SAlexNet: Superimposed AlexNet using Residual Attention Mechanism for Accurate and Efficient Automatic Primary Brain Tumor Detection and Classification

Results in Engineering 25 (2025) 104025

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Research paper

SAlexNet: Superimposed AlexNet using residual attention mechanism for accurate and efficient automatic primary brain tumor detection and classification

Qurat-ul-ain Chaudhary [a], Shahzad Ahmad Qureshi [a,b,*], Touseef Sadiq [c,**], Anila Usman [a], Ambreen Khawar [d], Syed Taimoor Hussain Shah [e], Aziz ul Rehman [f,g]

[a] Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad 45650, Pakistan
[b] Centre for Mathematical Sciences, PIEAS, Islamabad 45650, Pakistan
[c] Centre for Artificial Intelligence Research (CAIR), Department of Information and Communication Technology, University of Agder, Jon Lil-letuns vei 9, Grimstad, Norway
[d] Department of Medical Sciences, PIEAS, Islamabad 45650, Pakistan
[e] PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin 10129, Italy
[f] Department of Physics and Astronomy, Macquarie University, Sydney, New South Wales 2109, Australia
[g] Agri & Biophotonics Division, National Institute of Lasers and Optronics College, PIEAS, Islamabad 45650, Pakistan

ARTICLE INFO

Keywords:
Brain tumor
Convolutional neural network
Feature map
Transfer learning
AlexNet
Deep learning

ABSTRACT

Accurate classification of brain tumors is crucial for informing clinical diagnoses and guiding patient treatment plans. It is one of the most aggressive tumors, leading to a short life expectancy. However, the classification of brain tumors is a challenging task due to the heterogeneity, complexity, and variability of brain tumors. In this work, we propose Superimposed AlexNet (SAlexNet-1 and its extension SAlexNet-2) to detect the malignancy of primary brain tumors (Glioma, Meningioma, and Pituitary) by incorporating three enhancements: (1) fusing Hybrid Attention Mechanism (HAM), (2) dense feature extraction by replacing initial convolution 11×11 layer with multiple convolution 3×3 layers for extra non-linearity alleviating parameter burden, and (3) pretraining the encoder path on a correlated dataset via semi-transfer learning (STL) enhancing model performance. HAM provides more comprehensive and accurate feature representations. In this study, we evaluated the performance of our proposed SAlexNet models on two publicly available extensive datasets for multi-class and binary classification tasks. Our results show that SAlexNet-1 achieved an accuracy of (98.78 ± 0.80 %) and (98.07 ± 0.02 %) on the multi-class and binary classification datasets, respectively. In comparison, SAlexNet-2 achieved outstanding accuracy of (99.69 ± 0.22 %) and (99.17 ± 0.00 %) on the multi-class and binary classification MRI datasets, respectively. The STL-based SAlexNet-2 surpassed previous literature with complex models and techniques, achieving an accuracy of (99.20 ± 0.01 %). Furthermore, we provided a comprehensive analysis of current state-of-the-art tumor classification methods on the same dataset, demonstrating the effectiveness of our approach. Enhanced tumor classification accuracy enables better diagnosis, treatment planning, and patient outcomes.

1. Introduction

The brain, with billions of cells, is affected by a problem that changes its typical structure and behavior, known as a brain tumor. It is one of the fatal forms of cancer among other cancer types, having an aggressive nature, heterogeneous characteristics, and low survival rate [1]. According to the latest statistical report of CBTRUS [2], about 86,010 new non-malignant and malignant brain tumors are estimated to be analyzed in the United States. There were 79,718 deaths recognized as malignant brain tumors between 2012 and 2016, with an annual average mortality

* Corresponding author at: Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad 45650, Pakistan.
** Corresponding author.
E-mail addresses: drsaqureshi@pieas.edu.pk (S.A. Qureshi), touseef.sadiq@uia.no (T. Sadiq).

https://doi.org/10.1016/j.rineng.2025.104025
Received 5 October 2024; Received in revised form 2 January 2025; Accepted 12 January 2025
Available online 13 January 2025

rate of 4.42. The incidence of brain tumor-related fatalities has risen among both adults and children. However, the exact origin and progression rate of brain tumors remain challenging. Brain tumors are generally categorized into two types: primary and secondary. Approximately 70 % of brain tumors originate within the brain itself, falling under the primary category. Notably, the majority of primary brain tumors exhibit malignant behavior, with gliomas being the most aggressive (accounting for 80 % of malignant cases, with only Grade I being non-malignant among Grades I-IV). Other primary brain tumors, such as meningioma and pituitary tumors, pose significant diagnostic and therapeutic challenges for healthcare professionals due to their elusive early detection [3]. Gliomas, arising from the brain's glial cells, are the most common type of primary brain tumor. In contrast, typically benign meningiomas occur within the skull and originate from the protective membranes surrounding the brain and spinal cord. Meanwhile, pituitary tumors develop on the pituitary gland, a crucial endocrine organ regulating hormonal balance. These tumors can exhibit both benign and malignant characteristics, and hormonal imbalances caused by pituitary tumors can lead to vision problems [4].

The implementation of multimodal brain tumor imaging paradigms, encompassing intraoperative magnetic resonance imaging (iMRI), advanced neuronavigation systems, intraoperative Raman spectroscopy (iRaman), intraoperative ultrasound (iUS), and real-time optical fluorescence imaging, has significantly enhanced the optimization of tumor resection procedures for both benign and malignant neoplasms. However, despite these technological advancements, radiologists often encounter diagnostic ambiguity when attempting to accurately classify the specific malignant tumor type due to inadequate spatial resolution, limited contrast-to-noise ratios, and insufficient molecular specificity, thereby hindering precise histopathological characterization and effective treatment planning [5,6]. Medical imaging repositories pose two significant challenges: an insufficient number of training instances and, more critically, the class-imbalance issue, particularly in datasets with multiple classes, which hinders practical training and model performance. The manual examination of MRI scans is labor-intensive for radiologists and physicians, particularly in complex cases [7]. MRI remains the preferred non-invasive scanning method due to its superior detection of subtle structural changes that are difficult to identify using CT-based imaging. Despite ongoing efforts in brain tumor detection and removal, existing solutions face challenges, including compromised accuracy, brain tumor heterogeneity, false positives, and operational difficulties during surgery. Effective treatment relies on early and precise diagnosis of brain tumors. However, diagnosing brain tumors poses significant challenges due to the diverse and intricate characteristics of tumors in images, such as varying sizes, shapes, locations, and intensities, requiring specialized expertise from neuroradiologists [8]. Therefore, a significant gap exists in addressing this crucial research problem, and the detection and classification of brain tumor malignancy and subtypes remains challenging. Experts are exploring novel approaches to enhance detection accuracy.

The seamless integration of artificial intelligence (AI) in radiology hinges on the transformative power of machine learning (ML) and deep learning (DL) algorithms. A crucial step in this process, feature extraction, unlocks valuable information from raw data, enabling informed decision-making. Gomroki et al. [9] introduced semi-transfer learning (STL) to improve the performance of their encoder-decoder network by using a pretrained model to initialize the encoder path of the proposed network. The Yolov7 network was pretrained as an encoder using the MS COCO dataset, a prominent object detection and segmentation benchmark. The pretrained model already learned general features from a diverse dataset, which helped better feature extraction when applied to the specific datasets used in the study. They used the modified UNet for the decoder path using variable-size kernels instead of fixed-size ones. This helped the complexity and diversity of urban buildings more effectively. Gomroki et al. [10] proposed an STL approach, denoted as EffV2 T-UNet, which integrates the feature extraction capabilities of EfficientNetV2 T (pretrained on ImageNet) as the encoder and the reconstruction proficiency of UNet's convolutional layers as the decoder. The proposed method was evaluated on two datasets. This research utilized an STL approach, where the encoder component (EfficientNet V2T) used pretrained weights from the ImageNet dataset. In contrast, the decoder component (UNet's convolutional layers) was initialized without prior training.

In our proposed superimposed AlexNet frameworks, SAlexNet-1 and SAlexNet-2, the term 'superimposed' is used to describe a novel architectural design that integrates DL techniques, using additional information onto the original model, enabling a better brain tumor detection system from MRI images. The proposed system uses the potential of channel- and spatial-attention mechanisms and the architectural design modification to the original model. The idea of an attention mechanism is exploited in different layers of AlexNet by superimposing the hybrid attention mechanism to provide improved feature space. Furthermore, in the case of STL, a model is pretrained on Task A (source task). It is fine-tuned on Task B (target task) and is similar to Task A. The critical difference is that the model is not fine-tuned from scratch on Task B but builds upon the knowledge learned from Task A. In other words, semi-transfer learning assumes that the model has already learned some general features applicable to tasks A and B and only needs to adapt to the specific differences between the tasks. The term 'semi' acknowledges that the model is still using some of the knowledge from the source task but also needs to learn new information specific to the target task. It is a way to indicate that the model is in an intermediate state between complete transfer learning and training from scratch. The significant contributions of this research effort are listed below.

- Two novice strategies for efficient feature extraction frameworks are presented for primary brain tumor malignancy detection and classification: Superimposed AlexNet-1 & 2 (SAlexNet-1 and its extension SAlexNet-2).
- The SAlexNet-2 architecture is introduced, which further improves the SAlexNet-1 by replacing the first large convolution layer with sequences of smaller-sized convolution kernels to enhance the efficacy of the receptive field with added non-linearity.
- The STL-based SAlexNet frameworks are introduced as variants of the predecessors to improve the classification performance of the correlated datasets. The pretrained networks obtained using DS-1 are further used for the training and testing of DS-2 with associated changes in the lateral layers of the network.
- The proposed frameworks are robust to inherent noise in the dataset and the consequent feature space due to the cascaded-attention mechanism used to discriminate the tumorous regions. They work as residual blocks in sequential convolution layers of SAlexNet architectures.
- An extensive comparison of published and self-adopted datasets is unleashed to highlight the challenges met by the cohorts active in the field.

The paper's structure is as follows: Section 2 reviews relevant literature, while Section 3 details the materials and methods. Results and discussion are presented in Section 4. Section 5 summarizes the conclusions, following a comparison with state-of-the-art techniques, limitation analysis, and future research directions.

2. Literature review

Brain tumor classification is a critical area of research due to its significant implications for patient diagnosis, treatment planning, and prognosis. Accurate classification can lead to more targeted therapies and improved outcomes. However, the complexity and variability of brain tumors pose considerable challenges to achieving precise classification. Over the past decade, numerous studies have explored various methods for brain tumor classification. Early approaches predominantly

relied on traditional imaging techniques and manual interpretation by radiologists. In this context, AI-powered ML and DL systems have been widely adopted for brain tumor detection using MRI scans. Sompong and Wongthanavasu [11] proposed a method for brain tumor segmentation based on a hybrid of fuzzy c-means algorithm and cellular automata. The hybrid approach may not be robust to noise or variations in image intensity. This could lead to a lowered segmentation performance score. Research has shown promising results in brain tumor classification. For instance, Zacharaki et al. [12] developed a system utilizing SVM and k-NN classifiers to categorize glioma grades, achieving 85 % accuracy in multi-class and 88 % in binary classification. The study relies on manual feature extraction, which can be prone to bias and may only capture some relevant information. Similarly, El-Dahshan et al. [13] proposed a method employing discrete wavelet transform (DWT) for feature extraction, principal component analysis (PCA) for dimensionality reduction, and artificial neural networks (ANN) and k-NN for classification. This approach yielded 97 % and 98 % accuracy for brain tumor-normal image classification. DWT hinders effectively capturing non-linear relationships within the tumorous data compared with DL-based methods or texture analysis, potentially improving performance. Another cohort, Cheng et al. [14], enhanced classification performance by dilating and splitting tumor regions. They explored three feature extraction techniques to improve brain tumor classification accuracy: intensity histograms, gray level co-occurrence matrices (GLCM), and bag-of-words (BOW). By combining ring-form partition with tumor region augmentation, they achieved an accuracy of 91.28 %. Our approach involved selecting the most distinctive features through representation learning and using an attention mechanism to focus on specific parts of the input data relevant to the task.

In their study, Paul et al. [15] investigated the efficacy of two neural network architectures, a fully connected neural network (FCNN) and a convolutional neural network (CNN), for classifying brain tumor scans. Notably, their CNN architecture, which integrated convolutional layers with max-pooling and fully connected layers, demonstrated an accuracy of 91.43 %. Nevertheless, their approach was effective for tumors requiring image dilation and the formation of ring-like sub-regions. Anaraki et al. [16] introduced models to classify brain tumors based on CNN and genetic algorithms (GA-CNN). Case study results showed 90.9 % accuracy for glioma grading and 94.2 % for tumor type classification. The reliance on genetic algorithms can lead to computational inefficiencies and potential convergence issues. Building on Hinton et al.'s [17] deep neural network, research advanced, and Litjens et al.'s [18] review demonstrated the viability of automatic feature extraction, replacing handcrafted features in intensive learning. DL architectures used for the segmentation of brain structures and brain lesions, as well as their performance, speed, and properties, are discussed in [19]. The analysis is biased towards specific architectures or datasets. Swati et al. [20] introduced an innovative feature extraction framework grounded in the VGG19 architecture. This framework integrated closed-form metric learning (CFML) to assess the similarity between query and database images while exploiting transfer learning and block-wise fine-tuning strategies to refine retrieval accuracy. However, limitations include potential mismatches between pretrained models and target tasks.

Moreover, feature extraction with GLCM and CNN was introduced to classify brain tumors and processed as DL to measure accuracy [21]. Badža et al. [22] introduced a CNN architecture for brain tumor classification that was tested on T1-weighted CE-MRI. The results were obtained from record-wise cross-validation, and the accuracy was 96.56 %. Our work met the objective using AlexNet architecture as the fundamental system subjected to performance enhancement, introducing a cascaded attention mechanism and feature boosting through dimension reduction strategies. The feature set was finely used to train and test potential ML classifiers.

Researchers in [23,24] extensively reviewed DL techniques, finding that many models demonstrate domain-specific effectiveness and can be trained through multiple methods. They conclude that hybrid DL models, combining conventional approaches, offer improved efficiency in addressing the limitations of traditional DL models. Latha et al. [25] used brain tumor MR images and segments using Otsu's threshold technique, where segmented images were transmitted to DWT to get the features. The extracted features were further subjected to the PCA for dimensionality reduction. Moreover, the synthetic minority over-sampling technique (SMOTE) alleviated the class imbalance problem. They tested the system using *k*-NN and SVM models to perform the classification task. Hsieh et al. [26] presented a comprehensive brain tumor classification methodology incorporating ROI extraction, feature extraction and selection, and classification. Their analysis of 107 glioma images (73 low-grade, 34 high-grade) demonstrated the efficacy of combining local texture and global histogram moments. The study's relatively small dataset cannot ensure generalizability to larger, more diverse populations. Additionally, the analysis focused solely on gliomas, leaving room for investigation into other brain tumor types. Sachdeva et al. [27] introduced a sophisticated CAD system for brain tumor diagnosis. By integrating color and textural feature extraction from segmented ROIs with GA-based feature selection, they attained accuracy rates of 91.70 % (GA-SVM) and 94.90 % (GA-ANN). The reliance on manual segmentation of regions of interest and the GA-based feature selection process can be computationally expensive, increasing computational complexity.

Soltaninejad et al. [28] presented a novel 3D supervoxel learning paradigm for tumor segmentation in multimodal MRI brain scans, utilizing random forest classification into tumor core, edema, and healthy brain tissues. However, the reliance on random forest classification needs to be more fair to complex feature relationships. Huang et al. [29] introduced a multi-task DL architecture for brain tumor segmentation, incorporating fusion units and distance-transform decoder modules. The potential overfitting risks are accompanied by the multi-task learning framework. Fan et al. [30] introduced RMAP-ResNet, a novel DL architecture for accurate brain tumor segmentation in OCT imaging. By integrating a residual multi-core attention pooling module, this model effectively captures tumor regions of diverse sizes through the synergistic combination of spatial attention and multiple receptive fields. Evaluation of RMAP-ResNet on a mouse brain tumor OCT image dataset yielded impressive results, with a Dice score of 94.78 % and Intersection over Union (IoU) of 90.58 %. The study's focus on the mouse brain creates challenging applicability to human brain tumor segmentation.

Islam et al. [31] investigated the efficacy of TL-based architectures in classifying brain tumors from MRI scans. Their research utilized four prominent pretrained models, VGG19, InceptionV3, DenseNet121, and MobileNet, to categorize brain tumors into four classes. Image augmentation techniques were employed to address data imbalance. Data augmentation techniques cannot address the issues related to dataset diversity and quality. Further, the evaluation of only four pretrained models may not provide a comprehensive understanding of TL potential. Aljohani et al. [32] proposed a novel automated framework using metaheuristic optimization for brain tumor diagnosis and classification, integrating CNNs with TL. The approach utilized the Manta-Ray Foraging Optimization (MRFO) algorithm to optimize CNN hyperparameters, enhancing model performance. The study employed a multimodal imaging approach, incorporating X-ray and MRI images for tumor classification. A two-stage process was implemented: initial diagnosis (healthy or brain tumor) using X-ray images, followed by tumor type classification using MRI scans. The study, however, is limited to a specific optimization technique, limiting the exploration of other potentially effective optimization techniques.

Several sophisticated DL models have been employed for brain tumor classification and segmentation tasks [23,33–38]. Recently, Krishnan et al. [34] developed the Rotation Invariant Vision Transformer (RViT) to tackle the longstanding issue of orientation variability in medical imaging, particularly pertinent to brain tumor classification using MRI scans. The RViT's innovative design incorporates rotated patch

embeddings, enabling robust handling of brain tumors across diverse orientations. Evaluation on the Brain Tumor MRI Dataset demonstrated exceptional results, with RViT achieving an accuracy of 98.6 %. However, manual annotation can be time-consuming, expensive, and prone to inter-observer variability, potentially limiting RViT's applicability. Agrawal et al. [35] proposed the Auto Contrast Enhancer, Tumor Detector, and Classifier for enhanced brain tumor diagnosis and classification from poor-quality MRI images. The classifier operates in two phases, first employing an optimized double threshold weighted constraints histogram equalization (ODTWCHE) technique to enhance image contrast for accurate tumor detection, and then using pretrained Inception V3 for refined tumor classification through deep transfer learning. Suboptimal tuning of ODTWCHE and Inception V3 leads to reduced accuracy or inconsistent results. RobU-Net [36] is a robust technique proposed to capture fuzzy border information of brain tumor MRI scans using 2D slices of a T1-weighted MRI dataset using an intense upside-down convolution approach that concentrates on the responsive patches. However, the clinical study needs to be addressed, which is challenging, requiring human subjects and ethical considerations. Aboussaleh et al. [37] introduced a hybrid 3D brain tumor segmentation model, combining V-Net and 3DU-Net's strengths. The model extracts features using encoders, adds 3D convolution and Transformer layers, fuses features at each decoder depth, and refines the output with a final convolution block. The hybrid model's layers raise significant computational requirements, potentially limiting its deployment in clinical settings. In another recent work, Priya et al. [38] introduced a hybrid framework for brain tumor classification, combining AlexNet with a Gated Recurrent Unit (GRU) to improve MRI diagnosis. The process began with denoising and enhancing MRI images using a non-local means filter. AlexNet was used for deep feature extraction, while the GRU addressed gradient vanishing issues. The SoftMax classifier then categorized tumors into glioma, meningioma, pituitary tumor, and normal tissue. This model achieved a 97.00 % accuracy and a 97.25 % F1-score, demonstrating its effectiveness. However, the study's reliance on a specific preprocessing technique limits exploration of other denoising methods leading to bias in image enhancement. Our study proposes novel DL frameworks, SAlexNet-1 and SAlexNet-2, for enhanced brain tumor detection from MRI images. The 'superimposed' design integrates channel- and spatial-attention mechanisms into the original AlexNet model. By STL, the frameworks use pretrained knowledge and adapt to specific task differences, streamlining the learning process and improving performance.

3. Materials and methods

This section presents the materials and methodology underlying the investigation of the proposed brain tumor classification approach. Two publicly available datasets, DS-1 and DS-2, are utilized to evaluate the efficacy of the developed architectures. The methodology comprises several key components: data preprocessing, data augmentation, attention-based DL, and performance evaluation. Specifically, the study employs advanced techniques such as wavelet denoising, bilinear interpolation, and hybrid attention mechanisms to enhance image quality and feature extraction. Two novel architectures, SAlexNet-1 and SAlexNet-2, are developed and trained using these methodologies, with their performance evaluated using six key metrics: recall, precision, F1-score, accuracy, and area under the curve (AUC) for receiver operating characteristics (ROC) and precision-recall (PR) curves. The proposed feature extraction framework is illustrated in Fig. 1, presenting the overall workflow of different processes via training and test phases. The complete proposed framework is depicted in Fig. 1(c). The dataset, constituting the training and test sets, is processed using pipeline steps depicted in the diagram. The proposed model is trained using the loss function and optimized using the training set. The trained model is then used to evaluate the test set. The evaluation phase is followed by extracting training and testing features (in the forward direction) and feature boosting. These discriminative features are further used to detect and classify brain tumor types.

The AlexNet architecture [39] comprises eight layers, divided into two primary segments: convolutional and fully connected layers. Initially, input images pass through five successive convolutional layers. The first convolutional layer applies 96 filters, each measuring 11×11 pixels, to the input image. The output is then fed into the second convolutional layer, which employs 256 kernels of size 5×5. Following this, pooling operations are performed on the output. This process is repeated for the third, fourth, and fifth convolutional layers, each applying distinct filter sizes and numbers. Subsequently, the output is flattened and passed through three fully connected networks (FCNs) [40].

3.1. Publicly available datasets

This study uses two publicly available datasets to evaluate the efficacy of the proposed brain tumor classification approach. The first dataset, Msoud (DS-1), is a hybrid collection of 7023 MRI scans, while the second dataset (DS-2) comprises binary classification MRI scans. A detailed analysis of these datasets provides valuable insights into the performance and sensitivity of the proposed architectures.

In this context, DS-1 merging three challenging datasets (FigShare.com, SARTAJ, and Br35H datasets) has four classes: glioma, meningioma, pituitary, and no-tumor (Notumor), and it is retrievable through the Kaggle database [41]. The use of DS-1 has been illustrated in Fig. 1 (a). In the DS-1, the contributors already define and delineate the training and test partition. The processing pipeline is used to produce the images for the forthcoming steps to training and test the unknown instances for their generalization score. The preprocessed training instances are then forwarded to the augmentation process. The augmented training set is used to optimize the DL model to produce a suboptimal solution. Finally, the test partition is passed over the optimized model to the label prediction, leading to the proposed model's performance measure. Table 1 depicts the individual class status associated with each of the datasets. The FigShare.com dataset consists of three classes (Glioma, Meningioma, and Pituitary) providing 3064 samples, constituting 44 % share in DS-1. The SARTAJ dataset consists of four classes (Glioma, Meningioma, Pituitary, and Notumor) providing 2459 samples, having a 35 % contribution in DS-1. Fig. 2 shows the visual representation of the DS-1 formation. The Br35H dataset has two classes but only the Notumor class is included in DS-1. It has a 21 % contribution in DS-1 providing 1500 images for the Notumor class. The total samples in DS-1 are 7023.

Similarly, Table 2 illustrates a comparison of multiple views of three classes of brain tumors and one class that represents the Notumor cases. Fig. 3 explores the instance's distribution for class imbalance in DS-1. A significant class imbalance can be observed in DS-1.

3.1.1. Dataset for binary classification (DS-2)

Another Kaggle brain tumor MRI scans dataset [42], DS-2 used for binary classification, through a 5-fold cross-validation scheme, as shown in Fig. 1(b). The SAlexNet-1 and SAlexNet-2 algorithms were also tested on the DS-2 dataset, to substantiate the results as a multi-centric study. The dataset consists of classes 'No' and 'Yes', each with 1500 images. The images are placed in two folders with names corresponding to their class. The image names in the folder 'No' and 'Yes' are divided into 5 equal splits. For each of the i^{th} folds, the images in the i^{th} split are taken for the validation set, and the images in sets j (not equal to i) are merged in one array as the training set. Hence, the size of the validation set in each split is 600 images, and for the training set, the size is 2400 images. Table 3 shows the visual representation of samples for the two classes. Since no class imbalance was present, no augmentation was required for the training instances.

