



OPEN

A novel deep neural architecture for efficient and scalable multidomain image classification

SM Nuruzzaman Nobel^{1,9}, Md All Moon Tasir^{2,9}, Humaira Noor^{3,9},
Muhammad Mostafa Monowar⁴, Md. Abdul Hamid⁴, Md. Shohel Sayeed⁵✉,
Md. Rajibul Islam⁶, M. F. Mridha⁷ & Nilanjan Dey⁸

Deep learning has significantly advanced the field of computer vision; however, developing models that generalize effectively across diverse image domains remains a major research challenge. In this study, we introduce DeepFreqNet, a novel deep neural architecture specifically designed for high-performance multi-domain image classification. The innovative aspect of DeepFreqNet lies in its combination of three powerful components: multi-scale feature extraction for capturing patterns at different resolutions, depthwise separable convolutions for enhanced computational efficiency, and residual connections to maintain gradient flow and accelerate convergence. This hybrid design improves the architecture's ability to learn discriminative features and ensures scalability across domains with varying data complexities. Unlike traditional transfer learning models, DeepFreqNet adapts seamlessly to diverse datasets without requiring extensive reconfiguration. Experimental results from nine benchmark datasets, including MRI tumor classification, blood cell classification, and sign language recognition, demonstrate superior performance, achieving classification accuracies between 98.96% and 99.97%. These results highlight the effectiveness and versatility of DeepFreqNet, showcasing a significant improvement over existing state-of-the-art methods and establishing it as a robust solution for real-world image classification challenges.

Keywords Deep learning, Hand sign, MRI tumor classification, Blood cells, Transfer learning, Vision transformer, Computer vision

The domain of computer vision has experienced a groundbreaking evolution with the introduction of deep learning techniques, which have reimagined the process of image classification by enabling the automated recognition of intricate and abstract patterns directly from unprocessed data¹. As a core area in both machine learning and computer vision, image classification faces significant hurdles, particularly in managing the complexity of heterogeneous datasets while ensuring consistent and precise performance across various application domains². Deep neural networks, particularly convolutional neural networks (CNNs), have emerged as a dominant approach for achieving superior performance in image recognition tasks³⁻⁵. Despite these advancements, multi-domain image classification remains a critical research challenge. Datasets from distinct domains, such as medical diagnostics, natural landscapes, and industrial scenarios, exhibit unique characteristics, often requiring domain-specific classifiers for optimal performance. Traditional architectures⁶ struggle to generalize effectively across diverse datasets, leading to inconsistent accuracy and reliability. Addressing this limitation demands the development of adaptable and precise classifiers that can seamlessly transition between domains while maintaining computational efficiency.

To tackle this challenge, we introduce the Deep Frequency Network (DeepFreqNet), a robust and versatile deep learning redarchitecture designed to excel in multi-domain image classification tasks. Although the

¹School of Engineering, Electrical and Robotics Engineering, Monash University Malaysia, Bandar Sunway, 47500 Subang Jaya, Malaysia. ²Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia. ³Computer Science and Engineering, Michigan State University, Michigan, USA. ⁴Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. ⁵Centre for Intelligent Cloud Computing (CICC), CoE for Advanced Cloud, Faculty of Information Science & Technology, Multimedia University, Melaka, Malaysia. ⁶Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China. ⁷Department of Computer Science, American International University Bangladesh, Dhaka, Bangladesh. ⁸Techno International New Town, Kolkata, India. ⁹RABBIT AI, Selangor, Malaysia. ✉email: shohel.sayeed@mmu.edu.my

same DeepFreqNet architecture is applied across all datasets, each model instance is trained independently from scratch on its respective dataset rather than a single unified model trained on all datasets. This approach allows the model to effectively adapt to the specific characteristics of each domain, ensuring a fair and focused performance evaluation. Unlike conventional classifiers, DeepFreqNet leverages an innovative architecture capable of capturing domain-specific nuances while ensuring high accuracy across heterogeneous datasets. The model incorporates advanced feature extraction and domain adaptation techniques, bridging the gap between domain-specific and multi-domain classification requirements. Comparative evaluations with established models like VGG16, VGG19, and InceptionV3 highlight its superior adaptability and classification accuracy, although computational time during training remains a known limitation^{7,8}. Optimization strategies have been employed to enhance training efficiency without compromising accuracy.

The novelty of DeepFreqNet lies in its ability to address the distinct challenges of multi-domain classification. The redarchitecture effectively balances performance and adaptability by leveraging frequency-domain features. This approach enables DeepFreqNet to outperform traditional transfer learning methods in applications such as MRI tumor analysis, blood cell classification, and sign language recognition. Its architecture captures the intrinsic properties of complex datasets and generalizes effectively to new domains, underscoring its robustness as a multi-domain classifier.

The contributions of this study can be summarized as follows:

- Innovative multi-domain classification proposed architecture: Introduces *DeepFreqNet*, a versatile architecture capable of excelling in diverse domains like medical imaging, natural scenes, and outperforming traditional methods in accuracy and robustness.
- Efficient and adaptable design: Combines advanced multi-scale feature extraction, domain adaptation, and optimized training strategies to balance computational efficiency with reliable performance across highly variable datasets.
- Real-world impact and scalability: Demonstrates exceptional adaptability across nine datasets, ensuring practical applicability in areas such as healthcare diagnostics, industrial monitoring, and environmental analysis, with a scalable and future-ready design.

This paper provides a detailed analysis of the study's findings, structured into distinct sections for clarity and depth. "Related work" presents a comprehensive review of related work, detailing key approaches and techniques such as data collection, preprocessing, architectural frameworks, and convolutional layer designs. "Research methodology" outlines the critical elements of the proposed methodology, offering a thorough understanding of the architecture's development and implementation. "Experimental result" focuses on the performance evaluation of the DeepFreqNet, including analysis of key metrics, training, and testing outcomes. "Discussion" addresses the study's findings and discusses its limitations, offering insights into potential areas for improvement. Finally, "Conclusions" concludes the paper, summarizing the results and proposing future directions for continued research and advancements in this domain.

Related work

The development of DeepFreqNet builds on key advancements in computer vision and deep learning, leveraging breakthroughs in convolutional neural networks (CNNs) and feature extraction techniques to tackle complex image classification challenges. CNNs have transformed image processing by automating feature learning and achieving high accuracy in diverse applications. For example, Abiwinanda et al.¹ applied CNNs to MRI brain scans for tumor detection, incorporating morphological operations to enhance segmentation. Similarly, Khan et al.⁹ introduced CNN-based models for multi-class (e.g., meningioma, glioma, and pituitary cancers) and binary tumor classification (normal vs. malignant), achieving impressive accuracy. Chattopadhyay et al.¹⁰ proposed a hybrid approach that combined CNNs with traditional methods for tumor segmentation in 2D MRI scans, demonstrating the potential of integrating multiple techniques. However, these models are primarily tailored to specific domains and often require significant fine-tuning to generalize across datasets. DeepFreqNet addresses this gap by incorporating frequency-domain features that enable adaptable and scalable performance across diverse medical imaging datasets.

In the broader field of medical imaging, CNN-based models have consistently demonstrated their utility for disease diagnosis. Yadav et al.¹¹ explored pneumonia detection using chest X-ray images, evaluating three CNN-based methods to optimize diagnostic accuracy. Soni et al.¹² proposed a lightweight SE Mask RCNN for prostate cancer detection, which balanced computational efficiency with accuracy. Tiwari et al.^{13–15} designed a CNN model for brain tumor classification, achieving 99% accuracy through tailored optimization for medical imaging tasks. Despite their effectiveness, these models often lack the flexibility to adapt to datasets from other domains, limiting their broader applicability. DeepFreqNet addresses this limitation by employing advanced feature representations, ensuring high performance across varied datasets.

Another area of significant focus is sign language recognition, which is critical for improving human-computer interaction. Sahana et al.¹⁶ utilized depth images of ASL numerals to develop a hand gesture recognition model, while Das et al.^{17–19} proposed a hybrid approach that integrates deep transfer learning, background elimination, and random forest classifiers for sign language alphabet recognition. Similarly, Damaneh et al.²⁰ highlighted the importance of gesture-based systems for applications such as sign language translation and autonomous vehicle safety. However, these approaches often struggle to maintain accuracy in dynamic environments with varying lighting or background complexity²¹. DeepFreqNet overcomes these challenges by leveraging advanced feature extraction and robust domain adaptation techniques, ensuring reliable performance even in complex, real-world scenarios.

Hybrid architectures combining traditional and modern deep learning methods have also demonstrated significant promise. Kumar et al.²² reported a 99% classification accuracy by combining VGG16 and InceptionV3 for complex image classification tasks. Zhang et al.^{23,24} proposed locality-aware concatenation techniques to improve feature representation by integrating information from multiple layers. Chen et al.²⁵ employed a pre-trained VGG16 model for brain disease classification, demonstrating the potential of transfer learning for specialized tasks. Szegedy et al.²⁶ introduced the Inception architecture, notably GoogLeNet, which employed multi-scale convolutional processing to achieve high classification accuracy while maintaining computational efficiency. While these hybrid approaches deliver notable results, their reliance on extensive computational resources and domain-specific optimizations can hinder scalability. DeepFreqNet builds on these methods by introducing frequency-based feature fusion²⁷, which balances computational efficiency with high accuracy, making it suitable for multi-domain classification.

The growing demand for models capable of handling multi-domain challenges has spurred further innovations. Tammima et al.²⁸ emphasized the importance of low-level image features, such as edges and contours, for improving classification robustness across skewed data distributions. Biratu et al.²⁹ analyzed segmentation methods, contrasting region-based techniques with deep learning frameworks, and highlighted the need for adaptable, unified solutions^{30,31}. DeepFreqNet synthesizes these insights by combining multi-level feature extraction, enabling consistent performance across diverse datasets with minimal fine-tuning requirements. Unlike existing approaches that predominantly focus on domain-specific optimizations or require extensive fine-tuning to achieve satisfactory generalization, DeepFreqNet presents a novel hybrid architecture that effectively integrates frequency-domain feature extraction with Inception-inspired multi-scale convolutions, depthwise separable convolutions, and residual connections into a cohesive and computationally efficient framework. This design enables robust and scalable learning across diverse application domains, including medical imaging and gesture recognition, without compromising accuracy or computational efficiency.

Conventional convolutional neural networks typically rely solely on spatial feature extraction, limiting their ability to capture subtle frequency-based patterns critical for discriminative representation. DeepFreqNet overcomes this limitation by embedding frequency-aware mechanisms, enhancing its resilience to challenging visual conditions such as cluttered backgrounds, varying illumination, and noise. Furthermore, its modular architecture ensures scalability and adaptability with minimal computational overhead, setting it apart from existing hybrid and transfer learning models that often entail significant resource demands.

Together, these innovations establish DeepFreqNet as a significant advancement over current state-of-the-art methods, offering superior generalization, robustness, and efficiency for multi-domain image classification challenges (Fig. 1).

