MethodsX 14 (2025) 103255

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Optimizing brain tumor detection in MRI scans through InceptionResNetV2 and deep stacked Autoencoders with SwiGLU activation and sparsity regularization ☆,☆☆

Vishal Awasthi [a], Mamta Tiwari [b], Amit Yadav [c], Gesu Thakur [d], Mamata Mayee Panda [b], Hemant Kumar [e,*], Shivneet Tripathi [b]

[a] Department of Electronics and Communication Engineering, Chhatrapati Shahu Ji Maharaj University, Kanpur, India
[b] Department of Computer Application, Chhatrapati Shahu Ji Maharaj University, Kanpur, India
[c] Department of Computer Application, PSIT College of Higher Education, Kanpur, India
[d] COER University, Roorkee, India
[e] Department of Information Technology, Chhatrapati Shahu Ji Maharaj University, Kanpur, India

ARTICLE INFO

Method name:
InceptionDSAE-SwiGLU Framework

Keywords:
Brain tumor classification
InceptionResNetV2
Deep stacked autoencoders (DSAE)
SwiGLU activation function
Sparsity regularization
MRI imaging

ABSTRACT

This study presents an automated framework for brain tumor classification aimed at accurately distinguishing tumor types in MRI images. The proposed model integrates InceptionResNetV2 for feature extraction with Deep Stacked Autoencoders (DSAEs) for classification, enhanced by sparsity regularization and the SwiGLU activation function. InceptionResNetV2, pre-trained on ImageNet, was fine-tuned to extract multi-scale features, while the DSAE structure compressed these features to highlight critical attributes essential for classification. The approach achieved high performance, reaching an overall accuracy of 99.53 %, precision of 98.27 %, recall of 99.21 %, specificity of 98.73 %, and an F1-score of 98.74 %. These results demonstrate the model's efficacy in accurately categorizing glioma, meningioma, pituitary tumors, and non-tumor cases, with minimal misclassifications. Despite its success, limitations include the model's dependency on pre-trained weights and significant computational resources. Future studies should address these limitations by enhancing interpretability, exploring domain-specific transfer learning, and validating on diverse datasets to strengthen the model's utility in real-world settings. Overall, the InceptionResNetV2 integrated with DSAEs, sparsity regularization, and SwiGLU offers a promising solution for reliable and efficient brain tumor diagnosis in clinical environments.

- Leveraging a pre-trained InceptionResNetV2 model to capture multi-scale features from MRI data.
- Utilizing Deep Stacked Autoencoders with sparsity regularization to emphasize critical attributes for precise classification.
- Incorporating the SwiGLU activation function to capture complex, non-linear patterns within the data.

☆ **Related research article:** None.
☆☆ ***For a published article:*** None.
* Corresponding author.
E-mail address: hemantkumar@csjmu.ac.in (H. Kumar).

https://doi.org/10.1016/j.mex.2025.103255
Received 30 November 2024; Accepted 5 March 2025
Available online 7 March 2025

Specifications table

Subject area:	Neuroscience
More specific subject area:	*Brain tumor, Deep Learning,*
Name of your method:	*InceptionDSAE-SwiGLU Framework*
Name and reference of original method:	*None*
Resource availability:	https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

Background

Brain tumors present a formidable challenge in medical diagnostics due to their complex nature and the critical importance of accurate and timely detection. Early and precise identification of brain tumors is crucial for improving treatment outcomes and enhancing patient survival rates. Magnetic Resonance Imaging (MRI) is widely regarded as the gold standard for visualizing brain abnormalities, as it provides high-resolution images that aid clinicians in diagnosing and planning treatments for various types of brain tumors. However, manual interpretation of MRI scans is a labor-intensive process prone to inter-observer variability, emphasizing the need for automated and reliable classification systems [1].

Traditional brain tumor classification methods often rely on manual analysis of MRI scans by radiologists. While effective to some extent, these approaches are inherently subjective, time-consuming, and susceptible to human error. The advent of deep learning (DL) has transformed the landscape of automated medical image analysis, offering promising solutions for brain tumor classification. DL models have demonstrated the ability to analyze high-dimensional data and extract hierarchical features directly from raw inputs, surpassing the limitations of traditional methods. Yet, challenges such as overfitting, limited interpretability, and dependency on large annotated datasets persist, highlighting the need for continued advancements in this domain [2].