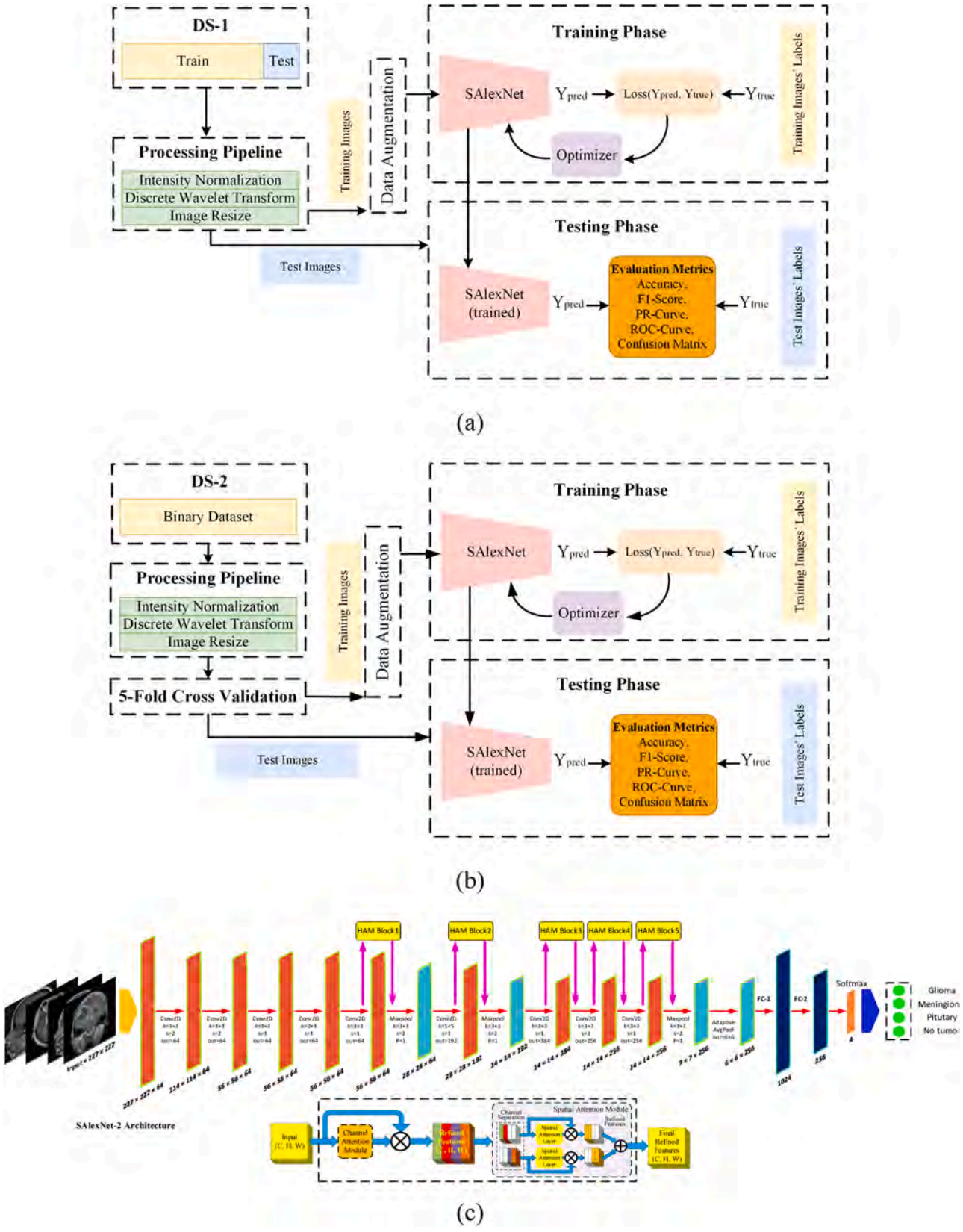


Fig. 1. The overall workflow of the proposed framework (SAlexNet-1 / SAlexNet-2) for preliminary feature extraction. After preprocessing, the dataset is divided into train and test sets by k-fold cross-validation. The transformed train set is used to train the proposed models. Before feature extraction, the test set is used to evaluate the optimized model to obtain the evaluation metrics (a) for DS-1, (b) for DS-2, and (c) the proposed SAlexNet-2 framework with HAM.

Table 1
Individual classes share a percentage associated with each dataset indicating the training, testing, and total number of instances of each class.

Dataset	Category	Samples				Share (%)
		Train	Test	Total	Dataset	
Figshare	Glioma	–	–	1426	3064	44 %
	Meningioma	–	–	708		
	Pituitary	–	–	930		
SARTAJ	Glioma	95	100	195	2459	35 %
	Meningioma	822	115	937		
	Pituitary	827	–	827		
	Notumor	395	105	500		
Br35H	No	–	–	1500	1500	21 %
	Yes	–	–	–		
DS-1 (Msoud dataset)	Glioma	1321	300	1621	7023	100 %
	Meningioma	1339	306	1645		
	Pituitary	1457	300	1757		
	Notumor	1595	405	2000		

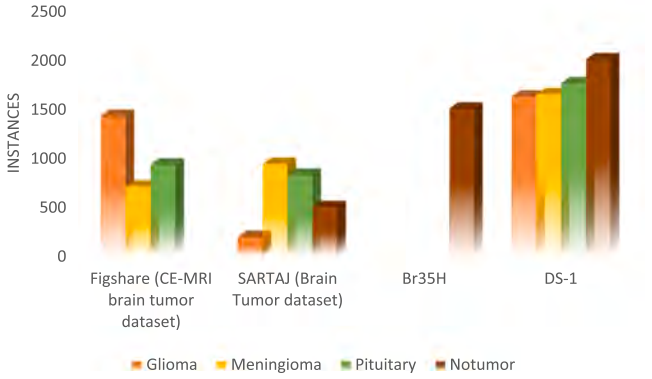


Fig. 2. Graphical view of the DS-1 highlighting the contribution of multiple brain tumor classes.

3.2. Preprocessing of datasets

This section describes the preprocessing techniques applied to the DS-1 and DS-2 datasets to ensure optimal image quality for brain tumor classification. The preprocessing pipeline consisted of image normalization, wavelet denoising, and resizing using bilinear interpolation.

DS-1 and DS-2 datasets consisted of grayscale images with intensity values between 0 and 255. The images were normalized for all experiments from the range [0, 1]. For image denoising, the Haar wavelet was applied sequentially using the original image's wavelet transform. The initial transformation yielded four matrices denoted LL1, LH1, HL1, and HH1, representing the approximation, horizontal, vertical, and diagonal detail components. Subsequently, a secondary wavelet transform was applied exclusively to the approximation component (LL1), resulting in four additional matrices denoted as LL2, LH2, HL2, and HH2. Notably, the horizontal and vertical detail matrices (LH2 and HL2) obtained from this secondary transform were set to zero matrices. Following this, a reconstruction wavelet (inverse wavelet transform) was executed, utilizing LL2, LH2, HL2, and HH2, with the constraint that LH2 and HL2 were null matrices. The resultant reconstructed image, derived from this modified transformation process, served as the noise-suppressed input for subsequent stages in the classification framework. Wavelets are defined by the function $\psi(t)$, the mother wavelet, and the scaling function $\varphi(t)$ (the father wavelet in the time domain). The wavelet function is a band-pass filter, scaling for each level half its bandwidth. The wavelet transform is defined as

$$Y_{(s,\tau)} = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-\tau}{s}\right)dt, \ a > 0, \tag{1}$$

where $\psi(t)$ represents the scaling factor and τ is a translation factor. Haar wavelet is constructed using the multiresolution analysis (MRA) generated by the scaling function as given by

$$\varphi(t) = \sum_n H(n)\sqrt{2}\varphi(2t-n), \tag{2}$$

where $H(n)$ is the scaling coefficient. This equation is also known as the refinement equation or multiresolution analysis equation.

For image resizing, we have used bilinear interpolation, where the position of the pixel in the enhanced image is transformed into the original image, and then the influence of the four-pixel points a, b, c, and d is calculated, as illustrated in Fig. 4. Suppose the coordinates a, b, c, and d are represented as (i, j), $(i, j+1)$, $(i+1, j)$, and $(i+1, j+1)$. The coordinates of Z are (u, v). Bilinear interpolation is carried out by calculating the influence of a and b and denoting it as X, as given by $f(i, j+v) = [f(i,j+1) - f(i,j)]v + f(i,j)$. Next, calculate the influence of C and D and denote it as Y, given by $f(i+1, j+v) = [f(i+1, j+1) - f(i+1, j)]v + f(i+1, j)$. Finally, the influence of X and Y is calculated and denoted as Z, as given by $f(i+u, j+v) = (1-u)(1-v)f(i,j) - (i-u)vf(i,j+1) + u(1-v)f(i+1,j) + uvf(i+1,j+1)$. This method performs interpolation in horizontal and vertical directions. It provides better results than the nearest neighbor interpolation and is computationally less intensive than bicubic interpolation [43,44].

3.3. Data augmentation

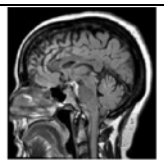
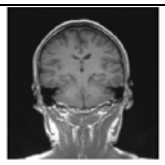
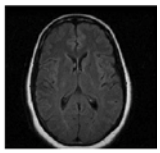
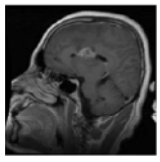
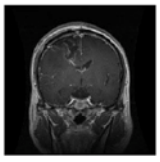
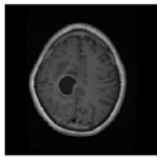
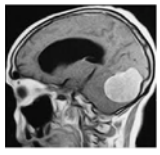
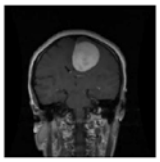
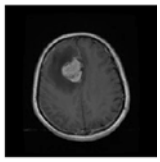
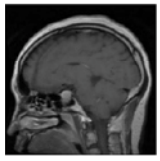
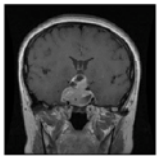
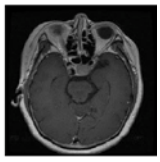
In this section, data augmentation was employed to address the class imbalance inherent in DS-1 and enhance the generalizability of the proposed DL techniques. A comprehensive augmentation strategy was developed, incorporating four primary techniques: horizontal and vertical flips, photometric distortion, Gaussian blur, and Gaussian noise injection. These techniques were carefully selected and combined to create a robust and diverse training dataset, ultimately improving the performance of the proposed SAlexNet-1 and SAlexNet-2 models.

The main augmentation techniques, horizontal and vertical flips, photometric distortion to create diverse images, Gaussian blur, and noise, have been illustrated in Fig. 5. In the horizontal/vertical flip augmentation, the image is rotated along the X-axis, with 0.5 probability, to flip the image horizontally / vertically. In the photometric distortion augmentation [45,46], hue, saturation, contrast, and brightness are randomly sampled from the different ranges defined in Table 4, and then the image is perturbed based on sampled parameters. More significant parameters give more challenging samples, whereas smaller parameters give identical samples. During training, a random value is selected for each image sample with 0.5 probability. In the Gaussian blur augmentation, the image is blurred using a Gaussian filter with kernel size randomly sampled from the set (5, 7, 9). The blurring technique helps the model focus on coarse features. In the Gaussian noise augmentation, the image is degraded by adding noise from the Gaussian/normal distribution, and the normal distribution with zero mean and standard deviation 0.01 was used for this operation. The probability represents the chance this operation might get applied to an image sample during training. The augmentation process is not used during testing for the images under trial. The augmentation pipeline evolved during the experimentation phase. The initial experiments were performed with flip operation variations, leading to sub-optimal results. Then, photometric distortion was added to improve the generalization of the model. Finally, Gaussian blur and noise were tried in the pipeline, improving the results. The exact final pipeline was used to train the proposed SAlexNet-1 and SAlexNet-2 models.

3.4. AlexNet architecture

The AlexNet architecture [39] is a well-established DL model, which our proposed SAlexNet-1 and SAlexNet-2 models build upon. This

Table 2
Multiple view brain tumor DS-1 representing instance cardinality along with their partitioning for training and test groups.

Class	View		
	Sagittal	Coronal	Transverse
Notumor			
Glioma			
Meningioma			
Pituitary			

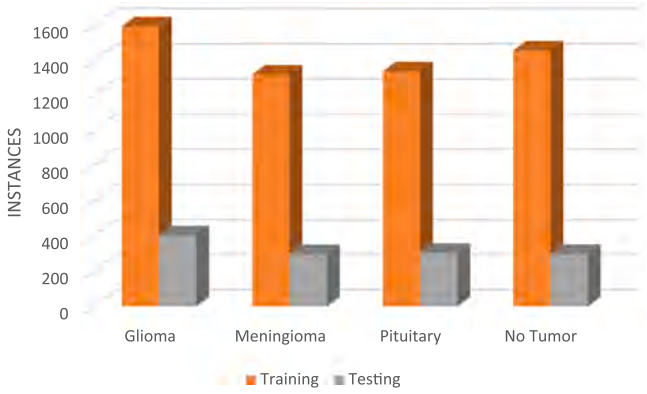


Fig. 3. DS-1, which is publicly available, shows a graphical view of its class imbalance.

section describes the original AlexNet architecture, highlighting its convolutional and fully connected layers.

The AlexNet consists of eight layers, consisting of five convolution layers and three fully connected (FC) layers. The input images go through the convolution layers, with ReLU and batch normalization after each convolution, followed by FC layers with ReLU after the first two FC layers and Softmax after the last FC layer. There are MaxPool layers after the first, second, and fifth convolution layers to make the model invariant to rotation, translation, and scale-like features. Images are convolved with 64 filters of size 11 × 11 in the first convolution layer, followed by MaxPool layer. The second convolutional layer transforms the features with 192 filters of size 5 × 5, followed by the MaxPool layer. Similarly, three convolution layers with 384, 256, and 256 kernels (3 × 3 sized) were applied, followed by a third MaxPool Layer. After the last MaxPool layer, the feature maps are passed through an Adaptive average-pooling (AvgPool) layer, which provides a fixed-size feature map of shape (6, 6, 256). These features are transformed using FC1 and FC2 layers of size 4096. The final classification layer gives class probability distribution using the SoftMax function [40].

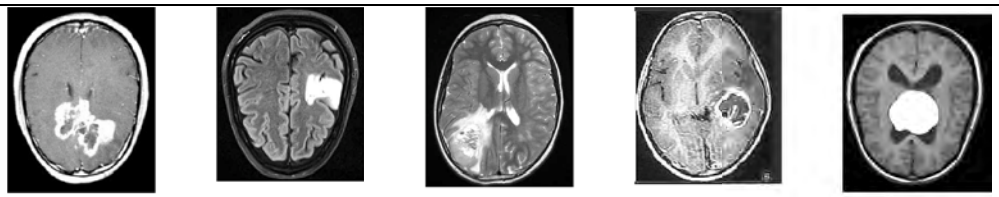
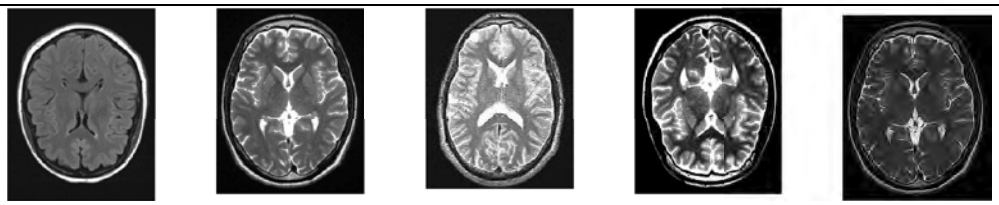
3.5. SAlexNet-1 method

This section introduces SAlexNet-1, a novel extension of the AlexNet architecture designed to enhance feature extraction and noise robustness. Along with the architecture modification, SAlexNet-1 integrates a cascaded attention mechanism, HAM, after each convolutional stage.

The proposed SAlexNet-1 extends the AlexNet by introducing the cascaded-attention mechanism, HAM, after each stage of the convolution layer. The proposed framework is an improvement to the original AlexNet architecture by integrating the HAM module after each convolution layer of AlexNet as shown in Fig. 6. The attention mechanism improves the feature space along channels and the spatial axis. The AlexNet is originally a shallow model, but our proposed SAlexNet-1 takes advantage of these attention blocks to reduce the effect of noise even with a shallowness approach. The HAM block processes the features using channel and spatial attention mechanisms, suppressing the redundant features along the channel and spatial axis respectively. In detail, our proposed SAlexNet consists of five Conv2d layers with a ReLU activation function. The MaxPool layer is applied after the 1st, 2nd, and fifth convolution layers. After the sequence of convolution, ReLU, and MaxPool layers, the features are passed through two FC layers of sizes 1024 and 256 neurons. The final layer transforms the features into 4 classes (DS-1)/2 classes (DS-2) based on the multiclass/binary class dataset.

Consider an input image of shape (H, W, 3). The input is passed through the first convolution layer, with 11 × 11 kernel size, to increase the channels to 64, and then a ReLU activation is applied. The first convolution is applied with a stride of 4 to drastically reduce the spatial

Table 3
DS-2 samples taken from the brain tumor dataset representing the binary MRI scans.

Class	MRI scans
Yes	
No	

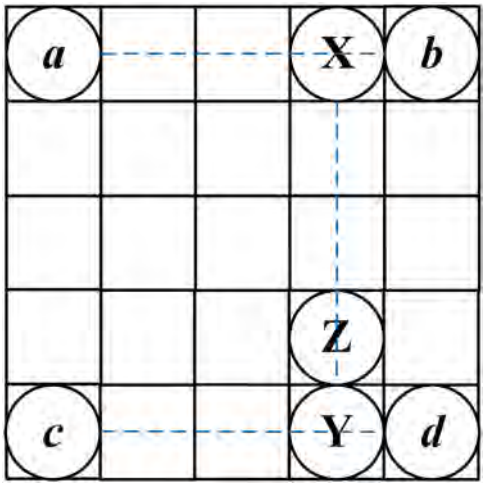


Fig. 4. MRI scan resizing using bilinear interpolation.

size and make the model invariant to noisy features and small changes in the input image. This sequence of convolution, ReLU, HAM, and MaxPool is repeated several times to produce rich discriminative features.

3.5.1. Hybrid attention mechanism

This section investigates the design of the HAM, a crucial component of SAlexNet-1. HAM combines channel attention and spatial attention layers to provide rich and focused features, enhancing feature representation and discrimination.

The sole objective of the HAM is to provide rich and focused features by processing the input through channel attention and spatial attention layers as a linear sequence. The input of shape (Channels, Height, Width) or (C, H, W) is passed through the channel attention module (CAM), which provides a vector of shape (1, 1, c) corresponding to the attention score value per feature channel as illustrated in Fig. 7. These attention scores are multiplied with original feature maps to reweight their importance [47].

An attention value of '0′ corresponds to a least important channel, and an attention value of 1 corresponds to a highly important channel. The feature channels are multiplied with attention scores to reweigh their importance. Channels with noisy information have low attention values, so they get suppressed as compared to channels with good discriminative features, which have high attention values. The spatial attention submodule within the channel attention map performs a channel-axis separation of the refined feature into two distinct groups. Consequently, two independent 2D spatial attention tensors, represented as $A_{S,1}$ and $A_{S,2}$ are generated. After channel attention, the channel refined features and channel attention scores are fed to the SAM that further enhances the feature discrimination by calculating the attention score per pixel location and weighing their importance by multiplying them with corresponding attention scores. The mathematical formulation of the HAM block is given by

$$F' = A_c(F) \otimes F, \tag{3}$$

$$F'_1 \oplus F'_2 = F', \tag{4}$$

$$F''_1 = A_{S,1}(F'_1) \otimes F'_1, \tag{5}$$

$$F''_2 = A_{S,2}(F'_2) \otimes F'_2, \tag{6}$$

$$F'' = F''_1 \oplus F''_2, \tag{7}$$

where $\otimes$ represents element-wise multiplication and $\oplus$ denotes element-wise summation. The multiplication process involves broadcasting channel attention values along the spatial dimension to obtain F', and broadcasting spatial attention values along the channel dimension to produce the spatially refined feature F''_i. Subsequently, during summation, the channel-refined feature F' is divided into two parts along the channel axis, yielding F'_1 and F'_2. Ultimately, the final output F'', is computed as the sum of F''_1 and F''_2, effectively integrating channel and spatial attention information.

3.5.2. Channel attention module

This section explains the design of CAM to efficiently calculate channel-wise attention scores, refining feature extraction in SAlexNet-1. CAM's architecture combines parallel global average and max pooling, adaptive mechanism blocks, and dynamic kernel-sized Conv1D layers.

CAM's central theme is finding each channel's importance (attention) score. Initially, the input feature tensor of shape (C, H, W) is applied with channel-wise GlobalAvgPooling and GlobalMaxPooling layers in parallel to produce two output vectors, F_{avg} and F_{max}, of shapes (1, 1, C). The GlobalAvgPooling layer computes the average value per channel, whereas the GlobalMaxPooling layer computes the maximum value per channel, as shown in Fig. 8.

The two vectors are combined using an adaptive mechanism block, which computes three vectors: (1) alpha times F_{avg}, (2) beta times F_{max},

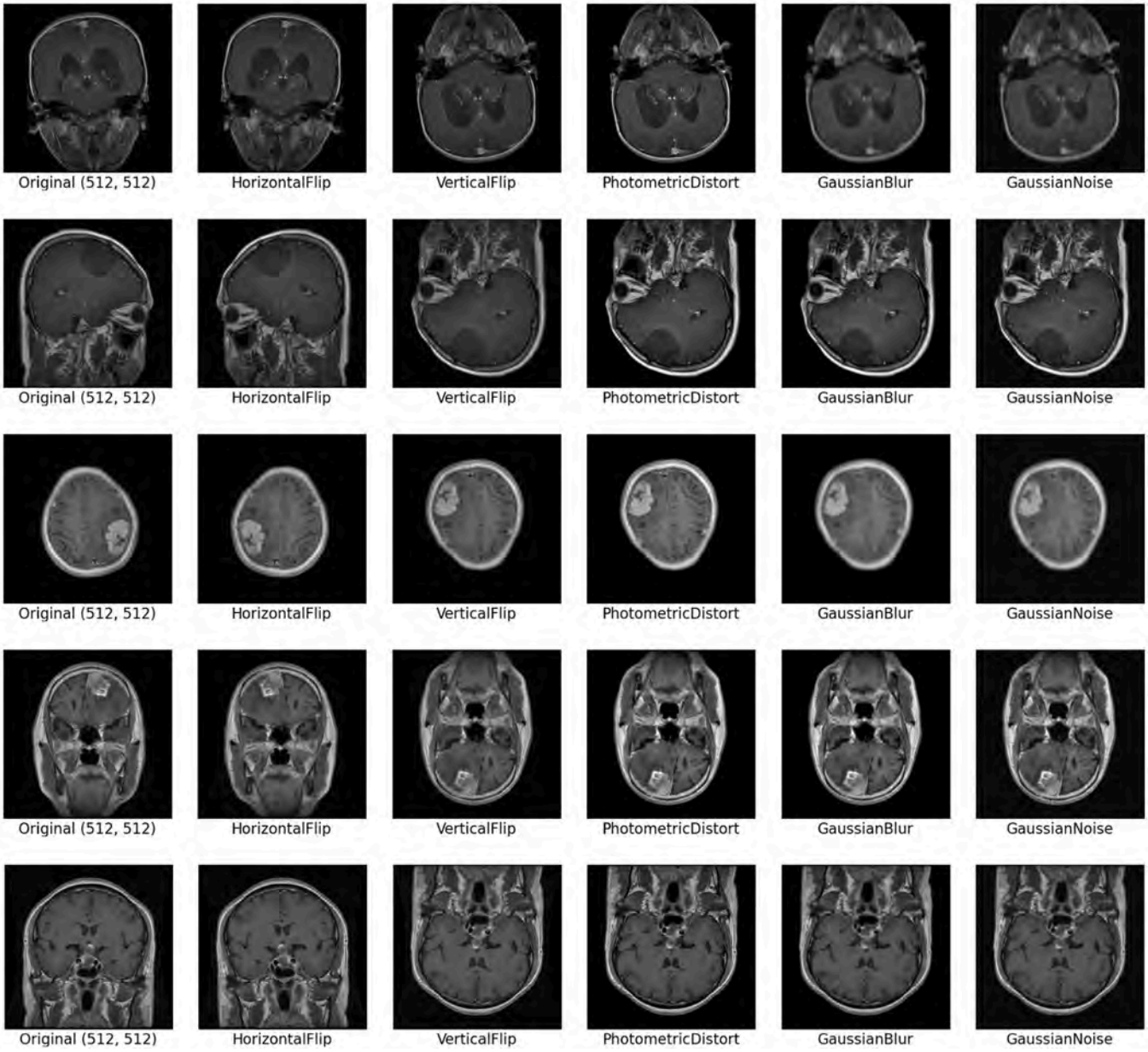


Fig. 5. Augmentation for DS-1 and on diverse image samples. The pipeline takes the original image (first column) and moves from left to right, producing horizontal flipped, vertical flipped, photometric distortion, blurred, and noisy images.

Table 4
Augmentation techniques used for DS-1 training instances with parameters and their corresponding range values.