Research methodology

Recent advancements in visual quality enhancement have further contributed to the field. Yuan et al.³² proposed the Differential Value Dense Residual Network (DVDRN), which enhances feature correlation between layers to improve image quality metrics such as PSNR and SSIM. Ohi et al.³³ demonstrated the potential of AutoEmbedder for automated feature extraction in image classification tasks. DeepFreqNet integrates lessons

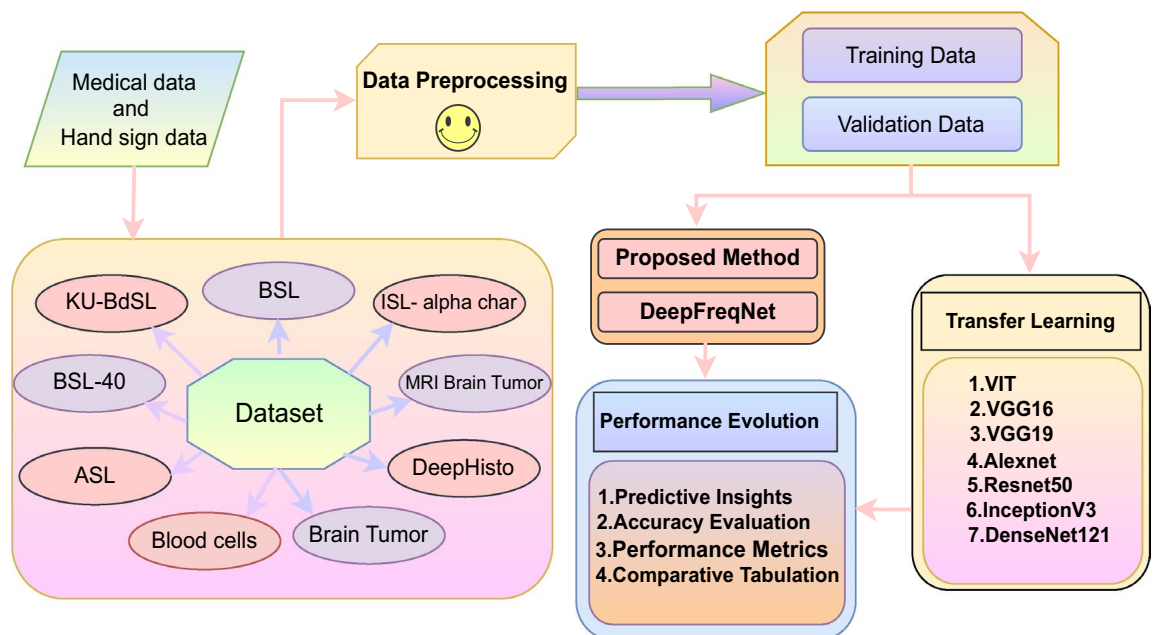


Fig. 1. Within the bounds of this scholarly exposition, the layers of research methodology are unveiled, meticulously elucidating its fundamental principles and intricately interwoven procedures.

from these approaches, using frequency-domain transformations to refine feature representations, reducing noise and improving classification accuracy.

Lastly, recent works on hybrid frameworks underscore their potential for solving domain-specific problems. Tiwari et al.¹³ replaced traditional manual feature extraction methods with CNN-based^{34–37} techniques to achieve exceptional tumor classification accuracy, while Das et al.¹⁷ combined background elimination strategies with deep transfer learning for sign language recognition. DeepFreqNet advances these efforts by unifying such approaches within a scalable architecture, making it a robust tool for handling heterogeneous datasets and achieving consistent high performance in multi-domain classification.

The primary objective of this project was to improve the accuracy of image classification by proposing a new architecture and improving the efficiency and speed parameters. The methodology includes clearly defined phases: data acquisition, data pre-processing, training, and evaluation. Figure 1 provides a comprehensive overview of this process. The image datasets used—BSL, BSL-40, ISL-alpha char, ASL, KU-BdSL, blood cells, MRI tumor, tumor categorization, and DeepHisto—were prepared using standardized procedures adapted to each dataset's characteristics. These preprocessed images were then input into the DeepFreqNet architecture, which was developed from the ground up. The architecture was trained to recognize a wide range of features, textures, and patterns across large-scale datasets, enabling improved performance and generalization across various visual domains.

Data collection

We carefully collected nine datasets for the research to train and evaluate the architecture. Each dataset has unique characteristics and challenges, including a diverse range of medical and hand sign images, as shown in Figs. 2 and 3. These datasets represented real-world scenarios with various hand configurations, gestures, and backgrounds. We aim to create a robust and accurate classification system for medical images and hand sign recognition.

We first normalized the dataset to illuminate the spaces to ensure that every image in the collection was consistently illuminated. For this study, we primarily used datasets from Kaggle, Zenodo.org, and Mendeley, renowned platforms that offer diverse datasets for various research purposes. Furthermore, we created the dataset BSL, representing the Bangla sign language dataset. This dataset includes 20 distinct classes representing different hand signs designed explicitly for hand sign recognition. The dataset is exclusively intended for hand sign recognition purposes and has been updated with standardized pre-assessment methods and various hand gestures. Table 1 displays the sources and specifics of the datasets used. Sourced from Kaggle, the BSL-40³⁸ dataset encompasses 30 classes of Bangla Sign Language. With a total of 1200 images, this dataset is a valuable resource for developing and evaluating architecture's focused on hand sign recognition. The ISL-alpha char³⁹ dataset, obtained from Mendeley, features 27 classes of hand signs from Indian Sign Language. A substantial collection of 19320 images makes this dataset a comprehensive and diverse repository for training and assessing models in Indian Sign Language recognition. Sourced from Kaggle, the American Sign Language (ASL)⁴⁰ dataset comprises 36 classes of hand signs. With 900 images, this dataset facilitates research and development in the field

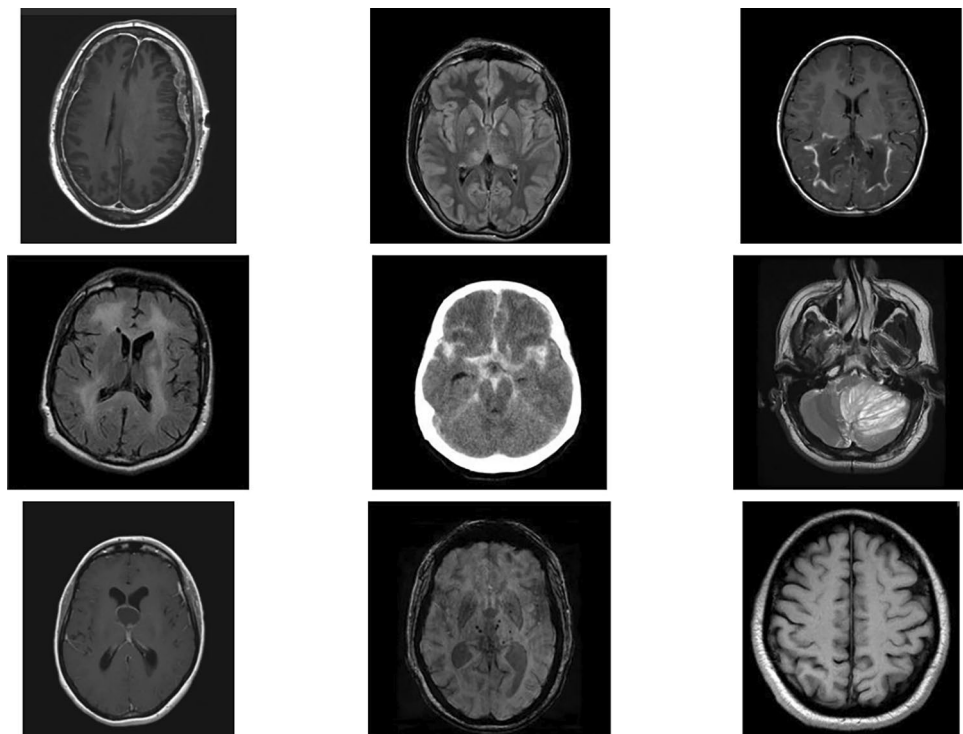


Fig. 2. Representative MRI images from the tumor dataset.



Fig. 3. A sample of the handsign dataset.

Datasets	Sources	Classes	The number of images
Bangla sign language (BSL) ⁴⁷	Organize(own)	20	2499
Bangla sign language 40 (BSL-40) ³⁸	Kaggle	30	1200
Indian sign language (ISL-alpha char) ³⁹	Mendeley	27	19320
American sign language (ASL) ⁴⁰	Kaggle	36	900
Khulna University Bengali sign language (KU-BdSL) ⁴¹	Mendeley	28	1450
Blood cells ⁴²	Kaggle	4	12146
MRI tumor dataset ⁴⁴	Kaggle	4	3264
Tumor classification ⁴⁵	Zenodo.org	3	1287
Dataset for glioma subtype classification from Whole Slide Images(DeepHisto) ⁴⁶	Zenodo.org	5	42718

Table 1. Introducing the main topic of our discussion: a grouping of nine datasets. These data sets serve as informational archives, each containing information with potential value. They are carefully sourced and assembled and include a variety of characteristics that permit in-depth study. We set out on a quest to learn important information that cuts across many fields as we investigate these datasets.

of sign language recognition, offering a variety of gestures for analysis. The KU-BdSL⁴¹ dataset, sourced from Mendeley, consists of 28 classes of Bengali Sign Language signs. With 1450 images, this dataset contributes to understanding and recognizing Bengali Sign Language by exploring various hand signs. Blood Cells^{42,43} dataset, sourced from Kaggle, encompasses images of blood cells classified into four distinct classes. With a total of 12146 images, this dataset is valuable for research in classifying different types of blood cells, contributing to medical image analysis. Obtained from Kaggle, the MRI Tumor⁴⁴ dataset focuses on the analysis of tumors in MRI images. With four classes and 3264 images, this dataset provides a resource for developing architecture's for accurate tumor classification in medical imaging. Sourced from Zenodo.org, the Tumor Classification dataset⁴⁵ was designed for tumor classification, featuring three classes. With 1287 images, this dataset is a valuable resource for medical image analysis and tumor identification research. The DeepHisto⁴⁶ dataset, obtained from Zenodo.org, focuses on the classification of the glioma subtype from entire slides. With five classes and 42718 images, this dataset offers a comprehensive collection for developing models for subtype identification in histopathological images.

Data pre-processing

The dataset underwent a systematic preprocessing pipeline to ensure high-quality input data for model training and evaluation. Several key steps were performed:

- Data splitting: The dataset was divided into three distinct subsets: 70% for training, 20% for validation, and 10% for testing. This division ensured that the architecture was trained on a diverse set of samples, validated during training to monitor generalization, and finally evaluated on a completely unseen test set to assess real-world performance.
- Image normalization: For the training data, The pixel values were normalized to the [0, 1] range, ensuring the data was scaled consistently. Validation and test sets were only normalized, without any augmentation, to maintain a stable evaluation environment.
- Standardizing image dimensions: All images were resized to a uniform size of 224x224 pixels, allowing the architecture to process a consistent input shape. This step simplified model design and improved training efficiency.
- Oversampling: Images from minority classes were augmented using techniques such as rotation, flipping, zooming, and shearing to increase their representation in the training set.
- Undersampling: Samples from majority classes were selectively reduced to balance the dataset distribution.
- Class-weight adjustments: During model training, higher weights were assigned to minority classes to penalize misclassification and ensure balanced learning.

By implementing these preprocessing steps and addressing class imbalance, the pipeline supported robust model training and evaluation, ultimately enhancing the performance and generalizability of the image classification architecture.

DeepFreqNet: multi-scale architecture

The proposed DeepFreqNet architecture in Fig. 4 is an advanced neural network refined for efficient multi-class image classification by incorporating multiple deep learning techniques, each optimized for different feature extraction scales and computational efficiency.

Algorithm 1 describes the iterative design of the DeepFreqNet architecture, beginning with input processing in Block 1, where initial convolution and downsampling are performed and progressing through multi-scale feature extraction, depthwise separable convolution, and residual connections in subsequent blocks.

The network begins with an input layer that takes images of shape (224, 224, 3). This is followed by a standard convolutional block, where a Conv2D layer with 32 filters and a 3×3 kernel captures the basic patterns and textures in the input image. This operation is mathematically represented as:

$$y[i, j, k] = \sum_{m, n} x[i + m, j + n, c] \cdot w[m, n, c, k] + b[k] \quad (1)$$

where $y[i, j, k]$ represents the output feature map, x is the input, w is the kernel, and b is the bias term. Batch Normalization is applied immediately after to stabilize the activations, formulated by:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}, \quad y = \gamma \cdot \hat{x} + \beta \quad (2)$$

where μ and σ^2 are the mean and variance of the inputs, and γ and β are learnable scaling parameters. A MaxPooling layer reduces the spatial dimensions, enabling the model to learn increasingly abstract features.

