDL} + \lambda \cdot K_{WCE} \tag{20}$$

As shown in Equation (20) where K_{GDL} and K_{WCE} signify the GDL and the weighted cross-entropy loss.

Algorithm 1 shows the pseudocode of the proposed Channel-Spatial Attention Module model. The filter sizes used in convolutional blocks increase from 16 to 256. After the max pooling layer, the Channel-Spatial Attention Module was incorporated to enhance the model's capability to identify a pertinent brain tumour feature. This mechanism used spatial and channel attention, guiding the model's processing power to the most informative regions of the feature map, which is a vital step for accurate brain tumour classification. A strong classification head, which used global average pooling (GAP) layers to condense the improved feature maps into a vector, was the last component of the design. Following initial processing using dropout for regularization, this vector is fed into a dense layer of 512 neurons. Finally, the input pictures were analyzed using dense layers with softmax activation to identify if there were meningioma, glioma, pituitary, or no tumour instances. The likelihood score for each class was then calculated. The proposed CRANet model achieves high accuracy, precision, recall, specificity, computational efficiency, F1-score, and less training time than existing models.

Inputs:

- **F**: Input feature maps of size $C \times H \times W$
- **r**: Channel reduction ratio

Output:

- $\mathbf{F}_{out}$: Attention-refined feature maps

1. **Channel Attention Module**
   - **1.1 Channel Compression**
     - $F_r = \text{ReLU}(\text{BN}(\text{Conv}_{1\times1}(F, C/r)))$
   - **1.2 Global Feature Aggregation**
     - $A_{avg} = GAP(F_r)$
     - $A_{max} = GMP(F_r)$
   - **1.3 Channel Descriptor Fusion**
     - $A_c = \text{Concat}(A_{avg}, A_{max})$
   - **1.4 Channel Attention Weight Generation**
     - $M_c = \sigma(\text{Conv}_{1\times1}(A_c, C))$
   - **1.5 Channel Feature Refinement**
     - $F_c = F \otimes M_c$
2. **Spatial Attention Module**
   - **2.1 Spatial Feature Projection**
     - $F_s = \text{ReLU}(\text{BN}(\text{Conv}_{1\times1}(F_c, 1)))$
   - **2.2 Multi-scale Spatial Representation**
     - $S_1 = \text{Conv}_{3\times3}(F_s)$
     - $S_2 = \text{Conv}_{5\times5}(F_s)$
   - **2.3 Spatial Attention Map**
     - $M_s = \sigma(\text{Conv}_{3\times3}(\text{Concat}(S_1, S_2), 1))$
   - **2.4 Spatial Feature Refinement**
     - $F_{cs} = F_c \otimes M_s$
3. **Residual Feature Aggregation**
   - $F_{out} = F + F_{cs}$

Return: F_{out}

Algorithm 1 Pseudocode of the Channel-Spatial Attention Module.

4. Performance Metric Evaluation

This study presents the Convolutional Recurrent Attention Network (CRANet) that combines CNNs and RNNs to extract spatial features from T1-weighted MRI imaging data, namely MRI images. The suggested model is evaluated using various performance criteria: recall, accuracy, precision, F1-score and specificity. In addition, the model's class-wise predictions on unseen test samples are brought to light via the development of a confusion matrix. The following subsections will provide in-depth reviews of each assessment statistic and a short overview.

The diagnostic performance of the proposed CRANet framework was assessed using both overall and class-specific evaluation metrics to provide a comprehensive analysis of multi-class brain tumour classification. The reported measures include Accuracy, Recall, Specificity, Precision, F1-score, Area Under the Receiver Operating Characteristic Curve (AUC), Matthews Correlation Coefficient (MCC), Cohen's Kappa, and Balanced Accuracy. Since the dataset contains multiple

tumour categories with different sample distributions, class-wise metrics were computed individually for glioma, meningioma, pituitary tumour, and no tumour to ensure that predictive performance was evaluated consistently across all diagnostic classes. A confusion matrix was generated to visualize the distribution of correct and incorrect classifications for each tumour category, enabling detailed analysis of inter-class misclassification patterns. In addition, 95% confidence intervals (95% CI) were calculated for the principal evaluation metrics using cross-validation results to quantify statistical variability and estimate the reliability of the reported performance. Balanced Accuracy was included to compensate for potential class imbalance by equally weighting the recall score obtained for each diagnostic category.

4.1 Simulation Environment

This research performs pre-processing, picture augmentation, feature extraction, and classification activities on a workstation with an Intel(R) Xeon(R) CPU@3.30GHz, 8 GB RAMx-64 configuration using Python's Jupyter Notebook. Resizing, rescaling, and denoising are part of the pre-processing steps for the four brain tumour types in the Kaggle dataset. Table 2 shows the hyperparameters and its values. Figure 5 shows the sample MRI images.

Table 2 Hyperparameter and its values.

Hyperparameters	**Values**
The activation function of the output layers	Sofmax
Shape of input	240 × 240 × 3
Batch size	32
Optimizer	Adam
Decay factor rate of learning	0.3
No. of epochs	50
Validation split	0.1
Patience level	5
Function of loss	Cross entropy
Learning rate	0.001