Augmentation	Description / Selected range	
Horizontal and vertical flip	**180**°	
Photometric-distortion	Hue	[−0.05 - 0.05]
	Saturation	[0.5 - 1.5]
	Contrast	[0.5 - 1.5]
	Brightness	[0.875 - 1.125]
Gaussian blur	Kernel = (5, 7, 9), sigma = [0.1 - 5.0]	
Gaussian noise	Mean (μ) = 0, standard deviation (σ) = 0.01	

and (3) an average of F_{max} and F_{avg}. These three vectors are added together to fuse the information effectively. The parameters α and β are trainable and tend to adapt in this fusion mechanism. Incorporating two trainable parameters enables a dynamic fusion mechanism between AvgPooled and MaxPooled features, augmenting the feature extraction process. Subsequently, the fused vector is subjected to a Conv1D operation with a kernel size (k) that scales logarithmically with the number of channels (C). For example, for a layer with input channels 64, log (C) will provide $k = 3$, and for a layer with input channels 384, log(C) will provide $k = 7$. The dependence of kernel size on the number of channels introduces another source of flexibility into the CAM. The output of Conv1D is applied with the sigmoid activation function to produce an attention score, a value between 0 and 1, for each feature channel. The mathematical formulation for the CAM [48] is given by

$$F_{C_n}^{avg} = AvgPool(F), \tag{8}$$

$$F_{C_n}^{max} = MaxPool(F), \tag{9}$$

$$F_{C_n}^{add} = \frac{1}{2} \otimes \left(F_{C_n}^{avg} \oplus F_{C_n}^{max}\right) \oplus \left(\alpha \otimes F_{C_n}^{avg}\right) \oplus \left(\beta \otimes F_{C_n}^{max}\right), \tag{10}$$

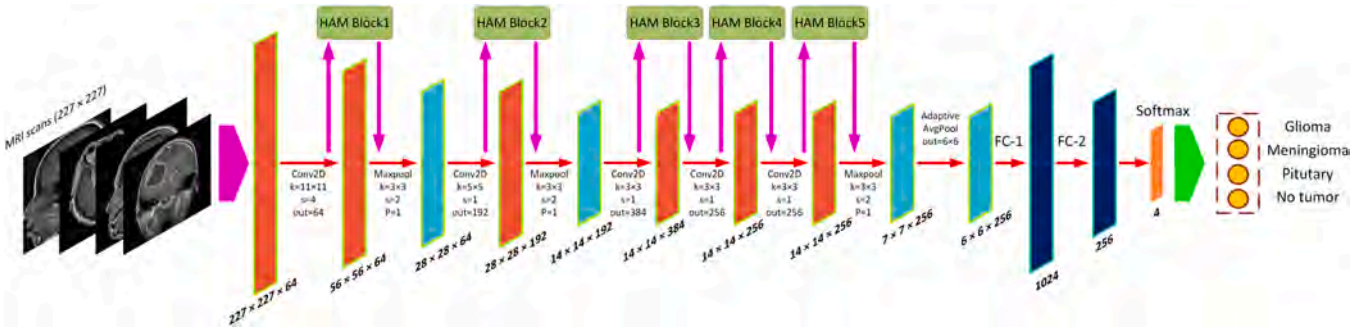


Fig. 6. The SAlexNet-1 with HAM blocks, consisting of channel- and spatial-attention modules, culminating in cascaded cross-domain fused feature space.

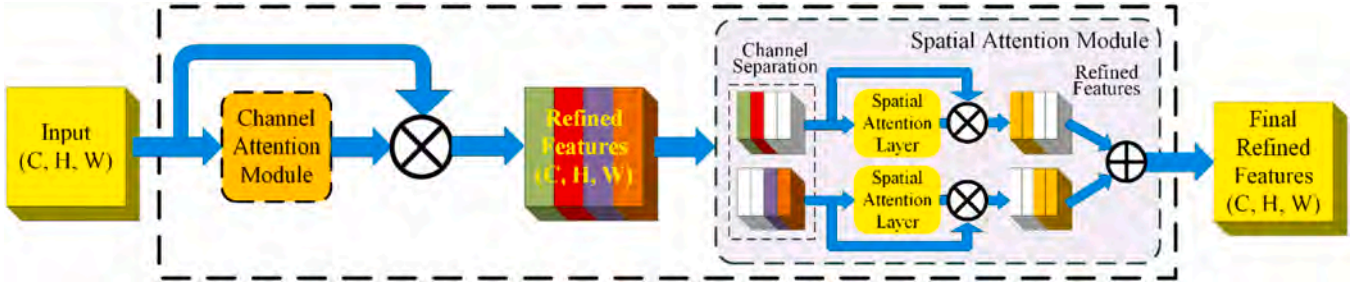


Fig. 7. HAM consists of CAM and SAM. CAM is responsible for enhancing feature representation along the channel axis. The SAM is responsible for improving features along the spatial axis.

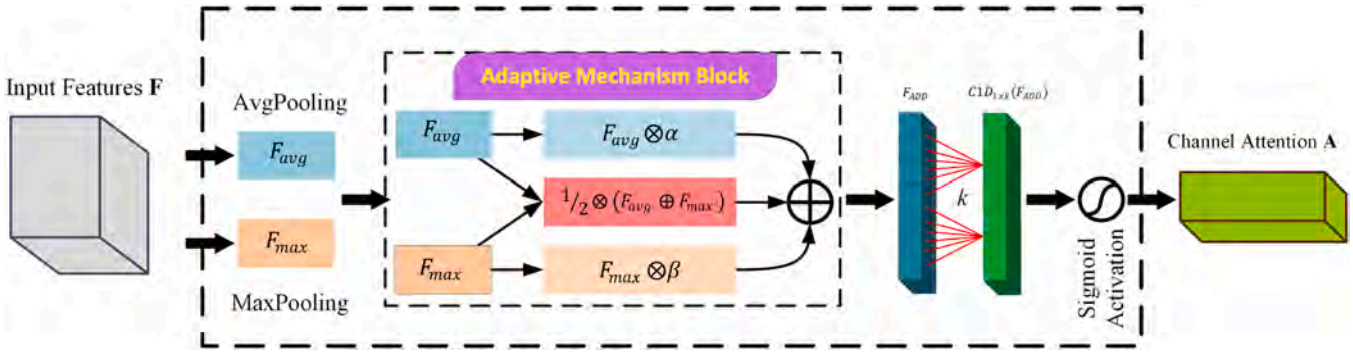


Fig. 8. CAM works by applying AvgPooling and MaxPooling operations on input tensor. It then calculates a linear combination of Favg and Fmax vectors, whereby the linear combination weights are learned during training. The output tensor is then applied with an adaptive kernel size Conv1d layer. Finally, the sigmoid is applied to calculate the importance score for each channel.

$$k = \varphi(C_n) = \left| \frac{log_2(C_n)}{\gamma} + \frac{b}{\gamma} \right|_{odd}, \tag{11}$$

where γ and b are both hyper-parameters. The mapping φ enables adaptive determination of kernel size k based on the number of channels, C_n. Subsequently, the feature map $F_{C_n}^{add}$ undergoes 1D convolution, followed by sigmoid activation, yielding output feature tensors, as given by

$$A_c(F) = \sigma\left(C1D_{1\times k}\left(F_{C_n}^{add}\right)\right), \tag{12}$$

where σ is the sigmoid function and $C1D_{1\times k}$ represents 1D-convolution with a kernel of size k.

3.5.3. Spatial attention module

This section introduces the SAM, a critical component of the HAM. SAM computes spatial importance scores by separating feature channels, applying parallel average and max pooling, and shared convolutional operations.

The input of this module is a feature tensor F' of shape (C, H, W) and channel attention vector of size c. There is a one-to-one correspondence between these inputs, e.g., the first value in the channel attention vector gives an importance score of the first channel in F'. The channel separation mechanism sorts the channel attention scores and splits the vector in half to produce upper and lower attention vectors, as illustrated in Fig. 9.

The feature channels in F', corresponding to the upper attention vector, have rich feature information and are separated into a tensor called F1′. Meanwhile, feature channels in F' corresponding to the lower attention vector have less feature information and are separated as F2′. Both F1′ and F2′ have shapes (C/2, H, W). Each feature tensor F1′ and F2′ undergoes the following procedure: Firstly, pixel-wise AvgPool and MaxPool layers are applied parallel to produce two feature channels F_{avg} and F_{max} of shape (1, H, W). The AvgPool layer computes the average value per pixel location, whereas the MaxPool layer computes the maximum value per pixel location. The two channels are concatenated

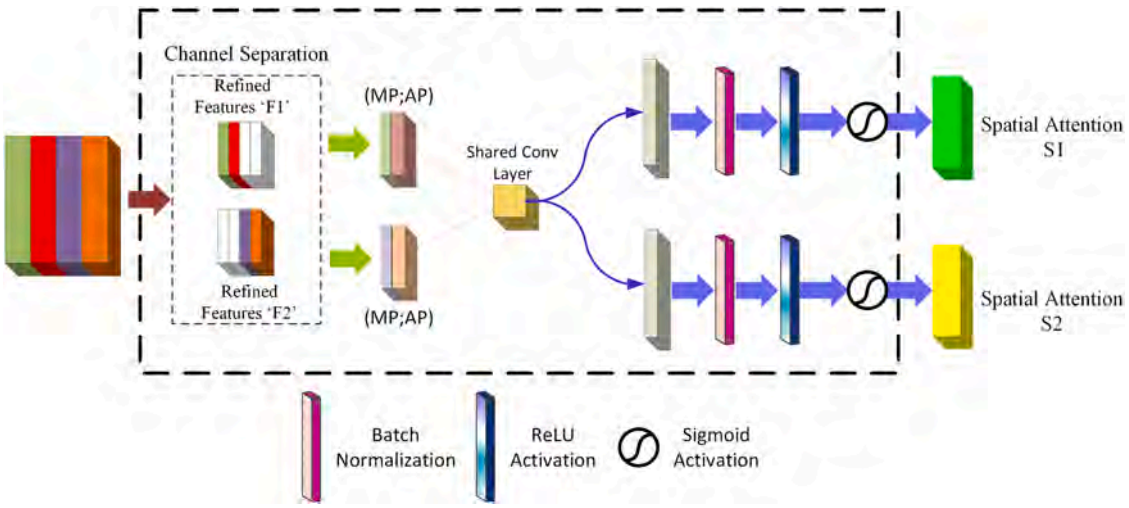


Fig. 9. SAM works by initially separating the features along the channel axis into two groups based on channel attention scores. Then, it applies MaxPool and AvgPool, to each group to obtain pixel-wise primary information. Then, shared convolution layers are applied to process spatial information. Finally, the sigmoid is applied to obtain an importance score for each pixel location in both groups. The output of both groups is further concatenated along the channel axis.

along the channel axis to produce a tensor of shape (2, H, W). Secondly, the concatenated tensor is processed using a shared Conv2D layer with kernel size 7, BatchNorm, ReLU, and Sigmoid layers in the same sequence. The convolution operation aims to fuse the information of MaxPool efficiently with AvgPool features. After applying these steps on $F1'$ and $F2'$, we get spatial attention scores $A1'$ and $A2'$, respectively. These spatial attention scores are applied to their corresponding tensors, $F1'$ and $F2'$, to enhance spatial information and reduce noisy information along spatial axes. The final output tensor is obtained by concatenating the spatially improved $F1'$ and $F2'$ tensors. The mathematical formulation of the spatial attention block [47,49] is given by

$$A_{S,1}(F_1') = \Phi\left(Conv2D_{7\times 7}\left(Concat\left(AvgPool(F_1');\ MaxPool(F_1')\right)\right)\right) = \Phi\left(Conv2D_{7\times 7}\left(Concat\left(F_{S,1}^{avg};\ F_{S,1}^{max}\right)\right)\right), \quad (13)$$

$$A_{S,2}(F_2') = \Phi\left(Conv2D_{7\times 7}\left(Concat\left(AvgPool(F_2');\ MaxPool(F_2')\right)\right)\right) = \Phi\left(Conv2D_{7\times 7}\left(Concat\left(F_{S,2}^{avg};\ F_{S,2}^{max}\right)\right)\right). \quad (14)$$

The proposed architecture incorporates a nonlinear transformation Φ, comprising three stages: batch normalization, ReLU activation, and sigmoid activation. This transformation is complemented by a shared 2D convolutional layer, $conv2D_{7\times 7}$, utilizing a 7×7 kernel to facilitate efficient feature extraction.

3.6. SAlexNet-2 method

This section presents SAlexNet-2, an optimized variant of SAlexNet-1, designed to improve computational efficiency and reduce overfitting. By replacing large kernels with sequences of smaller kernels, SAlexNet-2 enhances pattern representation and captures intricate tumor details.

The second proposed architecture, SAlexNet-2, further improves SAlexNet-1 by replacing the first large-sized kernel with a sequence of smaller-sized kernels. This makes the network more computationally efficient and reduces the risk of overfitting. Finally, using multiple smaller kernels in deeper layers provides a hierarchical structure that allows CNNs to learn complex patterns and representations by combining simpler ones [50]. Further, the low-level features (minor tumorous details) in an image, such as edges and blobs and high-frequency details, are assumed to be local and, thus, captured best by using small-sized kernels.

The initial layer of both original AlexNet and SAlexNet-1 employs an 11×11 convolutional kernel (Fig. 6), which, despite forming a large receptive field, proves inefficient in parameter utilization and ineffective in capturing intricate tumor details in MRI scans. In contrast, our proposed SAlexNet-2 architecture (Fig. 10) adopts a series of 3×3 convolutional layers, drawing inspiration from VGG's non-linear modeling capabilities. This design modification yields two significant advantages. Firstly, the repeated convolutional layers introduce enhanced non-linearity, facilitating the representation of complex functions. Secondly, the total parameter count is substantially reduced. For instance, a single 11×11 convolutional layer with 64 kernels requires 23,232 parameters, whereas five 3×3 convolutional layers necessitate merely 8640 parameters (assuming a single gray-level input channel). This strategic revision enables SAlexNet-2 to surpass the performance of SAlexNet-1. The proposed methodology, where channel and spatial attention mechanisms employed are illustrated in Fig. 11.

3.7. Training procedure

This section outlines the training procedure for SAlexNet-1 and SAlexNet-2, addressing challenges associated with neural network training, such as vanishing gradients and overfitting. Key strategies employed attention mechanisms, data augmentation, regularization, and batch normalization.

Neural network-based systems are complex and have several challenges and problems [51–53]. Neural networks employ online weight tuning to adapt and learn in real time, refining their performance as they receive new data. This adaptive process utilizes devised tuning laws, such as stochastic gradient descent or reinforcement learning algorithms, to adjust weights and minimize errors. By continuously updating weights based on incoming data, neural networks can learn from experience, respond to changing environments, and improve accuracy over time. One major issue is the vanishing or exploding gradient problem, where gradients become very small or large during back-propagation, causing weights to update improperly or not at all. In our proposed training method, the attention mechanism enables parallel processing of input sequences, eliminating the sequential back-propagation through time that aggravates vanishing gradients. Additionally, attention computes a weighted sum of input segments, focusing on relevant information and preserving gradient information by assigning greater importance to critical components.

The AlexNet-based models are prone to overfitting. The proposed systems mitigate this by adopting augmentation during training, regularization to penalize large weights using AdamW as an optimizer, and batch normalization. Furthermore, finding optimal hyperparameters

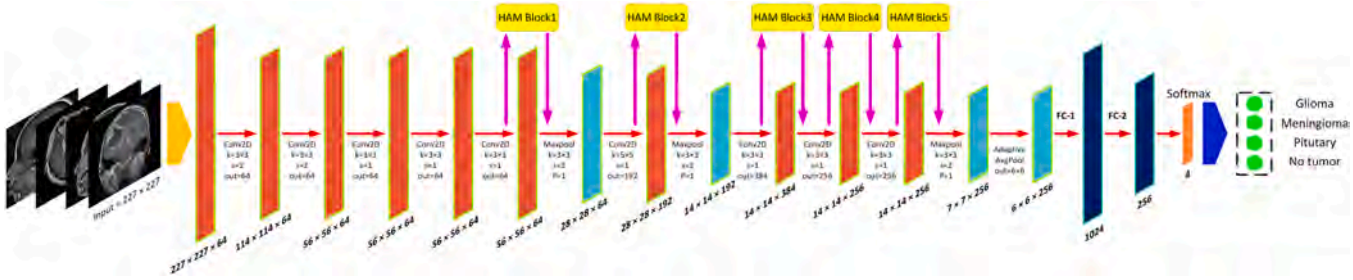


Fig. 10. The architectural view of the proposed SAlexNet-2 that extends SAlexNet-1 by replacing kernel of 11 × 11 with a sequence of smaller sized (3 × 3) kernels.

	Algorithm: Proposed methodology using Hybrid Attention Mechanism
	Input: **M** = Feature maps of dimension (C×H×W) **R** = Ratio factor in Channel Attention
	Output: M'' = Refined feature maps after applying HAM on SAlexNet-1 and SAlexNet-2
1	Channel Attention Channel reduction $C_R = ReLU\left(BN\left(conv\left(M, max\frac{C}{R}, 1\right), 1\times 1\right)\right)$, where, $ReLU(x) = max(0, x)$ Batch Normalization (BN) is defined for the given batch x_i; $\mu_k = \frac{1}{m}\sum_{i=1}^{m} x_i$, where, μ_k is mini batch and m represents mini batch size, Average pooling $A_p = G_{AP}(C_R)$, Max pooling $MP = G_{MP}(C_R)$ Reshaping $A_R = Reshape\left(concat(AP, MP), \left[1, 1, 2 \times \max\left(\frac{C}{R}, 1\right)\right]\right)$ Scaling $S = \sigma(conv(A_R, C, 1\times 1))$, where, sigmoid is defined as $\sigma(x) = \frac{1}{1+e^{-x}}$ $M' = M \oplus (M \otimes S)$
2	Spatial Attention Compress to single channel $M_c = ReLU\left(BN\left(conv(M', 1, 1\times 1)\right)\right)$ Multiscale convolution operation $conv_1 = conv(M_c, 1, 3\times 3), conv_2 = conv(M_c, 1, 7\times 7)$ Attention Map $M_A = sigmoid\left(conv(concat(conv_1, conv_2),\ 1,\ 3\times 3)\right)$ $M'' = M' \oplus (M' \otimes M_A)$
3	Combining with input we get, $M'' = M \oplus M''$

Fig. 11. Illustration of the proposed methodology's algorithm, highlighting the key steps involving channel and spatial attention mechanisms in SAlexNet-1 and SAlexNet-2.

can be challenging, and scalability remains a concern as large networks require substantial resources.

The SAlexNet-1 algorithm is presented in Fig. 12 using dataset DS-1. The algorithm begins with dataset preparation and model definition, followed by hyperparameter initialization. The training process consists of two loops: an outer loop iterating over epochs and an inner loop iterating over batches of the dataset. In each batch iteration, images are passed through the model to obtain predictions, and the loss is calculated by comparing predictions with true labels. The gradients of model parameters are then determined, and the optimizer updates the model parameters. Hyperparameter selection is detailed in Section 4.2.

The training model learns to classify brain tumor images into four categories: 'notumor', 'glioma', 'meningioma', and 'pituitary'. This process involves defining hyper-parameters such as epochs, learning

Algorithm: Run Complete Experiment

Input:
(a) Dataset consisting of images X and output labels Y. Each image x has a shape $(H, W, 3)$ and each label $y \in \{'notumor', 'glioma', 'meningioma', 'pituitary'\}$

Output: A directory containing the best model, plots, evaluation metrics, etc.

1. Define Hyper-Parameters: $epochs_{total}$, $Learning\_Rate$, $weight\_decay$, etc.
2. Define augmentation pipeline consisting of **{Resize, HorizontalFlip, VerticalFlip, PhotometricDistortion, GaussianBlur, GaussianNoise}**
3. Prepare and create $dataset_{train}$ and $dataset_{test}$.
4. Create $SAlexNet-1\ Model$, AdamW $Optimizer$, $Scheduler_{LR}$, and $Loss\_Function$ objects.
5. For $epoch = 1, \ldots, epochs_{total}$ do
    # train model for one epoch
    For $(x_{batch}, y_{batch}) = GET\_BATCH(dataset_{train})$ do
        $\hat{y}_{batch} = Model(x_{batch})$
        $loss = Loss\_Function(\hat{y}_{batch}, y_{batch})$
        $Optimizer(loss, Model)$
        Track Training Matrices
    End
    Calculate and save metrics on $dataset_{train}$.
    Calculate and save metrics on $dataset_{test}$.
    Keep track of $Model_{best}$.
End
6. Create and save plots for $ConfusionMatrix, PR_{Curve}, and\ ROC_{curve}$.

Fig. 12. The proposed training model, SAlexNet-1, outlines the critical steps for brain tumor detection and classification using DS-1.

rate, and weight decay. A data augmentation pipeline is created, consisting of transformations including resizing, horizontal and vertical flipping, photometric distortion, Gaussian blur, and Gaussian noise. The dataset DS-1 is then split into training and testing sets. The SAlexNet-1 model, optimizer, learning rate scheduler, and loss function are initialized. The training loop iterates through epochs, processing batches of images and labels. For each batch, the model predicts labels, calculates loss and updates weights using the optimizer. Training metrics are tracked, and metrics are calculated and saved for both training and testing datasets. The best-performing model is updated accordingly.

Upon completing the training loop, the algorithm evaluates the model's performance on test instances (without augmentation) by creating plots for the confusion matrix, precision-recall curve, and receiver operating characteristic curve. The evaluation metrics are saved and later used for visualization purposes. A similar procedure is applied to the SAlexNet-2 architecture, ensuring the last layer's neuron count matches the dataset classes.

3.8. Performance measures

This section outlines the performance evaluation framework for the proposed SAlexNet architectures, focusing on six key metrics: recall, precision, F1-score, accuracy, and AUC for ROC and PR curves.

Performance evaluation is a crucial step in the development of ML models. Effective evaluation metrics are crucial for assessing the model's capacity to generalize and generate predictions as labels for unseen data. Various performance indicators including accuracy, precision, recall, and F1-score, facilitate a comprehensive assessment of ML model efficacy. In the present study, the performance was evaluated using the following measures.

3.8.1. Recall

The true positive rate (TPR), also called sensitivity or recall, calculates the percentage of correctly identified diseased individuals within the actual diseased population. This metric represents the probability of accurately predicting a positive outcome, given that the patient has the disease. Essentially, TPR evaluates a model's effectiveness in detecting true positive instances. Mathematically, it is expressed as

$$Recall = \frac{T_{+ive}}{T_{+ive} + F_{-ive}}. \quad (15)$$

True positives (T_{+ive}) represent the accurately predicted positive instances, while false negatives (F_{-ive}) denote the missed positive instances misclassified as negative. Recall, ranging from 0 to 1, quantifies the model's proficiency in identifying positive cases, with higher values signifying better performance. Conversely, the false positive rate (FPR) calculates the ratio of misclassified negative events to the total number of actual negative events [54].

3.8.2. Precision

Precision, also known as positive predictive value (PPV), represents the likelihood that individuals testing positive actually have the disease [54]. Mathematically, PPV calculates the proportion of true positive predictions among all positive test results.

$$\text{Precision} = \frac{T_{+ive}}{T_{+ive} + F_{+ive}}, \quad (16)$$

where true positive (T_{+ive}) indicates a correct positive prediction, where the test result aligns with the positive outcome. Conversely, a false positive (F_{+ive}) represents an incorrect positive prediction, where the test yields a positive result indicating a negative outcome.

3.8.3. F1-score

The F1-score is a widely used metric for evaluating classifier performance. Despite its popularity, it has limitations due to its non-linear relationship with confusion matrix coefficients, making it non-decomposable. As a continuous and monotonically increasing function of the TPR, the F1-score is optimized by threshold-based classifiers. It represents the harmonic mean of precision and recall [54].

$$\text{F1} - \text{score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}, \quad (17)$$

3.8.4. Accuracy

The total accuracy of a ML model can be described as a metric that provides insight into the overall accuracy of the model. It considers both the accurate positive and negative predictions and represents the fraction of correctly classified instances, including both T_{+ive} and T_{-ive} considerations. The total accuracy is computed by

$$\text{Acc} = \frac{T_{+ive} + T_{-ive}}{T_{+ive} + F_{+ive} + T_{-ive} + F_{-ive}}, \quad (18)$$

where a true negative (T_{-ive}) represents the number of times a negative test result accurately identifies a negative condition [54].

3.8.5. AUC (ROC) curve

The area under the curve (AUC) of the receiver operating characteristic (ROC) curve serves as a comprehensive performance metric for evaluating classification models across diverse threshold settings [55]. ROC analysis provides a holistic assessment of classifier performance, illustrating efficiency across the entire operating spectrum. As a probability curve, ROC quantifies separability, while AUC measures the model's ability to discern between classes. Notably, AUC values are comparable to accuracy scores, offering a singular metric for evaluation. This metric is calculated by integrating the area under the ROC curve. The better the trained model, the higher its ROC curve and the larger its AUC, and vice versa.

3.8.6. AUC (PR) curve

Precision-recall (PR) curves provide a more informative performance summary for class-imbalanced datasets, where positive instances are scarce, than ROC curves [56]. By plotting precision against recall across various threshold values, PR curves illustrate the trade-off between accuracy and sensitivity. Notably, the PR curve's AUC offers a robust evaluation metric, accurately capturing model performance in class-imbalanced scenarios [57].

4. Results and discussion

In this section, the experimental results of this study have been thoroughly investigated. All the experimentation was carried out using a 64-bit operating system on a machine having specifications: (13th Gen Intel(R) Core(TM) i5–13420H, 2.10 GHz, NVIDIA GeForce RTX 3050 GPU with 6GB Memory, and 16GB RAM). The working environment was set up using the Conda package manager (version 23.7.4). The Python version used was 3.10.13. For DL purposes, PyTorch 2.1.2 was utilized with the 11.8 CUDA version.