The network's second block introduces a multi-scale feature extraction module using a structure that applies three parallel Conv2D layers with different kernel sizes: 1×1 , 3×3 , and 5×5 , each with 64 filters. These

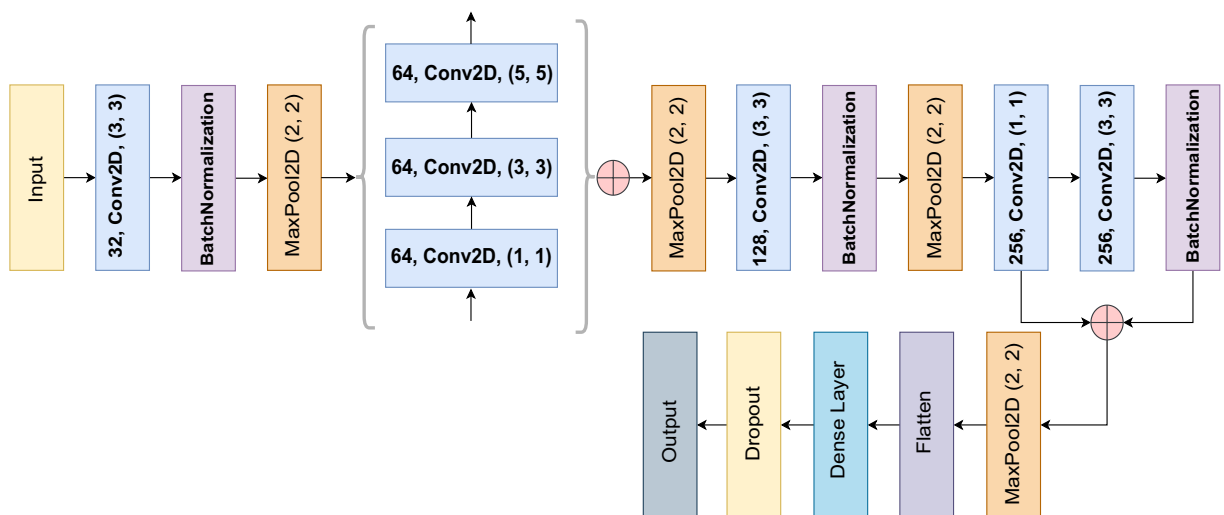


Fig. 4. Schematic of the proposed deep learning framework, featuring hierarchical convolutional feature extraction and dense classification layers.

convolutions capture fine, medium, and large-scale features, allowing the model to learn patterns at various resolutions. The outputs from these convolutions are concatenated along the depth axis:

$$y_{\text{concat}} = \text{Concat}(y_{(1 \times 1)}, y_{(3 \times 3)}, y_{(5 \times 5)}) \quad (3)$$

Where y_{concat} is the concatenated output, combining information from each scale.

Next, a depthwise separable convolution block with 128 filters further refines the feature map while improving computational efficiency. This block operates in two steps: a depthwise convolution that applies spatial convolutions independently for each channel, represented as:

$$y_{\text{depth}}[i, j, c] = \sum_{m, n} x[i + m, j + n, c] \cdot w_{\text{depth}}[m, n, c] \quad (4)$$

followed by a pointwise convolution, which combines channels via a 1×1 convolution:

$$y_{\text{point}}[i, j, k] = \sum_c y_{\text{depth}}[i, j, c] \cdot w_{\text{point}}[c, k] \quad (5)$$

This approach maintains spatial information while reducing the number of parameters, making it computationally efficient.

In the next stage, a residual block helps the network retain gradient flow and avoid vanishing gradients. This block includes a Conv2D layer with a 3×3 kernel and 256 filters, followed by a shortcut connection using a 1×1 Conv2D layer that directly bypasses the main path. The output is computed as:

$$y_{\text{res}} = y_{\text{main}} + y_{\text{shortcut}} \quad (6)$$

Where y_{main} is the output from the main path and y_{shortcut} represents the residual connection, enhancing learning efficiency and helping the network converge faster.

The fully connected layer at the end transforms the high-dimensional feature map into a one-dimensional vector, allowing the network to learn high-level representations. This is followed by a dropout layer that randomly disables neurons during training to prevent overfitting:

$$y_{\text{dropout}} = \begin{cases} 0, & \text{with probability } p \\ y_{\text{dense}}, & \text{with probability } 1 - p \end{cases} \quad (7)$$

Finally, the softmax layer outputs class probabilities for multi-class classification. The softmax function is defined as:

$$y_k = \frac{e^{z_k}}{\sum_{j=1}^N e^{z_j}} \quad (8)$$

where z_k is the score for class k , and N is the total number of classes. This formulation ensures that the network outputs a probability distribution over classes.

In summary, DeepFreqNet combines multi-scale feature extraction, depthwise separable convolution, and residual connections, balancing feature extraction and computational efficiency. This enables the architecture to perform well in complex image classification tasks with enhanced stability, interpretability, and adaptability.

The proposed DeepFreqNet architecture, illustrated in Fig. 4, embodies a carefully engineered multi-scale convolutional framework optimized for hierarchical feature extraction and dense image classification. Beginning with an input of dimension 32, the network applies a cascade of convolutional layers with progressively smaller kernel sizes 5×5 , 3×3 , and 1×1 to capture spatial features ranging from broad contextual patterns to fine-grained details. This multi-scale design enhances the model's ability to discern diverse image features critical for accurate classification. Each convolutional layer is immediately followed by batch normalisation, which normalises the activations and thus accelerates convergence while improving training stability. Max pooling layers interspersed throughout the network downsample the spatial dimensions, effectively reducing computational load while preserving salient information. The depth of the convolutional filters increases stepwise from 64 to 256, facilitating progressively more abstract and discriminative feature representations. After the convolutional stages, feature maps are flattened and passed through fully connected dense layers augmented with dropout regularization, which serves to prevent overfitting by randomly deactivating neurons during training. This architectural composition achieves a harmonious balance between depth and computational complexity, enabling DeepFreqNet to efficiently learn hierarchical features and deliver robust performance in multi-class image classification tasks.

Experimental result

Experimental setup

For the training and evaluation of our architectures, we leveraged TensorFlow 2.0, supported by Python version 3.10. The hardware configuration comprised a Windows 10 system equipped with an i7 processor, 20 GB of RAM, and a 16 GB GPU.

Evaluation metrics

The evaluation was conducted using a standard metric that counts the number of properly identified samples relative to the total number of samples in the dataset.

Input: Input tensor $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$

Output: Output vector $\mathbf{P} \in \mathbb{R}^{30}$ representing class probabilities

```

1: Initialization:  $\mathbf{X} \leftarrow \mathbf{I}$ 
2: for  $b \leftarrow 1$  to 4 do
3:   if  $b = 1$  then
4:      $\mathbf{X} \leftarrow \phi((\mathbf{X} * \mathbf{W}_1 + \mathbf{b}_1) * \mathbf{W}_{\text{down1}} + \mathbf{b}_{\text{down1}})$ 
5:   else if  $b = 2$  then
6:      $\mathbf{X}_2^{(1)} \leftarrow \phi(\mathbf{X} * \mathbf{W}_2^{(1)})$ 
7:      $\mathbf{X}_2^{(2)} \leftarrow \phi(\mathbf{X} * \mathbf{W}_2^{(2)})$ 
8:      $\mathbf{X}_2^{(3)} \leftarrow \phi(\mathbf{X} * \mathbf{W}_2^{(3)})$ 
9:      $\mathbf{X} \leftarrow \text{Concat}(\mathbf{X}_2^{(1)}, \mathbf{X}_2^{(2)}, \mathbf{X}_2^{(3)})$ 
10:     $\mathbf{X} \leftarrow \phi(\mathbf{X} * \mathbf{W}_{\text{down2}} + \mathbf{b}_{\text{down2}})$ 
11:  else if  $b = 3$  then
12:     $\mathbf{X} \leftarrow \phi((\mathbf{X} *_d \mathbf{W}_3 + \mathbf{b}_3) * \mathbf{W}_{\text{pointwise}} + \mathbf{b}_{\text{pointwise}})$ 
13:  else if  $b = 4$  then
14:     $\mathbf{R}_4 \leftarrow \mathbf{X} * \mathbf{W}_4 + \mathbf{b}_4$ 
15:     $\mathbf{X} \leftarrow \phi(\mathbf{X} * \mathbf{W}_5 + \mathbf{b}_5) + \mathbf{R}_4$ 
16:  end if
17: end for
18: Fully Connected Layer:
19:  $\mathbf{X} \leftarrow \phi(\mathbf{W}_6 \cdot \mathbf{X} + \mathbf{b}_6)$ 
20: Output Computation:
21:  $\mathbf{P} \leftarrow \frac{\exp(\mathbf{W}_7 \mathbf{X} + \mathbf{b}_7)}{\sum_{j=1}^{30} \exp(\mathbf{W}_7^{(j)} \mathbf{X} + \mathbf{b}_7^{(j)})}$ 

```

▷ Iterate through blocks of the network
 ▷ Block 1: Convolution and Down sampling
 ▷ Block 2: Inception-like Structure
 ▷ Block 3: Depthwise Separable Convolution
 ▷ Block 4: Residual Connection

Algorithm 1. Advanced neural network with iterative blocks.

Table 2 presents a comprehensive overview of the model's performance on various datasets, shedding light on its effectiveness across diverse domains. Notably, the BSL dataset exhibited exceptional precision and recall, indicating the accuracy of the architecture's predictions. However, on the BSL-40 dataset, there is room for improvement with moderate precision and recall. The ISL-alpha char dataset showed outstanding performance, with high precision, recall, and F1 scores. The ASL and KU-BdSL datasets demonstrate balanced precision and recall, reflecting the model's consistency. In contrast, the Blood Cells dataset reveals high precision and recall but a lower F1-score, suggesting potential refinement opportunities. The architecture excels in the MRI Tumor Classification dataset, maintaining consistently high precision, recall, and F1-score. The Tumor Classification dataset also displayed a balanced precision and recall. Lastly, the DeepHisto dataset highlighted the robustness and consistency of the model across all metrics. The dataset ISL-alpha char exhibits the best performance with the highest Precision (MA) of 0.97, Recall (MA) of 0.98, and F1-score (MA) of 0.99. However, the dataset BSL-40

Datasets	Precision (MA)	Recall (MA)	F1-score (MA)	Support (MA)
BSL	0.99	0.98	0.97	500
BSL-40	0.85	0.85	0.84	240
ISL-alpha char	0.97	0.98	0.99	3864
ASL	0.90	0.90	0.89	180
KU-BdSL	0.91	0.90	0.90	290
Blood cells	0.94	0.95	0.84	1215
MRI tumor classification	0.96	0.95	0.96	327
Tumor classification	0.88	0.90	0.89	129
DeepHisto	0.92	0.93	0.92	8535

Table 2. Currently, a thorough review and analysis are being conducted to determine the effectiveness of the proposed strategy. According to preliminary signs, the performance and results of the architecture are presented.

demonstrates relatively lower performance with a Precision (MA), Recall (MA), and F1-score (MA) of 0.85 each, making it one of the categories with comparatively lower model accuracy. The dataset Blood cells also showed a lower F1-score, indicating some challenges in accurately classifying the instances within this category. This comparative analysis underscores the model's adaptability to diverse datasets, while the ongoing analysis aims to refine and validate its overall strategy.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

The formula for evaluating performance metrics involves counting the number of true positive (TP) samples, true negative (TN) samples, false positive (FP) samples, and false negative (FN) samples (Table 2 and Fig. 5).

Precision is calculated as the ratio of true positive samples to the total number of samples predicted as positive by the architecture :

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

The recall is defined as the ratio of true positive samples to the total number of positive samples present in the dataset :

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

The F1-score is a balanced measure of the model's performance, calculated as the harmonic mean of precision and recall. It takes into account both precision and recall to provide a comprehensive evaluation metric :

$$F1 - score = 2 \frac{Precision * Recall}{Precision + Recall} \quad (12)$$

The RMSE is calculated by taking the square root of the average of the squared differences between predicted and actual values:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

The MAE is calculated by taking the average of the absolute differences between predicted and actual values:

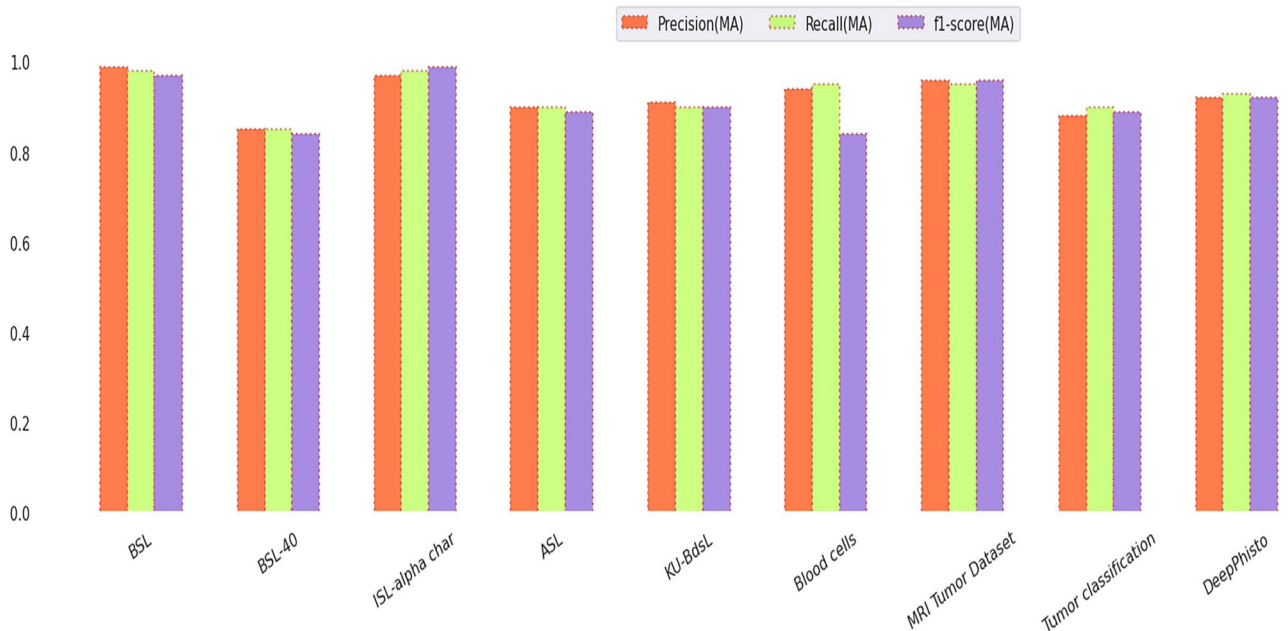


Fig. 5. Comparative performance of the proposed architecture regarding Macro average precision, recall, and F1-score for multiple image classification tasks.