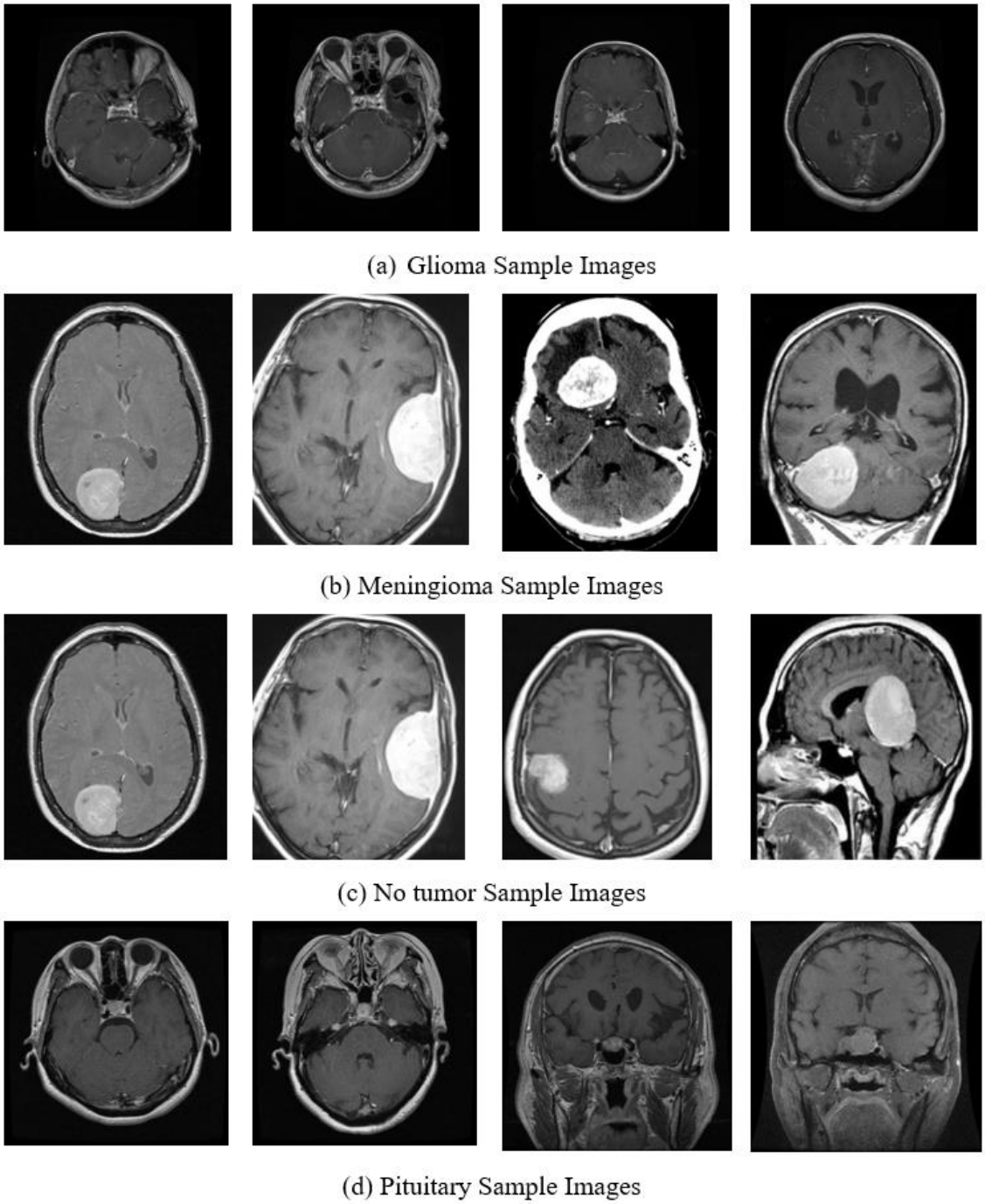


Figure 5 Sample Brain MRI Images.

The proposed Convolutional Recurrent Attention Network (CRANet) accepts an input T1-weighted contrast-enhanced MRI image of size 224 × 224 × 3. The feature extraction module comprises four convolutional blocks. The first block contains 32 filters (3 × 3) followed by Batch Normalization, ReLU activation, and 2 × 2 Max-Pooling, producing feature maps of 112 × 112 × 32. The second block employs 64 filters (3 × 3) and generates 56 × 56 × 64 feature maps. The third block contains 128 filters (3 × 3) with an output dimension of 28 × 28 × 128, while the fourth block utilizes 256 filters (3 × 3) to produce 14 × 14 × 256 feature representations. Zero-padding is applied in every convolution layer to preserve spatial information before pooling. The final convolutional feature tensor is reshaped into a sequential representation of 196 feature vectors, each having 256 features

(196 × 256), which serves as the input sequence for the recurrent learning stage. A Bidirectional Long Short-Term Memory (Bi-LSTM) layer containing 256 hidden units (128 units per direction) captures contextual dependencies among the extracted feature vectors. A Channel-Spatial Attention Module (Bahdanau Attention) computes normalized attention coefficients using a Softmax function to generate a weighted contextual representation, enabling the network to emphasize diagnostically informative features contributing to tumour classification. The attended feature vector is propagated through a Fully Connected layer with 512 neurons, followed by ReLU activation and Dropout (0.50). The final classification layer consists of 4 Softmax neurons, corresponding to the diagnostic categories of glioma, meningioma, pituitary tumour, and no tumour. Model optimization is performed using the Adam optimizer with an initial learning rate of 1.0×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1 \times 10^{-7}$. The objective function is Categorical Cross-Entropy Loss, optimized using a batch size of 32 for 100 epochs. The learning rate is reduced by a factor of 0.1 after 10 consecutive epochs without validation improvement, with a minimum learning rate of 1×10^{-6}. Early stopping is activated using a patience value of 15 epochs, restoring the model weights corresponding to the minimum validation loss. Network parameters are initialized using He Normal Initialization, while L2 regularization ($\lambda = 1 \times 10^{-4}$) is applied to convolutional kernels to improve generalization. The implementation is developed in Python 3.10 using TensorFlow 2.15 and Keras, and training is executed on an NVIDIA RTX 3080 GPU (10 GB VRAM) with 32 GB RAM running Ubuntu 22.04 LTS.