Machine learning (ML) methods have traditionally provided structured approaches to analyzing brain tumor data, often relying on hand-engineered features to achieve classification outcomes. Sandhiya and Kanaga Suba Raja employed optimized learning machines that integrated radiomic properties, achieving an accuracy of 97.92 % using benchmark datasets [3]. Amin et al. [4], combined DenseNet169-based feature extraction with classical ML classifiers, including Random Forests and Support Vector Machines, attaining an accuracy of 95.10 %. Zulfiqar et al. [5], introduced a lightweight classification model using EfficientNet with feature fusion techniques, which demonstrated computational efficiency and accuracy critical factors in real-world clinical applications. While effective, these ML approaches often face limitations when dealing with unstructured and high-dimensional medical imaging data.

Deep learning, with its ability to automatically extract features from raw data, has brought revolutionary improvements to medical image classification. Pacal et al. [6], proposed a Swin Transformer model with Hybrid Shifted Windows Multi-Head Self-Attention (HSW-MSA) and Residual-based Multi-Layer Perceptron (ResMLP), achieving an accuracy of 99.92 % on public MRI datasets. Khan et al. [7] developed a 23-layer Convolutional Neural Network (CNN), achieving 100 % accuracy in binary tumor classification and demonstrating the model's ability to distinguish subtle variations in tumor characteristics. Moreover, Montalbo [8], employed Vision Transformers enhanced with residual attention mechanisms, which improved both classification performance and model interpretability, critical factors for clinical adoption. Despite these advancements, challenges related to data dependency, generalizability, and model interpretability persist, leaving room for further innovation.

Hybrid approaches, combining deep learning and machine learning techniques, have recently gained significant attention for brain tumor classification. Nassar et al. [9], developed a hybrid ensemble model integrating outputs from five different architectures, achieving an accuracy of 99.31 % on T1-weighted contrast-enhanced MRI images. These ensemble methods effectively leveraged diverse feature representations, highlighting their robustness in clinical settings. Hossain et al. [10] explored explainable AI techniques through an ensemble of Vision Transformers and CNNs, achieving high accuracy alongside interpretability. Such hybrid approaches demonstrate the potential to address the limitations of standalone methods by combining their strengths.

In this research, we propose a novel system that integrates InceptionResNet-V2 as a feature extractor with a Deep Stacked Autoencoder, enhanced by the SwiGLU activation function in the classification layer. The SwiGLU activation function, which combines Swish activation with gating mechanisms, improves the network's ability to model complex, non-linear relationships within the data. This integration aims to enhance the classification accuracy of brain tumors from MRI images by leveraging advanced feature extraction and refined classification techniques.

Our proposed methodology addresses key limitations of prior approaches, such as overfitting and insufficient generalization capabilities when applied to unseen data. By incorporating sparsity regularization and weight decay, the model mitigates overfitting risks, ensuring better generalization. Moreover, the introduction of the SwiGLU activation function enhances non-linearity and gating mechanisms, enabling the model to capture intricate patterns essential for accurate classification. This framework aims to provide a robust, efficient, and interpretable solution for automated brain tumor diagnosis, contributing to improved clinical decision-making and patient outcomes.

Method details

The proposed system integrates InceptionResNet-V2 as a feature extractor with Deep Stacked Autoencoders for classification. Each MRI image undergoes preprocessing to standardize the data, after which features are extracted using InceptionResNet-V2. Leveraging its deep architecture and advanced modules, this network captures critical patterns relevant for tumor classification. The extracted features are then fed into the DSAE, which compresses them and performs classification. Fig. 1 illustrates the overall proposed methodology for brain tumor classification.