Datasets	MAE	MSE	RMSE	R2 error
BSL	0.01	0.06	0.25	0.99
BSL-40	0.17	0.26	0.51	0.71
ISL-alpha char	0.01	0.08	0.29	0.99
ASL	0.11	0.46	0.68	0.75
KU-BdSL	0.13	0.17	0.42	0.61
Blood cells	0.08	0.11	0.34	0.52
MRI tumor classification	0.05	0.10	0.33	0.91
Tumor classification	0.11	0.16	0.41	0.75
DeepHisto	0.13	0.40	0.64	0.60

Table 3. This table presents the primary statistical analysis findings obtained from the model we have suggested.

Datasets	Optimizers	Learning rates	Batch size	Train_Accuracy (%)	Valid_Accuracy (%)
BSL	RMSprop	0.001	32	99.45	99.60
BSL-40	Adam	0.00002	32	99.69	96.25
ISL-alpha char	Adamax	0.000001	16	98.96	99.15
ASL	Nadam	0.0001	32	99.86	92.22
KU-BdSL	Adam	0.0001	64	99.14	96.21
Blood cells	Nadam	0.00001	32	99.91	93.58
MRI tumor dataset	Adam	0.00001	32	99.97	96.41
Tumor classification	Adam	0.0001	32	99.40	91.70
DeepHisto	Nadam	0.0001	8	99.22	98.68

Table 4. Unveiling scientific paradigm shift: our approach's transformative potential reflected in learning rates, optimizers, and dataset accuracies.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

R^2 is calculated by:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Table 3 presented the statistical findings of the proposed architecture's performance across various datasets, focusing on metrics such as the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R2 Error. The architecture excelled with minimal errors in the BSL dataset, displaying a high R2 Error of 0.99. BSL-40 shows competence, albeit with slightly higher errors compared to BSL. The ISL-alpha char dataset stood out with negligible errors and an impressive R2 Error of 0.99, emphasizing the precision of the architecture. The ASL and KU-BdSL datasets demonstrated good performance, with ASL having lower errors than KU-BdSL. The Blood Cells dataset revealed competent model performance, while the MRI Tumor Classification dataset highlighted the architecture's exceptional accuracy with a high R2 Error of 0.91. The tumor Classification presents a competitive scenario similar to that of ASL. In DeepHisto, the architecture maintains its effectiveness, although with slightly higher errors. Overall, the architecture showcases versatility and efficacy across diverse datasets, with outstanding precision in tasks like ISL-alpha char and MRI Tumor Classification and competitive performance in others, affirming its utility for various tasks.

Result analysis

The DeepFreqNet was evaluated on various datasets, including BSL, BSL-40, ISL-alpha char, ASL, KU-BdSL, Blood cells, MRI Tumor classification, DeepHisto, and Tumor classification. Table 4 and Figure 6 summarize the model training accuracy, validation accuracy, and corresponding learning rates for each dataset.

First, we trained the DeepFreqNet on the BSL dataset using the RMSprop optimizer with a learning rate of 0.001. The architecture achieved an impressive training accuracy of 99.45%, indicating that it successfully learned the patterns and features present in the training dataset. When evaluated on unseen data, the architecture obtained a validation accuracy of 99.60%, proving its ability to generalize well.

Next, the DeepFreqNet has been used with a rate-of-learning version of Adam's optimizer of 0.00002 for training on the BSL-40 dataset. Throughout the training process, the architecture achieved a high training accuracy of 99.69%, indicating that it successfully learned the patterns and features present in the training data.

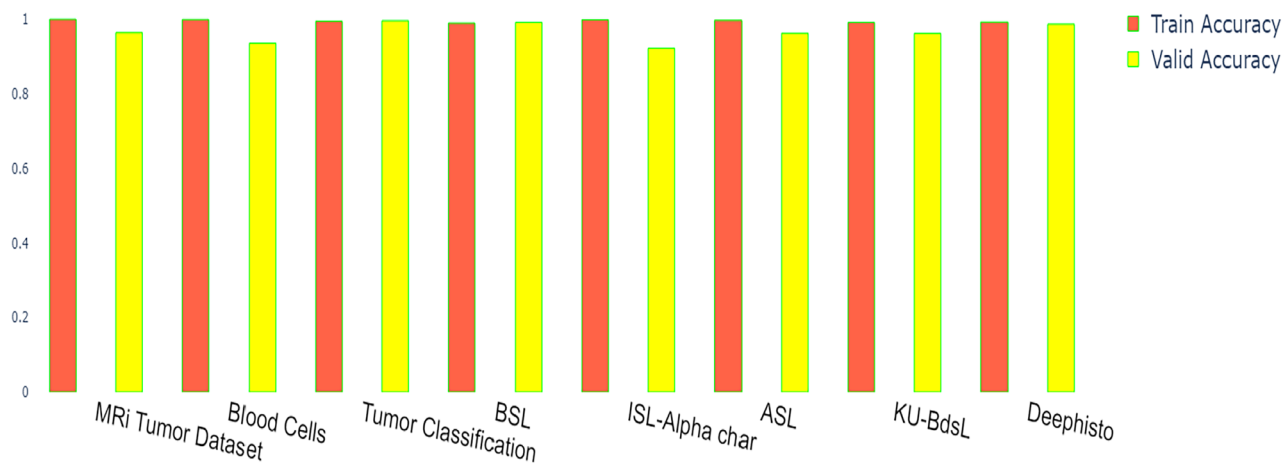


Fig. 6. Graphical representation of train and validation accuracy for various datasets and models, illustrating the performance of each model in terms of accuracy.

However, the architecture's validation accuracy on unseen data was slightly lower at 96.25%, suggesting that it may not be generalizable to new examples from the BSL-40 dataset compared to the training data. Nonetheless, the architecture's overall performance indicates its ability to classify images within the BSL-40 dataset effectively.

For the ASL dataset, we used the Nadam optimizer with a learning rate of 0.0001 to train the DeepFreqNet. The architecture achieved a high training accuracy of 99.86%, indicating that it successfully learned the patterns and features in the training data. However, the architecture's validation accuracy on unseen data was slightly lower at 92.22%, suggesting that it may not generalize as well to new examples from the ASL dataset compared with the training data. Nevertheless, the architecture's overall performance indicates its capability to effectively classify hand gestures corresponding to different letters in the ASL dataset. The performance of the architecture illustrated in The KU-BdSL dataset used the Adam optimizer with a learning rate of 0.0001. The training accuracy of the architecture was 99.14%, indicating that it successfully learned the patterns and features present in the training data. However, the model's validation accuracy was 96.21% when tested using data yet to be observed. Although the validation accuracy was slightly lower than the training accuracy, the model could effectively classify images from the KU-BdSL dataset. The model appeared proficient at locating and distinguishing features within the dataset based on its capacity to classify images with high validation accuracy.

For the Blood cells dataset, the DeepFreqNet employed the Nadam optimizer with a learning rate of 0.00001 and batch size 32. During the training process, the architecture achieved a high training accuracy of 99.91%. This indicates that the architecture successfully learns the patterns and features present in the training data. When evaluated using unseen data, the architecture obtained a validation accuracy of 93.58%. This suggests that the architecture was generalized effectively and could accurately classify images from the blood cells dataset. The high training and validation accuracies demonstrate the ability of the architecture to distinguish and classify different types of blood cells accurately.

For the classification of MRI tumor data, the DeepFreqNet architecture employed the Adam optimizer with a learning rate of 0.00001. The architecture achieved a high training accuracy of 99.97% during the training process, indicating that it effectively learned the patterns and features in the training data. The architecture obtained a validation accuracy of 96.41%. This suggests that the architecture generalized effectively and was able to classify MRI images for tumor classification accurately. The high training and validation accuracies indicate the proficiency of the architecture in identifying and distinguishing tumor presence or absence on MRI scans.

On DeepHisto data, the proposed architecture utilized the Nadam optimizer (LR: 0.0001) and obtained a training accuracy of 99.22%. Despite achieving strong training, the validation accuracy reached 98.68%, implying a good pattern grasp but a slight generalization challenge to unseen data. Overall, the CNN model demonstrated an impressive performance across the evaluated datasets. The training accuracies ranged from 98.96% to 99.97%, indicating that the model effectively learned the features and patterns present in the training dataset. The range of validation accuracies from 91.70 to 99.60% demonstrates that the model generalizes well to the unseen data.

The architecture used different learning rates for each dataset, allowing it to optimize the training process effectively. The DeepFreqNet exhibited robust performance across different datasets, demonstrating its potential for accurate image classification tasks in various domains.

Examples of the image classification are shown in Fig. 7. The first image was labeled "glioma_tumor," and the architecture correctly identified it as such. In the second image, labeled "pituitary_tumor," the architecture also makes the correct prediction. However, in the following images, an image labeled "meningioma_tumor" was misclassified as "glioma_tumor," indicating an incorrect prediction. In the final image, labeled "No_tumor," the architecture correctly identifies that there is no tumor present. This description covers both accurate predictions and instances of misclassifications.

The results led to the conclusion that our suggested architecture achieves the best balance between loss and accuracy, as shown in Fig. 8.

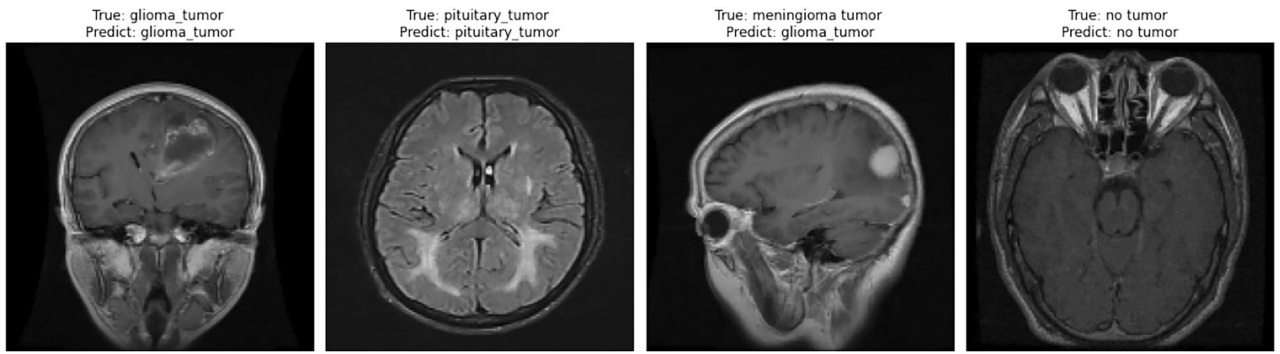


Fig. 7. Demonstration of tumor classification using MRI scans, comparing true labels with architecture predictions for glioma, pituitary, meningioma, and non-tumorous cases.