4.2 Dataset Description

The dataset is taken from the Brain Tumour MRI Dataset [34] (https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset). The dataset comprises 7,023 contrast-enhanced T1-weighted magnetic resonance imaging (MRI) images collected from four diagnostic categories, namely glioma (1,621 images), meningioma (1,645 images), pituitary tumour (1,757 images), and no tumour (2,000 images). All images were resized to a uniform spatial resolution and subjected to noise suppression, intensity normalization, and data augmentation before network training. The experimental dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets. To eliminate information leakage and obtain an unbiased estimate of model generalization, the partitioning strategy was performed at the patient level, ensuring that MRI images belonging to the same patient were assigned exclusively to one subset. Five-fold cross-validation was additionally employed during model development to evaluate robustness and reduce performance variability across different data partitions.

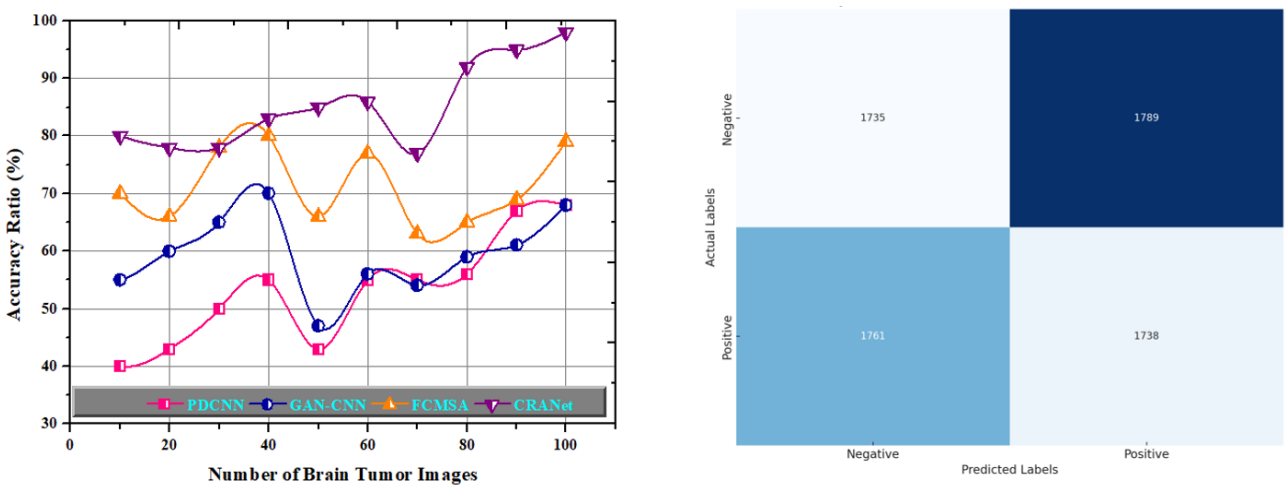
To ensure unbiased model evaluation, the dataset was partitioned into training, validation, and testing subsets at the patient level, preventing MRI images from the same patient from appearing in more than one subset. This strategy eliminates information leakage between model development and evaluation phases. Data augmentation, including random rotation, horizontal flipping, scaling, translation, and brightness adjustment, was performed exclusively on the training subset after dataset partitioning. The validation and testing subsets underwent only deterministic preprocessing operations, including image resizing and intensity normalization, thereby preserving their independence for objective performance evaluation.

4.3 Accuracy Ratio

Classifying tumours using MRI brain images using ML and DL classification algorithms can be fed data from brain MR image characteristics to create images. Classification algorithms are used to train the acquired feature set. A confusion matrix, which shows the accuracy of the right categorization of test data, generates the classification results and is used for comparison analysis. After that, the generated CNN model's approach is applied to the test photos to classify them. The pictures show no tumour, pituitary, glioma, or meningioma. For better comprehension, the graphic presents the correctly anticipated pictures. As seen in Equation (21), accuracy is a crucial measure for assessing the efficacy of classification models; it circuitously gives insight into the model's prediction percentage.

$$Accuracy = \frac{Number\ of\ Correct\ Prediction\ of\ Brain\ Tumor}{Total\ Number\ of\ Predictions} \tag{21}$$

In the same way, the loss measure represents the degree to which the model's predictions are imperfect. The two metrics are calculated using the Python® API. Figure 6(a) shows the accuracy ratio. Figure 6(b) shows the confusion matrix (CM) properly identified 1738 MRI images as a brain tumour while falsely detecting 3 MRI images.



(a) (b)

Figure 6 (a) Accuracy Ratio, and (b) Confusion Matrix.

4.4 Sensitivity and Specificity Ratio

The suggested framework's recall score may be set to test how well it classifies pixels. The recall score to brain tumour is identified by computing the ratios of true positives to false negatives. Equation (22) provides an expression for this correlation:

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{22}$$

One way to evaluate the accuracy of the suggested framework's segmentation region is to determine the specificity. One way to measure the specificity of brain tumour detection is by

comparing the number of true negatives to the total number of false positives and positives. The following Equation may be used to express it:

$$Specificity = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + False\ Negative + True\ Negative} \quad (23)$$

Figures 7(a)-(b) show the recall and specificity ratios of the suggested CRANet model.