4.1. General discussion of problem and solution

The proposed study focuses on the problem of brain tumor classification. For this task, the AlexNet model was inspected and modified diligently to incorporate an attention mechanism. The SAlexNet-1 architecture was initially proposed to incorporate the HAM after each convolution layer to improve feature representation at each stage. Secondly, the proposed solution was further improved by utilizing a kernel size 3×3 sequence in place of kernel size 11×11 in the first layer. These modifications endowed the AlexNet with the ability to capture contextually discriminative features. The proposed frameworks were trained and tested extensively using various augmentations and weight initialization schemes with hyper-parameterization. The results of the proposed solution on multi-class and binary-class datasets showed its ability to extract domain-relevant features for medical image classification. The results show the model's resilience against problems like class imbalance, overlapping, and noise impregnation.

4.2. Experimental setup

For this study, the AdamW optimizer was selected with a learning rate 0.001 and weight decay of 0.0005. The CosineAnnealingLR scheduler was utilized to gradually reduce the learning rate as the training progressed [58]. The models are trained for 50 epochs. Table 5 summarizes the salient hyper-parameters used for this experiment. Following the original AlexNet architecture, we have utilized image size 227×227. AdamW optimizer is a stochastic optimization algorithm that improves the Adam optimizer by modifying its typical implementation of weight decay. AdamW specifically decoupled the weight decay step from the gradient update step [59].

Table 5
Initial parameterization used in the investigation.

Option	Selection
Optimizer	AdamW
Activation Function	ReLU
Loss function	Binary cross entropy
Learning Rate	0.001
Weight decay	0.0005
Epochs	25
Scheduler (LR)	CosineAnnealingLR with eta_min=0.0001
Image matrix at the framework input	(227×227)

4.3. Convolution kernels' cardinality analysis

All convolutional layers in a CNN for classification using fully connected layers and SMC play an important role in feature extraction. However, the importance of each layer may vary depending on the architecture and application. The earlier convolutional layers (closer to the input) tend to extract low-level features such as edges, textures, and shapes. These layers are important for the basic visual processes required for recognition. We have critically analyzed the alteration of the receptive field for the first layer using different kernel sizes for problem-specific classification.

Numerous experiments were conducted by altering the kernel size of the first convolution layer of SAlexNet-1, with models using kernel sizes of 11×11, 9×9, 7×7, 5×5, and 3×3. The study investigated the impact of these variations on brain tumor classification (**Appendix** I for kernels other than 11×11). Five models were trained and tested for fixed partitions of DS-1, as explained in Section 3.1, using convolution sizes ranging from 3×3 to 11×11. All models demonstrated competitive performance, with high F1-scores and AUC (ROC) values, indicating a strong ability to distinguish between tumor types and healthy tissue. However, the 11×11 convolution model emerged as the top performer, achieving the highest average F1-score and AUC (ROC).

Table 6 presents the results of 25 epochs using DS-1 with an 11×11 kernel. The SAlexNet-1 model achieved an average F1-score of 0.9744 ± 0.0183 using a softmax classifier (SMC), indicating high accuracy in classifying brain tumors. The balance between precision (0.9752 ± 0.0183) and recall (0.9737 ± 0.0248) suggests effective identification of true positives with minimal false positives. The AUC (ROC) of 0.9984 ± 0.0017 demonstrates the model's exceptional class distinction ability. The close alignment between testing accuracy (98.78 ± 0.80)% and training accuracy (99.20 ± 0.61)% indicates slight overfitting. The marked improvement of SAlexNet-1 can be attributed to HAM Blocks 1, 2, 3, 4, and 5, as illustrated in Fig. 6. Figs. 13, 14, and 15 illustrate the ROC curves, confusion matrices, and PR curves for different kernel sizes over 25 epochs. Fig. 13 shows class-wise ROC curves for kernel sizes (a: 11×11, b: 9×9, c: 7×7, d: 5×5, e: 3×3). Fig. 14 presents the corresponding confusion matrices, and Fig. 15 displays the PR curves following the same class-wise pattern. These visualizations validate the superior performance of SAlexNet-1.

4.4. Sensitivity analysis of AlexNet performance

In this experiment, we trained the AlexNet architecture from scratch using the DS-1 dataset, initializing the weights randomly for 25 epochs. The results in Table 7 show that our model faced challenges due to random weight initialization, leading to a prolonged training process. The F1-score for meningioma classification was relatively low due to structural similarities with other tumor types, increasing false positives. This highlights the need for data augmentation, transfer learning, and attention mechanisms to improve generalizability and accuracy.

A comparative performance analysis between the pretrained AlexNet architecture, using ImageNet-based weights, and the randomly initialized AlexNet (Table 7) for DS-1 reveals that the TL-based AlexNet outperforms its randomly initialized counterpart. However, the results are not exceptional due to the omission of HAM blocks, which are crucial for focusing on relevant regions. Using pretrained weights mitigates the underfitting problem, allowing the model to benefit from the knowledge gained from the ImageNet dataset and adapt more effectively to the DS-1 dataset.

In the same experimental setup, the DS-2 dataset was used to evaluate the performance of AlexNet with and without Transfer Learning (TL). The results obtained from training AlexNet from scratch, where the network was trained on the entire dataset without any pretraining for 25 epochs, are presented in Table 7. The binary classification task required significantly more epochs to converge to a satisfactory performance level compared to the TL case. The results also indicate slight overfitting, with the model performing well on training data but not as well on unseen data.

In contrast, the results of AlexNet using TL, where pretrained weights were fine-tuned on the DS-2 dataset, show a substantial improvement in performance. A notable enhancement in performance is observed compared to the from-scratch case, with test accuracy closely matching training accuracy. However, the F1-score remains relatively low, indicating room for improvement in balancing precision and recall.

4.5. Sensitivity analysis of tumor classification using vision transformer

The Vision Transformer (ViT), proposed in [60,61], extends the self-attention mechanism to images, capturing local and global features. Unlike CNNs, which excel at extracting local neighborhood features hierarchically, ViT competes by delivering results in medical image classification [62], object detection, and segmentation [63,64]. The model transforms images into 16×16 patches using a convolution layer with a 16 kernel size. These patches are encoded with positional information and a class embedding vector, as transformers lack the positional inductive bias of CNNs. The encoded vectors pass through transformer encoder blocks, where self-attention calculates the importance of each patch relative to others, encoding global information and identifying relevant or redundant patches. Finally, the class embedding vector is processed through fully connected layers for classification tasks.

We trained the ViT from randomly initialized weights to compare with SAlexNet-1 (Table 6). The results, illustrated in Table 8, show that ViT requires a large amount of data for proper training due to the lack of CNN-like inductive bias. Consequently, ViT struggled with underfitting, as it has more parameters to learn and insufficient data. Despite the inherent attention mechanism, ViT could not outperform SAlexNet-1 in terms of F1-score. Interestingly, test results were better than training results, possibly due to initial good weight settings or less memorization of training data, leading to better generalization. In another experiment, we used a pretrained ViT from ImageNet, as shown in Table 8 for DS-1 with TL. While the TL-based ViT performed better than the randomly initialized version, it did not achieve exceptional results. The classification performance of TL-based ViT was superior to that of the randomly initialized weights.

A customized sensitivity analysis was conducted for ViT using DS-2 to investigate the model's behavior from scratch. The outcomes, presented in Table 8, show that ViT outperformed AlexNet when both were trained from scratch. No overfitting was observed in ViT, indicating generalizable feature learning. However, the low F1-score suggests the need for additional training epochs for optimal accuracy. Further

Table 6
Results using kernel size 11×11 for SAlexNet-1 based on SMC for DS-1 and 25 epochs (σ represents the standard deviation of the performance scores of Glioma, Meningioma, Pituitary, and Notumor classes).

Class	$Acc_{(train)}$ (%)	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
Glioma	98.76	0.9433	0.9792	0.9970	0.9610	98.25
Meningioma	98.60	0.9641	0.9486	0.9968	0.9562	97.94
Pituitary	99.65	0.9900	0.9900	0.9997	0.9900	99.54
Notumor	99.81	0.9975	0.9830	0.9999	0.9902	99.39
Average ± σ	**99.20 ± 0.61**	**0.9737 ± 0.0248**	**0.9752 ± 0.0183**	**0.9984 ± 0.0017**	**0.9744 ± 0.0183**	**98.78 ± 0.80**

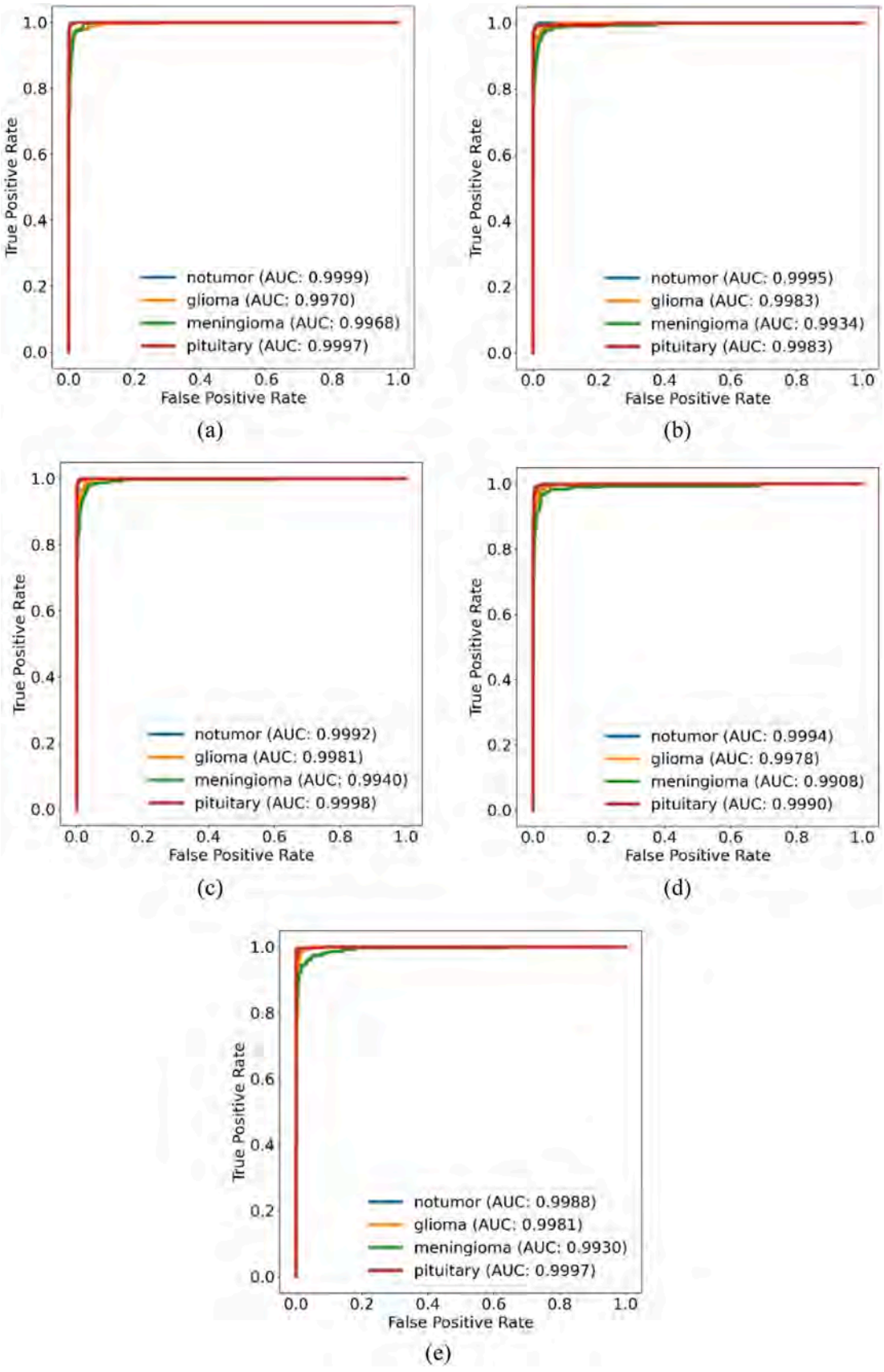


Fig. 13. Performance using AUC (ROC) curves for kernel sizes (a)11 × 11, (b) 9 × 9, (c) 7 × 7, (d) 5 × 5, and (e) 3 × 3, for SAlexNet-1 using HAM and DS-1. The legends show the AUC-ROC value for each class.

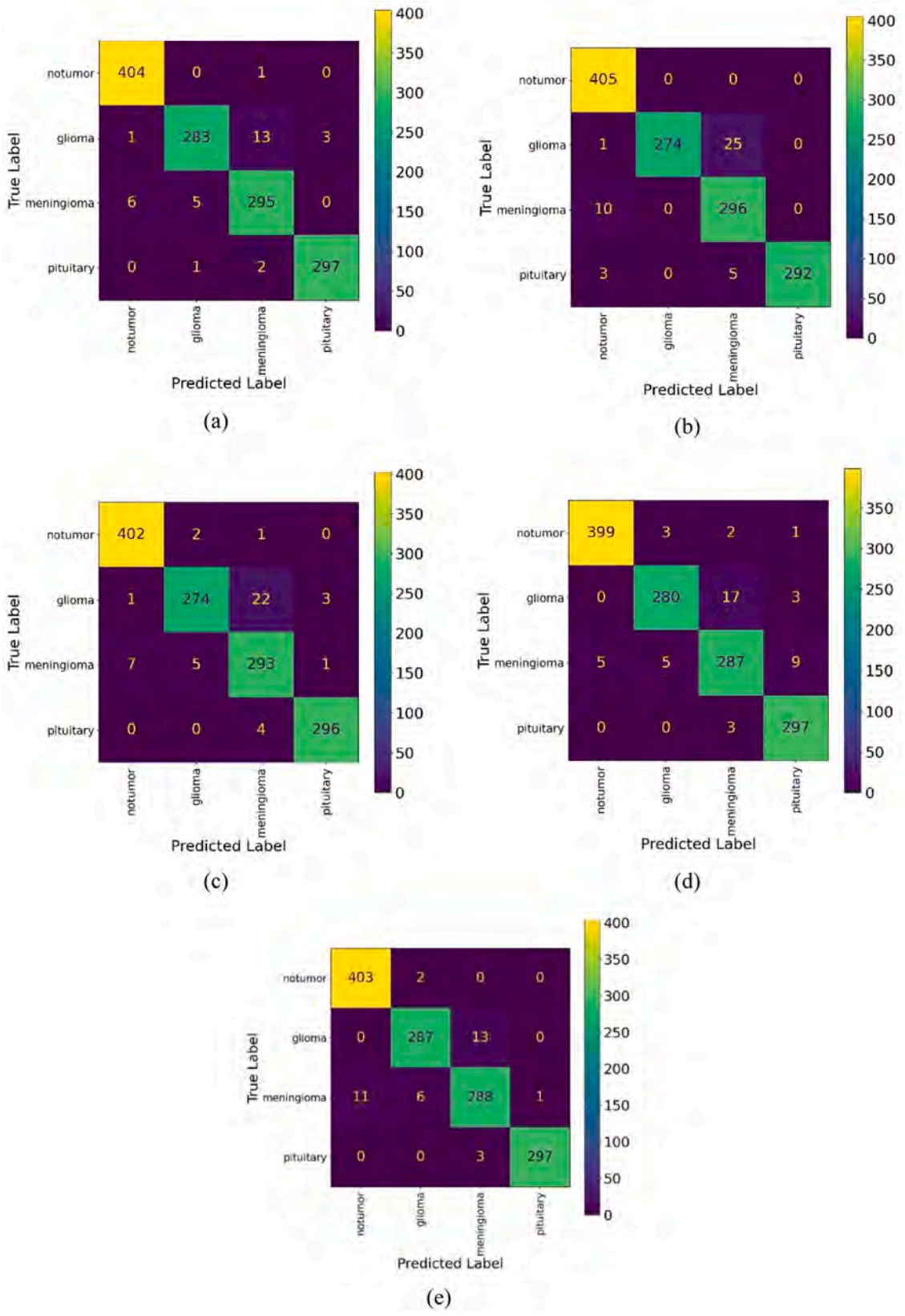


Fig. 14. Confusion matrices for kernel sizes (a)11 × 11, (b) 9 × 9, (c) 7 × 7, (d) 5 × 5, and (e) 3 × 3, for SAlexNet-1 using HAM and DS-1.

experiments with a pretrained ViT model, also shown in Table 8, demonstrated better performance compared to TL-based AlexNet, with outstanding F1-scores indicating successful adaptation to the new task.

4.6. Comparative analysis of SAlexNet-1 and SAlexNet-2 using DS-1

In this section, we have replaced the first convolution layer (11 × 11) of SAlexNet-1 with 5 convolution layers of size (3 × 3) of the proposed

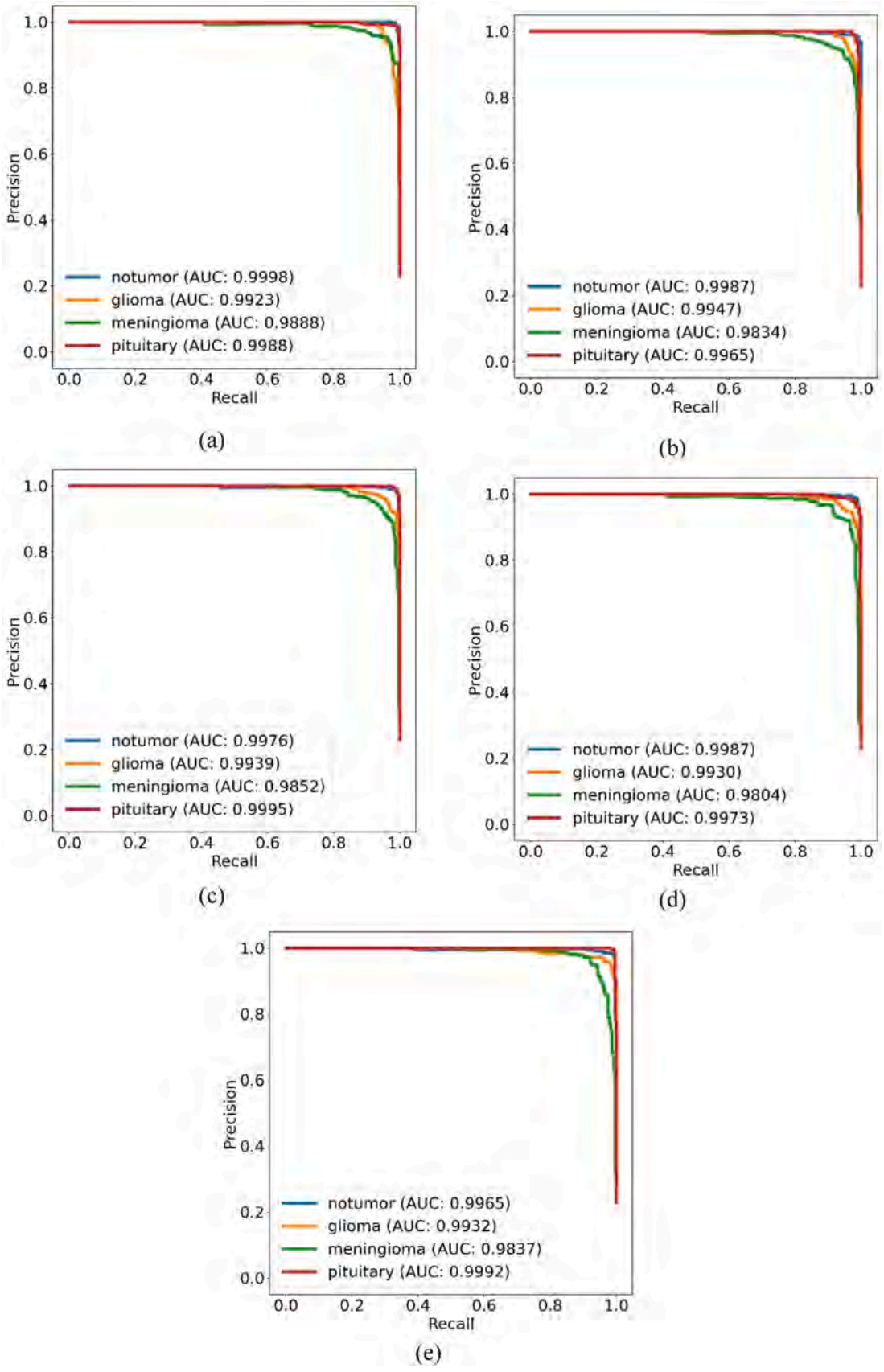


Fig. 15. PR curves for kernel sizes (a)11 × 11, (b) 9 × 9, (c) 7 × 7, (d) 5 × 5, and (e) 3 × 3, for SAlexNet-1 using HAM and DS-1. The legends show AUC-PR value for each class.

Table 7
Results of AlexNet (from scratch, by random weight initialization and TL) using SMC for DS-1 and DS-2 (σ represents the standard deviation of the performance scores; 25 epochs used for training).

Dataset	Status	Class	$Acc_{(train)}$ (%)	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
DS-1	Without TL	Glioma	86.82	0.9033	0.6407	0.9524	0.7497	86.19
		Meningioma	81.29	0.3007	0.5576	0.7954	0.3907	78.11
		Pituitary	92.23	0.8733	0.8086	0.9686	0.8397	92.37
		Notumor	94.63	0.8346	0.8471	0.9603	0.8408	90.24
		Average ± σ	**88.74 ± 5.94**	**0.7280 ± 0.2863**	**0.7135 ± 0.1373**	**0.9192 ± 0.0828**	**0.7052 ± 0.2140**	**86.73 ± 6.29**
	With TL	Glioma	96.57	0.8900	0.9303	0.9924	0.9097	95.96
		Meningioma	94.54	0.8529	0.8788	0.9777	0.8657	93.82
		Pituitary	97.67	0.9733	0.9125	0.9969	0.9419	97.25
		Notumor	98.72	0.9728	0.9681	0.9973	0.9704	98.17
		Average ± σ	**96.88 ± 1.78**	**0.9223 ± 0.0606**	**0.9224 ± 0.0372**	**0.9911 ± 0.0092**	**0.9219 ± 0.0450**	**96.30 ± 1.88**
DS-2	Without TL	NO	83.64	0.7527	0.7472	0.8166	0.7449	74.17
		YES	83.64	0.7920	0.7278	0.8105	0.7562	74.50
		Average ± σ	**83.64 ± 0.01**	**0.7723 ± 0.0597**	**0.7375 ± 0.0486**	**0.8136 ± 0.0210**	**0.7505 ± 0.0189**	**74.33 ± 0.03**
	With TL	NO	91.21	0.8653	0.9089	0.9647	0.8836	88.57
		YES	91.21	0.9060	0.8744	0.9647	0.8867	88.57
		Average ± σ	**91.21 ± 0.01**	**0.8857 ± 0.0629**	**0.8917 ± 0.0481**	**0.9647 ± 0.0085**	**0.8852 ± 0.0254**	**88.57 ± 0.02**

Table 8
Results of ViT (from scratch, by random weight initialization) using DS-1 and DS-2 trained for 25 epochs (σ represents the standard deviation of the performance).

Dataset	Status	Class	$Acc_{(train)}$ (%)	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
DS1	Without TL	Glioma	73.30	0.8000	0.4075	0.8196	0.5399	68.80
		Meningioma	72.97	0.0490	0.2830	0.5818	0.0836	74.90
		Pituitary	73.72	0.7367	0.5567	0.8442	0.6341	80.55
		Notumor	73.35	0.4765	0.7096	0.8328	0.5702	77.80
		Average ± σ	**73.34 ± 0.30**	**0.5156 ± 0.3411**	**0.4892 ± 0.1847**	**0.7696 ± 0.1256**	**0.4570 ± 0.2520**	**75.51 ± 5.03**
	With TL	Glioma	87.76	0.7267	0.7032	0.9290	0.7148	86.73
		Meningioma	81.57	0.4608	0.5423	0.7693	0.4982	78.34
		Pituitary	88.94	0.7833	0.8453	0.9652	0.8131	91.76
		Notumor	89.46	0.8543	0.7473	0.9176	0.7972	86.58
		Average ± σ	**86.93 ± 3.64**	**0.7063 ± 0.1718**	**0.7095 ± 0.1263**	**0.8953 ± 0.0864**	**0.7058 ± 0.1450**	**85.85 ± 5.55**
DS-2	Without TL	NO	83.13	0.8147	0.8475	0.9079	0.8293	82.71
		YES	83.13	0.8480	0.8209	0.9079	0.8327	82.71
		Average ± σ	**83.13 ± 0.04**	**0.8313 ± 0.0592**	**0.8342 ± 0.0507**	**0.9079 ± 0.0344**	**0.8310 ± 0.0452**	**82.71 ± 0.04**
	With TL	NO	92.58	0.8940	0.9413	0.9608	0.9136	91.83
		YES	92.58	0.9427	0.9065	0.9608	0.9219	91.83
		Average ± σ	**92.58 ± 0.05**	**0.9183 ± 0.0693**	**0.9239 ± 0.0577**	**0.9608 ± 0.0360**	**0.9178 ± 0.0518**	**91.83 ± 0.05**

SAlexNet-2. The advantage of five convolution layers is extended non-linearity at the start of the model, which helps in extracting fine-level details more efficiently, as explained in Section 3.6.

