



Review

Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging

Anh T. Tran ¹, Tal Zeevi ² and Seyedmehdi Payabvash ^{1,*} ¹ Department of Radiology, Columbia University Irving Medical Center, New York-Presbyterian Hospital, Columbia University, New York, NY 10032, USA; at4049@cumc.columbia.edu² Department of Biomedical Engineering, Yale University, New Haven, CT 06520, USA; tal.zeevi@yale.edu

* Correspondence: sp4479@cumc.columbia.edu

Abstract: Artificial Intelligence (AI) and deep learning models have revolutionized diagnosis, prognostication, and treatment planning by extracting complex patterns from medical images, enabling more accurate, personalized, and timely clinical decisions. Despite its promise, challenges such as image heterogeneity across different centers, variability in acquisition protocols and scanners, and sensitivity to artifacts hinder the reliability and clinical integration of deep learning models. Addressing these issues is critical for ensuring accurate and practical AI-powered neuroimaging applications. We reviewed and summarized the strategies for improving the robustness and generalizability of deep learning models for the segmentation and classification of neuroimages. This review follows a structured protocol, comprehensively searching Google Scholar, PubMed, and Scopus for studies on neuroimaging, task-specific applications, and model attributes. Peer-reviewed, English-language studies on brain imaging were included. The extracted data were analyzed to evaluate the implementation and effectiveness of these techniques. The study identifies key strategies to enhance deep learning in neuroimaging, including regularization, data augmentation, transfer learning, and uncertainty estimation. These approaches address major challenges such as data variability and domain shifts, improving model robustness and ensuring consistent performance across diverse clinical settings. The technical strategies summarized in this review can enhance the robustness and generalizability of deep learning models for segmentation and classification to improve their reliability for real-world clinical practice.

Keywords: robustness; generalization; neuroimaging; deep learning; segmentation; classification



Academic Editor: James C. L. Chow

Received: 11 February 2025

Revised: 3 April 2025

Accepted: 7 April 2025

Published: 14 April 2025

Citation: Tran, A.T.; Zeevi, T.; Payabvash, S. Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging.

BioMedInformatics **2025**, *5*, 20.<https://doi.org/10.3390/biomedinformatics5020020>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Machine learning and artificial intelligence have revolutionized medical imaging workflows and applications in recent years [1]. These tools are applied before or during image acquisition for purposes such as denoising, radiation dose reduction, image reconstruction, and workflow optimization—including scheduling exams, triaging patients, and prioritizing imaging studies. The downstream applications of artificial intelligence include image analysis, computer-assisted diagnosis, radiology report generation, and clinical decision support. Machine learning models used in medical imaging have shown remarkable potential for improving diagnosis and treatment planning [2], by harnessing imaging patterns that are imperceptible to human eyes. The main downstream applications of these tools in medical image analysis can be categorized into segmentation and classification tasks, which are the focus of this review. However, their successful deployment in clinical

practice depends on ensuring consistent performance across diverse real-world scenarios. This challenge highlights the need for models that are both robust to imaging variations and generalizable across different clinical settings.

Robustness refers to a model's ability to maintain performance despite the variability in medical imaging environments [3]. This variability often arises from multiple sources, including differences in scanner manufacturers and types, scan acquisition protocols, patient positioning within the imaging machine, image artifacts, and noise [4–6]. Without robust models, even minor changes in image quality or acquisition parameters can result in substantial classification errors or imprecise segmentation boundaries [7,8].

Generalizability, on the other hand, extends beyond robustness and focuses on a model's ability to perform effectively on entirely new, unseen datasets [9–11]. This property is essential for translating research into clinical practice, ensuring reliable performance across diverse patient populations, and maintaining accuracy across healthcare settings. Models lacking generalizability often fail to capture universally essential features, instead relying on spurious correlations in the training data (overfitting), which limits their practical utility in real world scenarios.

To address these challenges, researchers have developed a variety of strategies to enhance both robustness and generalizability. Data augmentation techniques simulate realistic variations in medical image acquisition by applying controlled changes to contrast, resolution, orientation, and noise levels, reflecting differences in imaging protocols and scanner types [12,13]. Adversarial training improves a model's resilience by exposing it to the potential noise and distortions that are encountered in clinical settings. Transfer learning leverages pre-training on large-scale medical imaging datasets, followed by fine-tuning for specific clinical applications, while domain adaptation minimizes systematic differences between images acquired at different medical centers [14,15]. Inverse supervised learning [16], which complements traditional supervised learning by focusing on the inverse mapping between inputs and outputs, can also reduce overfitting to specific patterns in the dataset and enhances interpretability by highlighting the causal factors behind predictions. These approaches succeeded across various neuroimaging tasks, including tumor detection, brain structure segmentation, and neurocognitive disease classification [17,18], ensuring consistent clinical performance across different clinical settings and patient populations [19,20].

While previous articles have summarized the methods used to increase the robustness and generalizability of machine learning models in general or with a narrow focus on a specific task [21–24], our review provides an overview of strategies for improving the robustness and generalizability of deep learning models in neuroimaging to achieve a balance between accuracy and multi-modal adaptability for clinical applications. In addition, segmentation and classification are core tasks in the downstream application of artificial intelligence tools in neuroimaging. Segmentation requires pixel-level robustness, classification requires feature-level robustness, and they are most strongly affected by domain shifts in neuroimaging. We summarize the previous studies on improving the robustness and generalization ability of deep learning models in neuroimaging segmentation and classification tasks, such as transfer learning, regularization, and adversarial training; we further emphasize the important role of evaluation metrics and uncertainty estimation.

2. Methods

This comprehensive review follows a structured protocol [25] to retrieve and summarize the current state of robustness and generalizability in deep learning models for neuroimaging. Our methodology includes a systematic literature search and summarizes various strategies. The review addresses three key aspects of deep learning models in brain imaging:

- The current state and challenges in model robustness and generalizability.
- Strategies for enhancing and monitoring these attributes.
- Barriers in transitioning models from research into clinical practice.

2.1. Search Strategy

Our search strategy targeted three major databases: Google Scholar, PubMed, and Scopus. The key search terms were categorized as follows: (i) Primary: “brain images” OR “neuroimaging” OR “brain imaging” OR “neuro-imaging”; (ii) Task-specific: (“segmentation” OR “classification”) AND “deep learning”. (iii) Model-focused: “robustness”, “generalizability”. These terms were systematically combined using Boolean operators to achieve comprehensive yet focused search results.

2.2. Selection Criteria

We included studies that met the following quality and relevance criteria: peer-reviewed, English-language publications with accessible full texts, focusing on original research in neuroimaging. The exclusion criteria were survey studies, duplicate studies, and research not directly addressing model robustness or generalizability. Figure 1 depicts the flowchart of our search strategy and the final number of articles that are referenced in our review. While our study was not conducted as a formal systematic review, we applied PRISMA principles to guide our search strategy [26].