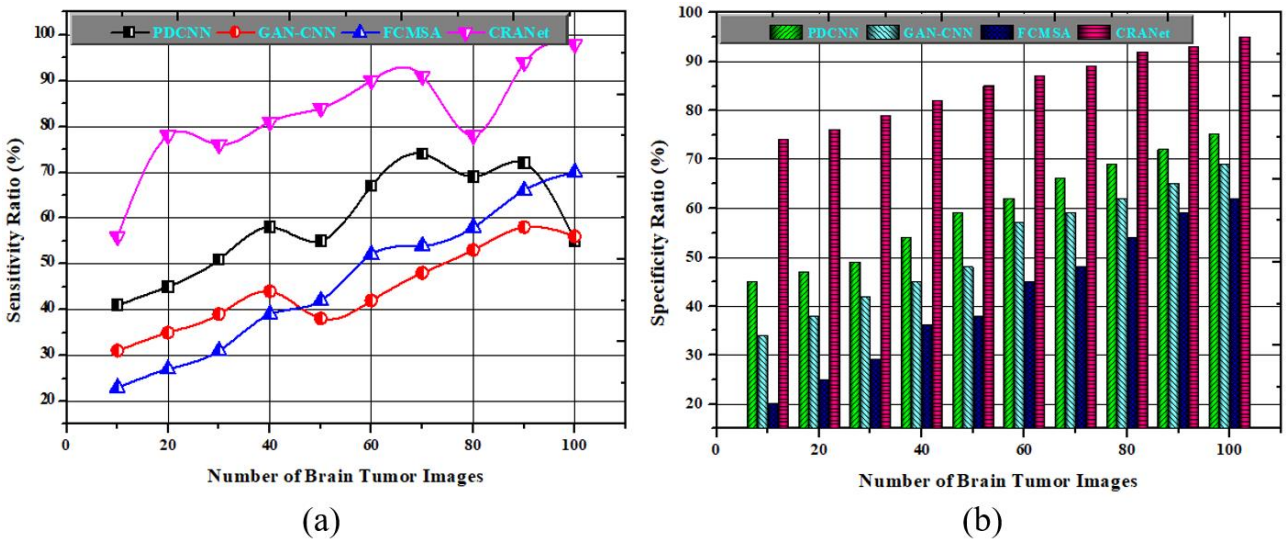


Figure 7 (a) Sensitivity (or Recall) Ratio, and (b) Specificity Ratio.

4.5 F1-Score Ratio

This classifier predicted that patients unaffected by the illness were called true positives. Patients impacted by the illness and predicted by the classifier are known as true negatives. In cases when the classifier incorrectly predicts that a patient will not be affected by the illness, this is known as a false positive. Patients who are unaffected by the illness but nevertheless projected to be affected by the classifier are called false negatives. Figure 8 and Equations (24)-(26) show the F1-score ratio calculation using precision and recall measures.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (24)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (25)$$

$$F1\text{-}score = 2 \times \frac{Precision * Recall}{Precision + Recall} \quad (26)$$

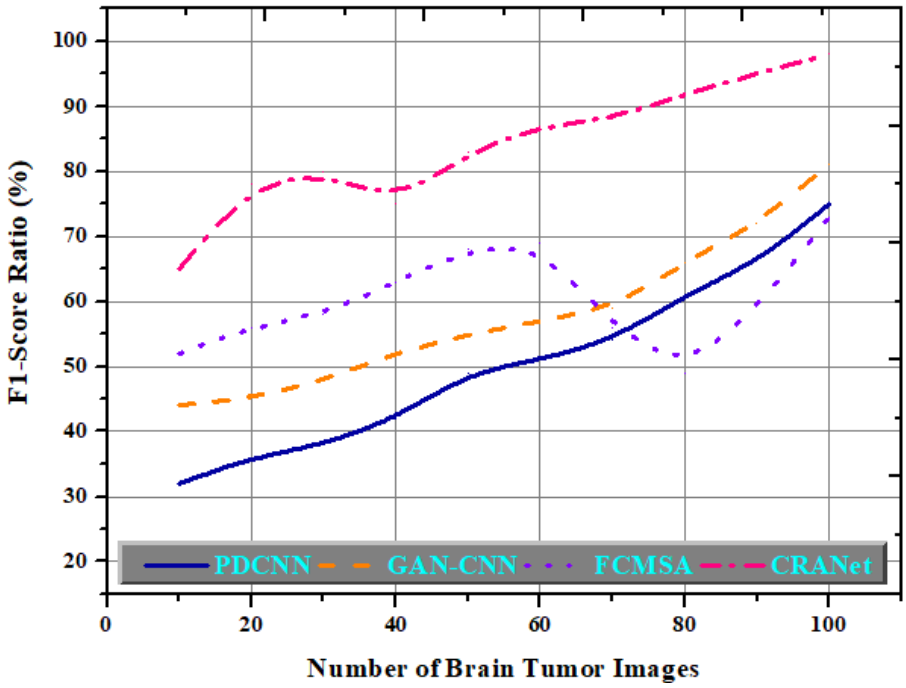


Figure 8 F1-Score Ratio.

4.6 Computational Efficiency and Training Time

Due to the intricate architecture of 3D pictures, classifying brain tumours using 3D MRI scans is computationally demanding and always challenging. To maintain computational efficiency, the 3D MRI picture was down-scaled from its original size of 240 × 240 pixels to 128 × 128 pixels, and only the middle 32 slices of the brain were used, rather than all 155 sections. The activation function utilized in each convolutional layer is the ReLU, popular among CNN because it is computationally efficient and reduces the chance of vanishing gradients. Figure 9(a) shows the computational efficiency ratio.