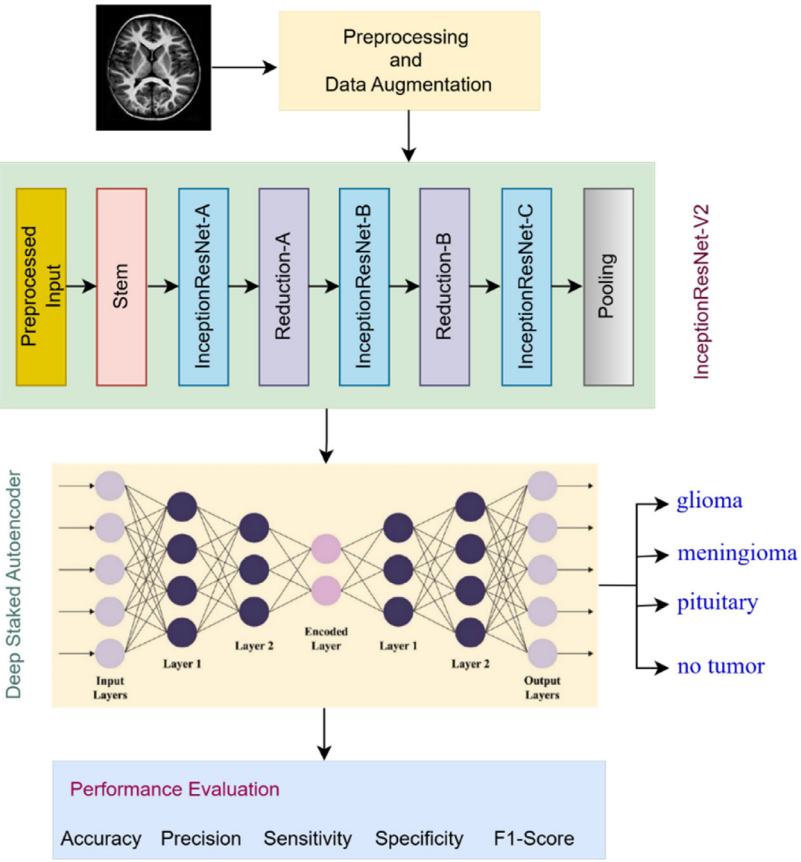

Fig. 1. Architecture of proposed methodology.

Preprocessing

Preprocessing is a crucial step to ensure uniformity in image quality and to prepare the data for effective feature extraction. Initially, the MRI images are resized to match the input requirements of InceptionResNet-V2, specifically to dimensions of 224×224 pixels. The images are then normalized to have zero mean and unit variance across the dataset, which aids in accelerating the convergence of the network during training. To reduce noise and enhance image quality, we apply a Gaussian filter with a kernel size of 3×3 and a standard deviation of 0.5. This step smooths the images and eliminates irrelevant high-frequency information that could adversely affect feature extraction.

Feature extraction using InceptionResNetV2

InceptionResNet-v2 [11] is a convolutional neural network (CNN) that combines the efficiency of Inception modules with depth-enabling residual connections, making it highly effective for image recognition and classification tasks. The architecture leverages Inception modules to capture multi-scale features through parallel convolutions and pooling operations, allowing the network to process features at varying levels of granularity simultaneously.

Residual connections, originally introduced in ResNet architectures, are integrated into InceptionResNet-V2 to address the vanishing gradient problem, a common challenge in training deep neural networks. These residual connections enable the model to train effectively at greater depths by preserving gradients, allowing InceptionResNet-V2 to achieve enhanced performance without the risk of degrading accuracy as the network depth increases.

Let X represent the input feature map, where $X \in \mathbb{R}^{\{h \times w \times d\}}$, h, w, and d are the height, width, and depth, respectively. The feature extraction process starts by passing X through the Inception modules. Each module performs convolutions at different kernel sizes to capture the different levels of abstraction.

The fundamental equation underlying the residual block, pivotal to InceptionResNet-v2 is

$$y = F\left(x, \{W_i\}\right) + x \tag{1}$$

where y is the output of the residual block, $F(x, \{W_i\})$ represents the transformation function learned by the InceptionResNet block, and x is the input to the block. $\{W_i\}$ refers to the learnable parameters within the transformation. The addition operation combines the output of the transformation $F(x)$ with the input x, allowing the network to learn residual mappings more efficiently.

Inception modules form the backbone of InceptionResNet-v2, employing parallel convolutional layers of varying filter sizes

$$f_{inception}(X) = \left[f_1(X),\ f_3(X),\ f_5(X), f_{pool}(X)\right] \tag{2}$$

where $f_1(X)$, $f_3(X)$, $f_5(X)$, and $f_{pool}(X)$ represent the 1×1 convolution, 3×3 convolution, 5×5 convolution, and pooling operations, respectively. These operations are concatenated along the depth dimension

$$f_{concat}(\mathrm{X}) = Concat\left(f_1(X),\ f_3(X),\ f_5(X), f_{pool}(X)\right) \tag{3}$$

InceptionResNet-V2 introduces residual connections by adding the original input X to the output of the Inception module

$$Y = X + f_{concat}(X) \tag{4}$$

where Y is the output feature map obtained after applying the residual connection. The residual addition allows the network to learn the residual mapping, $f_{concat}(X)$, which simplifies the optimization of deeper networks and ensures efficient feature propagation across layers.