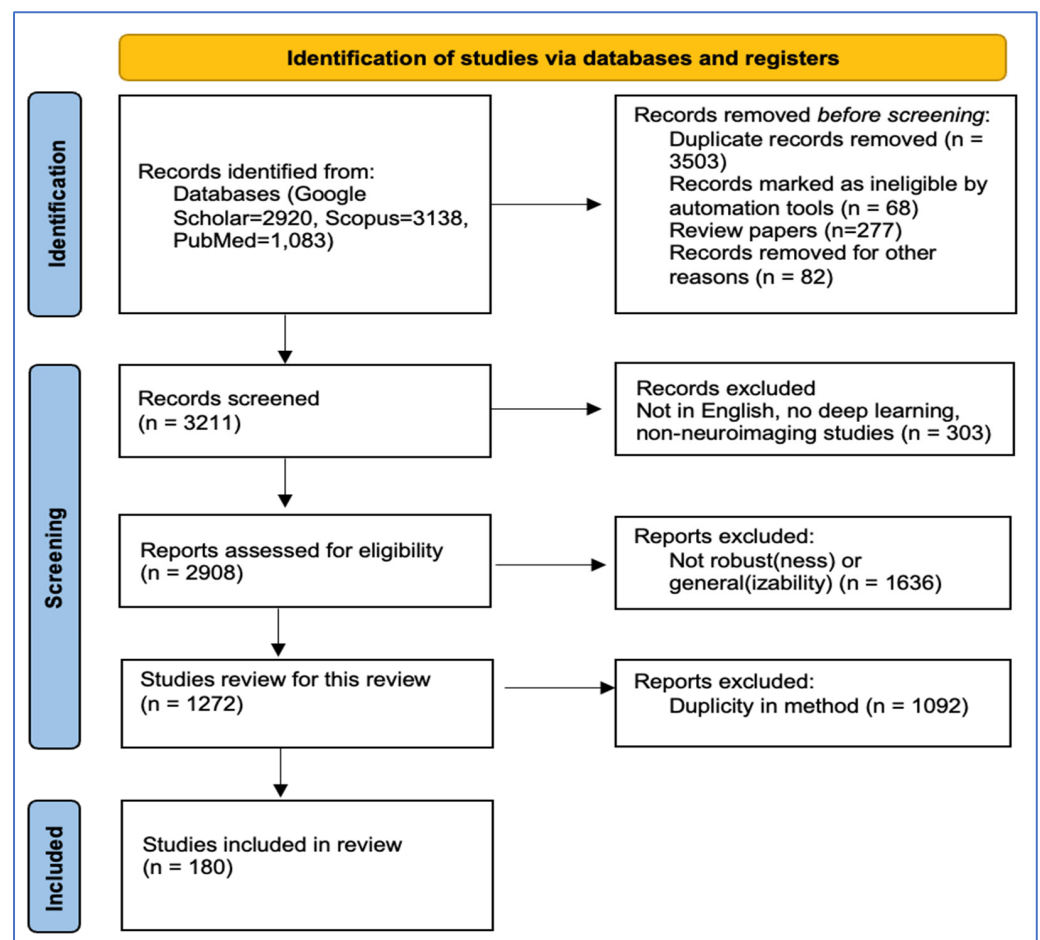


Figure 1. Flowchart of the search and selection of studies.

2.3. Data Extraction

Relevant information was systematically collected by carefully reviewing each study. The strategies were categorized into three subcategories: definition, training usage, and evaluation methods. The study parameters were organized using a spreadsheet, with details such as the objectives, deep learning models, network architectures, publication year, journal, datasets, performance metrics, and challenges placed in separate columns. Each study was listed as a separate row for clarity.

3. Strategies for Improving Robustness and Generalizability

Figure 2 summarizes the strategies and key methods for improving the robustness and generalizability of segmentation and classification deep learning models.

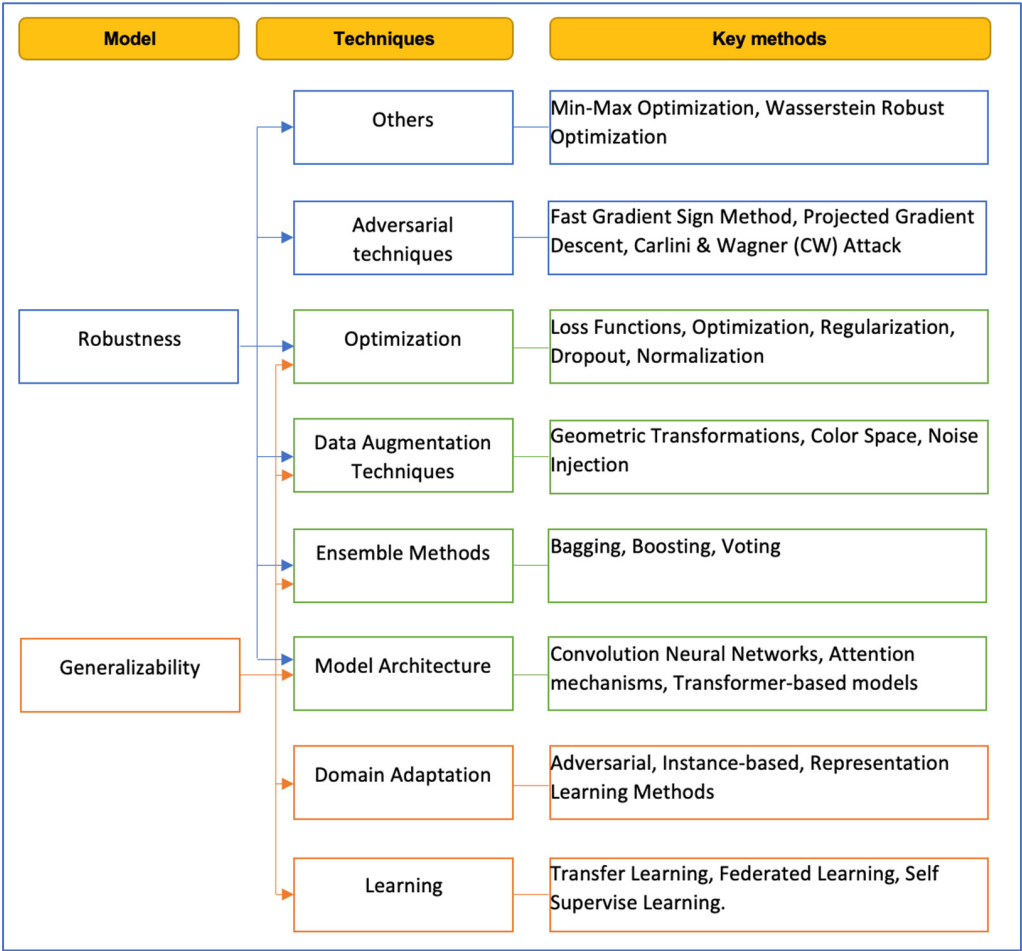


Figure 2. Strategies for improving robustness and generalizability.

3.1. Shared Approaches Improving Both Robustness and Generalizability

3.1.1. Optimization Techniques

The loss function [27] quantifies how well the model’s predictions match the ground truth labels and is used during optimization to update the model parameters. Dice loss [28], which is widely used in segmentation tasks, measures the overlap between the predicted segment and the actual segment, ensuring high-quality segmentation, supporting both robustness and generalization. For classification, in the case of imbalanced data, using weighted cross-entropy loss [29] helps the model to focus on underrepresented classes, improving its performance on diverse data.

Adaptive optimization methods: techniques such as Adam [30] dynamically adjust the learning rate to stabilize the training process and improve convergence, especially in noisy or incomplete data.

Regularization helps prevent overfitting by introducing constraints to the learning process, ensuring that models capture general patterns rather than features limited to training data. Several key regularization approaches have been established in the field. L1 Regularization (Lasso) [31] adds penalties based on absolute coefficient values, promoting model sparsity, while L2 Regularization (Ridge) [32] applies penalties based on squared coefficient values to encourage even weight distribution. In neural networks, Dropout [33] randomly deactivates neurons during training to prevent over-reliance on specific pathways, and Batch Normalization [34] normalizes layer inputs to stabilize training to enhance reliability. Early Stopping [35] prevents overfitting by monitoring the validation performance and halting training at the optimal point.

Feature size reduction, or dimensionality reduction, is a crucial step in the preprocessing pipeline for deep learning models in neuroimaging. The most popular techniques using feature reduction in neuroimaging are Principal Component Analysis (PCA) [36], Independent Component Analysis (ICA) [37] and feature selection techniques such as LASSO [38]. The PCA and AutoFeat techniques led to increased accuracy for the models in EEG-based emotional state classification [39]. Each technique has its strengths and limitations, and selecting the appropriate method depends on the specific neuroimaging task. Summaries of different feature size reduction strategies and their applications are included in Supplemental Material Section S1.

3.1.2. Data Augmentation

Data augmentation improves model performance by diversifying the training data without the need to collect additional samples [40]. This strategy includes a range of transformation approaches. Geometric transformations involve operations such as rotation, flipping, scaling, and cropping [41,42], while color space augmentation adjusts brightness, contrast, and saturation [43]. Noise injection introduces various types of noise to improve model resilience [44,45], and random erasing selectively occludes regions of an image to mimic real-world variability [46]. Advanced methods such as Mixup and CutMix combine images to create novel training examples, further enriching the training dataset [47].

3.1.3. Ensemble Learning Approaches

Ensemble learning improves model robustness and generalizability by combining multiple models into a stronger predictive system, leveraging the principle that diverse models can collectively overcome individual limitations. Each ensemble technique offers unique advantages for enhancing model reliability in medical imaging applications.

Key ensemble techniques include the following: bagging (Bootstrap Aggregating) [48,49], i.e., the independent training of multiple models on random subsets of data (using bootstrapping), and aggregating predictions through averaging or majority voting; boosting [50,51], which trains models sequentially, with each model focusing on correcting errors made by its predecessors by assigning higher weights to misclassified samples; stacking (Stacked Generalization) [52,53], which trains multiple models on the same dataset, using their predictions as input features for a meta-model that produces the final output; and voting ensembles [54], a technique that combines predictions from multiple models through either majority voting (hard voting) or probability-weighted voting (soft voting).