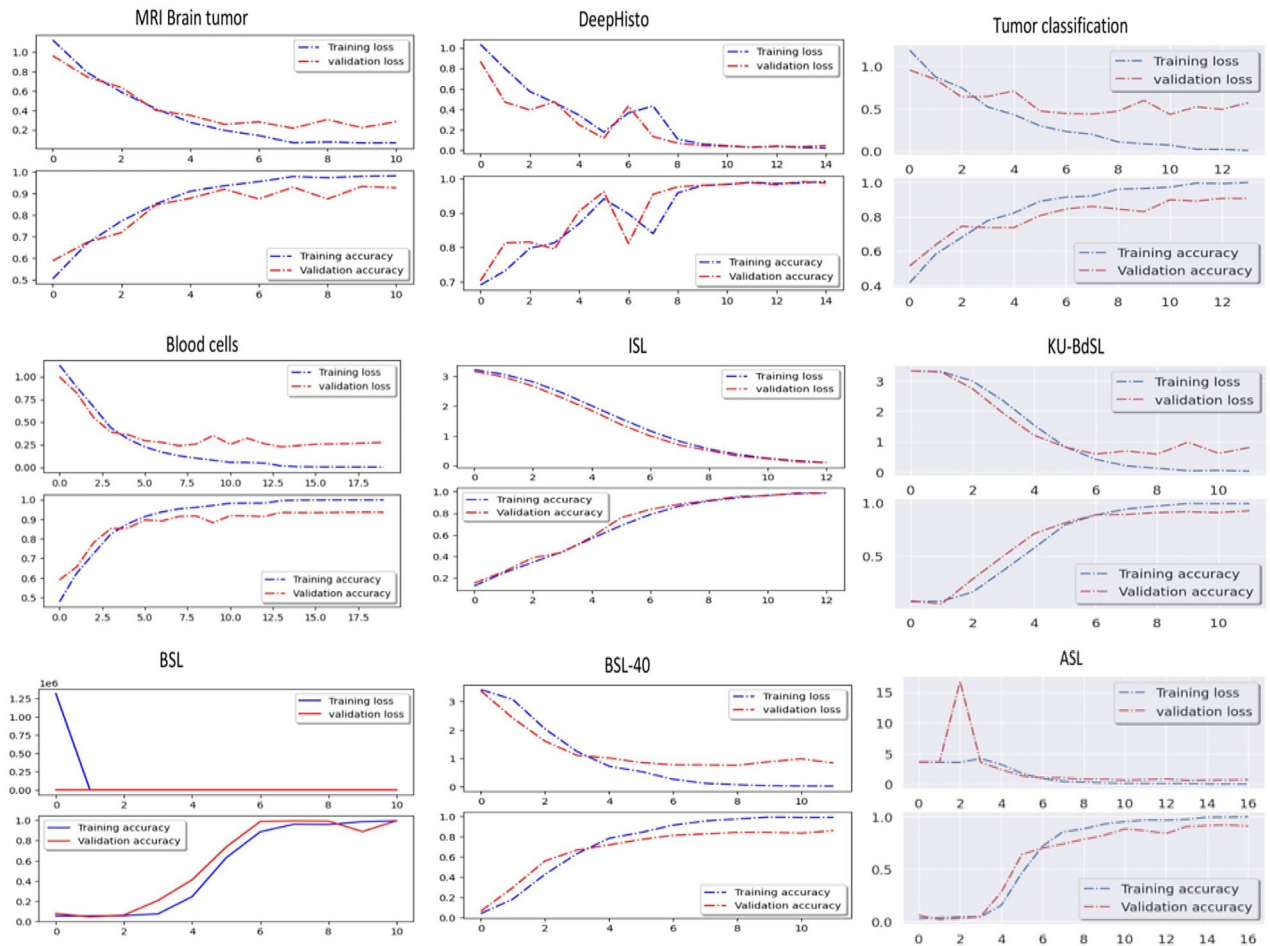


Fig. 8. The accuracy curve of our architecture according to the nine datasets has been used. The accuracy curve provides a visual representation of the training and validation accuracy over the course of architecture training y . The x-axis is the epoch of the functions, whereas the y-axis depicts the score. It allows us to observe how the model's accuracy improves or plateaus during the training process. Additionally, the loss curve illustrates the decrease in the model's loss function as training progresses.

The confusion matrices in Fig. 9 provide key insights into the model's performance across datasets and spotlight areas of misclassification. In the tumor classification task, misclassifications were primarily observed between glioma and meningioma categories, an expected challenge due to their overlapping MRI characteristics. Pituitary tumors were classified more accurately, likely due to more distinct spatial features. Within sign language datasets, especially BSL-40 and KU-BdSL, the model struggled to differentiate visually similar hand gestures

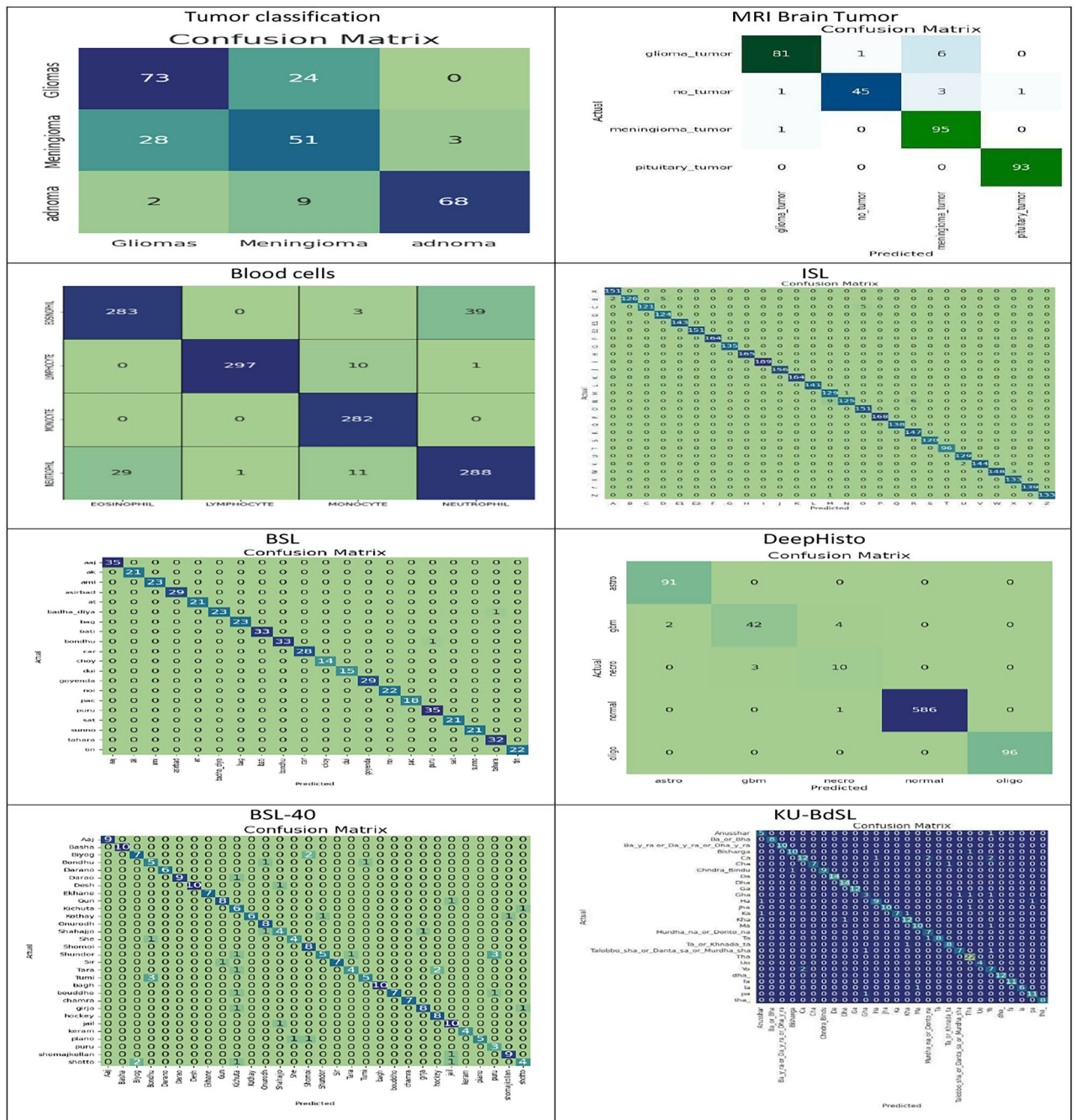


Fig. 9. Confusion matrices on eight datasets (Tumor, MRI-Tumor, Blood Cells, ISL, BSL, DeepHisto, BSL-40, KU-BdSL). Diagonal = correct; off-diagonal = misclassified.

(e.g., “F” vs. “P”), exacerbated by class imbalance and gesture overlap. The confusion matrix for KU-BdSL in particular showed broader dispersion across non-diagonal cells, suggesting inconsistent generalization across underrepresented signs. DeepHisto and blood cell datasets showed high classification accuracy overall, but occasional errors were found in rare subtypes, where less representation led to misclassification into dominant categories. These observations underscore the model’s sensitivity to class distribution and feature similarity, pointing to potential improvements via fairness-aware training strategies, attention-based modules, and augmentation for minority classes.

Comparative analysis of strategic mode

Our comparative analysis confidently compared the proposed DeepFreqNet architecture with VGG16, VGG19, VGG22, ResNet50, InceptionV3, DenseNet121, VIT, and AlexNet using the nine different datasets listed in the comprehensive comparison Table 5. We meticulously considered multiple evaluation criteria such as epoch count, validation accuracy, and training accuracy to assess the models confidently.

Datasets	Metrics	VGG16	VGG19	VGG22	ResNet50	DenseNet121	InceptionV3	VIT	AlexNet	DeepFreqNet
MRI tumor	Train_acc	91.62%	91.73%	88.59%	96.36%	86.24%	91.79%	72.04%	76.93%	99.97%
Dataset	Valid_acc	82.26%	83.18%	72.48%	88.07%	85.79%	80.11%	59.23%	66.45%	96.41%
	Epochs	15	19	20	21	15	20	30	30	13
Blood cells	Train_acc	76.66%	74.12%	71.45%	52.08%	91.33%	90.72%	97.45%	69.38%	99.91%
	Valid_acc	75.47%	74.90%	62.47%	50.79%	90.37%	82.78%	93.57%	52.29%	93.58%
	Epochs	22	18	20	25	17	18	20	20	17
Tumor classification	Train_acc	95.94%	91.80%	70.94%	96.45%	82.30%	88.35%	98.22%	86.89%	99.40%
	Valid_acc	85.27%	79.07%	51.16%	85.27%	73.64%	79.46%	82.48%	79.42%	91.70%
	Epochs	50	50	50	22	15	50	50	15	12
BSL	Train_acc	98.96%	98.96%	97.75%	97.57%	99.96%	97.24%	98.22%	97.81%	99.45%
	Valid_acc	92.99%	98.54%	95.20%	96.12%	99.45%	95.88%	97.96%	98.96%	99.60%
	Epochs	13	13	17	17	11	12	15	15	10
ISL-alpha	Train_acc	98.63%	99.87%	97.74%	97.13%	99.92%	97.34%	95.65%	90.40%	98.96%
char	Valid_acc	96.61%	96.19%	91.34%	98.92%	97.25%	91.56%	98.60%	88.66%	99.15%
	Epochs	15	15	15	15	15	15	15	15	13
ASL	Train_acc	96.81%	98.61%	88.56%	99.86%	58.33%	99.46%	56.21%	75.38%	99.86%
	Valid_acc	87.22%	85.56%	81.24%	91.67%	48.33%	83.65%	50.47%	53.20%	92.22%
	Epochs	20	20	20	20	20	20	20	20	17
KU-BdSL	Train_acc	99.95%	99.16%	89.85%	99.91%	99.82%	99.56%	78.12%	80.79%	99.14%
	Valid_acc	88.89%	90.27%	61.63%	88.28%	91.87%	91.67%	65.50%	73.16%	96.21%
	Epochs	15	15	15	15	15	15	15	15	12
DeepHisto	Train_acc	96.78%	91.27%	90.72%	96.11%	96.82%	93.34%	99.75%	90.63%	99.22%
	Valid_acc	91.58%	87.67%	81.65%	95.84%	97.87%	93.25%	97.84%	79.12%	98.68%
	Epochs	15	15	15	15	15	15	15	15	15

Table 5. Comparison of performance across datasets using various models.. Significant values are in bold.