Additionally, the scaling operation, introduced in InceptionResNet-v2, is critical in stabilizing training when deeper models are employed. The scaling factor, denoted as α, is applied to the residual branch of each block

$$y = \alpha \cdot F\left(x, \left\{W_i\right\}\right) + x \tag{5}$$

The value of α, typically set to a small constant, ensures that the learned residual functions are scaled down before being added to the input. This prevents drastic updates to the parameters, which could destabilize the learning process.

The feature extraction in InceptionResNet-v2 is completed by global average pooling at the end of the network, where all the spatial dimensions of the feature maps are averaged, producing a compact and global feature representation

$$y = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{i,j} \tag{6}$$

where H and W represent the height and width of the feature map, respectively, and x_{ij} is the feature value at position (i, j). This global average pooling ensures that the extracted features are spatially invariant, a desirable property for downstream tasks like classification.

The InceptionResNet-V2 model efficiently combines the multi-scale feature extraction capabilities of inception modules with the depth and gradient-flow improvements brought about by residual connections. Through its hybrid architecture, it performs robust feature extraction, capturing spatial hierarchies and fine-grained details across various scales.

Classification using deep stacked Autoencoder with SwiGLU activation

The classification phase employs a Deep Stacked Autoencoder (DSAE) [12], enhanced with the SwiGLU activation function in the classification layer, to classify brain tumors based on the features extracted from InceptionResNet-V2. DSAEs are powerful extensions of Autoencoders (AEs), specifically designed for compressing and learning useful representations from complex high-dimensional data.

Initially, the AE is trained in an unsupervised manner to reduce irrelevant noise and map input data into a reduced-dimensional space. Once the AE learns a suitable encoding, it provides a compressed version of the data. The DSAE, which stacks multiple AEs, aims to progressively refine the feature representations at each layer, learning higher-order features that become more abstract as the network depth increases [13].

In the DSAE architecture, the encoder-decoder pairs play a critical role. At each layer, the encoder compresses the input feature set into a smaller representation, while the corresponding decoder attempts to reconstruct the original feature set from this compressed space. The stacked arrangement of AEs enables the DSAE to learn hierarchical feature representations. Specifically, the input features from InceptionResNet-V2 pass through multiple encoding layers, each extracting more refined and compressed information until reaching the deepest hidden layer

$$h^{(l)} = ReLU\left(W^{(l)}\ h^{(l-1)} + b^{(l)}\right) \tag{7}$$

where, $h^{(l)}$ refers to the output at layer l, $W^{(l)}$ is the weight matrix, $b^{(l)}$ the bias term. This process enables the DSAE to concentrate on the most important features that the initial model, InceptionResNet-V2, might have missed.

A crucial enhancement in this work is the incorporation of the SwiGLU activation function in the classification layer. SwiGLU combines the advantages of the Swish activation function with gating mechanisms, improving the model's capacity to capture complex patterns in the data. The SwiGLU function is

$$SwiGLU(x) = x_1 \cdot Swish\left(x_2\right) \tag{8}$$

where $x = [x_1,\ x_2]$is split along the feature dimension, and Swish is

$$Swish(x) = x \cdot \sigma(x) \tag{9}$$

By applying the SwiGLU activation in the classification layer, the network can model more complex, non-linear relationships, which is particularly beneficial for distinguishing between different categories of brain tumors.

Additionally, sparsity regularization [14] is added in the encoder. Sparsity ensures that only a few neurons are active at any given time, leading to more robust feature extraction. The regularization term controlling sparsity is expressed

$$\Omega_{\text{spaarsity}} = \sum_{i=1}^{n} \mathbf{KL}(\rho \mid\mid \hat{\rho}_i) \tag{10}$$

$KL(\cdot)$ represents the Kullback-Leibler (KL) divergence [15], which is used to measure the difference between the desired activation level ρ and the average activation $\hat{\rho}_i$ of neuron i.