3.1.4. Model Architecture

Model architecture improvements refer to strategies that enhance the structure and design of machine learning models to improve performance, robustness, and generalization.

These improvements often involve changes in the organization of layers, the use of specialized mechanisms, or the introduction of innovative training strategies. U-Net [55] with skip connections is widely used for segmentation tasks in medical images. Variants such as Attention U-Net [56] further improve feature extraction and spatial consistency, enhancing both robustness and generalization. Transformer-based Networks: Vision Transformers (ViT) [57,58] and hybrid Convolutional Neural Network (CNN)-Transformer models [59] are capable of capturing long-range dependencies in brain images, improving both robustness and generalization by focusing on important spatial relationships.

3.2. Robustness Improvement Methods

3.2.1. Adversarial Training

Adversarial Training [60–62] enhances model robustness by defending against adversarial attacks, i.e., carefully designed perturbations intended to cause model failure. This approach employs different methods to create adversarial examples of original scans: The Fast Gradient Sign Method (FGSM) [63–65] creates adversarial examples by perturbing inputs along the gradient direction of the loss function. Projected Gradient Descent (PGD) [63–65] extends this by generating stronger adversarial examples through iterative optimization. The Carlini and Wagner (CW) Attack [63–65] iteratively optimizes perturbation that minimizes the perceived change while maximizing the model's misclassification probability. The effectiveness of these strategies is systematically evaluated using specific attack scenarios and defense efficacy metrics [62,66–68].

3.2.2. Other Methods

Advanced optimization techniques can further improve model robustness. Min–Max Optimization [69] trains models under worst-case scenarios, effectively preparing them for adversarial conditions. Wasserstein Robust Optimization [70] addresses distribution shifts by employing the Wasserstein distance metric in the optimization process. These methods complement traditional approaches by targeting specific vulnerabilities that standard training procedures may not fully address.

3.3. Generalizability Improvement Methods

3.3.1. Domain Adaptation and Invariant Learning

Domain Adaptation [71] focuses on improving model performance across different domains or data distributions: it includes Feature alignment and matching the statistical properties of datasets. For example, in brain tumor segmentation, domain adaptation can address variations caused by different MRI protocols. Augmented domain adaptations, such as style transfer methods, further mitigate domain mismatches by transforming source data to mimic target domain characteristics, ensuring robust performance across scanners [72]. In addition, Karthik Gopinath et al. [73] trained neural networks on a vastly diverse array of synthetically generated images with random contrast properties.

Invariant learning emphasizes the extraction of features unaffected by domain-specific variations to ensure consistency across environments. Methods such as Invariant Risk Minimization and causal representation learning eliminate unwanted correlations, focusing instead on causal relationships that generalize well [74]. Contrastive learning has also been applied, particularly in tasks such as stroke lesion segmentation in the brain, by enhancing within-domain similarities and minimizing cross-context differences [75].

3.3.2. Model Training Strategies

Transfer Learning [76,77] enables models to leverage knowledge from one task to improve performance in related tasks. In medical imaging, it is particularly valuable for addressing limited labeled data while maintaining high performance. Key strategies in-

clude: feature extraction, where pre-trained models are used to extract relevant imaging characteristics, as demonstrated in brain MRI analysis [78]; fine-tuning, where pre-trained models are adapted to specific tasks through continued training with adjusted learning rates [79,80]; and frozen layers, where early-layer weights are preserved while adapting later layers in neural networks, optimizing computational efficiency and reducing overfitting [81,82]. These methods have proven effective in reducing training times and data requirements [76,83,84].

Federated learning [85] also allows collaborative model training across institutions without the exchange of raw data, ensuring privacy and addressing data governance concerns. By aggregating updates from locally trained models, federated learning can create robust models that are capable of generalizing across varied datasets. Techniques such as federated averaging and differential privacy enable learning from diverse data distributions while safeguarding data privacy [86,87].

Self-supervised learning [88] is a machine learning paradigm that leverages large amounts of unlabeled data to learn useful representations without relying on manual labels. By designing tasks where the dataset itself provides supervision, self-supervised learning enables models to learn underlying patterns and structures that generalize well to many downstream tasks.

3.4. Evaluation and Monitoring

Evaluating the effectiveness of robustness and generalization techniques requires comprehensive metrics and systematic monitoring to ensure that models maintain reliable performance across various scenarios, patient populations, and imaging conditions post-deployment.

3.4.1. Key Performance Metrics and Statistical Results

For segmentation tasks, spatial overlap metrics such as Intersection over Union (IoU) and the Dice–Sørensen coefficient (DSC) quantify the accuracy of region delineation across different datasets and conditions [89,90]. Another metric, the Hausdorff distance (HD) [91], measures the maximum distance of a surface set to the nearest point in the other set. These metrics are widely used to assess models' accuracy and consistency in the presence of anatomical variations and differences in image quality.

Classification tasks are evaluated using complementary metrics. Accuracy provides an overall measure of correct predictions, while Precision and Recall offer insights into model reliability for different classes. The F1 Score, as the harmonic mean of precision and recall, balances these aspects to assess overall robustness. Additionally, the Area Under the Curve of Receiver Operating Characteristic (AUC-ROC) and the Confusion Matrix are especially useful for evaluating model performance across different operating thresholds and class distributions [92–94], making them crucial for assessing generalization across diverse patient populations.

Statistical significance and confidence intervals: several studies report statistical significance and confidence intervals to assess the reliability of their results. Common approaches include paired t-tests [95] or Wilcoxon signed-rank tests [96] to compare the performance of different models, and bootstrap methods [97], which estimate the variability of performance metrics. Results are typically presented with 95% confidence intervals, ensuring that the reported performance metrics are reliable and generalizable across different datasets. The list of assessment metrics is included in Supplemental Material Section S2.

Recently, Suhan You et al. [98] proposed SaRF, a novel method that takes salient information through two self-supervised loss terms during training. It improves sequence classification in terms of the F1 score, AUC, and accuracy (ACC), especially for T1 and

post-contrast T1 MRI sequences. Eman Younis et al. [99] presented a novel hybrid approach for improved brain tumor classification by combining CNNs and EfficientNetV2B3 for feature extraction, followed by K-nearest neighbors for classification. Table 1 summarizes the techniques for improving robustness and generalizability in neuroimaging, using key performance metrics from notable studies. This overview helps identify techniques that are suitable for different scenarios based on their previous examples.

Table 1. Examples of the main strategies used to improve robustness and generalizability in neuroimaging segmentation and classification using common performance metrics.

Techniques	Studies	Dataset	Performance	Conclusion
Loss Function [27]	Brain hemorrhage (ICH), intraventricular extension (IVH), and peripheral edema (PHE) segmentation [100].	Huashan Hospital, Fudan University	DSC = 0.92, 0.79, 0.71 and Sen = 0.93, 0.88, 0.81 for ICH, IVH, PHE in segmentation tasks.	DSC loss is essential for segmentation.
	ICH, IVH, PHE segmentation from non-contrast CT [101].	TICH-2	Improved average DSC by 0.02	Focal loss is valuable for class imbalance.
Regularization (L1/L2, Dropout) [31–33]	Regularized feature learning improves MRI sequence classification [98].	Swiss-First study	Improvement in mean accuracy by 4.4% (from 0.935 to 0.976), mean AUC by 1.2% (from 0.9851 to 0.9968), and mean F1-score by 20.5% (from 0.767 to 0.924).	Regularization is critical for training and improves robustness.
	Input-level dropout model for brain metastases segmentation [102].	Oslo University Hospital and Stanford	Improve DSC (0.795 ± 0.104 vs. 0.774 ± 0.104 , $p = 0.017$), and IoU (0.561 ± 0.225 vs. 0.492 ± 0.186 , $p < 0.001$). Tested on 6 datasets.	
Batch Normalization [34]	Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI [103].	BRATS 2013, 2014, 2015, 2016, 2017 and ISLES 2015.	Improves model convergence and boosts 0.9778 Acc, 0.9754 DSC, 0.9770 Spec, 0.9789 Sen on BRATS dataset 2017	Dependence on batch size Can increase computational cost but help models achieve higher accuracy and generalization.
	the combination of convolution, batch normalization and ReLU activation enhances the network's ability to discriminate and capture relevant information [104]	Kaggle (Brain Tumor MRI Dataset)	Improves with an accuracy of 99.88%	
Data Augmentation [40]	Data Augmentation improve Tumor Classification Using MRI Images [99].	Tianjin Medical University General, Nan Fang Hospital, BR35H	Improvement in precision = 0.9951, recall = 0.9947, F1-score = 0.9944, spec = 0.9977.	Essential for improving robustness, especially in limited datasets.
	StyleGANv2-ADA is proposed for augmenting brain MRI slices [105]	Gazi University Faculty of Medicine, BR35H	BraTS 2021 = 75.18%, and Gazi Brains 2020 datasets = 99.36%, BR35H dataset= 98.99%	
Ensemble Methods [48,49]	Enhancing brain tumor classification through ensemble attention mechanism [106].	BraTS 2019	Improves acc = 0.9894, rrecision = 0.9891, recall = 0.9893, F1-Score = 0.9891, AUC = 0.984	Effective in improving model reliability for classification and segmentation tasks.
	An optimized triplanar (2.5D) model ensemble to generate accurate segmentation with fewer parameters [107]	BraTS 2020	Improving Dice with enhancing tumor = 0.713, whole tumor = 0.873, and tumor core = 0.778	