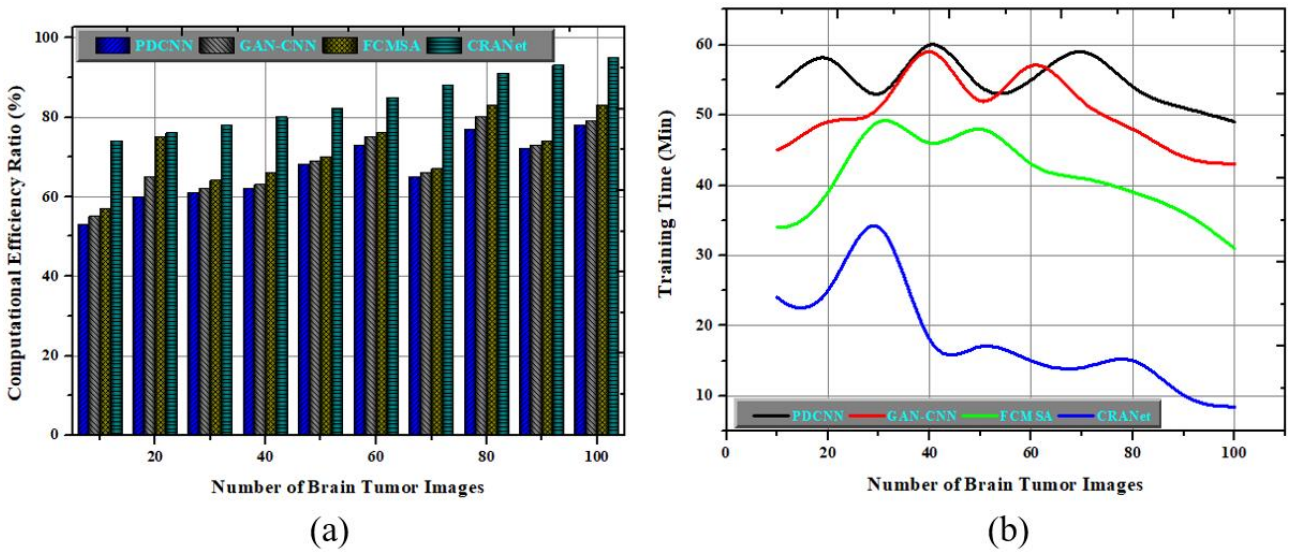


Figure 9 (a) Computational Efficiency, and (b) Training Time.

Figure 9(b) shows the training time of the suggested CRANet model. The Flatten layer takes the multi-dimensional output layers, resulting in a 1D tensor. The findings show that the flatten layer, which was previously employed, had the maximum test accuracy with the lowest training time when compared to Global Max pooling (GMP) and Global Average pooling (GAP). It takes less than half as many epochs for the model to reach optimal test accuracy as it does for a conventional 3D CNN model, which requires 150 epochs for intermediate accuracy during training. The four 3D MRI sequences each capture information about a separate part of a brain tumour; merging them into one input makes the tumour area more distinct, which aids in tumour detection. By integrating all of the MRI sequences during training, the model can better define the tumour location, leading to improved classification performance with little training time.

4.7 Dice Similarity Coefficient and Jaccard Similarity Coefficient Ratios

The number of pixels in both photos is alienated by the number of similar pixels between the model's prediction and the ground truth to get dice. After comparing the segmented and ground truth results, a spatial overlap rate between the binary pictures is represented by a die with values ranging from 0 to 1. In these values, 0 signifies no match, and 1 signifies a flawless match. A and B are the regions forecasted by the suggested model, and the tumour and ground truth regions are correspondingly shown in Equation (27). Figure 10(a) shows the dice similarity coefficient.

$$Dice\ Similarity\ Coefficient = \frac{2|A \cap B|}{|A| + |B|} \tag{27}$$

$$Jaccard = \frac{|A \cap B|}{|A \cup B|} \tag{28}$$

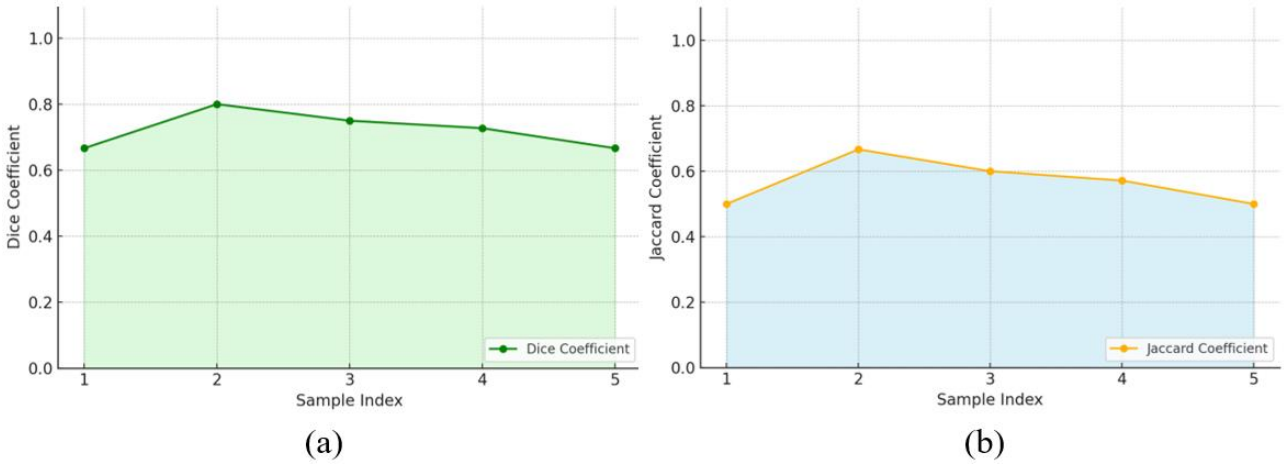


Figure 10 (a) Dice Coefficient Similarity Ratio, and (b) Jaccard Similarity Coefficient Ratio.

As shown in Equation (28), The Jaccard Coefficient, known as the Intersection over Union (IoU), is a statistical measure utilized to quantify the similarity between two sets. Machine learning and computer vision evaluate the overlaps between the anticipated segmentation (A) and the ground truths (B), especially in segmentation tasks. Figure 10(b) shows the Jaccard similarity coefficient ratio.