Furthermore, to mitigate the risk of overfitting, a weight decay term is incorporated into the model's loss function

$$\Omega_{\text{weight}} = \frac{\lambda}{2} \sum_{l=1}^{L} ||W^{(l)}||^2 \tag{11}$$

where λ is the regularization constant, and L denotes the number of layers. By controlling the magnitude of weights, this term ensures that the learned features generalize well to unseen data.

After encoding the refined features, the classification layer employs the SwiGLU activation function to enhance the model's expressive power. The output of the classification layer is

$$o = SwiGLU\left(W_o\, h^{(L)} + b_o\right) \tag{12}$$

where W_o and b_o are the weight matrix and bias vector of the output layer, and $h^{(L)}$ is the output from the last hidden layer of the encoder. The SwiGLU activation introduces non-linearity and gating mechanisms that improve the model's ability to capture intricate patterns in the data.

For final classification into brain tumor categories, the outputs from the SwiGLU-activated layer are directly used to determine the predicted class. Specifically, the class corresponding to the neuron with the highest activation in o is selected as the predicted category

$$\hat{y} = arg\max_j o_j \tag{13}$$

where o_j represents the j-th element of the output vector o and $\hat{y}$ is the predicted class label.

The optimization of the DSAE was carried out using the scaled conjugate gradient (SCG) algorithm, which is known for its efficiency in handling large datasets. The SCG minimizes the overall loss function, incorporating the reconstruction error, classification loss, and regularization terms. Although Softmax activation is not used, the model employs an appropriate classification loss function, such as the multi-class hinge loss, to compute the discrepancy between the predicted outputs and the true labels during training. The cost function is

$$J(W,b) = \frac{1}{N}\sum_{i=1}^{N}(||x^{(i)} - \hat{x}^{(i)}||^2 + \gamma \mathcal{L}_{class}\left(o^{(i)}, y^{(i)}\right)) + \beta\, \Omega_{\text{spaarsity}} + \lambda\, \Omega_{\text{weight}} \tag{14}$$

where N is the number of training examples, $x^{(i)}$ represents the feature set from InceptionResNet-V2, and $\hat{x}^{(i)}$is the reconstructed input. γ is a weighting factor balancing reconstruction and classification loss, $\mathcal{L}_{class}(o^{(i)}, y^{(i)})$ is the classification loss function.β and λ are the regularization parameters for sparsity and weight decay, respectively.

By leveraging the pre-extracted features from InceptionResNet-V2 and integrating the SwiGLU activation function in the classification layer, the DSAE achieves a highly refined and compact representation of the MRI data. The use of SwiGLU enhances the model's ability to capture complex, non-linear relationships within the data, which is essential for accurate brain tumor classification. The overall architecture ensures that the model performs well not only on the training data but also generalizes effectively to unseen cases, leading to superior classification accuracy and performance. This approach highlights the importance of integrating advanced activation functions like SwiGLU in deep learning architectures for complex medical imaging tasks

Method validation

The dataset utilized in this study comprises a total of 7023 MRI images categorized into four distinct classes: glioma, meningioma, pituitary tumor, and no tumor. Fig. 2, shows the sample image of brain MRI dataset of each class.

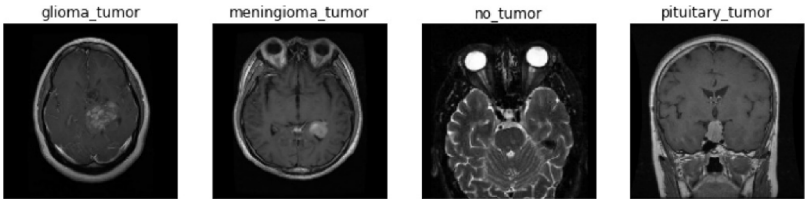

Fig. 2. Sample image from each class of the dataset.

Table 1
Dataset description.

Model	Training set	Validation set	Testing SET	Total
glioma	1321	180	120	1621
meningioma	1339	183	123	1645
pituitary tumor	1457	180	120	1757
no tumor	**1595**	**243**	**162**	**2000**
Total	**571**	**786**	**525**	**7023**

Table 2
Comparative performance of feature extractors across different metrics.