Table 1. *Cont.*

Techniques	Studies	Dataset	Performance	Conclusion
Model Architecture Improvements	DeepLabV3 + Bayesian optimization for segmentation and classification of brain tumor in MRI scans [108].	Brats 2021	Improves acc = 97.0%, recall = 0.966, spec = 0.988, F1-Score = 0.96, precision = 0.966	Advanced architectures such as SwinUNETR and GNNs can improve performance but have a high computational demand.
	Swin transformers for semantic segmentation of brain tumors [109].	BRATS 2021	DSC and HD in this approach are better than nnU-Net, SegResNet, TransBTS.	
Adversarial Training [60–62]	Robust influence-based training methods for noisy brain MRI [110]	BRATS 2017	Increases robustness, ACC = 89.52 ± 2.61	Effective for improving robustness but computationally intensive.
	Improving robustness in predicting hematoma expansion [111]	ATACH-2, YALE	AUC = 0.8 is the same but increases robustness	
Domain Adaptation [71]	Improving the whole-brain neural decoding of fMRI with domain adaptation [112]	OpenfMRI	The best Acc improvement is 10.47% (from 77.26% to 87.73%)	Highly recommended for multi-site datasets with distribution shifts.
	An unsupervised domain adaptation segmentation model is trained across modalities and diseases [113]	Decathlon medical segmentation challenge, RSNA	+11.55% DSC	
Transfer Learning [76,77]	Transfer learning for accurate brain tumor detection [80]	Brain tumor dataset. Figshare	Highest acc of 99.75%	Worth implementing for tasks with limited labeled data, especially in classification.
	Classification of Alzheimer’s disease using DenseNet-201 based on deep transfer learning techniques [114]	AD5C dataset	Acc = 98.24	
Federated Learning [85]	Integrated approach of federated learning with transfer learning for the classification and diagnosis of brain tumors on MRI [115]	Figshare, Br35H, SARTAJ	High precision (0.99 for glioma, 0.95 for meningioma, 1.00 for no tumor, and 0.98 for pituitary), recall, and F1-scores in classification, outperforming existing methods.	Promising multi-institutional collaborations, balancing performance and privacy.
	Enhancing Alzheimer’s disease classification through split federated learning [116]	Kaggle	Acc = 84.53%	
Self-Supervised Learning [88]	Improves the performance of classification in task-based functional MRI [117].	Human Connectome Project	Acc improves to $80.2 \pm 4.7\%$	Reliable but heavily reliant on large, labeled datasets.
	Contrastive self-supervised learning for neurodegenerative disorder classification [118]	Alzheimer’s Disease Neuroimaging Initiative (ADNI), Australian Imaging, Biomarker and Lifestyle Flagship Study of Aging (AIBL), Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI)	For AD vs. CN, acc= 82% test subset and acc = 80% independent holdout dataset	

Details of datasets are included in the Supplementary Table S1 [98,99,102,105,119–126].

3.4.2. Computational Complexity Analysis

Computational complexity analysis is an essential step in evaluating deep learning models, especially in neuroimaging, where large and high-dimensional datasets are prevalent. The goal of computational complexity analysis is to understand the time and space requirements of a model, ensuring that it can handle the scale of neuroimaging data without sacrificing performance or efficiency.

Time complexity refers to the amount of time a deep learning model takes to process a given input. In neuroimaging, inputs typically consist of high-dimensional data such as 3D MRI volumes, 4D fMRI data, or multi-modal imaging. The size and complexity of these inputs can significantly impact the training and inference time of deep learning models. Beside batch size and data augmentation, the time complexity is primarily driven by the network architecture. Based on the architecture's complexity, computational cost, parameter count, and memory usage, we categorize the deep learning models into three main groups:

- Low-complexity models, such as Multilayer Perceptron and basic CNNs, are suitable for small datasets and simple classification tasks.
- Moderate-complexity models such as ResNet [15] and VAEs [127] balance feature learning efficiency and computational cost.
- High-complexity models such as GANs [128] and ViTs [57] achieve state-of-the-art performance but require high computational resources.

Space complexity refers to the amount of memory, especially high-dimensional data on neuroimaging, that a model requires during training and inference. Some of the key aspects that contribute to space complexity include model parameters, activations, and multimodal data. For example, when using 3D U-Net for volumetric medical image segmentation, handling space complexity is a significant challenge due to the high-dimensional input data. Instead of processing entire 3D scans, 3D U-Net splits large volumetric data into smaller patches (e.g., $64 \times 64 \times 64$ voxels).

There are several optimization strategies that reduce computational complexity in deep learning, such as transfer learning, data parallelism, and model parallelism. By using pre-trained models on similar tasks, transfer learning reduces the need to train a model from scratch, which can be computationally expensive [129]. Fine-tuning a pre-trained model requires fewer resources and can still achieve high performance on neuroimaging tasks. For large-scale models and datasets, parallelism techniques can be employed. Data parallelism involves splitting the data across multiple processors, while model parallelism involves splitting the model across processors [130]. This helps speed up both the training and inference times.

3.4.3. Cross-Validation Strategies

Cross-validation (CV) can provide a systematic method for evaluating model generalizability across different data distributions [27,131]. K-Fold CV divides the dataset into K subsets, using K-1 folds for training and one for validation, rotating through all combinations [132,133]. Stratified K-Fold CV extends K-Fold by maintaining proportions (e.g., pathological conditions or comorbidities) across folds, ensuring balanced evaluation [134,135]. In Leave-One-Out CV, each sample serves as the validation set once, which is particularly useful in small datasets [136]. This approach can be extended to evaluate generalizability across different data sources (e.g., Leave-One-Hospital-Out), to assess robustness to institutional imaging protocol variations [137]. Nested CV incorporates two validation loops, providing unbiased estimates of model performance and hyperparameter optimization [138].

3.4.4. Validation Framework

Comprehensive validation allows for the assessment of model performance across multiple dimensions. When conducting validation with out-of-distribution data and adversarial samples, models are evaluated on their ability to maintain performance under previously unseen variation [7,66,139]. Defense efficacy tests robustness against variations in imaging protocols [62,66–68]. The Augmentation effectiveness is assessed using metrics like the Fréchet Inception Distance and Inception Score [140,141], which evaluate the diversity and realism of augmented samples. Ensemble stability, determined by cross-validation performance variance, reflects consistency across data subsets. Training efficiency metrics evaluate models' adaptability and feasibility across different clinical settings [76,83,84]. Such multi-faceted validation frameworks ensure a comprehensive evaluation of models' generalizability and robustness.

3.5. Pros and Cons of Different Robustness and Generalizability Improvement Methods

The choice of strategy to improve generalizability and robustness depends on the specific requirements and constraints of neuroimaging applications. Each approach offers unique advantages and limitations that must be carefully considered for clinical deployment. These techniques can be broadly categorized into training-time methods, including regularization and data augmentation techniques which enhance model performance during training, and inference-time approaches, including uncertainty estimation and ensemble techniques that improve reliability and robustness during prediction. Validation strategies, including cross-validation and adversarial testing, also evaluate model performance at the time of inference. Each category involves trade-offs in terms of computational demands, implementation complexity, and effectiveness in improving model robustness and generalizability. Table 1 provides a comprehensive overview of these methods, highlighting their strengths, limitations, and notable applications in neuroimaging tasks. This highlights an important concern regarding the practical implementation of complex AI techniques in neuroimaging, especially for resource-constrained settings. While adversarial training, advanced architectures, and ensemble models have shown promising results in improving model robustness and performance, they often incur increased computational costs. These approaches can require more GPU resources, longer training times, and higher memory consumption, creating barriers for smaller research institutes and clinics that do not have such infrastructure.