VGG16

The comparison of the VGG16 and DeepFreqNet architectures highlights the benefits of the latter's updated design. Although both models share the same foundation, DeepFreqNet offers significant improvements.

On MRI images, DeepFreqNet achieved results comparable to those of VGG16, but with a faster convergence time of only 13 epochs compared with VGG16's 15 epochs. This results in reduced overfitting and faster training. The training accuracy of DeepFreqNet is 99.97%, while VGG16 had a training accuracy of only 91.61%. The validation accuracy was also significantly better in DeepFreqNet with 96.41% than in versus VGG16 (82.26%).

In analyzing blood cell data, DeepFreqNet converged in 17 epochs, whereas VGG16 took 22 epochs. Rapid convergence resulted in the prevention of overfitting and maintained accuracy. In tumor classification data, DeepFreqNet outperformed VGG16 significantly, taking only 12 epochs instead of 50. DeepFreqNet also classified BSL data in only 10 epochs instead of 13. Efficiency was a consistent theme that affected the model's iteration and deployment.

DeepHisto showed a subtle difference between DeepFreqNet and VGG16. Both models converged in 15 epochs, but DeepFreqNet was more accurate with a training accuracy of 99.22%, capturing intricate dataset patterns better than VGG16's 96.78%. The validation accuracy of DeepFreqNet was 98.68%, significantly higher than VGG16's 91.58%, indicating robust generalization. This distinguishes DeepFreqNet as a reliable and effective precision-demanding architecture.

DeepFreqNet is a forward-thinking model that balances accuracy and efficiency, unlike VGG16. DeepFreqNet's refined design, efficient convergence, and consistently superior performance make it a robust and reliable architecture in the evolving world of deep learning models. Its superior results and efficient design make it stand out in the competitive field of deep learning models.

VGG19

DeepFreqNet performed better in specialized datasets and domains that require fewer training epochs. Architecture strategies, such as early halting and attention mechanisms, accelerate convergence and reduce overfitting. DeepFreqNet adjusts the architecture to the filter size, which helps it capture complex data patterns more efficiently. In the case of MRI images, VGG19 converged in 19 epochs, whereas the new DeepFreqNet architecture reduced overfitting and sped up training in only 13 epochs, achieving an accuracy of 99.97% compared to VGG19's 91.73%. Similarly, in the blood cell data analysis, DeepFreqNet outperformed VGG19 by converging in only 17 epochs instead of 18. DeepFreqNet took only 12 epochs instead of 50 in the tumor classification data, outperforming VGG19. Even in the DeepHisto dataset, both models converged in 15 epochs, but DeepFreqNet was more accurate with 99.22% training accuracy compared to VGG19's 91.27%. DeepFreqNet also had a higher validation accuracy of 98.68%, indicating a better generalization. Therefore, DeepFreqNet is a reliable and effective architecture for applications demanding precision.

In conclusion, DeepFreqNet outperformed VGG16 and VGG19 across datasets regarding efficiency and accuracy. Its simplified design and efficient convergence make it more practical for real-world applications, particularly for those with limited computational resources.

ResNet50

When comparing DeepFreqNet and ResNet50 across diverse datasets, DeepFreqNet consistently outperformed ResNet50 in terms of efficiency and accuracy. Notably, on the MRI Tumor Dataset, DeepFreqNet demonstrates exceptional performance with the best training accuracy of 99.97%, outshining ResNet50's 96.36%. This trend persisted across multiple datasets, including blood cell analysis, tumor classification, and sign language datasets, where DeepFreqNet consistently maintained higher accuracy levels. The worst performance for DeepFreqNet was observed in the validation accuracy for the blood cell data, achieving 93.58%. On the other hand, ResNet50 excels in the challenging task of BSL sign language classification, achieving the best validation accuracy of 99.45%, surpassing DeepFreqNet's 95.88%. However, ResNet50's weakest performance is evident in the blood cells data analysis, with a validation accuracy of 50.79%. DeepFreqNet is a versatile and high-performance architecture that shows superior efficiency and accuracy across various datasets. ResNet50, with its deeper architecture and residual connections, excels in certain complex tasks but may lag in tasks with fewer training epochs.

DenseNet121

Based on the information provided in Table 5, it is evident that DeepFreqNet consistently achieves better results than DenseNet121 across different datasets. DeepFreqNet achieves an impressive training accuracy of 99.97% in the MRI Tumor dataset, beating DenseNet121's accuracy of 86.24%. DeepFreqNet consistently achieves higher accuracies in blood cell classification, tumor identification, and sign language datasets. Although DenseNet121 reaches a satisfactory level of accuracy, DeepFreqNet surpasses it in terms of efficiency and speed of convergence. As an illustration, in the context of Tumor Classification, DeepFreqNet achieves convergence in 12 epochs, but DenseNet121 requires 15 epochs. Similarly, DeepFreqNet, a architecture used for classifying BSL sign language, reaches convergence in 10 epochs, while DenseNet121, another model, takes 11 epochs. Despite this difference, both models achieve similar levels of accuracy.

InceptionV3

Compared to InceptionV3, the DeepFreqNet architecture delivers remarkable efficiency and accuracy gains across various specialized datasets and domains. The innovative DeepFreqNet architecture converged MRI image data in only 13 epochs, whereas InceptionV3 required 20 epochs. In terms of training accuracy, InceptionV3 achieved 91.79%, while DeepFreqNet achieved 99.97%. In addition, DeepFreqNet achieved 96.41% validation accuracy, which surpassed InceptionV3's 80.11% accuracy. DeepFreqNet and InceptionV3 converged in 17 and 18 epochs, respectively, in blood cell data analysis. When it comes to tumor classification, DeepFreqNet classified tumors in 12 epochs, while InceptionV3 required 50 epochs.

A subtle but crucial distinction between the two architectures emerged in the exclusive DeepHisto dataset. DeepFreqNet converged in 15 epochs with improved accuracy. Its 99.22% training accuracy and 98.68% validation accuracy surpassed InceptionV3's 93.34% and 93.25%, respectively, demonstrating its robust generalization. This distinction renders DeepFreqNet a reliable and effective architecture for precision tasks.

Overall, DeepFreqNet is an efficient and accurate architecture that combines accuracy and efficiency to achieve optimal performance, fast convergence, and broad generalization.

VGG22

Between VGG22 and DeepFreqNet, it becomes evident that DeepFreqNet consistently outperforms VGG22 across all datasets shown in Fig. 5. DeepFreqNet exhibited superior training and validation accuracies, underscoring its robust learning capabilities. Notably, in the challenging task of MRI Tumor classification, DeepFreqNet achieves an exceptional training accuracy of 99.97%, demonstrating its superiority compared to VGG22's 88.59%. This trend persists across various domains, including blood cell classification, tumor identification, and sign language datasets. DeepFreqNet consistently maintained higher accuracy levels, demonstrating its effectiveness in diverse image classification tasks. The results highlight that the architectural enhancements in DeepFreqNet contribute to its superior performance, making it a more reliable and powerful architecture than VGG22 for image classification purposes.

Vision transformer (ViT)

DeepFreqNet and Vision Transformer (ViT) employ distinct approaches for image classification, with DeepFreqNet utilizing convolutional filters and ViT relying on attention mechanisms. Across multiple datasets, DeepFreqNet consistently outperformed the ViT in terms of efficiency and accuracy. In the MRI dataset, DeepFreqNet achieves convergence in 13 epochs, which is significantly faster than ViT's 30 epochs. Similar trends were observed in blood cell and tumor classification, where DeepFreqNet converged in fewer epochs, 17 and 12, respectively, compared with ViT, which were 20 and 50. Notably, DeepFreqNet exhibits superior accuracy on small datasets such as ASL and maintains efficiency in BSL sign language classification, requiring only 10 epochs compared with ViT's 15. In the DeepHisto dataset, while ViT has a higher training accuracy, DeepFreqNet exhibits a validation accuracy of 98.68%. Overall, DeepFreqNet stands out for its accuracy, efficiency, and rapid convergence, making it a more sophisticated and reliable choice for diverse deep-learning tasks than ViT.

AlexNet

In the evaluation across various datasets, the performance metrics for the AlexNet and DeepFreqNet models were as follows. For the MRI Tumor Dataset, AlexNet achieved a training accuracy of 76.93% and a valid accuracy of 66.45% after 30 epochs. DeepFreqNet attained significantly higher accuracy with a training accuracy of 99.97% and a valid accuracy of 96.41% in only 13 epochs. This notable difference in accuracy demonstrates the superior performance of DeepFreqNet compared to AlexNet. In the Blood Cells dataset, AlexNet achieved a training accuracy of 69.38% and a valid accuracy of 52.29% after 20 epochs. In comparison, DeepFreqNet outperformed with a training accuracy of 99.91% and a valid accuracy of 93.58% in 17 epochs. The Tumor Classification dataset saw AlexNet with a training accuracy of 86.89% and a valid accuracy of 79.42% after 15 epochs. In contrast, DeepFreqNet achieved higher accuracy with a training accuracy of 99.40% and a valid accuracy of 91.70% in 12 epochs. Similar trends continued across the BSL, ISL-alpha char, ASL, BSL-40, KU-BdSL, and DeepHisto datasets, with DeepFreqNet consistently outperforming AlexNet in both training and validation accuracies across fewer epochs, showcasing its efficacy in image classification tasks.

Upon scrutinizing the results, it was evident that the proposed DeepFreqNet consistently outperformed the other models in terms of accuracy. Notably, it exhibited the highest average accuracy among all tested models. Furthermore, our architecture demonstrated enhanced efficiency, implying that fewer training rounds are required to achieve optimal performance. The training process utilized early halting, a technique that permits the specification of many training epochs, and the termination of training when the model's performance on a distinct validation dataset stops improving.

Evaluating Contemporary Global Strategy Approaches

We conducted a thorough and comprehensive review of recent studies on magnetic resonance imaging (MRI) interpretation, hand sign recognition, and blood cell classification. Throughout our investigation, we utilized our proposed method named DeepFreqNet, and our comparative analysis revealed that it significantly outperformed other methods. As shown in Table 6, the proposed architecture, the DeepFreqNet architecture, achieved remarkable results across various domains and outperformed several other models and methodologies. In the recognition tasks domain, utilizing DeepFreqNet, achieved an outstanding accuracy of 99.91% in 2023, surpassing the accuracy of Acevedo et al.⁴⁸ and Hemalatha et al.⁴⁹, who employed CNN and Enhanced CNN respectively. Our method outperformed other models and methodologies across diverse domains, including MRI and hand sign recognition. In the MRI domain, our method achieved a remarkable accuracy of 99.97% in 2023, surpassing the accuracy of Alrashedy et al.⁵⁵, who combined the GAN architectures and CNN models. Our proposed method achieved 99.86% accuracy for tasks such as hand sign recognition and hybrid models. Overall, this comprehensive comparison highlights the effectiveness and versatility of the DeepFreqNet in outperforming various methodologies and models for diverse recognition tasks. Our results demonstrate that the proposed method, utilizing DeepFreqNet, is a highly effective approach for achieving remarkable accuracy in recognition tasks across diverse domains.

Explainable AI (XAI) methodologies

We implemented explainable AI (XAI) methodologies, specifically GradCAM and GradCAM++ in Figure 10. These methodologies offer insights into complex deep learning models, enhancing model transparency, interpretability, and reliability.

Reference	Year	Methodology	Performance (%)	Domain
Acevedo et al. ⁴⁸	2019	CNN	90.5	–
Hemalatha et al. ⁴⁹	2022	Enhanced CNN	95.0	Blood cell
Nobel et al. ⁵⁰	2024	MCT	97.35	–
Ozdemir et al. ⁵¹	2025	ConvNeXtV2	93.48	–
Our method	2025	DeepFreqNet	99.91	–
Ozkaraca et al. ⁵²	2023	Dense CNN	99.0	–
Ozdemir et al. ⁵³	2025	InceptionNeXt	99.54	–
Aurna et al. ⁵⁴	2022	CNN	99.13	–
Alrashedy et al. ⁵⁵	2022	GAN & CNN	99.09	MRI
Gomez-Guzman et al. ⁵⁶	2023	CNN	97.12	–
Pacal et al. ⁵⁷	2024	HSW-MSA + ResMLP	99.92	MRI
Nobel et al. ⁵⁰	2024	MCT	97.35	–
Our method	2025	DeepFreqNet	99.97	–
Katoch et al. ⁵⁸	2022	SVM & CNN	99.17 & 99.64	–
Das et al. ⁵⁹	2024	CNN	96.88	Hand sign
Pacal et al. ⁶⁰	2024	MaxCerVixT	99.48	–
Tyagi et al. ⁶¹	2022	Hybrid FiST_CNN	97.89	–
Our method	2025	DeepFreqNet	99.86	–

Table 6. Performance comparison of our proposed architecture (DeepFreqNet) with existing methodologies.. Significant values are in bold.