Model	CNN	LSTM	GRU	Transformer	DASE
VGG19	89.12	90.34	90.78	91.23	92.45
EfficientNetB0	92.34	93.44	93.89	94.56	95.12
MobileNetV2	90.89	91.78	92.23	92.96	93.45
ResNet50	91.56	92.67	93.17	93.90	93.51
InceptionV2	93.67	94.15	94.57	95.28	95.97
InceptionResNetV2	**93.14**	**94.21**	**95.63**	**95.88**	**96.12**

Before augmentation, the dataset was meticulously divided into training, validation, and test sets to ensure robust model training and unbiased evaluation. Training set consists of 5412 images, validation set consists of 786 images and test set consists of 525 images, shown in Table 1. This distribution reveals a slight imbalance, particularly in the "No Tumor" category, which is more prevalent in both the training and validation sets. To address this imbalance and enhance the model's generalization capabilities, data augmentation techniques such as rotation, flipping, and scaling were applied to the underrepresented classes. Rotation involved turning images at various angles to introduce rotational invariance, allowing the model to recognize tumors regardless of their orientation. Flipping included both horizontal and vertical flips to simulate different anatomical perspectives, enhancing the model's ability to generalize across diverse orientations. Additionally, scaling was applied by resizing images to different dimensions, which improved the model's capability to detect tumors of varying sizes. These augmentation strategies collectively ensured a more balanced and diverse training dataset, thereby enhancing the model's overall performance and robustness.

Implementing the proposed InceptionResNet-V2 with DSAEs model was carried out in a high-performance computing environment; leveraging an NVIDIA RTX 4060 GPU with CUDA, alongside 32 GB RAM, and an Intel i9 processor which handled the computational demands of deep learning on MRI data. The software environment included Python (v3.8), PyTorch for model construction, OpenCV and NumPy for image preprocessing, and Scikit-learn for data utilities.

The training procedure combined feature extraction with InceptionResNet-V2 and classification via Deep Stacked Autoencoders (DSAEs), utilizing an optimized set of parameters to enhance model performance. Training spanned over 100 epochs ensuring effective learning for both stages. A ReduceLROnPlateau callback was applied to adjust the learning rate dynamically; reducing it when performance plateaued to avoid overfitting and speed up convergence. Initial learning rates were set at 0.0001 for fine-tuning InceptionResNet-V2 and 0.001 for the DSAE layers, providing stable conditions for both feature learning and classification tasks.

For feature extraction, InceptionResNet-V2 was pre-trained on the ImageNet dataset and fine-tuned on MRI images resized to 224×224 pixels to ensure compatibility with the pre-trained weights. Key parameters included a learning rate of 0.0001, a batch size of 32, and a dropout rate of 0.3 which mitigated overfitting. The resulting feature vectors, capturing detailed multi-scale spatial features, were then passed to the DSAE for classification.

In the classification stage, Deep Stacked Autoencoders were employed to compress the feature vectors, learning hierarchical and compressed representations of the extracted features. The DSAE included three encoder-decoder layers; progressively reducing dimensionality from 512 to 256 and then to 128 in the latent space. A learning rate of 0.0001, L2 regularization of 0.001 to control model complexity and a sparsity coefficient of 0.1 were applied, enhancing focus on the most significant features while reducing noise.

The final classification was performed using the SwiGLU activation function, effectively distinguishing among the target classes: glioma, meningioma, pituitary tumor, and non-tumor. By incorporating SwiGLU in the classification layer, the model enhanced its capacity to capture complex nonlinear relationships within the compressed latent features.

In our experimental investigation, we initially evaluated several deep learning models to determine the most effective architecture for classifying brain tumor types from MRI images. We employed InceptionResNetV2 as the feature extractor due to its superior performance in preliminary experiments. Various deep learning classifiers were integrated with this feature extractor to identify the optimal combination. Each model was trained for 50 epochs using a learning rate of 0.001 with the Adam optimizer.

Table 2, shows that InceptionResNetV2 combined with the Deep Stacked Autoencoder (DSAE) achieved the highest accuracy of 96.12 %. Therefore, we selected InceptionResNetV2 as the feature extractor and proceeded with further experiments using the DSAE classifier.

Subsequently, we applied selected model, InceptionResNetV2 with DSAE on the brain tumor dataset for 100 epochs. The learning rate was set to 0.0001, and a weight decay of 0.01 was employed with the Adam optimizer. To mitigate overfitting, we utilized a

Table 3
Performance metrics of InceptionResNetV2 with DSAE.