To address this concern, users can apply strategies that balance performance with computational efficiency. Techniques such as knowledge distillation, in which a smaller model emulates the behavior of a larger, more complex model, can effectively reduce resource demands while maintaining robust performance. Additionally, methods such as quantization, which compresses model weights to a lower precision, and pruning, which removes redundant network connections, are effective in reducing model size and accelerating inference. Incorporating low-rank approximation or a neural architecture search (NAS) can further optimize model design for efficiency. By integrating these lightweight strategies into the discussion, the authors provide a more comprehensive perspective on practical AI implementations, especially for organizations with limited computational resources.

Recently, Barati et al. [142] evaluated the impact of optimizers and loss functions on brain tumor type prediction accuracy. Their study shows that the Adam optimizer combined with either the Categorical Cross-Entropy (CCE) or Binary Cross-Entropy (BCE) loss function outperforms other combinations. Moreover, Nadam and RMSprop outperform other optimizers. The strengths and limitations of techniques used to improve the model's robustness and generalizability in neuroimaging are shown in Table 2.

Table 2. Overview of the strengths and limitations of techniques used to improve the model’s robustness and generalizability in neuroimaging.

Technique	Strengths	Limitations	Implementation Considerations	Examples
Loss function (for example, Dice loss)	Often used for segmentation tasks by directly optimizing the overlap (e.g., the Dice coefficient) between the predicted mask and the ground-truth.	Less sensitive to small structures	Used in conjunction with other losses such as cross entropy for better performance on imbalanced datasets.	[142,143]
Regularization (L1/L2/Dropout)	Controls model complexity Reduces overfitting Computationally efficient	Uniform penalty across features May oversimplify important patterns Hyperparameter sensitivity	Balance with domain-specific constraints Considers anatomical priors	[144,145]
Batch Normalization	Stabilizes training Reduces internal covariate shifts Enables higher learning rates	Batch size dependency Memory requirements Inference stability issues	Consider batch size constraints Address multi-site variations	[146,147]
Data Augmentation	Increases effective dataset size Improves generalization Addresses class imbalance	May introduce unrealistic variations Risk of violating anatomical constraints Computational overhead during training	Ensures clinically plausible transformations Validates augmented samples with experts	[148,149]
Ensemble Methods	Robust predictions Uncertainty quantification Handles different aspects of data	Increased computational cost Storage requirements Inference time overhead	Balances diversity and accuracy Considers clinical time constraints	[143,150]
Model architecture improvements	Improved feature extraction: advanced architectures combining CNNs and transformer-based models capture complex patterns in neuroimaging data. Scalability: Modularly designed architectures (e.g., nnU-Net) adapt to different neuroimaging modalities (e.g., MRI, fMRI, PET) Multimodal processing: Models such as multimodal CNNs integrate different types of neuroimaging data, improving robustness Better temporal modeling: attention-based or periodic components efficiently process temporal neuroimaging data such as fMRI and EEG	Increased computational demands, especially for architectures such as transformers and deep CNNs. Potential for overfitting when dealing with small datasets, as seen in neuroimaging. Complex hyperparameter tuning is required for architectures such as attention mechanisms	For segmentation tasks, architectures such as U-Net and its variants (3D U-Net, nnU-Net) are specifically designed for volumetric neuroimaging data Considers Graph Neural Networks (GNNs) for connectivity studies, as they model relationships between brain regions. Uses self-supervised pretraining with architectures like Vision Transformers (ViT) to improve performance on limited labeled data Uses model ensembling or dropout models to reduce overfitting and improve generalization	[109,143]
Adversarial Training	Improves robustness to perturbations Handles image artifacts Better generalization	Computationally intensive May reduce standard accuracy Complex hyperparameter tuning	Use clinically relevant perturbations Balance robustness and accuracy	[151,152]
Domain Adaptation	Addresses scanner variations Handles protocol differences Improves cross-site generalization	Requires data from target domain May not capture all domain shifts Complex implementation	Validates on multiple scanner types Considers temporal domain shifts	[153,154]
Transfer Learning	Leverages knowledge from larger datasets Reduces required training data Accelerates convergence	Source-target domain mismatch can degrade performance May preserve unwanted biases from source domain Requires careful layer-specific fine-tuning	Validates anatomical consistency Adjusts learning rates per layer based on domain similarity	[155,156]

The strengths and limitations of each strategy with representative work are summarized and cited.

4. Challenges in Translating Robust and Generalizable Models to Clinical Settings

Deep learning models for neuroimaging segmentation and classification tasks face unique challenges that affect their reliability and adaptability in clinical practice. These challenges arise from the nature of neuroimaging data, as well as the complexity of clinical environments, including population variability, task-specific demands, and workflow constraints. Below, we discuss these challenges using examples from classification and segmentation tasks, focused on Alzheimer's disease, traumatic brain injury, stroke, and intracerebral hemorrhage (ICH).

4.1. Data Quality and Standardization

Neuroimaging data quality is highly variable due to scanner artifacts, acquisition protocols, and preprocessing methods [157]. For example, artifacts such as patient motion during fMRI or DWI acquisition can cause blurring or misalignment, leading to errors in stroke lesions or ICH segmentation [158]. In Alzheimer's classification, inconsistent intensity normalization across multi-site T1-weighted MRI datasets can negatively impact feature extraction, such as cortical thickness estimations, degrading model performance [159].

Imbalanced datasets also present significant challenges. For instance, brain aging classification models often favor younger adults due to the scarcity of labeled data for older populations, reducing accuracy in predicting age-related neurodegeneration [160]. Similarly, in ICH segmentation, smaller hemorrhages are frequently underrepresented in the training data, leading to overfitting on larger lesions and poor generalization in subtle cases [161].

Proposed solutions: robust preprocessing pipelines tailored to specific tasks, such as motion correction for DWI in stroke imaging or intensity harmonization for Alzheimer's studies, are essential [162]. Addressing data imbalance through oversampling underrepresented cases or generating synthetic data using GANs has shown promise [163]. For instance, GANs have been used to simulate infarct lesions to segment stroke lesions, improving segmentation accuracy in noisy settings [164]. Caihua Wang et al. [165] proposed a hybrid framework consisting of multiple CNNs, and a linear SVM to make robust final predictions from limited data.

4.2. Population Variability and Cross-Site Generalization

Deep learning models often struggle to generalize across diverse populations and imaging sites due to domain shifts. For example, Alzheimer's disease classification models trained on data from a single scanner or region may perform poorly when tested on datasets from other regions, reflecting differences in demographics, genetic factors, or scanner properties [166]. In stroke lesion segmentation, differences in imaging protocols (e.g., different b-values in DWI) across institutions can cause domain mismatches, reducing model accuracy [167].

Proposed solutions: Federated learning frameworks enable training across multiple sites without sharing sensitive patient data, thereby exposing models to a broader demographic and scanner variability while preserving data privacy [87,168]. Transfer learning has also proven effective, allowing models pre-trained on one dataset to adapt to specific conditions, such as ICH or Alzheimer's progression [169].

4.3. Task-Specific Reliability in Segmentation and Classification

Segmentation and classification tasks in neuroimaging pose distinct reliability challenges. In segmentation, accurately delineating small or subtle lesions, such as small hematomas in ICH or small ischemic strokes, remains difficult due to the low contrast

between pathological and normal tissue [170]. For classification, models may rely on spurious correlations, such as scanner-specific noise, to predict conditions such as Alzheimer's disease or brain age [171]. Emergency settings exacerbate these issues, with low-quality scans reducing reliability in stroke segmentation. Moreover, models often fail to generalize to atypical stroke presentations, such as chronic infarcts with diffuse boundaries [172].

Proposed Solutions: Uncertainty-aware frameworks can identify subjects where predictions are less reliable, allowing clinicians to focus on areas of high confidence. For instance, Bayesian neural networks have been applied to ICH segmentation to estimate uncertainty in hemorrhage boundaries [173]. For classification, ensemble methods have reduced reliance on spurious correlations, improving robustness in Alzheimer's diagnosis across multi-site datasets [174].