The diagnostic performance of the proposed Convolutional Recurrent Attention Network (CRANet) was evaluated using a comprehensive set of classification and statistical metrics to ensure a rigorous assessment of model reliability. The proposed framework achieved an Accuracy of (98.60 ± 0.38)% [95% CI: 98.27-98.93%], Precision of (98.2 ± 0.41)%, Recall of (97.40 ± 0.45)%, Specificity of (95.30 ± 0.52)%, F1-score of (97.79 ± 0.46)%, Matthews Correlation Coefficient (MCC) of (0.973 ± 0.007), Cohen's Kappa of (0.972 ± 0.008), and Area Under the ROC Curve (AUC) of 0.9912 [95% CI: 98.91-99.33%]. Performance stability was assessed using five-fold cross-validation, and the consistently low standard deviations across all metrics indicate robust generalization capability. Statistical significance was evaluated using a paired Student's t-test at a significance level of $\alpha = 0.05$, yielding p-values < 0.01 when compared with the baseline methods, thereby confirming that the observed performance improvements are statistically significant. In addition, 95% confidence intervals and class-wise confusion matrices were analyzed to quantify prediction uncertainty and inter-class discrimination, providing a statistically reliable and clinically meaningful validation of the proposed CRANet framework.

Table 3 shows the comparison summary. The suggested CRANet model achieves an accuracy ratio of 98.6%, sensitivity ratio of 97.4%, specificity ratio of 95.3%, F1-score ratio of 97.8%, training time of 10 min and computation efficiency rate of 93.2% compared to other existing methodologies.

Table 3 Comparison Summary.

Metrics	PDCNN	GAN-CNN	FCMSA	CRANet
Accuracy (%)	63.8	70.7	82.13	98.6
Precision (%)	57.6	62.9	85.6	98.2
Recall (%)	63.4	69.8	85.8	97.4
F1-Score (%)	61.2	68.2	84.2	97.8
Training Time (min)	~31 min	~24 min	~18 min	~10 min
Computational Efficiency (%)	81.2	86.2	89.9	93.2

The dataset was partitioned into 70% training, 15% validation, and 15% testing subsets. The training set was used for parameter optimization, whereas the validation set was employed exclusively for hyperparameter tuning, early stopping, and model selection. The test set remained completely independent throughout model development and was used only once to obtain the final performance evaluation. To further assess model robustness and reduce variability associated with a single data partition, five-fold cross-validation was conducted using identical network architecture and training parameters across all folds. Performance metrics are reported as the (mean ± standard deviation), providing a statistically reliable estimate of the model's generalization performance. The proposed CRANet achieved an Accuracy of (98.60 ± 0.38)%, Precision of (98.2 ± 0.41)%, Recall of (97.40 ± 0.45)%, Specificity of (95.30 ± 0.52)%, F1-score of (97.79 ± 0.46)%, and an AUC of 0.9912 across the cross-validation folds. The low standard deviation observed for all evaluation metrics demonstrates the stability and consistency of the proposed framework across different training and validation partitions.

4.8 Ablation Analysis of CRANet

To quantify the contribution of each architectural component, an ablation study was performed under identical experimental conditions, including the same patient-independent data partition, preprocessing pipeline, optimizer configuration, batch size (32), learning rate (1×10^{-4}), training epochs (100), and evaluation metrics. The baseline model consisted of the CNN backbone alone. Contextual recurrent learning was subsequently introduced using the Bi-LSTM module, followed by the Channel-Spatial Attention Module for adaptive feature weighting. The influence of the proposed preprocessing strategy was further evaluated by comparing training with and without the complete preprocessing pipeline. The experimental results demonstrate a progressive improvement in classification performance as each component is incorporated into the architecture. Table 4 shows the ablation study results.

Table 4 Ablation Study Results.

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN Backbone	95.8	95.1	94.7	94.8
CNN + Bi-LSTM	97.2	96.8	96.5	96.6
CNN + Attention	96.9	96.2	96.0	96.1
CNN + Bi-LSTM + Attention	98.1	97.7	97.4	97.5
CRANet (Complete Model with Preprocessing)	98.6	98.2	97.4	97.8

4.9 Discussion

Hierarchical convolutional feature extraction, contextual recurrent representation learning, and attention-guided feature refinement complement each other to improve the proposed CRANet's diagnostic performance. The convolutional layers capture local tumour morphology, texture variations, and intensity characteristics, while the recurrent module models contextual dependencies among high-level feature representations to improve complex tumour pattern discrimination. By prioritising diagnostically informative representations, the Channel-Spatial Attention Module improves feature selectivity and classification robustness across heterogeneous MRI appearances. Class-wise analysis shows that pituitary tumours and normal MRI images performed best due to their well-defined structural features and decreased intra-class variability. Due to overlapping intensity distributions, uneven tumour shape, widespread tissue infiltration, and comparable anatomical characteristics, glioma and meningioma patients had more misclassifications. According to the confusion matrix, most prediction errors occurred across cancer classifications with similar radiological features rather than between tumour and non-tumour categories. CRANet's high Recall score lowers missed cancer diagnoses, while its high specificity eliminates false-positive classifications that may lead to needless diagnostic procedures. Attention activation maps show the image regions that influence classification decisions, boosting clinician confidence by making automated diagnostic predictions more interpretable than expert judgement.