Model	Training accuracy	Validation accuracy	Precision	Recall	Specificity	F1-score
InceptionResnetV2 + DSAE	96.12	92.19	93.45	94.10	94.21	93.77
InceptionResnetV2 + DSAE + Sparsity	98.56	94.34	96.01	96.83	96.22	96.42
InceptionResnetV2 + DSAE + Sparsity + SwiGLU	**99.53**	**97.11**	**98.27**	**99.21**	**98.73**	**98.74**

Table 4
Confusion matrix for InceptionResNetV2 + DSAE with sparsity and SwiGLU activation.

Actual\ Predicted	Glioma	Meningioma	Pituitary	NoTumor	Total
Glioma	*118*	*1*	*1*	*0*	*120*
Meningioma	*1*	*121*	*1*	*0*	*123*
Pituitary	*0*	*2*	*117*	*1*	*120*
NoTumor	*0*	*0*	*2*	*160*	*162*
Total	*119*	*124*	*121*	*161*	*525*

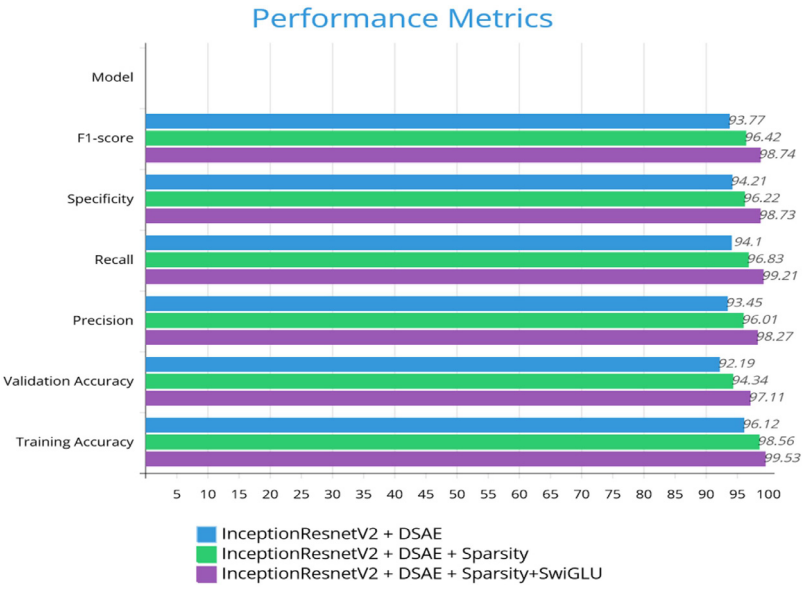

Fig. 3. Performance matrics.

dropout rate of 0.3 and implemented early stopping based on validation loss. we evaluated the performance of the proposed models on brain tumor classification using MRI images. The models tested included InceptionResNetV2 combined with a Deep Stacked Autoencoder (DSAE), with and without additional techniques such as sparsity regularization and the SwiGLU activation function. Table 3 summarizes the performance metrics for each model.

Fig. 3, shows the effectiveness of proposed model. The baseline model without sparsity regularization achieved a training accuracy of 96.12 % and a validation accuracy of 92.19 %. Introducing sparsity regularization significantly enhanced the model's performance, increasing training and validation accuracies to 98.56 % and 94.34 %, respectively. The precision and recall also saw improvements, indicating more accurate and reliable classifications.The most substantial improvement was observed upon incorporating the SwiGLU activation function. The final model achieved an impressive training accuracy of 99.53 % and a validation accuracy of 97.11 %. Precision, recall, specificity, and F1-score increased to 98.27 %, 99.21 %, 98.73 %, and 98.74 %, respectively. These enhancements underscore the effectiveness of SwiGLU in capturing complex, non-linear relationships within the data, thereby boosting the model's classification capabilities.

The Confusion Matrix in Table 4 illustrates that the model correctly classified 118 out of 120 glioma cases, 121 out of 123 meningioma cases, 117 out of 120 pituitary tumor cases, and 160 out of 162 no tumor cases. Misclassifications were minimal, with only a few instances where one tumor type was incorrectly predicted as another.

The results obtained from our experiments clearly demonstrate the effectiveness of the proposed model architecture. The initial combination of InceptionResNetV2 and DSAE achieved a respectable training accuracy of 96.12 % and validation accuracy of 92.19 %.

However, by introducing sparsity regularization, the model's ability to generalize improved significantly, with training and validation accuracies rising to 98.56 % and 94.34 %, respectively.