The experimental results for SAlexNet-2 architecture are illustrated in Table 9, with DS-1 using SMC. The model achieved excellent overall performance, with an average F1-score of (0.9935 ± 0.0050), indicating an excellent performance score in classifying different types of tumors. The balance between precision (0.9937 ± 0.0088) and recall (0.9933 ± 0.0090) suggests the model effectively identifies true positives without predicting excessive false negatives. The average AUC (ROC) of (0.9994 ± 0.0008) demonstrates the model's outclass performance in distinguishing between classes, even when their features overlap. The close alignment between testing accuracy (99.69 ± 0.22) and training accuracy (99.96 ± 0.03) shows that the model can generalize well to the unseen data.

A visual performance comparison is presented in Fig. 16 for SAlexNet-1 and SAlexNet-2 using Tables 6 & 9, respectively, illustrating their performance on dataset DS-1 after 25 epochs. The key evaluation metrics, F1-score, and test instance-based accuracy demonstrate the exceptional performance of both models. The outstanding results can be attributed to incorporating HAM and data augmentation techniques, which enhance the models' ability to learn robust features and generalize well to unseen data.

Moreover, the modifications made to the initial layers of the architecture (as depicted in Fig. 10) have significantly contributed to the improvement in performance scores. The careful tuning of hyperparameters and regularization techniques has enabled both algorithms to achieve good generalization, effectively avoiding overfitting. This is evident from the optimal balance between training and test accuracy, indicating that the models have learned to capture the underlying patterns in the data without memorizing the training instances.

4.7. Comparative analysis of SAlexNet-1 and SAlexNet-2 using DS-2

This section presents a comparative performance analysis between SAlexNet-1 and SAlexNet-2, conducted using dataset DS-2 over 25

Table 9
Results with SAlexNet-2 model using five convolution layers of kernel size 3 × 3 and DS-1 (25 epochs).

Class	$Acc_{(train)}$ (%)	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
Glioma	99.98	0.9800	1.0000	0.9998	0.9899	99.54
Meningioma	99.93	0.9967	0.9807	0.9982	0.9887	99.47
Pituitary	99.93	0.9967	0.9967	1.0000	0.9967	99.85
Notumor	99.98	1.0000	0.9975	0.9996	0.9988	99.92
Average ± σ	**99.96 ± 0.03**	**0.9933 ± 0.0090**	**0.9937 ± 0.0088**	**0.9994 ± 0.0008**	**0.9935 ± 0.0050**	**99.69 ± 0.22**

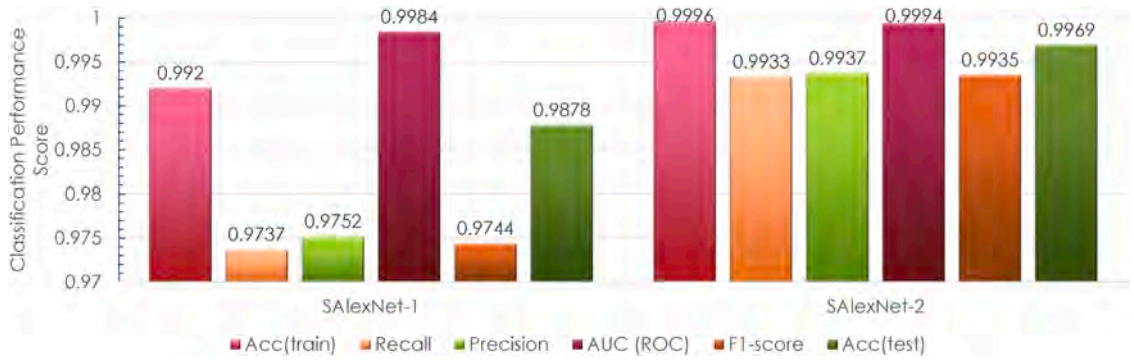


Fig. 16. Comparison of SAlexNet-1 and SAlexNet-2 using 25 epochs (using the results from Tables 6 & 19) for the sets of respective classification performance scores using DS-1.

training epochs. The class separation for SAlexNet-1, illustrated in Table 10, demonstrates exceptional performance. This success is attributed to using a large kernel size in the initial convolutional layer, enabling the model to capture complex patterns and features in the data, thus allowing for more comprehensive feature extraction and improved performance.

Conversely, the results for SAlexNet-2 using DS-2, also presented in Table 10, show significant improvement in binary classification. This enhancement is due to the use of multiple smaller-sized kernels in the initial layer, which allows the model to capture discriminative features more effectively. The smaller kernel size facilitates more focused feature extraction, resulting in better performance in binary classification tasks.

Fig. 17 visually represents a comparative analysis of SAlexNet-1 and SAlexNet-2, explaining their performance on the DS-2 dataset after 25 epochs. The results demonstrate a notable enhancement in SAlexNet-2′s performance, indicating improved overall accuracy and generalizability. This suggests that SAlexNet-2 has achieved superior performance compared to SAlexNet-1, likely due to its modified architecture, optimized hyperparameters, and effective training strategy. The enhanced performance of SAlexNet-2 can be attributed to its ability to learn more robust features and adapt to the complexities of the DS-2 dataset.

4.8. Features space conception for SAlexNet-2

The heterogeneity of the feature space generated by the proposed SAlexNet-2 model is evaluated to clarify the intricate processes underlying tumor detection and classification. The model's discriminative power is intrinsically linked to the inherent attributes of the feature set. By enhancing the differentiation between class-specific features, the model's learning efficacy is augmented, thereby fortifying its robustness across diverse instances. The SAlexNet-2 framework significantly amplifies feature space diversity, advancing the precision and accuracy of brain tumor recognition and categorization.

Fig. 18 presents a visualization of four diverse brain MRI classes using the SAlexNet-2 model, enabled by a variant of Stochastic Neighborhood Embedding (t-SNE), providing an intuitive representation of the data. This visualization is rendered as a scatterplot for the test instances of DS-1, wherein each feature vector is non-linearly mapped to a two-dimensional plane. This methodology enhances visual perception by counteracting data point congestion toward the central axis, optimizing data dispersion, and alleviating visual occlusion through strategic distribution, enabling more discerning insights and improved pattern recognition capabilities. Notably, the discriminative capability is occasionally compromised by false events, particularly within the Glioma cluster. Nevertheless, our system significantly improves the discrimination for complex decision hyperplanes between the various classes of DS-1.

Similarly, Fig. 19 illustrates the visual performance analysis of the SAlexNet-2 model using the DS-2 dataset for binary classification, where the algorithm's robust delineation capability, free from institutional biases, is evident.

4.9. STL-based approach using DS-1 and DS-2

To assess the STL capability of SAlexNet-1 and SAlexNet-2, a comprehensive evaluation was conducted on the DS-2 dataset (Section 3.1.1). The models were initialized with pretrained weights from DS-1 through transfer learning to adapt to the new dataset (DS-2). Since the test set was not explicitly provided on the Kaggle site for DS-2, a 5-fold cross-validation strategy was employed to evaluate the models' performance in a multi-institutional setting. The output layers of the DL models were modified to consist of two neurons, representing two-class probability vectors for binary classification. Both models were trained for 25 epochs, considering the correlation between the two datasets. The remaining hyperparameters, including learning rate, weight decay, optimizer, activation functions, and loss function, were kept consistent throughout the training process. The 5-fold cross-validation pipeline for SAlexNet-1 is illustrated in Fig. 20. The images within each folder were divided into five equal splits, with each split serving as the validation set iteratively. When the i^{th} split was used for validation, the remaining splits were utilized for model training, ensuring a robust evaluation of the models' performance.

The 5-fold cross-validation results for SAlexNet-1 and SAlexNet-2 on DS-2, using 25 epochs each, are presented in Table 11. SAlexNet-1 achieved an average F1-score of (0.9629 ± 0.0113) across all folds and classes. In contrast, SAlexNet-2 demonstrated a significantly improved average F1-score of (0.9880 ± 0.0071), attributed to architectural modifications in the initial layers described in Section 3.6.

Table 10
Binary classification results with SAlexNet-1 and SAlexNet-2 models.

Model	Class	$Acc_{(train)}$ (%)	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
SAlexNet-1	NO	99.45	0.9813	0.9815	0.9951	0.9809	98.07
	YES	99.45	0.9800	0.9820	0.9950	0.9804	98.07
	Average ± σ	**99.45 ± 0.00**	**0.9807 ± 0.0293**	**0.9818 ± 0.0272**	**0.9950 ± 0.0054**	**0.9806 ± 0.0180**	**98.07 ± 0.02**
SAlexNet-2	NO	99.97	0.9947	0.9890	0.9981	0.9917	99.17
	YES	99.97	0.9887	0.9947	0.9983	0.9916	99.17
	Average ± σ	**99.97 ± 0.00**	**0.9917 ± 0.0106**	**0.9918 ± 0.0103**	**0.9982 ± 0.0029**	**0.9917 ± 0.0068**	**99.17 ± 0.01**

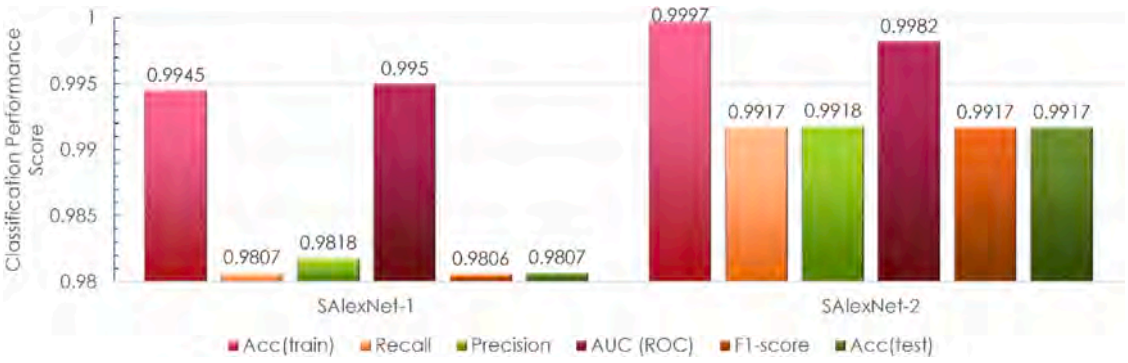


Fig. 17. Comparison of SAlexNet-1 and SAlexNet-2 for the sets of respective classification performance scores using DS-2 and 25 epochs.

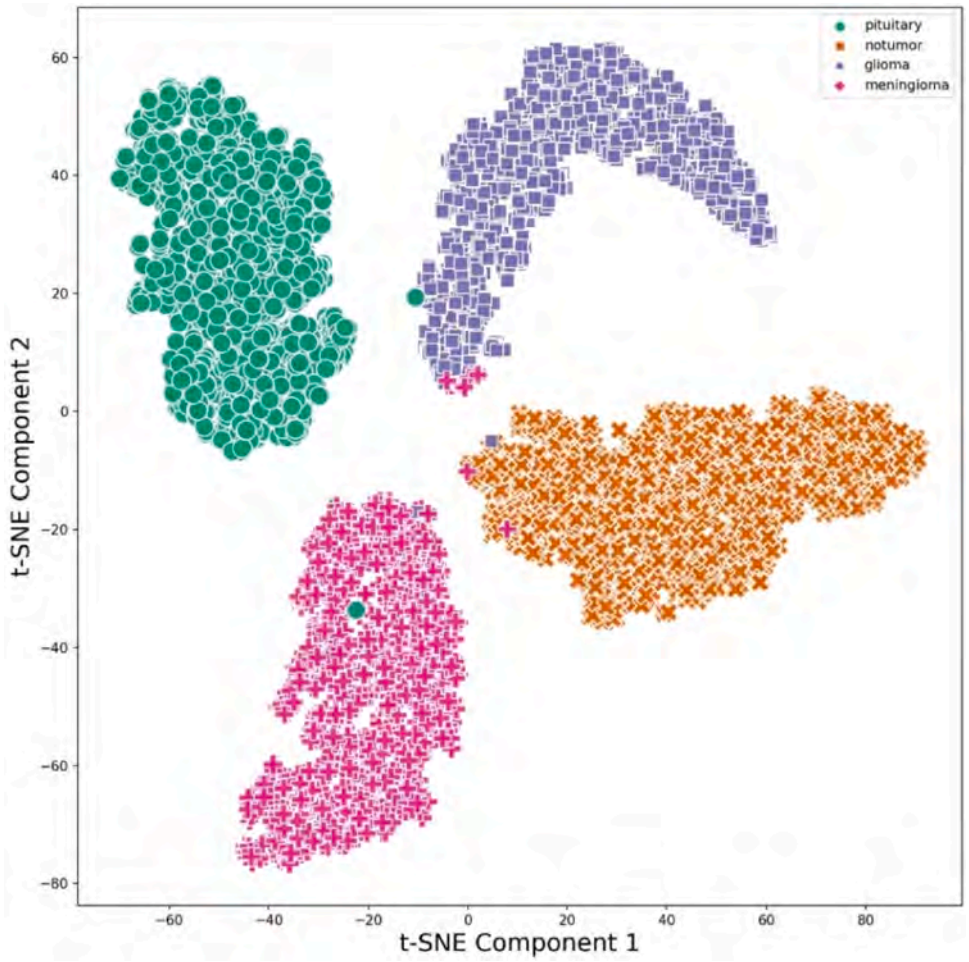


Fig. 18. Feature space visualization using t-SNE for performance analysis of proposed SAlexNet2 for DS-1 using 25 epochs.

Notably, SAlexNet-1 exhibited overfitting, with a high training accuracy of (99.58 ± 0.00)% and a relatively lower testing accuracy of (96.30 ± 0.01)%. Conversely, SAlexNet-2 addressed this issue, achieving a testing accuracy of (98.80 ± 0.00)%, showcasing superior generalization capabilities.

The classification performance analysis of the STL-based SAlexNet-1 model is presented in Table 11. Increasing the number of training epochs significantly improved the performance score. Pretraining on the DS-1 dataset for 50 epochs allowed for extensive fine-tuning and adaptation to DS-2, which was trained using STL for 25 epochs. This prolonged training enabled the model to thoroughly learn the spatial and temporal features of the data, leading to a performance score of (96.93 ± 0.0086)%. The high number of epochs in the STL approach improved performance by accommodating the domain shift between DS-1 and DS-2, enabling the model to recognize patterns at a higher level and improve classification accuracy.

Similarly, the STL-based SAlexNet-2 model was pretrained on DS-1 for 50 epochs to address the domain shift, allowing sufficient adaptation to DS-2 using 25 training epochs. The results, presented in Table 11, show an outstanding performance of (99.20 ± 0.0106)% for DS-2 using SAlexNet-2.

A comparative analysis of performance metrics, illustrated in Fig. 21, reveals a notable enhancement in the performance score of semi-transfer learning (STL)-based SAlexNet-2 over its predecessor, SAlexNet-1. This

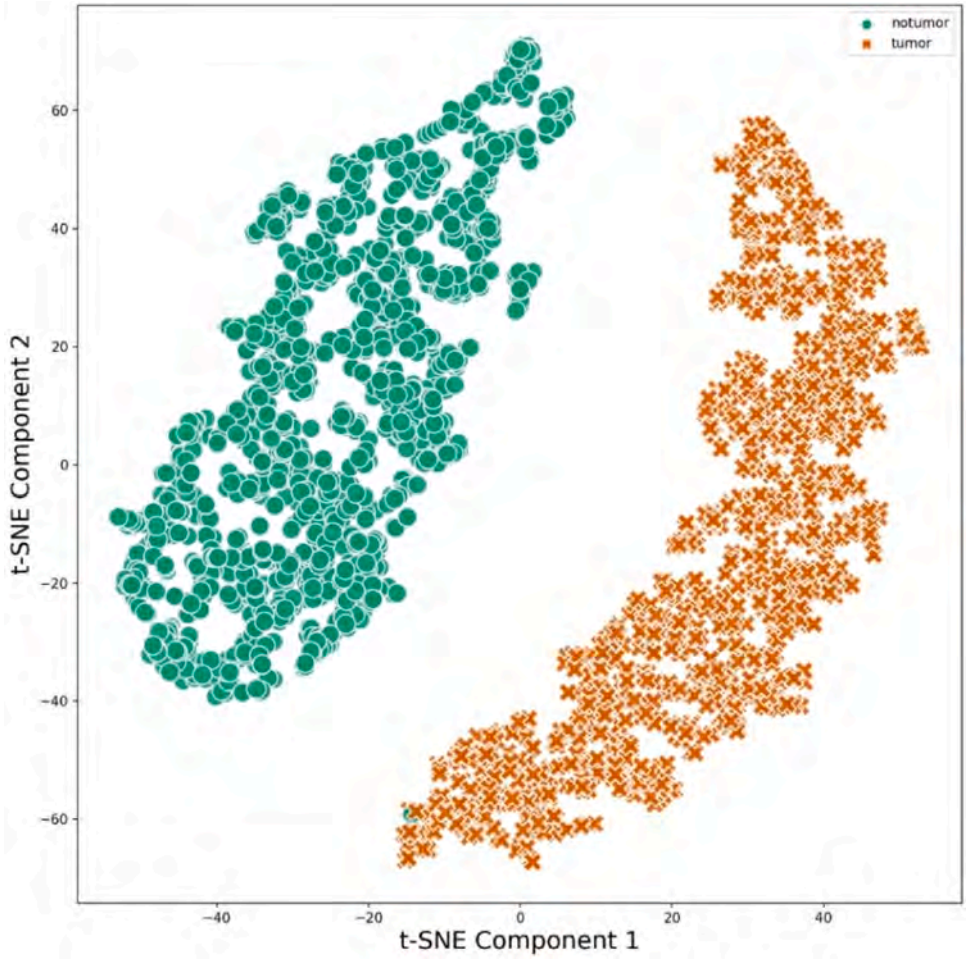


Fig. 19. Feature space visualization using t-SNE for performance analysis of proposed SAlexNet2 for DS-2 using 25 epochs.

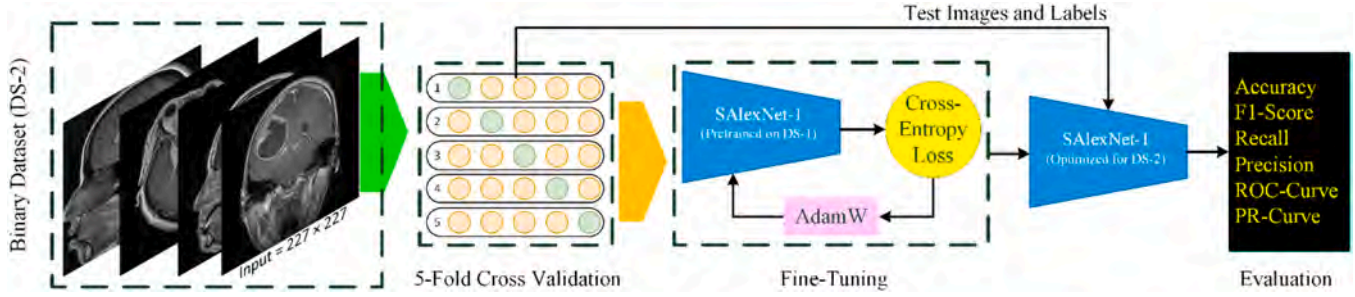


Fig. 20. Flowchart for STL using DS-1 (pretraining) and SAlexNet-1 trained and tested on DS-2 (binary dataset) using 5-fold cross-validation after the last layer modifications in the associated architecture (SAlexNet-2 has also been subjected to the STL strategy in a similar manner).

visual representation, derived from the data presented in Table 11, underscores the superior performance of the STL-based SAlexNet-2 framework, which utilizes the benefits of transfer learning to adapt pretrained models to new tasks. Furthermore, a careful examination of the results indicates a marginal yet significant improvement in the performance of STL-based SAlexNet-2 compared to the standard SAlexNet-2 architecture, which utilizes a conventional DL approach without transfer learning. This observation substantiates the efficacy of the STL approach, particularly in scenarios where datasets exhibit strong correlations, enabling the model to capture intricate patterns and relationships within the data. The STL-based SAlexNet-2 framework's improved performance can be attributed to its ability to fine-tune pretrained weights, allowing for a more effective adaptation to the new task, and its capacity to use domain-specific knowledge from the source dataset to improve performance on the target dataset. This leads to the conclusion that the STL approach is more pronounced in cases of correlated datasets, where the transfer of knowledge from the source domain to the target domain is more effective.

4.10. Comparative analysis with state-of-the-art techniques

This section presents a comparative analysis of the efficacy of the proposed SAlexNet-based models in brain tumor detection and classification, contrasting their performance with state-of-the-art techniques. Our results demonstrate notable improvements in performance scores compared to existing methods, with significant implications for clinical

Table 11
STL-based binary classification results of SAlexNet-1 and SAlexNet-2 with 5-fold cross-validation for DS-2 after 25 epochs for each run. PTE represents the pretraining epochs used for DS-1 with SAlexNet-1 and SAlexNet-2 (σ represents the standard deviation of the performance scores).

PTE	Model	Class	$Acc_{(train)}$ (%)	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
25	SAlexNet-1	NO	99.58	0.9953	0.9354	0.996	0.9643	96.30
		YES	99.58	0.9307	0.9951	0.996	0.9616	96.30
		Average ± σ	**99.58 ± 0.00**	**0.9630 ± 0.0152**	**0.9653 ± 0.0138**	**0.9960 ± 0.0034**	**0.9629 ± 0.0113**	**96.30 ± 0.01**
	SAlexNet-2	NO	99.77	0.9927	0.9836	0.9978	0.9881	98.80
		YES	99.77	0.9833	0.9926	0.9979	0.9879	98.80
		Average ± σ	**99.77 ± 0.00**	**0.9880 ± 0.0095**	**0.9881 ± 0.0092**	**0.9978 ± 0.0036**	**0.9880 ± 0.0071**	**98.80 ± 0.00**
50	SAlexNet-1	NO	99.86	0.9980	0.9441	0.9972	0.9702	96.93
		YES	99.86	0.9407	0.9979	0.9976	0.9684	96.93
		Average ± σ	**99.86 ± 0.00**	**0.9693 ± 0.0101**	**0.9710 ± 0.0095**	**0.9974 ± 0.0030**	**0.9693 ± 0.0076**	**96.93 ± 0.01**
	SAlexNet-2	NO	99.98	0.9940	0.9903	0.9976	0.9921	99.20
		YES	99.98	0.9900	0.9940	0.9978	0.9919	99.20
		Average ± σ	**99.98 ± 0.00**	**0.9920 ± 0.0121**	**0.9921 ± 0.0119**	**0.9977 ± 0.0042**	**0.9920 ± 0.0106**	**99.20 ± 0.01**

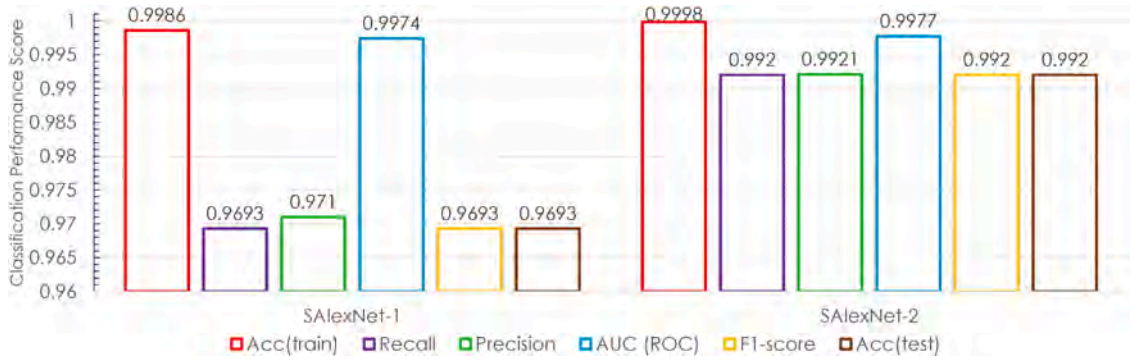


Fig. 21. Comparison of STL-based SAlexNet-1 and SAlexNet-2 for the sets of respective classification performance scores using DS-1 (pretraining,50 epochs) and DS-2 (25 training epochs on the DS-1 based pretrained model).

practice and future research. We have implemented various DL architectures (AlexNet, ViT, Xception, GoogLeNet, ResNet18, ResNet50, SqueezeNet, VGG-19, InceptionV3, and VGG-16) as listed in Tables 12 and 13 for DS-1 and DS-2, respectively, to compare their performance with our proposed SAlexNet-based frameworks.

The input dimensions for the ViT, GoogLeNet, ResNet18, VGG19, ResNet50, and VGG16 CNN architectures are uniformly set at 224 × 224 pixels with three color channels (RGB), totaling 672,768 pixels per input image. Conversely, SqueezeNet and AlexNet require slightly larger input sizes of 227 × 227 × 3, while Xception and InceptionV3 models need an input size of 299 × 299 × 3. To facilitate the extraction of intensity-based information and reduce the impact of chromatic variability, all input images undergo a grayscale transformation, effectively reducing the number of channels from three (RGB) to a single one. This dimensionality reduction is achieved through a weighted averaging process, combining the contributions from the three color channels to yield a single, intensity-based representation of the input image.