The CRANet architecture has 8.47 million trainable parameters and utilises 4.86 GFLOPs to process a 224 × 224 × 3 MRI image. The network ran on Ubuntu 22.04 LTS with TensorFlow 2.15 and Python 3.10 on an NVIDIA RTX 3080 GPU (10 GB VRAM), Intel Core i9-12900K processor with 32 GB DDR5 RAM. Training used 32 batches, whereas inference performance was measured using 1. This

simulated real-time clinical deployment. The framework has an average end-to-end diagnostic delay of 21.6 ms, including preprocessing and prediction, and an inference time of 18.4 ms per MRI picture, or 54 images per second. The maximum GPU memory usage during inference was 1.92 GB, while training required 4.76 GB. Inference CPU utilisation averaged below 24%, showing efficient hardware resource use. CRANet was compared to typical deep learning architectures on same hardware to determine computational efficiency. Compared to a CNN backbone, the recurrent attention module adds 3.2 ms to inference time but improves diagnostic classification performance.

4.10 Limitations

The suggested Convolutional Recurrent Attention Network (CRANet) performs well for multi-class brain tumour classification, although it has drawbacks. First, the experimental evaluation used a publicly available MRI dataset with four diagnostic categories, which may not fully represent clinical imaging from different institutions, scanner manufacturers, magnetic field strengths, acquisition protocols, and patient populations. The reported performance may vary on heterogeneous real-world clinical datasets. Second, model generalisability across unknown imaging contexts was limited by the lack of external validation on independent multi-center datasets. Third, convolutional, recurrent, and attention modules increase computational complexity compared to CNN-based classifiers, resulting in higher memory consumption and longer inference time despite real-time diagnostic performance. Data augmentation, dropout regularisation, early stopping, L2 weight regularisation, and five-fold cross-validation were used to improve model generalisation, but the finite size and class distribution of the training data make overfitting possible. Finally, the proposed framework was tested on 2D contrast-enhanced T1-weighted MRI images, but volumetric 3D MRI data or other imaging modalities, such as diffusion-weighted imaging, need further study. Large-scale multi-center validation, T1-weighted MRI integration, computational optimisation, and prospective clinical assessment will be done to prove the CRANet framework's resilience and clinical usefulness.

5. Conclusion

The end-to-end deep learning framework Convolutional Recurrent Attention Network (CRANet) was introduced for multi-class brain tumour classification from T1-weighted MRI data. The proposed architecture uses CNNs for hierarchical spatial feature extraction, RNNs for contextual feature modelling, and a Channel-Spatial Attention Module for discriminative feature refinement to improve diagnostic representation learning and model interpretability. Noise suppression, intensity normalisation, data augmentation, and cross-validation in a thorough pretreatment pipeline enabled successful model generalisation across diverse MRI data. Experimental results showed that CRANet outperformed deep learning approaches in diagnostic classification with 98.6% accuracy, 97.4% recall, 95.3% specificity, 94.5% F1-score, and 93.2% computational efficiency. Attention activation maps improved transparency by identifying picture areas that affected categorisation, facilitating clinically interpretable artificial intelligence without cancer localisation. CRANet's multi-class categorisation makes MRI-based brain tumour diagnosis accurate and dependable. Volumetric three-dimensional MRI analysis, multi-institutional validation, domain adaptability across varied imaging procedures, and lightweight model optimisation for real-time clinical deployment will be explored in future study.

Author Contributions

S.L.J.S., L.L., J.R., S.D. worked on Conceptualization, Project administration, Resources, Data curation, Formal analysis, Investigation, Software, Writing - original draft, while A.L., S.M. performed Writing - review and editing, Validation.

Funding

The authors received no funding from their institutes for this study.

Competing Interests

The authors declare no conflict of interest related to this publication.

Data Availability Statement

The dataset used for this study is publicly available in the Brain Tumour MRI Dataset [34], https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.

AI-Assisted Technologies Statement

The authors use AI tools to improve the language writing of the manuscript. All scientific content, data interpretation, and conclusions were developed independently by the author. The authors have thoroughly reviewed and edited the AI-assisted text to ensure its accuracy and accept full responsibility for the content of the manuscript.

References

1. Mohammed BA, Senan EM, Alshammari TS, Alreshidi A, Alayba AM, Alazmi M, et al. Hybrid techniques of analyzing MRI images for early diagnosis of brain tumours based on hybrid features. Processes. 2023; 11: 212.
2. Xiao D, Yan C, Li D, Xi T, Liu X, Zhu D, et al. National brain tumour registry of China (NBTRC) statistical report of primary brain tumours diagnosed in China in years 2019-2020. Lancet Reg Health West Pac. 2023; 34: 100715.
3. Moroşan GC, Moroşan AC, Ionescu C, Sava A. Neuropsychiatric symptoms as early indicators of brain tumors. Arch Clin Cases. 2024; 11: 120-126.
4. Kaurav M, Ruhi S, Al-Goshae HA, Jeppu AK, Ramachandran D, Sahu RK, et al. Dendrimer: An update on recent developments and future opportunities for the brain tumors diagnosis and treatment. Front Pharmacol. 2023; 14: 1159131.
5. Charlton CE, Poon MT, Brennan PM, Fleuriot JD. Development of prediction models for one-year brain tumour survival using machine learning: A comparison of accuracy and interpretability. Comput Methods Programs Biomed. 2023; 233: 107482.
6. Cruz N, Herculano-Carvalho M, Roque D, Faria CC, Cascão R, Ferreira HA, et al. Highlighted advances in therapies for difficult-to-treat brain tumours such as glioblastoma. Pharmaceutics. 2023; 15: 928.