The incorporation of the SwiGLU activation function further enhanced the model's performance, leading to a remarkable training accuracy of 99.53 % and validation accuracy of 97.11 %. This improvement is attributed to SwiGLU's capability to model complex, non-linear relationships within the data, thereby refining the feature representations learned by the DSAE. The high precision, recall, specificity, and F1-score across all classes indicate that the model is highly effective in accurately classifying brain tumor types and distinguishing them from non-tumor cases.

Limitations

None.

Ethics statements

Publicly available dataset, thus no permission required.

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.

Data availability

I have linked the open source dataset which is used in this research.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Md.A. Talukder, et al., An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning, Expert Syst. Appl. 230 (Nov. 2023) 120534, doi:10.1016/j.eswa.2023.120534.
[2] A. Younis, L. Qiang, C.O. Nyatega, M.J. Adamu, H.B. Kawuwa, Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches, Appl. Sci. 12 (14) (Jan. 2022) Art. no. 14, doi:10.3390/app12147282.
[3] A.B. Subba, A.K. Sunaniya, Computationally optimized brain tumor classification using attention based GoogLeNet-style CNN, Expert Syst. Appl. 260 (Jan. 2025) 125443, doi:10.1016/j.eswa.2024.125443.
[4] J. Amin, M. Sharif, M. Raza, M. Yasmin, Detection of brain tumor based on features fusion and machine learning, J. Ambient Intell. Humaniz. Comput. 15 (1) (Jan. 2024) 983–999, doi:10.1007/s12652-018-1092-9.
[5] F. Zulfiqar, U.Ijaz Bajwa, Y. Mehmood, Multi-class classification of brain tumor types from MR images using EfficientNets, Biomed. Signal Process. Control 84 (Jul. 2023) 104777, doi:10.1016/j.bspc.2023.104777.
[6] I. Pacal, A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images, Int. J. Mach. Learn. Cybern. 15 (9) (Sep. 2024) 3579–3597, doi:10.1007/s13042-024-02110-w.
[7] Md.S.I. Khan, et al., Accurate brain tumor detection using deep convolutional neural network, Comput. Struct. Biotechnol. J. 20 (2022) 4733–4745, doi:10.1016/j.csbj.2022.08.039.
[8] F.J.P. Montalbo, TUMbRAIN: a transformer with a unified mobile residual attention inverted network for diagnosing brain tumors from magnetic resonance scans, Neurocomputing 611 (Jan. 2025) 128583, doi:10.1016/j.neucom.2024.128583.
[9] S.E. Nassar, I. Yasser, H.M. Amer, M.A. Mohamed, A robust MRI-based brain tumor classification via a hybrid deep learning technique, J. Supercomput. 80 (2) (Jan. 2024) 2403–2427, doi:10.1007/s11227-023-05549-w.
[10] S. Hossain, A. Chakrabarty, T.R. Gadekallu, M. Alazab, Md.J. Piran, Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification, IEEE J. Biomed. Health Inform. 28 (3) (Mar. 2024) 1261–1272, doi:10.1109/JBHI.2023.3266614.
[11] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-ResNet and the impact of residual connections on learning, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI Press, 2017, pp. 4278–4284. AAAI'17.
[12] M.S. Ullah, M.A. Khan, N.A. Almujally, M. Alhaisoni, T. Akram, M. Shabaz, BrainNet: a fusion assisted novel optimal framework of residual blocks and stacked autoencoders for multimodal brain tumor classification, Sci. Rep. 14 (1) (2024) 1–20, doi:10.1038/s41598-024-56657-3.
[13] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, Y. Fan, A deep learning model integrating FCNNs and CRFs for brain tumor segmentation, Med. Image Anal. (2018), doi:10.1016/j.media.2017.10.002.
[14] Y. Yan, H. Chu, Y. Jin, Y. Huan, Z. Zou, L. Zheng, Backpropagation with sparsity regularization for spiking neural network learning, Front. Neurosci. 16 (2022), doi:10.3389/fnins.2022.760298.
[15] S. Ji, Z. Zhang, S. Ying, L. Wang, X. Zhao, Y. Gao, Kullback–leibler divergence metric learning, IEEE Trans. Cybern. 52 (4) (2022) 2047–2058, doi:10.1109/TCYB.2020.3008248.