Each of Tables 12 and 13 is divided into three sections: the first section presents the implementation results of state-of-the-art popular architectures run from scratch using randomly initialized weights, the second section covers the same architectures using TL with pretrained weights, and the innovative architectures are illustrated in the last section. The preprocessing for implementing the diverse state-of-the-art DL architectures using DS-1 included resizing, contrast enhancement using DWT, horizontal and vertical flips, photometric distortions, and Gaussian blur and noise. All state-of-the-art implementations from scratch and transfer learning, including the proposed methods, were trained for 25 epochs unless otherwise stated, as shown in Table 12 using the DS-1 dataset.

The performance score of AlexNet [39] (86.73 ± 6.29)% is better than ViT [61] (75.51 ± 5.03)%. AlexNet performs better in cases where local features and spatial relationships are essential, as its convolution with pooling layers preserves spatial information, capturing local patterns and edges of the detected tumor. In contrast, ViT's global self-attention mechanism increases algorithmic complexity, affecting generalization and leading to overfitting, especially with limited training data and epochs. Xception [65] (91.80 ± 4.99)% outperforms AlexNet and ViT due to depthwise separable convolution (DSC) and residual connections, which improve feature extraction and convergence speed. GoogLeNet [66] (95.50 ± 2.34)% excels due to inception modules that capture multi-scale features, leading to superior classification performance.

ResNet18 [67] (89.05 ± 5.40)% underperforms due to a limited number of epochs and shallow architecture, while ResNet50 [67] (84.40 ± 5.49)% struggles with poor generalization due to its deep residual learning-based architecture. InceptionV3 [70] (84.32 ± 4.49)% requires more epochs to match other DL architectures, as its complex architecture with multiple branches and auxiliary classifiers leads to overfitting. SqueezeNet [68] (83.91 ± 5.43)% shows limited performance due to its compact architecture and reduced parameters. VGG-19 (94.05 ± 3.67)% and VGG-16 (93.71 ± 3.55)% [69] excel in tumor detection with sequential convolution and pooling layers, enhancing boundary detection between healthy and tumorous regions.

The TL-based implementations of diverse architectures, shown in the second section of Table 12, demonstrate remarkable performance compared to their from-scratch versions. VGG-16 (98.86 ± 0.64) %, InceptionV3 (98.13 ± 0.85)%, and VGG-19 (98.05 ± 1.16)% achieved outstanding results using 25 epochs. These models, pretrained on ImageNet, provide a strong foundation for feature extraction. Fine-tuning these models on the DS-1 dataset uses their learned knowledge. VGG-19 and VGG-16's homogeneous architecture and InceptionV3's modular design facilitate adaptation to new tasks. AlexNet, being relatively shallow, may not capture complex features as effectively. ViT's attention-based architecture requires more data and computational

Table 12
Problem-specific sensitivity analysis of traditional and innovative DL models to compare SAlexNet-1 and SAlexNet-2 using DS-1 (no validation due to the already defined test partition provided with the benchmark dataset; 25 epochs used for each implementation; Scratch results are with weights in the network without any pretraining; the values after '±' represents the standard deviation; TL stands for transfer learning with pretrained weights).

References	Method	Classification Performance		
		AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
Popular Traditional DL Methods from Scratch (Random Weight Initialization, 25 epochs)				
Krizhevsky et al. [39]	AlexNet	0.9192 ± 0.0828	0.7052 ± 0.2140	86.73 ± 6.29
Dosovitskiy et al. [61]	ViT	0.7696 ± 0.1256	0.4570 ± 0.2520	75.51 ± 5.03
Cholett [65]	Xception	0.9660 ± 0.0364	0.8305 ± 0.1096	91.80 ± 4.99
Szegedy et al. [66]	GoogLeNet	0.9856 ± 0.0164	0.9070 ± 0.0494	95.50 ± 2.34
Kaiming et al. [67]	ResNet18	0.9436 ± 0.0590	0.7704 ± 0.1352	89.05 ± 5.40
Kaiming et al. [67]	ResNet50	0.8721 ± 0.1027	0.6685 ± 0.1703	84.40 ± 5.49
Iandola et al. [68]	SqueezeNet	0.8612 ± 0.1020	0.6635 ± 0.1517	83.91 ± 5.43
Simonyan and Zisserman [69]	VGG-19	0.9806 ± 0.0210	0.8720 ± 0.0931	94.05 ± 3.67
Szegedy et al. [70]	InceptionV3	0.8761 ± 0.1105	0.6553 ± 0.1755	84.32 ± 4.49
Simonyan and Zisserman [69]	VGG-16	0.9767 ± 0.0273	0.8648 ± 0.0909	93.71 ± 3.55
Transfer Learning-based Implementations (25 epochs)				
Krizhevsky et al. [39]	AlexNet	0.9911 ± 0.0092	0.9219 ± 0.0450	96.30 ± 1.88
Dosovitskiy et al. [61]	ViT	0.8953 ± 0.0864	0.7058 ± 0.1450	85.85 ± 5.55
Cholett [65]	Xception	0.9734 ± 0.0240	0.8569 ± 0.0773	93.17 ± 3.15
Szegedy et al. [66]	GoogLeNet	0.9903 ± 0.0098	0.9205 ± 0.0477	96.22 ± 1.96
Kaiming et al. [67]	ResNet18	0.9944 ± 0.0068	0.9451 ± 0.0324	97.37 ± 1.38
Kaiming et al. [67]	ResNet50	0.9951 ± 0.0077	0.9455 ± 0.0343	97.77 ± 1.38
Iandola et al. [68]	SqueezeNet	0.9953 ± 0.0051	0.9559 ± 0.0278	97.90 ± 1.21
Simonyan and Zisserman [69]	VGG-19	0.9979 ± 0.0022	0.9595 ± 0.0266	98.05 ± 1.16
Szegedy et al. [70]	InceptionV3	0.9966 ± 0.0032	0.9613 ± 0.0205	98.13 ± 0.85
Simonyan and Zisserman [69]	VGG-16	0.9988 ± 0.0014	0.9761 ± 0.0151	98.86 ± 0.64
Innovative DL Architectures (published using DS-1)				
Albalawi et al. [71]	Multi-layer customized CNN with an optimized and advanced layer architecture	0.9850	99.00	0.9925
Rasheed et al. [72]	Combination of hybrid attention mechanism with CNN	0.9820	98.33	×
Sarkar et al. [73]	AlexNet-Naïve Bayes	0.9630	98.15	×
Celik and Inik [74]	EfficientNetB0-SVM	×	9 7.93	×
Bansal et al. [75]	Hybrid CNN-SVM model, utilizing CNN as a feature extractor and SVM for classification	0.9800	98.00	×
Özkaraca et al. [76]	Augmented CNN with extra layers after TL with VGG-16, DenseNet, and basic CNN.	0.9650	94.00	×
Proposed Architecture 1	SAlexNet-1	0.9988 ± 0.0016	0.9862 ± 0.0100	99.35 ± 0.44
Proposed Architecture 2	SAlexNet-2	0.9994 ± 0.0008	0.9935 ± 0.0050	99.69 ± 0.22

resources, and its performance may degrade with fewer epochs. Xception's depthwise separable convolutional layers need careful hyperparameter tuning. GoogLeNet's inception modules, while efficient, can complicate optimization. ResNets may experience performance saturation with fewer epochs. SqueezeNet's compact architecture may not capture complex features as effectively. Performance differences can be dataset-dependent, and variations in implementation, such as data augmentation, normalization, or optimization algorithms, can affect performance.

In the third section, Table 12 compares various innovative DL methods used for brain tumor classification, evaluated based on AUC (ROC), F1-score, and test accuracy. Among the potential innovative models, Albalawi et al. [71] multi-layer customized CNN achieved the highest AUC of 0.9850 and an impressive F1-score of 99.00, along with a test accuracy of 99.25 %. This suggests that their approach, which utilizes an advanced layer architecture, is highly effective at distinguishing between brain tumor classes. Other notable models include Rasheed et al. [72] hybrid CNN with an attention mechanism, which achieved a strong AUC(ROC) of 0.9820 and a high F1-score of 98.33, although test accuracy data was not available. AlexNet-Naïve Bayes model by Sarkar et al. [73] and the EfficientNetB0-SVM model by Celik and Inik [74] also showed compact performance, with AUC(ROC) values of 0.9630 and 0.9650, respectively. Bansal et al. [75] hybrid CNN-SVM model, which uses CNN as a feature extractor and SVM for classification, achieved an AUC(ROC) of 0.9800 and an F1-score of 98.00. Özkaraca et al. [76] augmented the CNN model with additional layers after transfer learning, producing an AUC(ROC) of 0.9650 and an F1-score of 94.00.

The marked improvement in the proposed SAlexNet architectures is attributed to the modification of the layered architecture. Capable of learning complex representations due to its multi-layered deep network, the proposed network captures features at various aspect ratios. SAlexNet-1 surpassed the previous best F1-score of (0.9613 ± 0.0205) achieved by InceptionV3, attaining an impressive F1-score of (0.9862 ± 0.0100). This significant improvement demonstrates the effectiveness of our novel architecture in capturing complex tumor patterns and relationships within different views of a tumor type (sagittal, coronal, and axial). SAlexNet-2, built upon the strengths of SAlexNet-1 with a smaller kernel size, achieved an even higher F1-score of (0.9935 ± 0.0050), solidifying its position as the top-performing model. SAlexNet-2 also excelled in other key metrics, including accuracy and AUC(ROC),

Table 13
Problem-specific sensitivity analysis of existing DL architectures to compare SAlexNet-1 and SAlexNet-2 using DS-2 (5-fold cross-validation used in all implementations; the values after '±' represents the standard deviation; 25 epochs run for each of the potential models, unless stated otherwise).

References	Method	Classification Performance		
		AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
Implementation Results from Scratch (Random Weight Initialization, 25 epochs)				
Krizhevsky et al. [39]	AlexNet	0.8136 ± 0.0210	0.7505 ± 0.0189	74.33 ± 0.02
Dosovitskiy et al. [61]	ViT	0.9079 ± 0.0344	0.8310 ± 0.0452	82.71 ± 0.03
Cholett [65]	Xception	0.9342 ± 0.0052	0.8691 ± 0.0162	86.75 ± 0.01
Szegedy et al. [66]	GoogLeNet	0.9138 ± 0.0192	0.8362 ± 0.0417	83.75 ± 0.01
Kaiming et al. [67]	ResNet18	0.7571 ± 0.0383	0.7202 ± 0.0150	68.00 ± 0.01
Kaiming et al. [67]	ResNet50	0.7433 ± 0.0020	0.6911 ± 0.0051	67.33 ± 0.00
Iandola et al. [68]	SqueezeNet	0.7486 ± 0.0071	0.7299 ± 0.0187	69.83 ± 0.02
Simonyan and Zisserman [69]	VGG-19	0.9216 ± 0.0064	0.8313 ± 0.0051	83.33 ± 0.01
Szegedy et al. [70]	InceptionV3	0.8114 ± 0.0211	0.7475 ± 0.0508	72.92 ± 0.01
Simonyan and Zisserman [69]	VGG-16	0.9267 ± 0.0013	0.8321 ± 0.0068	83.33 ± 0.01
Transfer Learning-based Implementations (25 epochs)				
Krizhevsky et al. [39]	AlexNet	0.9647 ± 0.0085	0.8852 ± 0.0254	88.57 ± 0.02
Dosovitskiy et al. [61]	ViT	0.9608 ± 0.0360	0.9178 ± 0.0518	91.83 ± 0.05
Cholett [65]	Xception	0.9525 ± 0.0129	0.8866 ± 0.0199	88.67 ± 0.02
Szegedy et al. [66]	GoogLeNet	0.9746 ± 0.0080	0.9224 ± 0.0176	92.25 ± 0.01
Kaiming et al. [67]	ResNet18	0.9874 ± 0.0052	0.9525 ± 0.0007	95.25 ± 0.00
Kaiming et al. [67]	ResNet50	0.9972 ± 0.0001	0.9725 ± 0.0025	97.25 ± 0.00
Iandola et al. [68]	SqueezeNet	0.9928 ± 0.0066	0.9766 ± 0.0200	97.67 ± 0.02
Simonyan and Zisserman [69]	VGG-19	0.9913 ± 0.0043	0.9587 ± 0.0140	95.87 ± 0.01
Szegedy et al. [70]	InceptionV3	0.9902 ± 0.0020	0.9583 ± 0.0017	95.83 ± 0.00
Simonyan and Zisserman [69]	VGG-16	0.9951 ± 0.0012	0.9700 ± 0.0032	97.00 ± 0.00
SAlexNet Classification Performance (proposed frameworks)				
Proposed Architecture 1	SAlexNet1 (25 epochs)	0.9950 ± 0.0054	0.9806 ± 0.0180	98.07 ± 0.01
Proposed Architecture 2	SAlexNet2 (25 epochs)	0.9982 ± 0.0029	0.9917 ± 0.0068	99.17 ± 0.00
Proposed Architecture 3	STL-SAlexNet1 (50 epochs)	0.9974 ± 0.0030	0.9693 ± 0.0086	96.93 ± 0.00
Proposed Architecture 4	STL-SAlexNet2 (50 epochs)	0.9977 ± 0.0042	0.9920 ± 0.0106	99.20 ± 0.01

indicating its robustness and reliability in classification tasks. SAlexNet-1 and SAlexNet-2 surpass existing models in exploiting DL strengths, achieving remarkable results, and establishing a new standard on the DS-1 dataset.

Fig. 22 provides a visual representation of the performance analysis for the DS-1 dataset using the proposed models SAlexNet-1 and SAlexNet-2. The graph showcases the performance of each classifier as a colored box, where the height of the box indicates its corresponding performance. Notably, all classifiers demonstrated exceptional performance, indicating their ability to handle complex problem structures effectively. Our analysis revealed a harmonious alignment between the cross-domain feature space and the solution space of the classifiers, suggesting that our models have successfully captured the underlying patterns and relationships within the data, enabling accurate classification across different domains. This visualization highlights the effectiveness of our proposed architectures, SAlexNet-1 and SAlexNet-2, in generalizing well across different domains by capturing complex problem structures through their solution spaces.

For DS-2, the preprocessing involved resizing and DWT-based denoising with Haar wavelets. No class imbalance was present in this dataset. Since the training and test directories were not defined in the DS-2 dataset, 5-fold cross-validation was adopted to evaluate model performance accurately and reduce overfitting. The comparative analysis using the DS-2 dataset for state-of-the-art methods is illustrated in Table 13.

AlexNet (74.33 ± 0.02)% and ViT (82.71 ± 0.03)% were implemented from scratch using pretrained methodology. ViT's superior results are likely due to its attention mechanism, which tracks element positions using index allocation. InceptionV3 (72.92 ± 0.01)% outperformed ResNet50 (67.33 ± 0.00)%. The empirical evaluation revealed that Xception achieved superior performance (86.75 ± 0.01)% on the DS-2 dataset, outperforming other cutting-edge architectures assessed in this study. For TL, state-of-the-art methods were implemented after modifying the output layers according to the DS-2 dataset. Pretrained weights after 25 epochs resulted in marked improvement due to training on an extensive dataset like ImageNet. TL provided better feature extraction and robustness to learn variations. Additionally, TL from related tasks improved results on target tasks with slight variations in the state-of-the-art algorithms. A smaller number of epochs improved performance by fine-tuning the model for the target dataset. SqueezeNet (97.67 ± 0.02 %), ResNet50 (97.25 ± 0.00)%, and VGG-16 (97.00 ± 0.00)% have a more suitable architecture for the binary classification of DS-2. It may be because SqueezeNet (97.67 ± 0.02)% is designed for mobile and embedded vision applications, making it efficient and lightweight. These architectures are crucial in DL as they enable efficient deployment on resource-constrained devices while maintaining performance [24,33]. Whereas ResNet50's residual connections help with gradient flow and feature learning, VGG-16's simplicity and a smaller number of parameters make it easier to fine-tune. AlexNet (88.57 ± 0.02)%, ViT (91.83 ± 0.05)%, Xception (88.67 ± 0.02)%, GoogLeNet (92.25 ± 0.01)%, ResNet18 (95.25 ± 0.00)%, VGG-19 (95.87 ± 0.01)%, and InceptionV3 (95.83 ± 0.00)% with TL appeared to need more epochs to reach the same performance level as achieved for the rest of the three architectures. Some of these models may have overfitted to the pretraining dataset and needed to generalize better to the target DS-2 dataset. Over and above, the limited number of epochs used might not be sufficient for some of the more complex models to converge or fine-tune effectively.

We conducted a problem-specific comparative analysis for the four proposed approaches (SAlexNet-1, SAlexNet-2, and the two respective STL-based approaches), the results of which are illustrated in Table 13. The first two models are unifold, using either SAlexNet-1 (98.07 ± 0.01 %) or SAlexNet-2 (99.17 ± 0.00 %). The rest of the proposed models are related to the STL approach, where the dataset DS-1 trains the model for 50 epochs, and then DS-2 is used for another 25 epochs by adjusting its input and output layers. Our analysis revealed that the STL-based approaches, STL SAlexNet-1 (96.93 ± 0.00 %) and STL SAlexNet-2 (99.20 ± 0.01 %), consistently delivered superior performance. The size of DS-1 is significantly smaller than the ImageNet dataset, and we achieved a balance for lower-sized datasets by using highly correlated data in the target domain.

Fig. 23 presents a visual representation of the performance evaluation of DS-2-based implementations, specifically showcasing a sensitivity analysis of the efficacy metrics obtained using the proposed DL paradigms, SAlexNet-1 and SAlexNet-2. This graphical illustration employs a box-plot visualization, where the vertical extent of each colored box corresponds to the magnitude of the respective performance metric, facilitating an intuitive comparison of classifier efficacy. A notable observation from this analysis is the exemplary performance exhibited

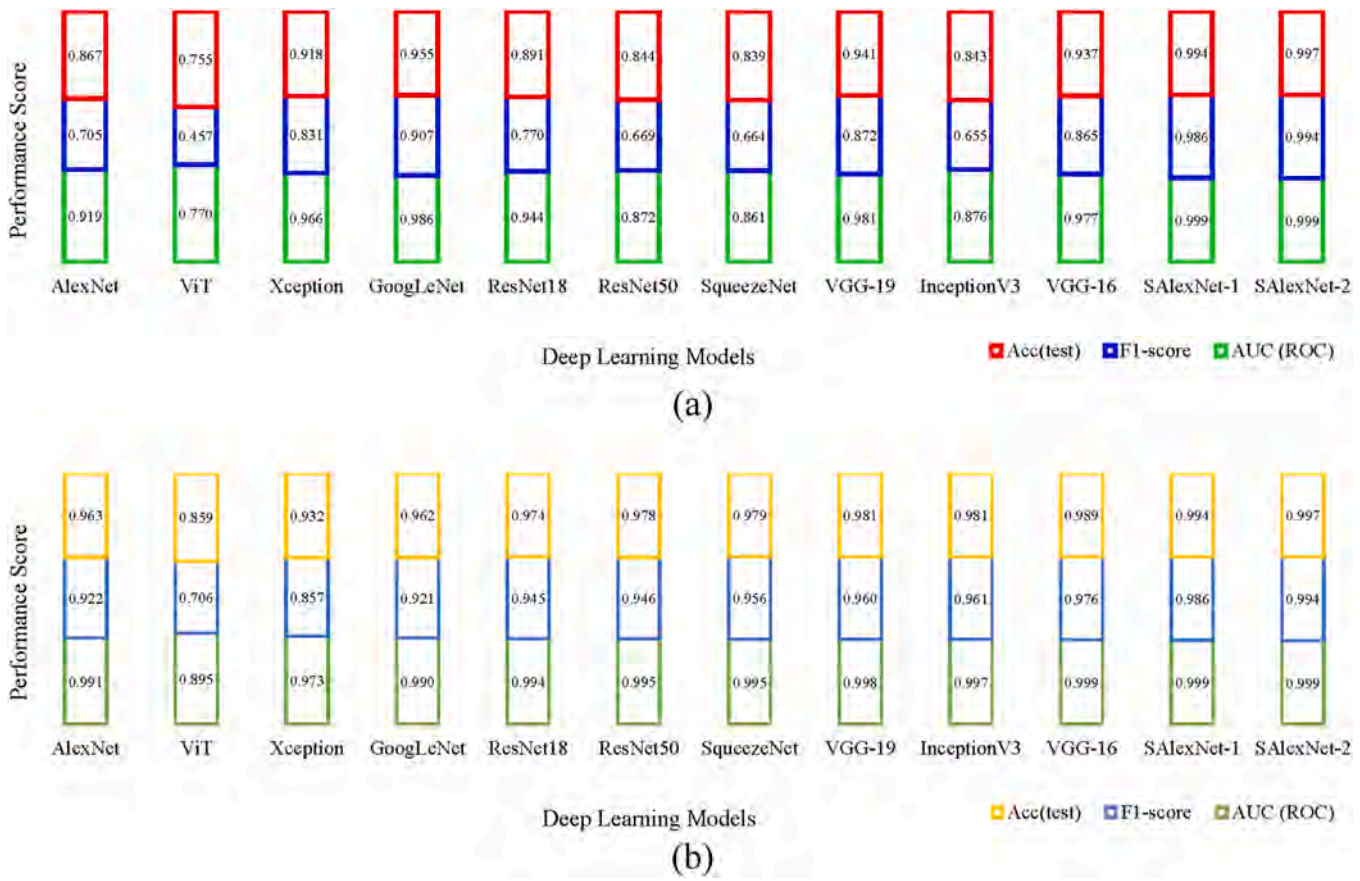


Fig. 22. A comparison of some state-of-the-art methods with SAlexNet-1 and SAlexNet-2 implemented for DS-1, (a) architectures from scratch, and (b) architectures with pretrained weights (TL). The percentage accuracy has been normalized for comparison purposes.

by all classifiers, underscoring their capacity to navigate intricate problem structures adeptly.

Our analysis reveals a remarkable harmony between the cross-domain feature space and the solution space of the classifiers, indicating a seamless alignment between the extracted features and the predictive models. This synergy suggests that the proposed models have successfully distilled the underlying patterns and interrelationships embedded within the data, enabling accurate classification across dissimilar domains. In the context of DL, this visualization underscores the efficacy of the proposed architectural frameworks, SAlexNet-1 and SAlexNet-2, and their STL-based variants using DS-2, in generalizing effectively across diverse domains. This is attributable to their capacity to capture complex problem structures through their solution spaces. More importantly, it demonstrates an enhanced ability to model and adapt to varied data distributions, instilling confidence in the versatility of our approach.

4.11. Main findings

This section provides a concise and objective account of the study's main outcomes on brain tumor classification framework, keeping the data separate from subjective analysis. In this context, we present the main findings of the SAlexNet models, a novel DL approach for brain tumor classification. Our proposed architectures, SAlexNet-1 and SAlexNet-2, introduce a paradigm-shifting, superimposed design methodology that integrates complementary information into the established AlexNet framework. This innovative approach uses the efficacy of channel- and spatial-attention mechanisms, facilitated through careful architectural modifications to the original model. By incorporating a hybrid attention mechanism at diverse layers within AlexNet, our framework harnesses the potential of attentional modulation to refine feature representations, culminating in enhanced brain tumor detection accuracy from MRI images. Ove and above, through the integration of hybrid attention mechanisms, dense feature extraction, and STL-based SAlexNet-1 and SAlexNet-2 demonstrate exceptional performance in tumor malignancy detection. The key highlights relate to the models' ability to capture salient features, extract rich information, and use prior knowledge to achieve state-of-the-art accuracy in multi-class and binary classification scenarios.

The HAM integrates spatial and channel-wise attention, selectively emphasizing key regional features and reweighting feature maps for enhanced representational power. Spatial attention focuses on relevant regions within feature maps, allowing the network to attend selectively to informative areas. Channel-wise attention emphasizes important feature channels, enhancing the discriminative power of the model. By integrating these attention mechanisms, SAlexNet-1 improves its ability to capture salient features related to tumor malignancy. Instead of a single large convolutional kernel, SAlexNet-1 employs multiple smaller kernels. This approach introduces extra non-linearity, enabling better feature extraction. The resulting feature maps are richer in information, contributing to improved classification performance.

The SAlexNet-2 improved the performance of SAlexNet-1 by changing the large-sized initial kernel with 5 convolution layers without any pooling and HAM involvement. It helped to boost the hidden details by learning complex representations combining simpler patterns. Due to their local focus, small kernels effectively captured subtle image details, like edges and blobs..