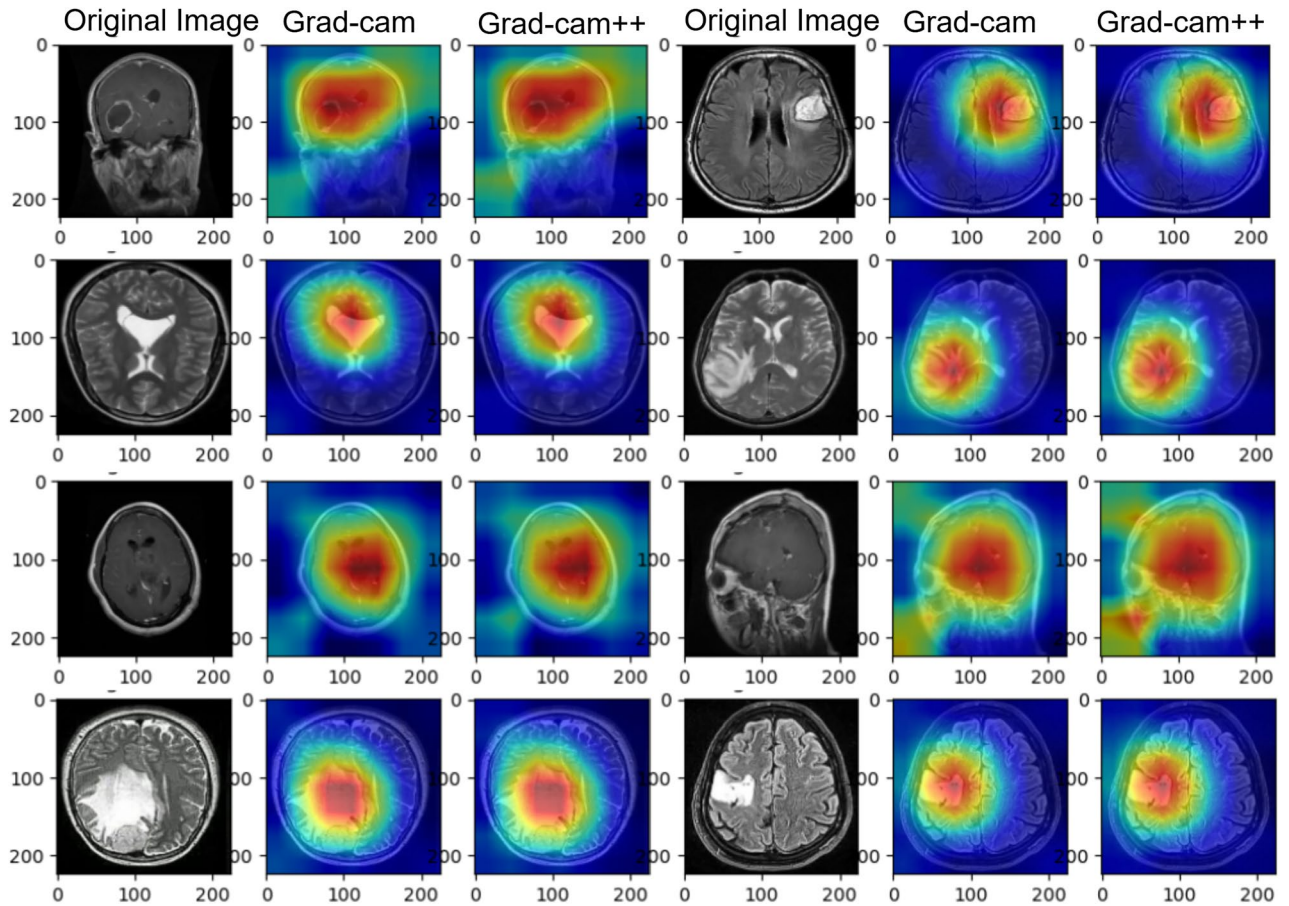


Fig. 10. GradCAM and GradCAM++ visualizations applied to brain MRI images for tumor localization and classification. The heatmaps utilize a color gradient, with brighter regions (yellow and red) indicating stronger neural activations critical to the model's decision-making. GradCAM++ provides enhanced visual interpretability by precisely localizing medically relevant tumor regions.

In 2017, Selvaraju et al. introduced Gradient-weighted Class Activation Mapping (GradCAM)⁶². GradCAM helps visualize crucial input areas that influence predictions made by convolutional neural networks (CNNs)⁶³. It uses gradients flowing into the final convolutional layers to highlight important regions.

Initially, the input image is passed into the CNN model to generate predictions. The model then calculates the gradient of the predicted class (y^c) concerning the feature map activation (A^k), defined as:

$$G_{ij}^k = \frac{\partial y^c}{\partial A_{ij}^k}$$

Subsequently, the neuron significance weights (α_k^c) are determined by performing a global average pooling of the gradients across spatial dimensions (i, j):

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

Here, Z is the total number of spatial elements in the feature map.

The final GradCAM heatmap (L_{GradCAM}) is obtained by summing the neuron weights and feature map activations, followed by applying the Rectified Linear Unit (ReLU) activation function:

$$L_{\text{GradCAM}} = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right)$$

Despite its utility, GradCAM struggles with accurately localizing objects when multiple instances or fragmented object parts exist in a single image. To overcome these limitations, Chattopadhyay et al. proposed an enhanced technique, GradCAM++⁶⁴, incorporating higher-order gradient calculations for precise localization.

GradCAM++ refines neuron weights (α_k^{c++}) by including second- and third-order derivatives, expressed mathematically as follows:

$$\alpha_k^{c++} = \sum_i \sum_j \frac{\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2}}{2 \frac{\partial^2 y^c}{(\partial A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \frac{\partial^3 y^c}{(\partial A_{ij}^k)^3}}$$

where, y^c : The model's output for class c . A_{ij}^k : Activation map at spatial location (i, j) in feature map k . $\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2}$:

Second-order gradient of the output concerning activations. $\frac{\partial^3 y^c}{(\partial A_{ij}^k)^3}$: Third-order gradient, capturing higher-order interactions.

This advanced formulation helps overcome GradCAM's shortcomings, ensuring better localization and clarity, especially when multiple instances of the object class exist or object boundaries are complex.

These visualization techniques significantly improve interpretability, helping clinicians and researchers trust and understand deep learning models, particularly for medical diagnosis and critical decision-making tasks.

Discussion

DeepFreqNet demonstrates significant promise as a multi-domain image classification framework by cohesively merging standard 3×3 convolutions, multi-scale inception-style filters (1×1 , 3×3 , 5×5), efficient depthwise-separable layers, and residual shortcuts into a single end-to-end trainable architecture. This holistic integration enables rich feature representations at multiple receptive fields while preserving computational tractability and enhanced gradient flow, distinguishing DeepFreqNet from modular or sequential designs that apply these components in isolation. In particular, this architecture uniquely blends inception-style multi-scale convolutions with depthwise-separable operations and a residual shortcut, yielding robust learning across scales without ballooning computation. Furthermore, by integrating batch normalization, strategic pooling, and a compact classifier head with strong dropout regularization, it achieves fast convergence and resilience to overfitting. The entire model comprises approximately 26.1 million trainable parameters.⁶⁵

Performance-wise, DeepFreqNet achieves strong results across diverse domains, including medical imaging (e.g., MRI tumors, histopathology slides, blood cell classification) and hand sign recognition (BSL, ISL, ASL), demonstrating not only high accuracy but also computational efficiency and rapid convergence. Crucially, the network maintains peak performance with relatively fewer epochs, highlighting its optimization-aware design.

The practical significance of this work extends beyond numerical benchmarks. In clinical and accessibility-oriented scenarios—such as early cancer detection or sign language translation—DeepFreqNet's consistency and speed make it suitable for real-world deployment where real-time decisions are required. Moreover, explainability via GradCAM and GradCAM++^{62,64} supports model transparency in high-stakes applications. However, these visualization techniques may not fully capture subtle or context-specific patterns, suggesting that complementary XAI methods will be important for deeper interpretability in future studies.

Despite its strengths, DeepFreqNet faces limitations related to dataset biases and domain shifts. Hematology datasets can under-represent certain staining variations, and sign-language collections often skew toward younger users, potentially limiting generalization to other demographics. Likewise, MRI tumor samples frequently come from a single institution, reducing scanner and protocol diversity. To address these issues, future studies should employ domain-adaptive augmentations, fairness-aware loss functions, and multi-institutional data pooling to ensure equitable performance across varied populations and acquisition settings.

In our implementation, dropout was disabled after ablation experiments showed that batch normalization combined with early stopping sufficed to control overfitting—enabling dropout at a rate of 0.5 only improved accuracy by 0.5% while slowing convergence by 10%. Nonetheless, integrating dynamic attention modules (e.g., channel/spatial attention) could further focus feature learning, and pruning or quantization strategies should be explored for efficient real-time, on-device inference.

DeepFreqNet offers a compelling blend of performance, interpretability, and adaptability, laying the groundwork for scalable, trustworthy AI systems capable of operating effectively across heterogeneous visual domains.

Conclusions

DeepFreqNet is a cohesive, domain-agnostic architecture that harmonizes multi-resolution convolutional pathways, depthwise-separable filters, residual shortcuts, and spectral feature fusion into a single backbone requiring no per-domain redesign. Through rigorous evaluation of magnetic resonance images, whole-slide histology, blood smear micrographs, and three distinct sign-language corpora (BSL, ISL, ASL), we demonstrate that DeepFreqNet simultaneously captures shared representations and domain-specific nuances without sacrificing accuracy or efficiency, thereby redefining the frontier of multi-domain image classification. Moving forward, we will enhance the model's focus and adaptability by embedding dynamic channel- and spatial-attention layers, strengthen performance on underrepresented classes via semi-supervised learning techniques (consistency regularization and pseudo-labeling), enable collaborative refinement across distributed healthcare sites through privacy-preserving federated optimization, and apply pruning and quantization for lightweight on-device inference. These concrete advancements aim to elevate DeepFreqNet from a research prototype to a practical tool for real-time clinical diagnostics and assistive communication systems.

Data availability

www.kaggle.com/datasets/smnuruzzaman/bangla-sign-language-bsl/data; github.com/Patchwork53/BdSL40_Dataset_AI_for_Bangla_2.0_Honorable_Mention data.mendeley.com/datasets/7tsw22y96w/1; www.kaggle.com/datasets/debashishsau/aslamerican-sign-language-alphabet-dataset; data.mendeley.com/datasets/scpvm2nbkm/4; www.kaggle.com/datasets/paultimothymooney/blood-cells/data; www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset; zenodo.org/records/7941080; figshare.com/articles/dataset/Machine_Vision_Approach_for_Brain_Tumor_Classification_using_Multi_Features_Dataset/19915306.