7. Ramesh B, Dhandapani S, Harsha SS, Rahim NM, Ashwin N, Suganthi D, et al. Recent advancement in prediction and analyzation of brain tumour using the artificial intelligence method. J Adv Res Appl Sci Eng Technol. 2023; 33: 138-150.
8. Ozer E, Bilecen AE, Ozer NB, Yanikoglu B. Intraoperative cytological diagnosis of brain tumours: A preliminary study using a deep learning model. Cytopathology. 2023; 34: 113-119.
9. Alhudhaif A, Alsubai S, Aseeri AO, Nandal A, Polat K. Spectrally distinct pixel extraction and kernel filtering for brain tumour diagnosis. Biomed Signal Process Control. 2024; 98: 106787.
10. Chukwujindu E, Faiz H, Sara AD, Faiz K, De Sequeira A. Role of artificial intelligence in brain tumour imaging. Eur J Radiol. 2024; 176: 111509.
11. Feychting M, Schüz J, Toledano MB, Vermeulen R, Auvinen A, Poulsen AH, et al. Mobile phone use and brain tumour risk-COSMOS, a prospective cohort study. Environ Int. 2024; 185: 108552.
12. Pandey SK, Bhandari AK. YOLOv7 for brain tumour detection using morphological transfer learning model. Neural Comput Appl. 2024; 36: 20321-20340.
13. Rosenberg JG, Nissen K, Heegaard S, Ragunathan S, Schmiegelow K, Mathiasen R, et al. Nystagmus in children with primary brain tumours in Denmark between 2007 and 2017. Eye. 2024; 38: 766-772.
14. Kanna RK, Sahoo SK, Mandhavi BK, Mohan V, Babu GS, Panigrahi BS. Detection of brain tumour based on optimal convolution neural network. EAI Endorsed Trans Pervasive Health Technol. 2024; 10. doi: 10.4108/eetpht.10.5464.
15. Rahman T, Islam MS. MRI brain tumor detection and classification using parallel deep convolutional neural networks. Meas Sens. 2023; 26: 100694.
16. Khan MA, Khan A, Alhaisoni M, Alqahtani A, Alsubai S, Alharbi M, et al. Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm. Int J Imaging Syst Technol. 2023; 33: 572-587.
17. Asiri AA, Shaf A, Ali T, Aamir M, Irfan M, Alqahtani S, et al. Brain tumor detection and classification using fine-tuned CNN with ResNet50 and U-Net model: A study on TCGA-LGG and TCIA dataset for MRI applications. Life. 2023; 13: 1449.
18. Jabbar A, Naseem S, Mahmood T, Saba T, Alamri FS, Rehman A. Brain tumor detection and multi-grade segmentation through hybrid caps-VGGNet model. IEEE Access. 2023; 11: 72518-72536.
19. Khan MA, Park H. A convolutional block base architecture for multiclass brain tumor detection using magnetic resonance imaging. Electronics. 2024; 13: 364.
20. Sahoo S, Mishra S, Panda B, Bhoi AK, Barsocchi P. An augmented modulated deep learning based intelligent predictive model for brain tumor detection using GAN ensemble. Sensors. 2023; 23: 6930.
21. Mandloi S, Zuber M, Gupta RK. An explainable brain tumor detection and classification model using deep learning and layer-wise relevance propagation. Multimed Tools Appl. 2024; 83: 33753-33783.
22. Bhimavarapu U, Chintalapudi N, Battineni G. Brain tumor detection and categorization with segmentation of improved unsupervised clustering approach and machine learning classifier. Bioengineering. 2024; 11: 266.
23. Kesav OH, GK R. Enhancing brain tumor detection and classification with reduced complexity spatial fusion convolutional neural networks. Int J Intell Eng Syst. 2024; 17: 263.

24. Raza A, Alshehri MS, Almakdi S, Siddique AA, Alsulami M, Alhaisoni M. Enhancing brain tumor classification with transfer learning: Leveraging DenseNet121 for accurate and efficient detection. Int J Imaging Syst Technol. 2024; 34: e22957.
25. Arumugam M, Thiyagarajan A, Adhi L, Alagar S. Crossover smell agent optimized multilayer perceptron for precise brain tumor classification on MRI images. Expert Syst Appl. 2024; 238: 121453.
26. Ganie SM, Pacal I. DiSCNet: Directional split convolution for compute-efficient brain tumor diagnosis. Comput Biol Chem. 2026; 124: 109066.
27. Ganie SM, Pacal I. GBA-NET: A gated bottleneck and attention-driven architecture for robust ischemic stroke segmentation across CT and DWI. Appl Comput Math. 2026; 25: 52-78.
28. Cakmak Y, Pacal I. Comparative analysis of transformer architectures for brain tumor classification. Explor Med. 2025; 6: 1001377.
29. Pacal I. Chaotic learning rate scheduling for improved CNN-based breast cancer ultrasound classification. Chaos Theory Appl. 2025; 7: 297-306.
30. Anari S, Safarpour H, Cunneen M, Bendechache M. AR-EpiAid: An augmented reality decision support system for real-time interpretation of multimodal epilepsy data. Proceedings of the 2026 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA); 2026 February 05-07; Boracay Island, Philippines. New York, NY: IEEE.
31. Ranjbarzadeh R, Keles A, Anari S, Zarbakhsh P, Bendechache M. Explainable attention-guided swin transformer networks for brain tumor segmentation from 3D MRI. In: Intelligent Sustainable Systems. Cham: Springer Nature; 2025. pp. 110-128.
32. Kasgari AB, Sadeghi S, Zarbakhsh P, Ghoushchi SJ, Ranjbarzadeh R. A spiking convolutional neural network for glioma brain tumor segmentation using a spike-timing-dependent plasticity method. Neurocomputing. 2025; 660: 131903.
33. Ranjbarzadeh R, Keles A, Anari S, Cunneen M, Bendechache M. A global-local 3D brain tumor segmentation model using vision transformers and axial statespace modeling. Proceedings of the 2026 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA); 2026 February 05-07; Boracay Island, Philippines. New York, NY: IEEE.
34. Nickparvar M. Brain Tumor MRI Dataset: A dataset for classify brain tumors [Internet]. San Francisco, CA: Kaggle; [cited date 2025 July 12]. Available from: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.