5. Ablation Study: Robustness and Generalizability of Intracranial Hemorrhage Segmentation and Classification from Non-Contrast Head CT

Intracranial hemorrhage (ICH) is a life-threatening condition, as the accumulation of blood within the brain tissues can increase intracranial pressure, potentially leading to irreversible brain injury or death if not diagnosed and treated quickly. Computed tomography (CT) scans are the gold standard for initial diagnosis, as they provide rapid images of the brain and are highly effective in visualizing acute ICH. The early detection of a hemorrhage, as well as its location and its subtype, is crucial in preventing mortality and morbidity in patients with intracerebral hemorrhage. We evaluated the impact of various strategies on improving the segmentation and classification performance for five types of intracerebral hemorrhage (epidural hemorrhage (EDH), subdural hemorrhage (SDH), subarachnoid hemorrhage (SAH), intraventricular hemorrhage (IVH), and intraparenchymal hemorrhage (IPH)), as shown in Table 3.

To enhance the robustness and generalizability of deep learning models for ICH segmentation and classification, several key strategies have demonstrated significant improvements. Augmentation techniques, such as stochastic rotation, elastic deformation, and noise injection, have improved model generalization by simulating the realistic deformations commonly seen in clinical images. This approach has effectively reduced overfitting and improved performance for rare hemorrhage types such as EDH. Meanwhile, optimization strategies incorporating hybrid loss functions (such as Dice + Focal loss) have enhanced boundary delineation and improved model convergence in imbalanced datasets. Regularization has further stabilized the training process, especially when applied to models trained on CT datasets. Additionally, cross-validation using a 5-layer hierarchical scheme promotes model stability across different imaging centers, improving the detection of rare hemorrhage types by ensuring balanced data representation during training. Ensemble learning techniques improve performance in difficult cases characterized by ambiguous boundaries or subtle hemorrhage patterns, enhancing the model's resilience to noise and artifacts.

Finally, nnUNet and attention-based models demonstrated superior performance by leveraging all these mechanisms to capture long-range dependencies and spatial contexts. This improved feature aggregation significantly enhanced segmentation accuracy and classification accuracy, especially for complex hemorrhage subtypes such as subarachnoid hemorrhage and intraventricular hemorrhage. Together, these strategies enhance the robustness of the model and ensure improved performance across diverse clinical scenarios.

However, many promising techniques have yet to be deployed. Methods such as adversarial training, which enhances model resilience to perturbations, and domain adaptation, which mitigates performance degradation across different imaging centers and

scanner types, have yet to be applied in this context. Similarly, transfer learning—which leverages pretrained models to improve learning efficiency in data-limited situations—and self-supervised learning, which allows models to extract meaningful features from unlabeled data, have yet to be explored for ICH segmentation, and classification tasks. Combining these techniques could provide significant gains in model robustness, particularly for handling data variability, noise, and rare types of hemorrhage. Future research integrating these underutilized strategies could further advance the reliability of AI systems in clinical neuroimaging applications.

Table 3. A review of the robustness and generalizability of ICH segmentation and classification from non-contrast head CT.

Authors	Dataset	Results	Augmentation	Optimization	Cross-Validation	Ensemble Learning	Model Architectures
Segmentation (Dice as the main accuracy metric)							
Murat Yüce [175]	1508 CTs (QURE500+ RSNA 2019)	IPH = 0.59; IVH = 0.47; EDH = 0.35; SAH = 0.24; SDH = 0.34	✓	✓	✓		nnUNet
Zhegao Piao [176]	82,636 CTs, test 20%	IPH = 0.809; IVH = 0.742; EDH = 0.777; SAH = 0.545; SDH = 0.709	✓	✓			HarDNet based transformer
Chia Shuo Chang [177]	51 CTs, test 14.5%	IPH = 0.924; IVH = 0.858; EDH = 0.816; SAH = 0.567; SDH = 0.82	✓	✓			All Attention U-NET
Mayidili Nijati [178]	1157 CTs, test 200 CTs	IPH = 0.784; IVH = 0.680; EDH = 0.359; SAH = 0.337; SDH = 0.534	✓	✓			Sym-TransNet
Julia Kiewitz [179]	73 CTs, test 20 CTs	IPH = 0.743; IVH = 0.750; SAH = 0.686; SDH = 0.758	✓	✓	✓		nnUnet
Biao Wu [180]	192 CTs BHSD	IPH = 0.54; IVH = 0.51; EDH = 0.48; SAH = 0.215; SDH = 0.1523	✓	✓	✓		nnUnet
Classification (AUC as main outcome accuracy metric)							
Muhammad Asif [181]	13,334 CTs (CQ500 + RSNA), test 30%	IPH = 0.979; IVH = 0.977; EDH = 0.980; SAH = 0.976; SDH = 0.974	✓	✓			Res-Inc-LGBM
Snehalatha Umapathy [182]	133,709 slices (CQ500 + RSNA), test 14,600 slices	IPH = 0.99; IVH = 0.98; EDH = 0.99; SAH = 0.99; SDH = 0.99	✓	✓		✓	SE-ResNeXT, LSTM
Shanu Nizarudeen [183]	CQ500, 10%	IPH = 0.98; IVH = 0.98; EDH = 0.96; SAH = 0.98; SDH = 0.98	✓	✓			Attention-based RaNet

6. Discussion

Despite significant advances in neuroimaging segmentation and classification tasks undertaken by deep learning models, achieving robustness and generalization across diverse conditions remains a significant challenge. Deep learning models are typically sensitive to variations in input data, such as differences in scanner types, noise, artifacts, and acquisition protocols. This lack of robustness can limit the generalization of these models across different datasets and clinical settings. When models are trained on one dataset, they often fail to perform well on others due to the domain shift. Addressing this heterogeneity requires harmonization techniques and a shift toward federated learning approaches.

The generalization capability of deep learning models is a key challenge, especially in neuroimaging. These models often perform well on training data but struggle when applied to unseen data. In medical applications like brain tissue segmentation or characterization, this is a critical issue, as models need to be reliable across diverse patient populations and data sources. Moreover, the scarcity of labeled medical imaging data exacerbates the issue of generalization, as large, annotated datasets are difficult to acquire.

In classification, neuroimaging datasets often suffer from class imbalances, whereby pathological regions are much smaller than healthy brain tissue. This imbalance poses a challenge for deep learning models, as they can easily be overfitted to the more common classes. Additionally, labeled data for medical segmentation are costly and time-consuming to obtain, which limits the availability of the large datasets necessary for training robust models. One promising direction is the use of self-supervised learning, where models learn useful representations from unlabeled data. In the context of neuroimaging, self-supervised techniques can help to leverage large numbers of unlabeled medical images to improve generalization when the labeled data are limited.

Deep learning models are also susceptible to adversarial attacks, which can compromise their reliability in critical applications. To improve robustness to perturbations and domain shifts, adversarial training techniques are being explored. These methods train models to defend against adversarial attacks or variations in input data, making them more reliable across different clinical scenarios. Combining deep learning models with traditional machine learning approaches or incorporating domain knowledge into the architecture could lead to more reliable and interpretable systems. For example, hybrid systems that combine neural networks with rule-based systems could offer both high accuracy and explainability.

It is notable that inherent differences in image acquisition, resolution, contrast mechanisms, and artifacts significantly influence data quality and standardization between various neuroimaging modalities such as MRI, CT or PET. As a result, strategies for improving the robustness and generalizability of machine learning models may need to be tailored to the specific modality. For instance, MRI data may require harmonization techniques to address variability across scanners and protocols, whereas CT images might demand pre-processing to normalize differences in contrast administration or reconstruction algorithms. Therefore, understanding the modality-specific requirements is critical for developing robust and generalizable models for segmentation and classification tasks in neuroimaging.

In addition, some emerging methods in machine learning and related fields may have untapped potential for neuroimaging applications. For example, retrieval-augmented generation (RAG) [184] is a hybrid machine learning framework that combines retrieval mechanisms with generative models. While RAG is typically applied to natural language processing tasks, its principles can be extended to brain image classification by integrating retrieval mechanisms into the decision-making pipeline. This approach can improve interpretability, generalizability, and accuracy in neuroimaging tasks. Hyperbolic CNN has also shown improved generalizability and robustness compared to CNN [185]. Other

methods proposed to improve model performance include exploiting the asymmetries of brain scans to detect pathologies [186,187].