Received: 30 January 2025; Accepted: 3 July 2025

Published online: 26 September 2025

References

1. Abiwinanda, N., Hanif, M., Hesaputra, S. T., Handayani, A. & Mengko, T. R. Brain tumor classification using convolutional neural network. In *World Congress on Medical Physics and Biomedical Engineering 2018: June 3-8, 2018, Prague, Czech Republic*. Vol. 1. 183–189 (Springer, 2019).
2. Tang, L. Image classification based on improved VGG network. In *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*. 316–320 (IEEE, 2021).
3. Chen, X. & Jing, R. Video super resolution based on deformable 3D convolutional group fusion. *Sci. Rep.* **15**, 9050 (2025).
4. Alsaif, H. et al. A novel data augmentation-based brain tumor detection using convolutional neural network. *Appl. Sci.* **12**, 3773 (2022).
5. He, W. et al. Neuromorphic-enabled video-activated cell sorting. *Nat. Commun.* **15**, 10792 (2024).
6. Chen, X. et al. Joint scene flow estimation and moving object segmentation on rotational lidar data. In *IEEE Transactions on Intelligent Transportation Systems* (2024).
7. Chen, L. et al. Hpda/zn as a creb inhibitor for ultrasound imaging and stabilization of atherosclerosis plaque. *Chin. J. Chem.* **41**, 199–206 (2023).
8. Kaur, T. & Gandhi, T. K. Automated brain image classification based on vgg-16 and transfer learning. In *2019 International Conference on Information Technology (ICIT)*. 94–98 (IEEE, 2019).
9. Khan, M. S. I. et al. Accurate brain tumor detection using deep convolutional neural network. *Comput. Struct. Biotechnol. J.* **20**, 4733–4745 (2022).
10. Chattopadhyay, A. & Maitra, M. MRI-based brain tumour image detection using cnn based deep learning method. *Neurosci. Inform.* **2**, 100060 (2022).
11. Yadav, S. S. & Jadhav, S. M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **6**, 1–18 (2019).
12. Soni, M. et al. Light weighted healthcare cnn model to detect prostate cancer on multiparametric MRI. *Comput. Intell. Neurosci.* **2022** (2022).
13. Tiwari, P. et al. CNN based multiclass brain tumor detection using medical imaging. *Comput. Intell. Neurosci.* **2022** (2022).
14. Yang, B., Xu, S., Yin, L., Liu, C. & Zheng, W. Disparity estimation of stereo-endoscopic images using deep generative network. *ICT Exp.* **11**, 74–79 (2025).
15. Wang, W., Yuan, X., Wu, X. & Liu, Y. Fast image dehazing method based on linear transformation. *IEEE Trans. Multimed.* **19**, 1142–1155 (2017).
16. Sahana, T., Paul, S., Basu, S. & Mollah, A. F. Hand sign recognition from depth images with multi-scale density features for deaf mute persons. *Proc. Comput. Sci.* **167**, 2043–2050 (2020).
17. Das, S., Imtiaz, M. S., Neom, N. H., Siddique, N. & Wang, H. A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier. *Exp. Syst. Appl.* **213**, 118914 (2023).
18. Shi, G. et al. One for all: A unified generative framework for image emotion classification. *IEEE Trans. Circuits Syst. Video Technol.* **34**, 7057–7068 (2023).
19. Zeng, Y. et al. Gccnet: A novel network leveraging gated cross-correlation for multi-view classification. *IEEE Trans. Multimed.* (2024).
20. Damaneh, M. M., Mohanna, F. & Jafari, P. Static hand gesture recognition in sign language based on convolutional neural network with feature extraction method using orb descriptor and gabor filter. *Exp. Syst. Appl.* **211**, 118559 (2023).
21. Wang, W., Yin, B., Li, L., Li, L. & Liu, H. A low light image enhancement method based on dehazing physical model. *CMES-Comput. Model. Eng. Sci.* (2025).
22. Kumar, S., Gupta, S. K., Kaur, M. & Gupta, U. Vi-net: A hybrid deep convolutional neural network using vgg and inception v3 model for copy-move forgery classification. *J. Vis. Commun. Image Represent.* **89**, 103644 (2022).
23. Zhang, R. et al. Mvmrl: A multi-view molecular representation learning method for molecular property prediction. *Brief. Bioinform.* **25**, bbae298 (2024).
24. Zhang, Z.-W., Liu, Z.-G., Martin, A. & Zhou, K. BSC: Belief shift clustering. *IEEE Trans. Syst. Man Cybern. Syst.* **53**, 1748–1760 (2022).
25. Chen, S., Lai, X., Yan, Y., Wang, D.-H. & Zhu, S. Learning an attention-aware parallel sharing network for facial attribute recognition. *J. Vis. Commun. Image Represent.* **90**, 103745 (2023).
26. Szegedy, C. et al. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–9 (2015).
27. Nobel, S. N. et al. A machine learning approach for vocal fold segmentation and disorder classification based on ensemble method. *Sci. Rep.* **14**, 14435 (2024).
28. Tammina, S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *Int. J. Sci. Res. Publ. (IJSRP)* **9**, 143–150 (2019).
29. Biratu, E. S., Schwenker, F., Ayano, Y. M. & Debelee, T. G. A survey of brain tumor segmentation and classification algorithms. *J. Imaging* **7**, 179 (2021).
30. Nobel, S. N., Sultana, S., Tasir, M. A. M., Mridha, M. & Aung, Z. Cancernet: A comprehensive deep learning framework for precise and intelligible cancer identification. *Comput. Biol. Med.* **193**, 110339 (2025).
31. Pacal, I. Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model. *Expert Syst. Appl.* **238**, 122099 (2024).
32. Yuan, F., Zhang, Z. & Fang, Z. An effective cnn and transformer complementary network for medical image segmentation. *Pattern Recognit.* **136**, 109228 (2023).
33. Ohi, A. Q., Mridha, M. F., Safir, F. B., Hamid, M. A. & Monowar, M. M. Autoembedder: A semi-supervised dnn embedding system for clustering. *Knowl.-Based Syst.* **204**, 106190 (2020).
34. Nobel, S. N., Sifat, O. F., Islam, M. R., Sayeed, M. S. & Amiruzzaman, M. Enhancing GI cancer radiation therapy: Advanced organ segmentation with ResECA-U-net model. *Emerg. Sci. J.* **8**, 999–1015 (2024).

35. Nuruzzaman Nobel, S. et al. CRT: A convolutional recurrent transformer for automatic sleep state detection. *IEEE J. Biomed. Health Inform.* (2025).
36. Nobel, S. N., Afroj, M., Kabir, M. M. & Mridha, M. Development of a cutting-edge ensemble pipeline for rapid and accurate diagnosis of plant leaf diseases. *Artif. Intell. Agric.* **14**, 56–72 (2024).
37. Mostafa Monowar, M. et al. Advanced sleep disorder detection using multi-layered ensemble learning and advanced data balancing techniques. *Front. Artif. Intell.* **7**, 1506770 (2025).
38. Munna, M. A. H., Tunny, Z. F., Adrita, S. F. K., Lipi, K. A. & Kabir, A. Bangla sign language 40. <https://doi.org/10.34740/KAGGLE/DSV/5048694> (2023).
39. Singh, A., Singh, S. K., Mittal, A. & Gupta, B. B. Static gestures of Indian Sign Language (ISL) for English Alphabet, Hindi Vowels and Numerals. <https://doi.org/10.17632/7tsw22y96w.1> (2022).
40. Duarte, A., Palaskar, S., Ventura, L., Ghadiyaram, D., DeHaan, K., Metze, F., Torres, J. and Giro-i-Nieto, X., How2sign: a largescalemultimodal dataset for continuous american sign language. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp. 2735–2744 (2021).
41. Jaid Jim, A. A., Rafi, I., Akon, M. Z. & Nahid, A.-A. KU-BdSL: Khulna University Bengali Sign Language dataset. <https://doi.org/10.17632/scpvm2nbkm.1> (2021).
42. Blood Cell Images—kaggle.com. <https://www.kaggle.com/datasets/paultimothymooney/blood-cells/data>. Accessed 06 Dec 2023 (2023).
43. Nobel, S. N. et al. Segx-net: A novel image segmentation approach for contrail detection using deep learning. *Plos one* **19**, e0298160 (2024).
44. SartajBhuvaji. Brain-tumor-classification-dataset. <https://github.com/sartajbhuvaji/brain-tumor-classification-dataset> (2020).
45. Qadri, S. *Machine vision approach for brain tumor classification using multi features dataset*. <https://doi.org/10.6084/m9.figshare.19915306.v1> (2022).
46. Mittelbronn, M. et al. DeepHisto: Dataset for glioma subtype classification from whole slide images. <https://doi.org/10.5281/zenodo.7941080> (2023) (this work was supported by the Luxembourg National Research Fund (FNR) grants C21/BM/15739125/DIOMEDES and PEARL P16/BM/11192868).
47. Bangla Sign Language (BSL)—kaggle.com. <https://www.kaggle.com/datasets/smnuruzzaman/bangla-sign-language-bsl/data>. Accessed 05 Dec 2023 (2023).
48. Acevedo, A., Alferez, S., Merino, A., Puigvi, L. & Rodellar, J. Recognition of peripheral blood cell images using convolutional neural networks. *Comput. Methods Programs Biomed.* **180**, 105020 (2019).
49. Hemalatha, B., Karthik, B., Reddy, C. K. & Latha, A. Deep learning approach for segmentation and classification of blood cells using enhanced cnn. *Meas. Sens.* **24**, 100582 (2022).
50. Nobel, S. N. et al. A novel mixed convolution transformer model for the fast and accurate diagnosis of glioma subtypes. *Adv. Intell. Syst.* **2400566** (2024).
51. Ozdemir, B. & Pacal, I. A robust deep learning framework for multiclass skin cancer classification. *Sci. Rep.* **15**, 4938 (2025).
52. Özkara, O. et al. Multiple brain tumor classification with dense cnn architecture using brain MRI images. *Life* **13**, 349 (2023).
53. Ozdemir, B., Aslan, E. & Pacal, I. Attention enhanced InceptionNeXt based hybrid deep learning model for lung cancer detection. *IEEE Access* (2025).
54. Aurna, N. F., Yousuf, M. A., Taher, K. A., Azad, A. & Moni, M. A. A classification of MRI brain tumor based on two stage feature level ensemble of deep cnn models. *Comput. Biol. Med.* **146**, 105539 (2022).
55. Alrashedy, H. H. N., Almansour, A. F., Ibrahim, D. M. & Hammoudeh, M. A. A. BrainGAN: Brain MRI image generation and classification framework using GAN architectures and cnn models. *Sensors* **22**, 4297 (2022).
56. Gómez-Guzmán, M. A. et al. Classifying brain tumors on magnetic resonance imaging by using convolutional neural networks. *Electronics* **12**, 955 (2023).
57. Pacal, I. A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. *Int. J. Mach. Learn. Cybern.* **15**, 3579–3597 (2024).
58. Katoch, S., Singh, V. & Tiwary, U. S. Indian sign language recognition system using surf with SVM and CNN. *Array* **14**, 100141 (2022).
59. Das, S., Biswas, S. K. & Purkayastha, B. Occlusion robust sign language recognition system for Indian sign language using cnn and pose features. *Multimed. Tools Appl.* 1–20 (2024).
60. Pacal, I. Maxcervixt: A novel lightweight vision transformer-based approach for cervical cancer classification using pap smear images. *Knowl.-Based Syst.* **289**, 111482. <https://doi.org/10.1016/j.knosys.2024.111482> (2024).
61. Tyagi, A. & Bansal, S. Hybrid fist_cnn approach for feature extraction for vision-based Indian sign language recognition. *Int. Arab J. Inf. Technol.* **19**, 403–411 (2022).
62. Selvaraju, R. R. et al. Grad-cam: Why did you say that? arXiv preprint [arXiv:1611.07450](https://arxiv.org/abs/1611.07450) (2016).
63. Ahmed, S., Nobel, S. N. & Ullah, O. An effective deep cnn model for multiclass brain tumor detection using MRI images and Shap explainability. In *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. 1–6 (IEEE, 2023).
64. Chattopadhyay, A., Sarkar, A., Howlader, P. & Balasubramanian, V. N. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 839–847 (IEEE, 2018).
65. Hardani, D. N. K., Nugroho, H. A. & Ardiyanto, I. Comparison performance of deep learning models for brain tumor segmentation based on 2d convolutional neural network. In *Proceeding of the 3rd International Conference on Electronics, Biomedical Engineering, and Health Informatics: ICEBEHI 2022, 5–6 October, Surabaya, Indonesia*. 333–355 (Springer, 2023).

Acknowledgements

The authors would like to thank the Rabbit AI & Advanced Machine Intelligence Research Lab (AMIR Lab) for its supervision and resources.

Author contributions

Conceptualization, Nobel, S. N.; Data curation, Nobel, S. N. and Tasir M.A.M. ; Formal analysis, Nobel, S. N.; Investigation, Tasir M.A.M., M.A. Hamid; Mridha, M.F.; Methodology, Nobel, S. N., Noor, H.; Software, Nobel, S. N., Noor, H.; Supervision, M.M. Monowar; Mridha, M.F.; Visualization, Nobel, S. N., Noor, H.; Writing - original draft, Noor, H.; Nobel, S. N.; Writing - review & editing, M.A. Hamid; M.M. Monowar; Mridha, M.F., Sayeed, M.S., Islam, M.R., and Dey, N.; Funding acquisition, Sayeed, M.S. and Islam, M.R.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.S.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025