The STL-based SAlexNet-1 and SAlexNet-2 were comprehensively evaluated on the DS-2 dataset through a 5-fold cross-validation strategy, using pretrained weights from DS-1 as initialization. The models were

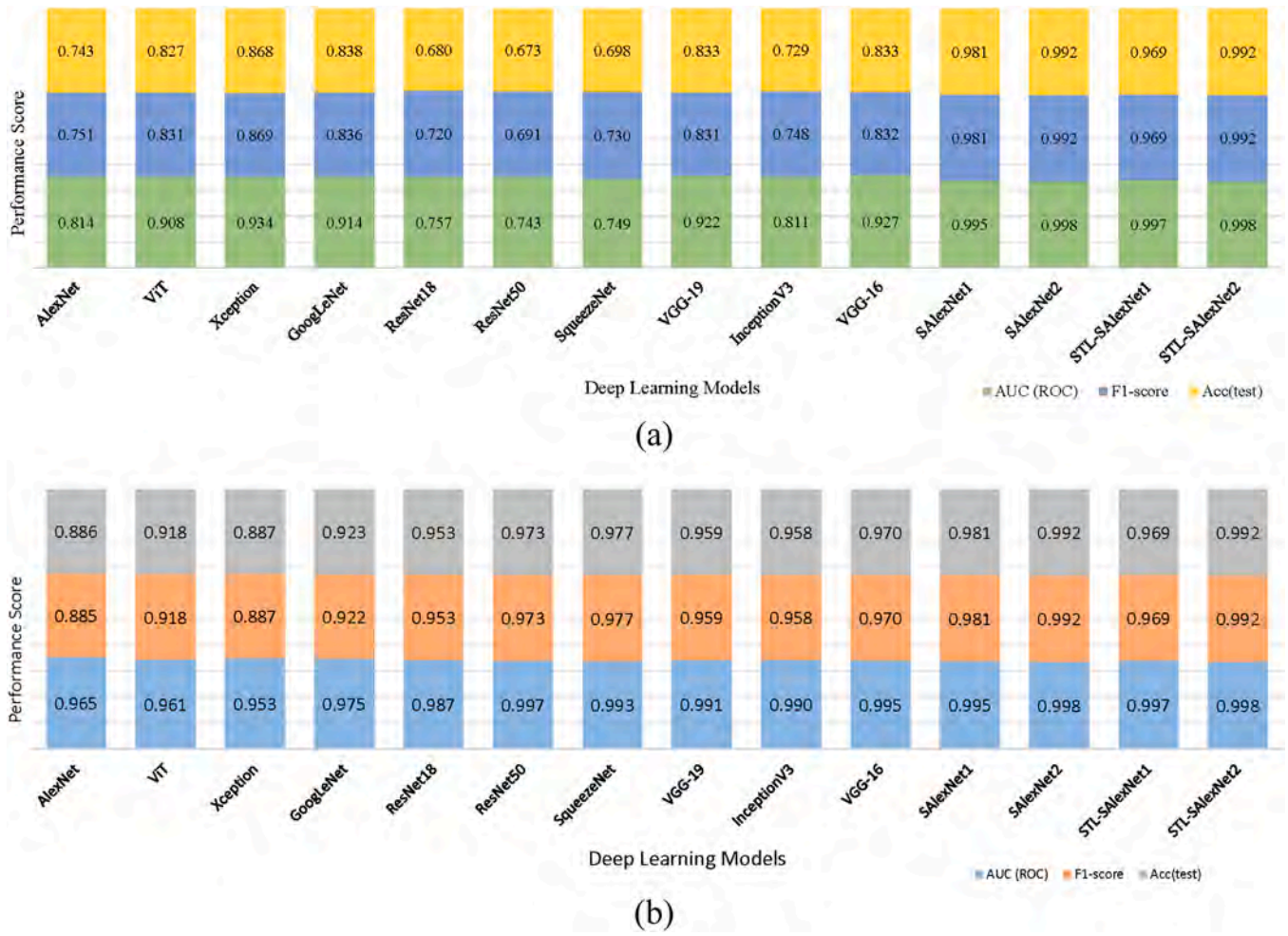


Fig. 23. A comparison of some state-of-the-art methods with SAlexNet-1, SAlexNet-2, and STL-based approaches implemented for DS-2, (a) architectures from scratch, and (b) architectures with pretrained weights (TL). The percentage accuracy has been normalized for comparison purposes.

trained for an optimum number of epochs with a consistent set of hyperparameters, including a fixed learning rate, weight decay, optimizer, activation functions, and loss function [50]. The output layers of the DL models were modified to consist of two neurons, facilitating binary classification through two-class probability vectors. The 5-fold cross-validation pipeline, illustrated in Fig. 20, ensured a robust assessment of the model's performance in a multi-institutional setting by iteratively dividing the images within each folder into five equal splits, utilizing each split as the validation set while training on the remaining splits. Furthermore, a marginal yet statistically significant improvement in the performance of STL-based SAlexNet-2 compared to the standard SAlexNet-2 architecture. This observation substantiates the efficacy of the STL approach, particularly in scenarios where datasets exhibit strong correlations, enabling the model to capture intricate patterns and relationships within the data. The STL-based SAlexNet-2 framework's enhanced performance can be attributed to its capacity to fine-tune pretrained weights, facilitating a more effective adaptation to the novel task, and its ability to use domain-specific knowledge from the source dataset to improve performance on the target dataset. This culminates in the conclusion that the STL approach is more pronounced in cases of correlated datasets, where the transfer of knowledge from the source domain to the target domain is more efficacious, thereby underscoring the utility of STL in harnessing the potential of pretrained models for improved performance in related tasks.

The proposed SAlexNet models hold promise for detecting primary brain tumor malignancy. Their high accuracy suggests potential clinical applications, aiding in early diagnosis, treatment planning, and patient management.

4.12. Cross-dataset evaluation for SAlexNet-1 and SAlexNet-2 robustness with diverse brain MRI scans

The robustness and generalizability of SAlexNet architectures were analyzed using external validation through cross-dataset testing on a distinct dataset. The efficacy of the proposed DL architectures, SAlexNet-1 and SAlexNet-2, was rigorously evaluated on publicly accessible datasets DS-1 and DS-2. To further assess their adaptability to diverse magnetic resonance imaging scans, the Brain Tumor MRI dataset, DS-3, was utilized. This dataset comprises 8764 instances, featuring two distinct classes, 'Yes' and 'No', publicly available at (https://www.kaggle.com/datasets/luluw8071/brain-tumor-mri-datasets). DS-3 is partitioned into training and testing directories, containing 4143 and 2869 instances for the 'Yes' and 'No' classes, respectively. For cross-dataset testing, the training instances of DS-2 were used for SAlexNet-1 and SAlexNet-2, whereas the performance metrics of SAlexNet-1, trained for 25 epochs, were tested on DS-3. The results of external testing, as illustrated in Table 14, exhibit a comparable trend consistent with previous findings. Utilizing SAlexNet-2 with 25 epochs on testing with the DS-3 yielded improved performance due to its enhanced encoder structure and HAM, as described in Section 3.5.1. The cross-dataset evaluation results have been illustrated in Table 14.

A comprehensive comparative analysis of the empirical results, as illustrated in Table 15, across two distinct datasets, DS-2 and DS-3, revealed negligible inter-dataset variability, with no statistically

Table 14
External validation results for SAlexNet-1 using 11 × 11 sized kernel for DS-2 / DS-3 (25 epochs, Sections 3.5 & 3.6).

Model	Class	$Acc_{(train)}$ (%)	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
SAlexNet-1	NO	98.41	0.9944	0.8901	0.9971	0.9393	94.74
	YES	98.90	0.9149	0.9957	0.9978	0.9536	94.74
	Average ± σ	**98.65 ± 0.00**	**0.9546 ± 0.0561**	**0.9429 ± 0.0747**	**0.9975 ± 0.0004**	**0.9465 ± 0.0100**	**94.74 ± 0.00**
SAlexNet-2	NO	99.79	0.9553	0.9771	0.9946	0.9661	97.26
	YES	99.79	0.9845	0.9695	0.9946	0.9769	97.26
	Average ± σ	**99.79 ± 0.00**	**0.9699 ± 0.0206**	**0.9733 ± 0.0053**	**0.9946 ± 0.0000**	**0.9715 ± 0.0076**	**97.26 ± 0.00**

Table 15
Relative sensitivity analysis of SAlexNet-1 and SAlexNet-2 generalization for cross-dataset validation.

Dataset Compared		DL Classification Architecture	
		SAlexNet-1 (DS-3, Table 14)	SAlexNet-2 (DS-3, Table 14)
DS-2 (%)	$Acc_{(test)}$	3.39	1.92
	F1-score	3.48	2.03
	Table involved	10	10

significant differences observed in test accuracy and F1-score metrics. This empirical evidence underscores the robustness, generalizability, and consistency of the proposed DL architectures, SAlexNet-1 and SAlexNet-2, in achieving superior performance across diverse dataset configurations and classification scenarios. The results demonstrate the architectures' ability to adapt to different datasets, reinforcing their reliability and effectiveness in real-world applications.

Table 14 for the SAlexNet architectures show evidence of overfitting. This phenomenon can be primarily attributed to the inherent complexity of these models, which often leads to memorization of training data and subsequently hinders the model's ability to generalize. Inadequate regularization techniques may also contribute to overfitting, as they fail to effectively prevent the model from fitting the noise in the training data.

Moreover, the degradation in performance observed in cross-dataset analysis can be attributed to the dataset shift phenomenon, where training and testing datasets exhibit different underlying distributions. This discrepancy arises from variations in data quality, resolution, and MRI machine parameters across datasets. Additionally, differences in demographics, anatomy, and pathology among the MRI scans in various datasets may further worsen this issue.

Another crucial factor contributing to performance degradation is the potential mismatch in feature representations between different brain tumor datasets. Specifically, label inconsistencies introduced during the annotation phase can lead to divergent feature spaces, making it challenging for the model to generalize across datasets. These findings underscore the importance of addressing dataset heterogeneity and developing robust models adapting to varying data distributions and characteristics.

4.12.1. STL-based cross-dataset analysis for SAlexNet-1 and SAlexNet-2

For external validation of STL-based models, pretraining is conducted using the DS-1 dataset. Subsequently, this pretrained model is employed for STL with 80 % of the instances from the DS-2 dataset after modifying the classification layer to match the number of classes in DS-3. Finally, instead of using the testing instances from DS-2, the testing instances from DS-3 are utilized to evaluate the generalization performance, following the cross-dataset analysis procedure.

The experimental outcomes have been illustrated in Table 16. The results of STL-SAlexNet-1 & -SAlexNet-2 models for brain tumor classification using datasets DS-1, DS-2, and DS-3 reveal notable differences in performance. SAlexNet-2 outperforms SAlexNet-1 across all metrics, particularly in the 'YES' class. Both models demonstrate consistent performance on the 'NO' class, with identical accuracy and similar recall, precision, and F1-score values. However, SAlexNet-2 significantly improves recall, precision, and F1-score for the 'YES' class, indicating its effectiveness in capturing relevant features.

The accuracy of SAlexNet-2 surpasses SAlexNet-1 on both training (99.79 %) and testing (97.26 %) sets. The excellent ROC-AUC values (>0.994) for both models suggest good discriminative power. Nonetheless, the 'YES' class appears to be the minority class, as evidenced by lower recall values compared to precision values. Both models generalize well from training to cross-testing sets, with minimal performance degradation. These findings imply that STL-SAlexNet-2′s architecture modifications are successful in enhancing brain tumor classification performance. Overall, STL-SAlexNet-2 demonstrates promising results, making it a potential candidate for brain tumor classification tasks.

4.13. Computational efficiency analysis

The computational analysis of the SAlexNet architectures has been carried out in four phases using DS-1 and DS-2 datasets. The investigation setup is illustrated in Table 17. The division of cost analysis of the system is based on algorithmic complexity, training time, memory usage, hardware requirements, and performance metrics. Both DL architectures, designed for brain tumor classification, employ convolutional layers, with SAlexNet-2 featuring more layers (9) than SAlexNet-1 (5). Regarding parameters, SAlexNet-2 has 13.9 million, which is only 1.4 % more than SAlexNet-1 (13.7 million).

Considering their architectural differences and training requirements, the computational cost of SAlexNet-1 and SAlexNet-2 is a critical aspect to evaluate. A key factor is the Floating Point Operations (FLOPS) required, with SAlexNet-2 demanding approximately 56 % more FLOPS (2.331 billion) compared to SAlexNet-1 (1.491 billion). This increase is largely attributed to the additional convolutional layers in SAlexNet-2. Although the difference in parameters is relatively small, the increased FLOPS in SAlexNet-2 likely results in longer training times.

The training time performance of SAlexNet-1 and SAlexNet-2 was evaluated on two publicly available datasets, DS-1 and DS-2. Notably, SAlexNet-2 demonstrated a higher total training time compared to

Table 16
Results with STL-based SAlexNet-1 and SAlexNet-2 models using DS-1, DS-2 and DS-3.

STL Model	Classes	$Acc_{(train)}$ (%)	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
SAlexNet-1	NO	98.41	0.9944	0.8901	0.9971	0.9393	94.74
	YES	98.90	0.9149	0.9957	0.9978	0.9536	94.74
	Average ± σ	**98.65 ± 0.00**	**0.9546 ± 0.0561**	**0.9429 ± 0.0747**	**0.9975 ± 0.0004**	**0.9465 ± 0.0100**	**94.74 ± 0.00**
SAlexNet-2	NO	98.41	0.9944	0.8901	0.9971	0.9393	94.74
	YES	99.79	0.9845	0.9695	0.9946	0.9769	97.26
	Average ± σ	**99.79 ± 0.00**	**0.9699 ± 0.0206**	**0.9733 ± 0.0053**	**0.9946 ± 0.0000**	**0.9715 ± 0.0076**	**97.26 ± 0.00**

Table 17
Computational cost analysis of SAlexNet-1 and SAlexNet-2 using DS-1 and DS-2.

Investigation domain	Option/Selection	Dataset			
		DS-1		DS-2	
		SAlexNet-1 (Table 6)	SAlexNet-2 (Table 9)	SAlexNet-1 (Table 10)	SAlexNet-2 (Table 10)
Algorithmic complexity	Type of layers	Convolutional	Convolutional	Convolutional	Convolutional
	Convolutional layers	5	9	9	9
	Number of parameters (million)	13.7	13.9	13.7	13.9
	Activations	ReLU	ReLU	ReLU	ReLU
	Input size	227 × 227	227 × 227	227 × 227	227 × 227
	Mini-batch size	16	16	16	16
	Data type	Float	Float	Float	Float
	Number of epochs	25	25	25	25
	Floating point operations (FLOPS)	1,491,055,888	2,331,447,568	1,491,055,888	2,331,447,568
Training Time (DS-2)	*Total training time, τ sec*	19,995	24,425	12,525	17,585
	Training Instances under trial	5712	5712	2400	2400
	Training time/ image, sec	3.5005	4.2760	5.2187	7.3270
	Learning rate	0.001	0.001	0.001	0.001
	Optimization algorithm	AdamW	AdamW	AdamW	AdamW
Memory Usage	Software frameworks	Python with core libraries: torch, torch.nn, torch.optim, torch.utils.data, torchvision. transfroms	Python with core libraries: torch, torch.nn, torch.optim, torch.utils.data, torchvision. transfroms	Python with core libraries: torch, torch.nn, torch.optim, torch.utils.data, torchvision. transfroms	Python with core libraries: torch, torch.nn, torch.optim, torch.utils.data, torchvision. transfroms
	Memory accesses (MHz)	2666	2666	2666	2666
Hardware Requirements	Energy requirements (W)	200	200	200	200
	Nvidia GeForce RTX 3050 GPU (6 GB)	Yes	Yes	Yes	Yes
	GPU cores	2304	2304	2304	2304
	RAM (GB)	16	16	16	16
	Processor details	13th Gen Intel(R) Core(TM) i5 13420H, 2.10 GHz	13th Gen Intel(R) Core(TM) i5 13420H, 2.10 GHz	13th Gen Intel(R) Core(TM) i5 13420H, 2.10 GHz	13th Gen Intel(R) Core(TM) i5 13420H, 2.10 GHz
Performance Metrics	Accuracy (test)	98.78 ± 0.80	99.69 ± 0.22	98.07 ± 0.02	99.17 ± 0.01
	F1-Score	0.9744 ± 0.0183	0.9935 ± 0.0050	0.9806 ± 0.0180	0.9917 ± 0.0068
	Precision	0.9752 ± 0.0183	0.9937 ± 0.0088	0.9818 ± 0.0272	0.9918 ± 0.0103
	Recall	0.9737 ± 0.0248	0.9933 ± 0.0090	0.9807 ± 0.0293	0.9917 ± 0.0106
	AUC (ROC)	0.9984 ± 0.0017	0.9994 ± 0.0008	0.9950 ± 0.0054	0.9982 ± 0.0029

SAlexNet-1 on both datasets, with 24,425 s on DS-2 and 17,585 s on DS-1. Despite this, the number of training instances under trial remained consistent across both models and datasets, with 5712 instances for DS-2 and 2400 instances for DS-1. Furthermore, SAlexNet-2 exhibited a higher training time per image compared to SAlexNet-1, with 7.3270 s on DS-1 and 5.2187 s on DS-2. This increased training time for SAlexNet-2 can be attributed to its enhanced architecture, incorporating additional convolutional layers and hybrid attention mechanisms, which ultimately contributed to its superior accuracy performance.

The training configuration for both architectures is identical, utilizing a mini-batch size of 16, float data type, and 25 epochs. The learning rate is set to 0.001, and the AdamW optimization algorithm is employed. This uniformity ensures a fair comparison between the two architectures. The input size for both is 227 × 227, and ReLU activation is used throughout.

In terms of hardware requirements, both architectures rely on the Nvidia GeForce RTX 3050 GPU (6 GB), with 2304 GPU cores, 16 GB RAM, and a 13th Gen Intel(R) Core(TM) i5 13420H processor clocked at 2.10 GHz. The energy requirement is 200 W for both setups. This consistency in hardware ensures that any performance differences can be attributed to the architectural designs rather than hardware variations.

The increased convolutional layers in SAlexNet-2 lead to better feature extraction and potentially improved performance. In this context, SAlexNet-2 consistently outperforms SAlexNet-1 across both datasets, with improvements in accuracy (DS-1: 99.69 % vs. 98.78 %, DS-2: 99.17 % vs. 98.07 %), F1-score (DS-1: 0.9935 vs. 0.9744, DS-2: 0.9917 vs. 0.9806), precision (DS-1: 0.9937 vs. 0.9752, DS-2: 0.9918 vs. 0.9818), and recall (DS-1: 0.9933 vs. 0.9737, DS-2: 0.9917 vs. 0.9807). These results suggest that the additional convolutional layers in SAlexNet-2 significantly enhance its ability to learn and generalize from the data.

Using Python with core libraries (torch, torch.nn, torch.optim, torch.utils.data, torchvision.transforms) as the software framework and the consistent memory access speed of 2666 MHz further ensure that the comparison is fair and unbiased. Overall, this setup provides a solid foundation for evaluating the effectiveness of SAlexNet-1 and SAlexNet-2 in brain tumor classification tasks.

As SAlexNet-2 demonstrates improved performance over SAlexNet-1, it's essential to acknowledge the trade-offs between performance gains and resource efficiency, particularly in practical healthcare environments. While the additional convolutional layers in SAlexNet-2 yield superior accuracy (testing) and F1-score, they also increase computational cost. Resource-constrained clinical environments may lead to prolonged processing times, increased costs, and reduced accessibility. Therefore, striking a balance between performance and efficiency is crucial. Clinicians and researchers must weigh the benefits of improved

diagnostic accuracy against the practical limitations of computational resources, considering factors such as hardware availability, power consumption, and potential delays in diagnosis.

4.14. Performance analysis of SAlexNet with poor contrast and low resolution

A comprehensive performance evaluation of SAlexNet-1 and SAlexNet-2 was conducted on contrast- and resolution-degraded images using the DS-2 dataset. The experimental protocol is detailed in Table 18. Specifically, the study's sample representations are outlined in the third column, while the subsequent columns display the images downsampled to 150 × 150 and then resized to 227 × 227. Resizing the original dataset to a fixed image size of (150 × 150) results in a consistent input shape, essential for the model to handle the data reliably. This resolution degradation process entailed a loss of texture and edge definition and increased noise levels. Different images were computed between the original and resized images to quantify the degradation, providing an objective measure of the impact.

The binary classification performance of SAlexNet-1 and SAlexNet-2 on downgraded DS-2 images, as illustrated in Table 19, reveals significant differences in their robustness to degraded image quality. SAlexNet-2 demonstrated exceptional performance, achieving an accuracy of 93.83 % and an AUC(ROC) of 0.9730, outperforming SAlexNet-1 (89.83 % accuracy on testing, 0.9398 AUC(ROC). Notably, SAlexNet-2′s recall and precision values exceeded 0.99 for the 'No' class, indicating effective TN identification. The standard deviation values were relatively low, indicating stable performance across classes. These results underscore SAlexNet-2′s enhanced ability to mitigate the effects of poor image quality, attributed to its hybrid attention mechanism and dense feature extraction capabilities. The substantial performance gap between SAlexNet-1 and SAlexNet-2 highlights the importance of robust feature extraction in brain tumor classification tasks.

4.15. Performance analysis of SAlexNet against hybrid noise impregnation

This study evaluates the robustness of SAlexNet methods against noise irregularities using a noise injection framework on the DS-2 image dataset. Specifically, we investigate the impact of Rician noise, salt-and-pepper noise, and a hybrid noise model combining both. The experimental setup and parameters are detailed in Table 20. Adding Rician noise to MRI images simulates low-contrast conditions with blurred boundaries and textures. Salt-and-pepper noise predominantly affects low-resolution images, where noise pixels dominate and obscure critical regions of interest, making accurate predictions challenging. The hybrid noise model, combining Rician and salt-and-pepper noise, further degrades image quality, increasing the difficulty for models in extracting accurate features. This hybrid noise introduces complexities such as reduced resolution, poor contrast, and edge blurring, thereby hindering the models' ability to learn discriminative patterns.

The robustness of SAlexNet-1 and SAlexNet-2 models against noise degradation was evaluated, as illustrated in Table 21, through binary classification tasks on images deteriorated by Rician noise, salt-and-

Table 18
Evaluation of dataset DS-2 on the effects of low contrast and limited resolution.

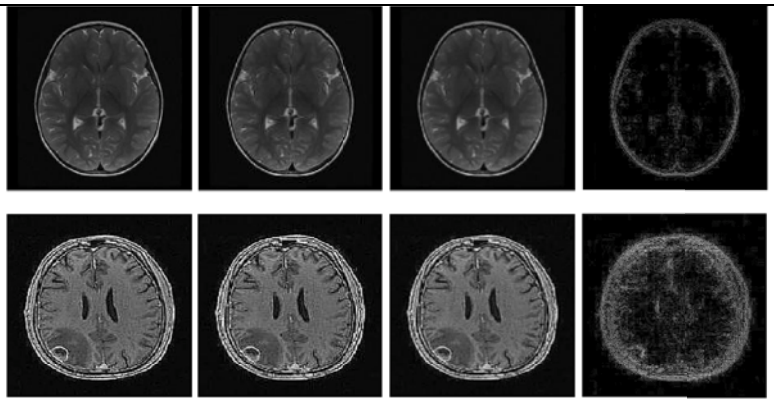
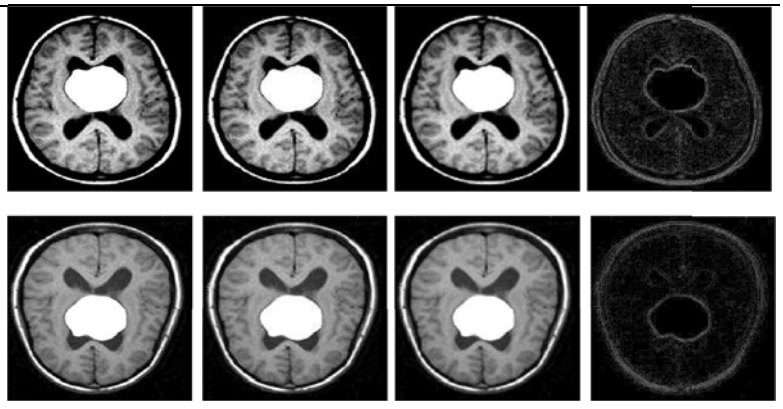
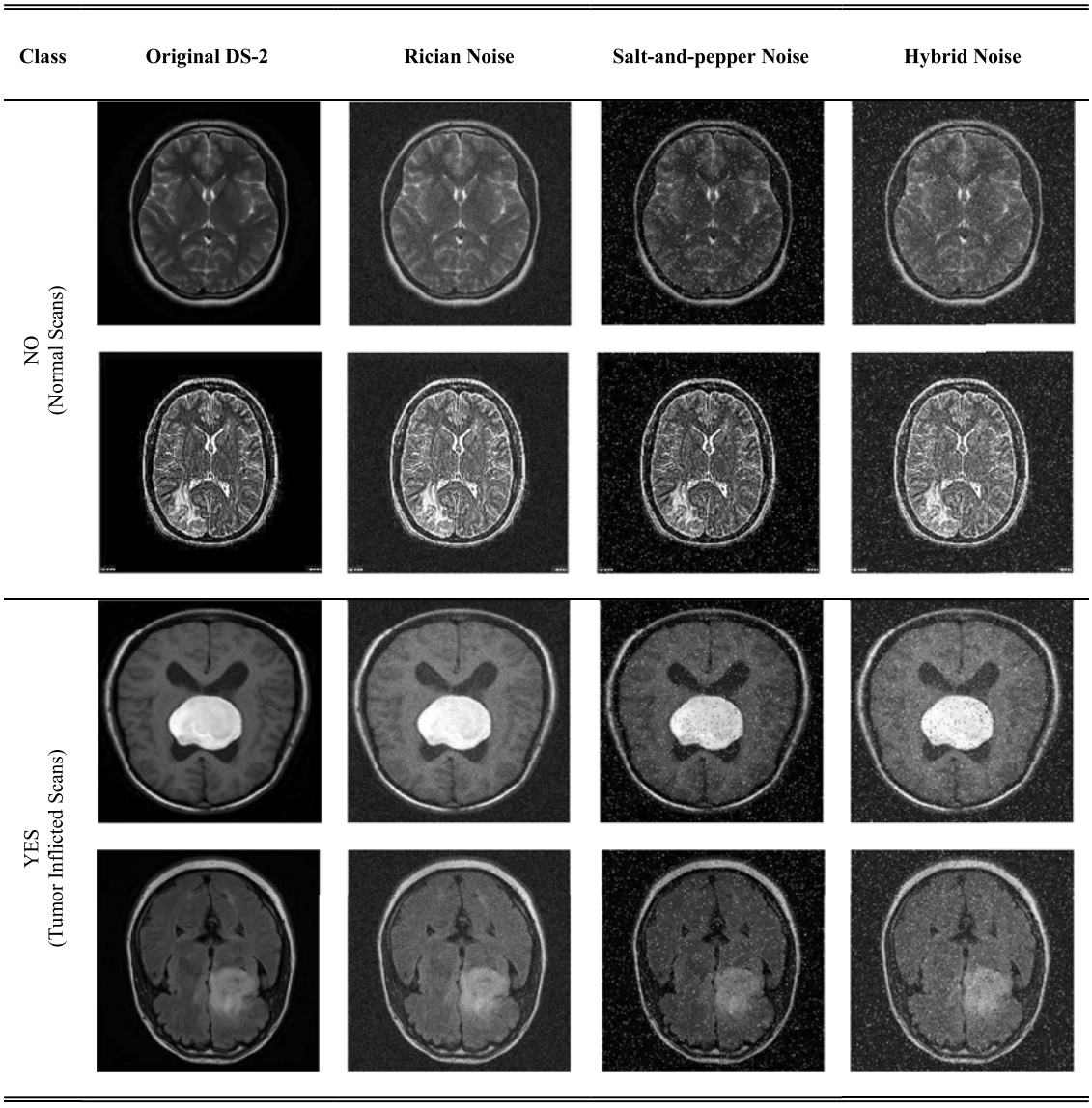
Class	Status	Original DS-2 (227^2)	Degraded images (150^2)	Resized images (227^2)	Degradation added
NO	Normal scans				
YES	Tumor inflicted scans				

Table 19
Binary classification results with SAlexNet models after training for 25 epochs (5 fold cross-validated) using resized DS-2 with downgraded images (150 × 150).

Model	Classes	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
SAlexNet-1	NO	0.9266	0.8769	0.9398	0.9011	89.83
	YES	0.8712	0.9222	0.9398	0.8953	89.83
	Average ± σ	**0.8989 ± 0.0391**	**0.8996 ± 0.0320**	**0.9398± 0.0000**	**0.8982 ± 0.0041**	**89.83 ± 0.00**
SAlexNet-2	NO	0.9966	0.8925	0.9730	0.9417	93.83
	YES	0.8811	0.9962	0.9730	0.9345	93.83
	Average ± σ	**0.9383 ± 0.0816**	**0.9443 ± 0.0733**	**0.9730 ± 0.0000**	**0.9381 ± 0.0050**	**93.83 ± 0.00**

Table 20
Binary classification results with SAlexNet models using noise-impregnated images from DS-2 after training for 25 epochs, whereas Rician noise ($\mu = 0$, $\sigma = 0.1$), Salt-and-pepper noise($p = 0.05$), and Hybrid noise (fused noise impregnation of Rician and salt-and-pepper noises).

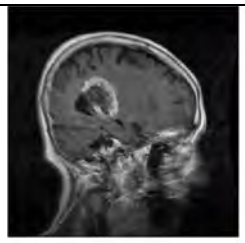
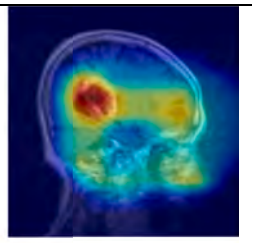
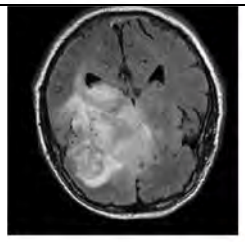
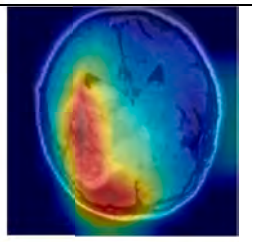
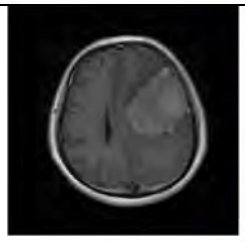
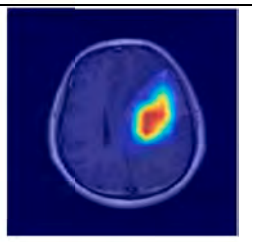
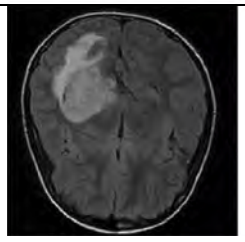
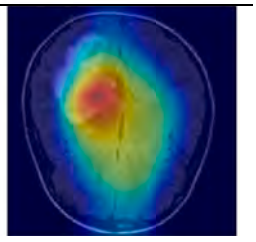


pepper noise, and hybrid noise. Notably, SAlexNet-2 consistently outperformed SAlexNet-1 across all noise types, achieving higher accuracy, precision, recall, and F1-score. Specifically, SAlexNet-2 demonstrated exceptional robustness against salt-and-pepper noise, with 97.66 % accuracy and 0.9973 AUC(ROC), and maintained high performance under hybrid noise degradation, with 97.21 % accuracy and 0.9972 AUC (ROC). In contrast, Rician noise had a more significant impact on SAlexNet-1′s performance, resulting in a 1.25 % accuracy drop compared to SAlexNet-2. These results underscore SAlexNet-2′s improved ability to mitigate noise degradation, attributed to its hybrid attention mechanism and dense feature extraction capabilities, making it a promising solution for reliable brain tumor classification in real-world scenarios where image quality may be compromised.

Table 21
Binary Classification results with SAlexNet models after training for 25 epochs based on images deteriorated by different noise models.

Variation	Arch.	Classes	Recall	Precision	AUC (ROC)	F1-score	$Acc_{(test)}$ (%)
Rician Noise	SAlexNet-1	NO	0.9933	0.9113	0.9712	0.9505	94.83
		YES	0.9033	0.9926	0.9708	0.9458	94.83
		Average ± σ	**0.9483 ± 0.0636**	**0.9519 ± 0.0574**	**0.9712 ± 0.0002**	**0.9481 ± 0.0033**	**94.83 ± 0.00**
	SAlexNet-2	NO	0.9933	0.9312	0.9791	0.9612	96.08
		YES	0.9266	0.9928	0.9789	0.9586	96.08
		Average ± σ	**0.9599 ± 0.0471**	**0.9620 ± 0.0435**	**0.9790 ± 0.0001**	**0.9599 ± 0.0018**	**96.08 ± 0.00**
Salt and Pepper Noise	SAlexNet-1	NO	0.9933	0.9226	0.9715	0.9566	95.50
		YES	0.9166	0.9927	0.9718	0.9532	95.50
		Average ± σ	**0.9549 ± 0.0542**	**0.9576 ± 0.0495**	**0.9716 ± 0.0002**	**0.9549 ± 0.0024**	**95.50 ± 0.00**
	SAlexNet-2	NO	0.9800	0.9735	0.9973	0.9767	97.66
		YES	0.9733	0.9798	0.9973	0.9765	97.66
		Average ± σ	**0.9766 ± 0.0047**	**0.9766 ± 0.0044**	**0.9973 ± 0.0000**	**0.9766 ± 0.0001**	**97.66 ± 0.00**
Hybrid Noise	SAlexNet-1	NO	0.9933	0.9400	0.9766	0.9659	96.50
		YES	0.9366	0.9929	0.9766	0.9639	96.50
		Average ± σ	**0.9649 ± 0.0400**	**0.9664 ± 0.0374**	**0.9766 ± 0.0000**	**0.9649 ± 0.0014**	**96.50 ± 0.00**
	SAlexNet-2	NO	0.9939	0.9519	0.9972	0.9705	0.9717
		YES	0.9512	0.9895	0.9972	0.9693	0.9726
		Average ± σ	**0.9725 ± 0.0301**	**0.9707 ± 0.0265**	**0.9972 ± 0.0000**	**0.9699 ± 0.0008**	**0.9721 ± 0.0006**

Table 22
Grad-CAM technique results for DS-1 and DS-2 using SAlexNet-1 and SAlexNet-2 architectures using HAM.

Model	Dataset	Original Image	Grad-cam Output
SAlexNet-1	DS-1		
	DS-2		
SAlexNet-2	DS-1		
	DS2		

4.16. Hybrid attention mechanism visualization using Grad-CAM

To enhance the interpretability and visualization of HAM, we employed the Gradient-weighted Class Activation Mapping (Grad-CAM) technique [77], a class-discriminative visualization method that highlights spatial regions contributing to the model's predictions. By computing the gradient-weighted importance of feature maps from the final convolutional layer (HAM Block5) in Figs. 6 & 10 for SAlexNet-1 and SAlexNet-2, respectively, which captures high-level semantic features essential for brain tumor detection, Grad-CAM generates attention maps quantifying pixel-wise importance in MRI scans. These maps provide insights into the model's decision-making process, revealing how attention is distributed across different regions of the input image. Preprocessing involved normalizing and resizing images to 227 × 227 pixels, followed by Grad-CAM's heatmap generation, where darker areas indicate stronger output influence and identify regions of interest. To refine heatmap quality and reduce spatial noise, we used eigen_smoothing, a smoothing option using the eigenvalues of the gradient covariance matrix. This technique enhances visualization by filtering out irrelevant activations and emphasizing critical pixels. As illustrated in Table 22, Grad-CAM visualizes the attention mechanism outcome for critical pixels, providing transparent insights into our HAM model's focus areas and supporting informed decision-making in medical diagnostics.

4.17. Limitations and future recommendations

Based on DL techniques, the proposed brain tumor classification has several drawbacks in terms of interpretability. One major concern is its black-box nature, making understanding the reasoning behind its predictions challenging. The model's complexity, with multiple layers and parameters, further hinders the ability to interpret its decisions. Additionally, the model's internal workings, including feature extraction and weighting, are not readily understandable, and it is difficult to determine which input features contribute most to the model's predictions.

To address these limitations, various techniques can be employed to explain the model's choices. Model-agnostic techniques, such as saliency maps, feature attribution methods like LIME (Local Interpretable Model-agnostic Explanations) [78] or SHAP (SHapley Additive exPlanations) [79], and partial dependence plots can provide valuable insights. Furthermore, Explainable AI (XAI) frameworks like TensorFlow Explainability and PyTorch Captum offer a range of interpretability techniques [80,81].

By implementing these techniques, the proposed model can provide insights into its decision-making process, increasing trust and understanding among clinicians, researchers, and patients. This transparency is crucial in healthcare applications, where model interpretability can directly impact patient outcomes and treatment plans.

Despite the promising results of the proposed SAlexNet brain tumor classification models, several clinical, practical, data protection, reimbursement, and implementation limitations necessitate further consideration to ensure its effective and responsible translation into real-world healthcare settings. From a clinical and practical perspective, a primary concern is the slight lack of clinical validation (Section 4.14), where the model's predictions are not correlated with patient outcomes, treatment responses, or survival rates. This gap makes it challenging to ascertain the model's actual impact on patient care and treatment planning. The use of sensitive medical imaging data raises risks related to patient data confidentiality, anonymity, and informed consent. Specifically, the model's training and testing datasets may contain identifiable patient information, potentially vulnerable to data breaches or unauthorized access. The integration of DICOM and PACS is essential for modern-day models. Regarding reimbursement and existing legislation, additional limitations emerge. The model's classification as a medical device may involve regulatory requirements under regional regulations. The reimbursement pathways for AI-driven diagnostic tools are still evolving and uncertain. The lack of clear guidelines on accountability in cases of model misdiagnoses also raises apprehensions. Over and above, future research should mitigate existing limitations and enhance the model's efficacy concerning clinical applicability. Additionally, explainability techniques should be integrated to provide insights into decision-making.

5. Conclusion

In conclusion, this study demonstrates the effectiveness of the proposed SAlexNet models, SAlexNet-1 and SAlexNet-2, in classifying primary brain tumors from MRI datasets with exceptional accuracy. Our models achieved state-of-the-art performance by incorporating HAM, dense feature extraction, and semi-transfer learning, surpassing existing complex models and techniques. Specifically, SAlexNet-2 attained outstanding accuracy rates of 99.69 % and 99.17 % in multi-class and binary classification tasks, respectively. These results underscore the potential of our approach to improve brain tumor diagnosis, treatment planning, and patient outcomes. The high accuracy and reliability of SAlexNet models make them valuable tools for clinical decision-making, ultimately contributing to better patient care.

CRediT authorship contribution statement

Qurat-ul-ain Chaudhary: Writing – original draft, Visualization, Software, Resources, Methodology, Data curation, Conceptualization. **Shahzad Ahmad Qureshi:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Touseef Sadiq:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition, Data curation. **Anila Usman:** Writing – review & editing, Validation, Investigation, Formal analysis. **Ambreen Khawar:** Writing – review & editing, Validation, Investigation, Formal analysis, Conceptualization. **Syed Taimoor Hussain Shah:** Visualization, Methodology, Investigation, Formal analysis. **Aziz ul Rehman:** Writing – review & editing, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was conducted with the support of the PIEAS IT & Telecom Endowment fund under the Pakistan Higher Education Commission (HEC). Registration ID: 03-7P1-006-2020.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rineng.2025.104025.

Data availability

Data will be made available on request.

References

[1] D.N. Louis, et al., The 2016 World Health Organization classification of tumors of the central nervous system: a summary, Acta Neuropathol. 131 (6) (2016) 803–820.
[2] Q.T. Ostrom, et al., CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016, Neuro-Oncology 21 (Supplement_5) (2019) v1–v100.

[3] M.L. Goodenberger, R.B. Jenkins, Genetics of adult glioma, Cancer Genet. 205 (12) (2012) 613–621.
[4] D.N. Louis, et al., The 2016 World Health Organization classification of tumors of the central nervous system: a summary, Acta Neuropathol. 131 (2016) 803–820.
[5] T.C. Hollon, et al., Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks, Nat. Med. 26 (1) (2020) 52–58.
[6] D. DePaoli, et al., Rise of Raman spectroscopy in neurosurgery: a review, J. Biomed. Opt. 25 (5) (2020), 050901-050901.
[7] H.H. Sultan, N.M. Salem, W. Al-Atabany, Multi-classification of brain tumor images using deep neural network, IEEE Access 7 (2019) 69215–69225.
[8] Rathi, V.P. and S. Palani, Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis. arXiv preprint arXiv: 1208.2128, 2012.
[9] M. Gomroki, M. Hasanlou, J. Chanussot, Automatic 3D multiple building change detection model based on encoder-decoder network using highly unbalanced remote sensing datasets, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2023).
[10] M. Gomroki, M. Hasanlou, P. Reinartz, STCD-EffV2T unet: semi transfer learning EfficientNetV2 T-unet network for urban/land cover change detection using sentinel-2 satellite images, Remote Sens. 15 (5) (2023) 1232.
[11] C. Sompong, S. Wongthanavasu, Brain tumor segmentation using cellular automata-based fuzzy c-means, in: Proceedings of the 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE, 2016.
[12] E.I. Zacharaki, et al., Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme, Magn. Reson. Med. 62 (6) (2009) 1609–1618. An Official Journal of the International Society for Magnetic Resonance in Medicine.
[13] E.S.A. El-Dahshan, T. Hosny, A.B.M. Salem, Hybrid intelligent techniques for MRI brain images classification, Digit. Signal Process. 20 (2) (2010) 433–441.
[14] J. Cheng, et al., Enhanced performance of brain tumor classification via tumor region augmentation and partition, PLoS ONE 10 (10) (2015) e0140381.
[15] J.S. Paul, et al., Deep learning for brain tumor classification, in: Proceedings of the Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging, International Society for Optics and Photonics, 2017.
[16] A.K. Anaraki, M. Ayati, F. Kazemi, Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms, Biocybern. Biomed. Eng. 39 (1) (2019) 63–74.
[17] G. Hinton, Y. LeCun, Y. Bengio, Deep learning, Nature 521 (7553) (2015) 436–444.
[18] G. Litjens, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[19] Z. Akkus, et al., Deep learning for brain MRI segmentation: state of the art and future directions, J. Digit. Imaging 30 (4) (2017) 449–459.
[20] Z.N.K. Swati, et al., Content-based brain tumor retrieval for MR images using transfer learning, IEEE Access 7 (2019) 17809–17822.
[21] W. Widhiarso, Y. Yohannes, C. Prakarsah, Brain tumor classification using gray level co-occurrence matrix and convolutional neural network, IJEIS (Indones. J. Electron. Instrum. Syst.) 8 (2) (2018) 179–190.
[22] M.M. Badža, M.Č. Barjaktarović, Classification of brain tumors from MRI images using a convolutional neural network, Appl. Sci. 10 (6) (2020) 1999.
[23] V. Rajasekar, et al., Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques, Results Eng. 18 (2023) 101111.
[24] S.A. Qureshi, et al., Intelligent ultra-light deep learning model for multi-class brain tumor detection, Appl. Sci. 12 (8) (2022) 3715.
[25] R. Latha, et al., Brain tumor classification using SVM and KNN models for smote based MRI images, J. Crit. Rev. 7 (12) (2020) 1–4.
[26] K.L.C. Hsieh, C.M. Lo, C.J. Hsiao, Computer-aided grading of gliomas based on local and global MRI features, Comput. Methods Programs Biomed. 139 (2017) 31–38.
[27] J. Sachdeva, et al., A package-SFERCB-"Segmentation, feature extraction, reduction and classification analysis by both SVM and ANN for brain tumors, Appl. Soft Comput. 47 (2016) 151–167.
[28] M. Soltaninejad, et al., Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels, Comput. Methods Programs Biomed. 157 (2018) 69–84.
[29] H. Huang, et al., A deep multi-task learning framework for brain tumor segmentation, Front. Oncol. 11 (2021) 690244.
[30] Y. Fan, et al., RMAP-ResNet: segmentation of brain tumor OCT images using residual multicore attention pooling networks for intelligent minimally invasive theranostics, Biomed. Signal Process. Control 90 (2024) 105805.
[31] M.M. Islam, et al., Transfer learning architectures with fine-tuning for brain tumor classification using magnetic resonance imaging, Healthc. Anal. 4 (2023) 100270.
[32] M. Aljohani, et al., An automated metaheuristic-optimized approach for diagnosing and classifying brain tumors based on a convolutional neural network, Results Eng. 23 (2024) 102459.
[33] S. Arvind, et al., Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest X-rays, Results Eng. 17 (2023) 100929.
[34] P.T. Krishnan, et al., Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer, Front. Neuroinform. 18 (2024) 1414925.
[35] M. Agarwal, et al., Deep learning for enhanced brain tumor detection and classification, Results Eng. 22 (2024) 102117.
[36] S.A. Qureshi, et al., RobU-Net: a heuristic robust multi-class brain tumor segmentation approaches for MRI scans, Waves Random Complex Media (2024) 1–51.
[37] I. Aboussaleh, et al., 3DUV-NetR+: a 3D hybrid semantic architecture using transformers for brain tumor segmentation with MultiModal MR images, Results Eng. 21 (2024) 101892.
[38] A. Priya, V. Vasudevan, Brain tumor classification and detection via hybrid alexnet-gru based on deep learning, Biomed. Signal Process. Control 89 (2024) 105716.
[39] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.
[40] S. Serte, A. Serener, F. Al-Turjman, Deep learning in medical imaging: a brief review, Trans. Emerg. Telecommun. Technol. 33 (10) (2022) e4080.
[41] Kadam, A., S. Bhuvaji, and S. Deshpande, Brain tumor classification using deep learning algorithms, 2024.
[42] Brain tumor detection MRI. Available from: https://www.kaggle.com/datasets/abhranta/brain-tumor-detection-mri/data, 2024.
[43] V. Patel, K. Mistree, A review on different image interpolation techniques for image enhancement, Int. J. Emerg. Technol. Adv. Eng. 3 (12) (2013) 129–133.
[44] S. Fadnavis, Image interpolation techniques in digital image processing: an overview, Int. J. Eng. Res. Appl. 4 (10) (2014) 70–73.
[45] L. Taylor, G. Nitschke, Improving deep learning with generic data augmentation, in: Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2018.
[46] S. Montaha, et al., MNet-10: a robust shallow convolutional neural network model performing ablation study on medical images assessing the effectiveness of applying optimal data augmentation technique, Front. Med. 9 (2022) 924979.
[47] G. Li, et al., HAM: hybrid attention module in deep convolutional neural networks for image classification, Pattern Recognit. 129 (2022) 108785.
[48] Q. Wang, et al., ECA-Net: efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
[49] X. Zhu, et al., An empirical study of spatial attention mechanisms in deep networks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.
[50] S. Zhang, Y. Gao, Hybrid multi-objective evolutionary model compression with convolutional neural networks, Results Eng. 21 (2024) 101751.
[51] Z. Wei, J. Du, Reinforcement learning-based optimal trajectory tracking control of surface vessels under input saturations, Int. J. Robust Nonlinear Control 33 (6) (2023) 3807–3825.
[52] Z.F. Xue, et al., Neural network-based knowledge transfer for multitask optimization, IEEE Trans. Cybern. (2024).
[53] P.N. Dao, Q.P. Nguyen, M.H. Vu, Adaptive optimal coordination control of perturbed Bilateral Teleoperators with variable time delays using Actor-Critic Reinforcement Learning algorithm, Math. Comput. Simul. (2024).
[54] A.P. Dhawan, Medical Image Analysis, John Wiley & Sons, 2011.
[55] Z.H. Hoo, J. Candlish, D. Teare, What is an ROC Curve? BMJ Publishing Group Ltd and the British Association for Accident, 2017, pp. 357–359, p.
[56] P. Flach, M. Kull, Precision-recall-gain curves: PR analysis done right, Adv. Neural Inf. Process. Syst. 28 (2015).
[57] J. Davis, M. Goadrich, The relationship between Precision-Recall and ROC curves, in: Proceedings of the 23rd International Conference on Machine learning, 2006.
[58] H. Feizi, M.T. Sattari, H. Apaydin, A comparative study of different optimization algorithms for the optimum operation of the Mahabad dam reservoir, Results Eng. 21 (2024) 101664.
[59] R. Llugsi, et al., Comparison between Adam, AdaMax and Adam W optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito, in, in: Proceedings of the IEEE Fifth Ecuador Technical Chapters Meeting (ETCM), IEEE, 2021.
[60] N. Parmar, et al., Image transformer, in: Proceedings of the International Conference on Machine Learning, PMLR, 2018.
[61] Dosovitskiy, A., et al., An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
[62] O.N. Manzari, et al., MedViT: a robust vision transformer for generalized medical image classification, Comput. Biol. Med. 157 (2023) 106791.
[63] R. Strudel, et al., Segmenter: transformer for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.
[64] Z. Li, et al., LViT: language meets vision transformer in medical image segmentation, IEEE Trans. Med. Imaging (2023).
[65] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[66] C. Szegedy, et al., Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[67] K. He, et al., Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[68] Iandola, F.N., SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360, 2016.
[69] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[70] C. Szegedy, et al., Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[71] E. Albalawi, et al., Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach, Front. Comput. Neurosci. 18 (2024) 1418546.
[72] Z. Rasheed, et al., Integrating convolutional neural networks with attention mechanisms for magnetic resonance imaging-based classification of brain tumors, Bioengineering 11 (7) (2024) 701.

[73] A. Sarkar, et al., An effective and novel approach for brain tumor classification using AlexNet CNN feature extractor and multiple eminent machine learning classifiers in MRIs, J. Sens. 2023 (1) (2023) 1224619.
[74] M. Celik, O. Inik, Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor Multi-Classification, Expert Syst. Appl. 238 (2024) 122159.
[75] Bansal, S., R.S. Jadon, and S.K. Gupta, A robust hybrid convolutional network for tumor classification using brain MRI image datasets. 2024.
[76] O. Özkaraca, et al., Multiple brain tumor classification with dense CNN architecture using brain MRI images, Life 13 (2) (2023) 349.
[77] R.R. Selvaraju, et al., Grad-cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017.
[78] D. Garreau, U. Luxburg, Explaining the explainer: a first theoretical analysis of LIME, in: Proceedings of the International Conference on Artificial Intelligence and Statistics, PMLR, 2020.
[79] L. Antwarg, et al., Explaining anomalies detected by autoencoders using Shapley Additive Explanations, Expert Syst. Appl. 186 (2021) 115736.
[80] S.T.H. Shah, et al., Data-driven classification and explainable-AI in the field of lung imaging, Front. Big Data 7 (2024) 1393758.
[81] R.K. Makumbura, et al., Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature, Results Eng. 23 (2024) 102831.