Despite these promising advances, many ethical challenges must be addressed to ensure the responsible use of artificial intelligence in medical imaging during data collection, development, and evaluation. Concerns about patient privacy, informed consent, and data ownership arise during data collection. Neuroimaging data often contain extremely sensitive information, making anonymization important to protect patient identities. Additionally, ensuring diverse and representative datasets is essential to preventing algorithmic bias, which may affect certain demographic groups. Informed consent procedures must also clarify how the data will be used, especially in cases where data uses may extend beyond the scope of the original study. During model development and evaluation, ethical concerns include data annotation integrity, model transparency, and performance fairness. Inconsistent labeling or informed annotation practices can reduce model generalizability. Developers must employ validation strategies to ensure that models perform reliably across diverse populations and clinical contexts. Furthermore, interpretability techniques such as SHAP, Grad-CAM, and feature attribution methods should be incorporated to enhance model interpretability, particularly for clinical decision making. Addressing these ethical challenges is key to ensuring that deep-learning-driven neuroimaging solutions are both effective and consistent with patient trust and societal benefit.

Future advances in neuroimaging depend on the development of transparent, reliable, and adaptable algorithms to ensure their integration into clinical workflows. Robust models must prioritize interpretability, allowing clinicians to trust and effectively use these tools. Equally important is adapting these algorithms to diverse patient populations, ensuring equitable healthcare outcomes. Multimodal imaging data fusion, combining insights from multiple imaging techniques, offers a path to significantly improving diagnostic accuracy and providing a more comprehensive understanding of neurological conditions.

To address challenges such as data privacy and heterogeneity, federated and distributed learning frameworks are becoming increasingly important. These approaches enable models to collaboratively learn from decentralized datasets without sharing sensitive information, ensuring data security while leveraging diverse sources. Furthermore, fostering collaboration between researchers, clinicians, and industry stakeholders is essential to linking technological advances to real-world needs. By addressing these challenges and leveraging collective expertise, neuroimaging can evolve into a more efficient, equitable, and patient-centered specialty.

One of the key challenges in evaluating and comparing various strategies aimed at improving robustness and generalizability is the variability in input datasets, performance metrics, and the specific tasks being addressed. These inconsistencies limit effective systematic reviews or the undertaking of meta-analyses. Therefore, in this article, we provided a comprehensive survey of the literature, highlighting notable approaches (Table 1) and summarizing their respective advantages and limitations (Table 2). Finally, our inclusion and exclusion criteria for articles may introduce bias in this review by limiting the scope to neuroimaging studies; meanwhile, many strategies applied to other body parts may also be translated to brain scans.

7. Conclusions

Robustness and generalizability in neuroimaging segmentation and classification are important challenges that directly impact the trustworthiness, reliability, and clinical applicability of these techniques. This review highlights a range of approaches to improving model performance in diverse and unpredictable real-world scenarios. However, several obstacles remain, including the computationally demanding nature of these strategies,

the need for the continuous monitoring of model performance in real-world settings, and the evolving nature of medical images, especially MRI sequences. By addressing these challenges and fostering interdisciplinary innovation, the field can move closer to realizing robust and generalizable neuroimaging tools with broad clinical impact.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedinformatics5020020/s1>, Supplementary Material Section S1: Feature size reduction techniques; Supplementary Material Section S2: Statistical assessment metrics; Supplementary Table S1. List of datasets used in studies cited in the article.

Funding: S.P. was supported by the Doris Duke Charitable Foundation (2020097), NIH (K23NS118056), and the NVIDIA Applied Research Accelerator Program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ViT	Vision transformers
CNN	Convolutional neural network
FGSM	Fast gradient sign method
PGD	Projected gradient descent
CW	Carlini and Wagner
CT	Computed tomography
MRI	Magnetic resonance imaging
fMRI	Functional magnetic resonance imaging
DWI	Diffusion-weighted imaging
IoU	Intersection over Union
DSC	Dice–Sørensen coefficient
HD	Hausdorff distance
AUC-ROC	Area under the curve of receiver operating characteristic
ICH	Intracerebral hemorrhage
IVH	Intraventricular hemorrhage
PHE	Perihematoma edema
Sen	Sensitivity
Acc	Accuracy
Spec	Specificity
BraTS	International brain tumor segmentation
SwinUNETR	Swin UNet TRansformers
CV	Cross-validation

References

1. Berson, E.R.; Aboian, M.S.; Malhotra, A.; Payabvash, S. Artificial Intelligence for Neuroimaging and Musculoskeletal Radiology: Overview of Current Commercial Algorithms. *Semin. Roentgenol.* **2023**, *58*, 178–183. [[CrossRef](#)] [[PubMed](#)]
2. Williams, K.S. Evaluations of artificial intelligence and machine learning algorithms in neurodiagnostics. *J. Neurophysiol.* **2024**, *131*, 825–831. [[CrossRef](#)]

3. Fernandez, J.-C.; Mounier, L.; Pachon, C.A. A Model-Based Approach for Robustness Testing. In *Proceedings of the IFIP International Conference on Testing of Communicating Systems, Montreal, QC, Canada, 31 May–2 June 2005*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2005; pp. 333–348.
4. Drenkow, N.; Sani, N.; Shpitser, I.; Unberath, M. A Systematic Review of Robustness in Deep Learning for Computer Vision: Mind the gap? *arXiv* **2021**, arXiv:2112.00639.
5. Zhu, Z.; Liu, F.; Chrysos, G.; Cevher, V. Robustness in deep learning: The good (width), the bad (depth), and the ugly (initialization). In *Proceedings of the Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 28 November–9 December 2022.
6. Freiesleben, T.; Grote, T. Beyond generalization: A theory of robustness in machine learning. *Synthese* **2023**, *202*, 109. [\[CrossRef\]](#)
7. Hendrycks, D.; Gimpel, K. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In *Proceedings of the International Conference on Learning Representations*, Toulon, France, 24–26 April 2017.
8. Kurakin, A.; Goodfellow, I.J.; Bengio, S. Adversarial examples in the physical world. *arXiv* **2016**, arXiv:1607.02533. [\[CrossRef\]](#)
9. Kawaguchi, K.; Kaelbling, L.P.; Bengio, Y. *Generalization in Deep Learning*; Cambridge University Press: Cambridge, UK, 2022.
10. Neyshabur, B.; Bhojanapalli, S.; McAllester, D.; Srebro, N. Exploring Generalization in Deep Learning. In *Proceedings of the Neural Information Processing Systems*, Long Beach, CA, USA, 4–9 December 2017.
11. Nagarajan, V. Explaining generalization in deep learning: Progress and fundamental limits. *arXiv* **2021**. [\[CrossRef\]](#)
12. Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding deep learning requires rethinking generalization. *arXiv* **2016**. [\[CrossRef\]](#)
13. Ying, X. An Overview of Overfitting and its Solutions. *J. Phys. Conf. Ser.* **2019**, *1168*, 022022. [\[CrossRef\]](#)
14. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
15. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
16. He, Y.; Guo, Y.; Lyu, J.; Ma, L.; Tan, H.; Zhang, W.; Ding, G.; Liang, H.; He, J.; Lou, X.; et al. Disorder-Free Data Are All You Need—Inverse Supervised Learning for Broad-Spectrum Head Disorder Detection. *NEJM AI* **2024**, *1*, A10a2300137. [\[CrossRef\]](#)
17. Ghassemi, M.; Oakden-Rayner, L.; Beam, A.L. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit. Health* **2021**, *1*, e745–e750. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Kelly, C.J.; Karthikesalingam, A.; Suleyman, M.; Corrado, G.; King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* **2019**, *17*, 195. [\[CrossRef\]](#)
19. Recht, B.; Roelofs, R.; Schmidt, L.; Shankar, V. Do ImageNet Classifiers Generalize to ImageNet? In *Proceedings of the International Conference on Machine Learning*, Long Beach, CA, USA, 9–15 June 2019; pp. 5389–5400.
20. Hendrycks, D.; Dietterich, T. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. *arXiv* **2019**, arXiv:1807.01697.
21. Barzamini, H.; Rahimi, M.; Shahzad, M.; Alhoori, H. Improving generalizability of ML-enabled software through domain specification. In *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*, Pittsburgh, PA, USA, 16–17 May 2022; pp. 181–192.
22. Degtiar, I.; Rose, S. A Review of Generalizability and Transportability. *Annu. Rev. Stat. Its Appl.* **2023**, *10*, 501–524. [\[CrossRef\]](#)
23. Fassia, M.K.; Balasubramanian, A.; Woo, S.; Vargas, H.A.; Hricak, H.; Konukoglu, E.; Becker, A.S. Deep Learning Prostate MRI Segmentation Accuracy and Robustness: A Systematic Review. *Radiol. Artif. Intell.* **2024**, *6*, e230138. [\[CrossRef\]](#)
24. Wang, S.; Veldhuis, R.; Brune, C.; Strisciuglio, N. A Survey on the Robustness of Computer Vision Models against Common Corruptions. *arXiv* **2023**, arXiv:2305.06024. [\[CrossRef\]](#)
25. Keele, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; School of Computer Science and Mathematics, Keele University: Keele, UK, 2007; pp. 1–2.
26. Rethlefsen, M.L.; Kirtley, S.; Waffenschmidt, S.; Ayala, A.P.; Moher, D.; Page, M.J.; Koffel, J.B.; Group, P.-S. PRISMA-S: An extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst. Rev.* **2021**, *10*, 39. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009.
28. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [\[CrossRef\]](#)
29. Ho, Y.; Wookey, S. The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. *IEEE Access* **2020**, *8*, 4806–4813. [\[CrossRef\]](#)
30. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2015**, arXiv:1412.698.
31. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [\[CrossRef\]](#)
32. Cortes, C.; Mohri, M.; Rostamizadeh, A. L2 regularization for learning kernels. In *Proceedings of the UAI '09: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, Montreal, QC, Canada, 18–21 June 2009; pp. 109–116.

33. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
34. Sergey Ioffe, C.S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
35. Yao, Y.; Rosasco, L.; Caponnetto, A. On Early Stopping in Gradient Descent Learning. *Constr. Approx.* **2007**, *26*, 289–315. [[CrossRef](#)]
36. Birgani, M.T.; Chegeni, N.; Birgani, F.F.; Fatehi, D.; Akbarizadeh, G.; Shams, A. Optimization of Brain Tumor MR Image Classification Accuracy Using Optimal Threshold, PCA and Training ANFIS with Different Repetitions. *J. Biomed. Phys. Eng.* **2019**, *9*, 189–198.
37. Nath, M.K.; Sahambi, J.S. Independent component analysis of functional MRI data. In Proceedings of the TENCON 2008—2008 IEEE Region 10 Conference, Hyderabad, India, 19–21 November 2008; pp. 1–6.
38. Abdumalikov, S.; Kim, J.; Yoon, Y. Performance Analysis and Improvement of Machine Learning with Various Feature Selection Methods for EEG-Based Emotion Classification. *Appl. Sci.* **2024**, *14*, 10511. [[CrossRef](#)]
39. Sadegh-Zadeh, S.A.; Sadeghzadeh, N.; Soleimani, O.; Ghidary, S.S.; Movahedi, S.; Mousavi, S.Y. Comparative analysis of dimensionality reduction techniques for EEG-based emotional state classification. *Am. J. Neurodegener. Dis.* **2024**, *13*, 23–33. [[CrossRef](#)]
40. Wang, J.; Perez, L. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv* **2017**. [[CrossRef](#)]
41. Hossain, T.; Zhang, M. MGAug: Multimodal Geometric Augmentation in Latent Spaces of Image Deformations. *arXiv* **2023**. [[CrossRef](#)]
42. Ramesh, J.; Dinsdale, N.; Yeung, P.H.; Namburete, A.I. Geometric Transformation Uncertainty for Improving 3D Fetal Brain Pose Prediction from Freehand 2D Ultrasound Videos. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2024, Marrakesh, Morocco, 6–10 October 2024.
43. Xiao, Y.; Decenciere, E.; Velasco-Forero, S.; Burdin, H.; Bornschlogl, T.; Bernerd, F.; Warrick, E.; Baldeweck, T. A New Color Augmentation Method for Deep Learning Segmentation of Histological Images. In Proceedings of the International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 886–890.
44. Akbiyik, M.E. Data Augmentation in Training CNNs: Injecting Noise to Images. *arXiv* **2023**. [[CrossRef](#)]
45. Dai, Y.; Qian, Y.; Lu, F.; Wang, B.; Gu, Z.; Wang, W.; Wan, J.; Zhang, Y. Improving adversarial robustness of medical imaging systems via adding global attention noise. *Comput. Biol. Med.* **2023**, *164*, 107251. [[CrossRef](#)]
46. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 13001–13008.
47. Zhang, X.; Liu, C.; Ou, N.; Zeng, X.; Zhuo, Z.; Duan, Y.; Xiong, X.; Yu, Y.; Liu, Z.; Liu, Y.; et al. CarveMix: A Simple Data Augmentation Method for Brain Lesion Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021; pp. 196–205.
48. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
49. Logan, R.; Williams, B.G.; da Silva, M.F.; Indani, A.; Scholnicov, N.; Ganguly, A.; Miller, S.J. Deep Convolutional Neural Networks with Ensemble Learning and Generative Adversarial Networks for Alzheimer's Disease Image Data Classification. *Front. Aging Neurosci.* **2021**, *13*, 720226. [[CrossRef](#)] [[PubMed](#)]
50. Breiman, L. *Bias, Variance, and Arcing Classifiers*; Statistics Department, University of California at Berkeley: Berkeley, CA, USA, 1996.
51. Nguyen, D.; Nguyen, H.; Ong, H.; Le, H.; Ha, H.; Duc, N.T.; Ngo, H.T. Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer's disease. *IBRO Neurosci. Rep.* **2022**, *13*, 255–263. [[CrossRef](#)]
52. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [[CrossRef](#)]
53. Rumala, D.J.; van Ooijen, P.; Rachmadi, R.F.; Sensusiaty, A.D.; Purnama, I.K.E. Deep-Stacked Convolutional Neural Networks for Brain Abnormality Classification Based on MRI Images. *J. Digit. Imaging.* **2023**, *36*, 1460–1479. [[CrossRef](#)] [[PubMed](#)]
54. Hosny, K.M.; Mohammed, M.A.; Salama, R.A.; Elshewey, A.M. Explainable ensemble deep learning-based model for brain tumor detection and classification. *Neural Comput. Appl.* **2024**, *37*, 1289–1306. [[CrossRef](#)]
55. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015.
56. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**. [[CrossRef](#)]
57. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
58. Zhao, L.; Wu, Z.; Dai, H.; Liu, Z.; Zhang, T.; Zhu, D.; Liu, T. Embedding Human Brain Function via Transformer. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2022, Singapore, 18–22 September 2022; pp. 366–375.

59. Zeineldin, R.A.; Karar, M.E.; Elshaer, Z.; Coburger, J.; Wirtz, C.R.; Burgert, O.; Mathis-Ullrich, F. Explainable hybrid vision transformers and convolutional network for multimodal glioma segmentation in brain MRI. *Sci. Rep.* **2024**, *14*, 3713. [[CrossRef](#)]
60. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and Harnessing Adversarial Examples. *arXiv* **2014**, arXiv:1412.6572. [[CrossRef](#)]
61. Kurakin, A.; Goodfellow, I.; Bengio, S. Adversarial Machine Learning at Scale. *arXiv* **2016**. [[CrossRef](#)]
62. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. Towards Deep Learning Models Resistant to Adversarial Attacks. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
63. Joel, M.Z.; Umrao, S.; Chang, E.; Choi, R.; Yang, D.X.; Duncan, J.S.; Omuro, A.; Herbst, R.; Krumholz, H.M.; Aneja, S. Using Adversarial Images to Assess the Robustness of Deep Learning Models Trained on Diagnostic Images in Oncology. *JCO Clin. Cancer Inform.* **2022**, *6*, e2100170. [[CrossRef](#)]
64. Liu, Z.; Zhang, J.; Jog, V.; Loh, P.-L.; McMillan, A.B. Robustifying Deep Networks for Medical Image Segmentation. *J. Digit. Imaging* **2021**, *34*, 1279–1293. [[CrossRef](#)] [[PubMed](#)]
65. Villegas-Ch, W.; Jaramillo-Alcázar, A.; Luján-Mora, S. Evaluating the Robustness of Deep Learning Models against Adversarial Attacks: An Analysis with FGSM, PGD and CW. *Big Data Cogn. Comput.* **2024**, *8*, 8. [[CrossRef](#)]
66. Carlini, N.; Wagner, D. Towards Evaluating the Robustness of Neural Networks. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2017; pp. 39–57.
67. Athalye, A.; Carlini, N.; Wagner, D. Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. In Proceedings of the International Conference on Machine Learning Conference, Stockholm, Sweden, 10–15 July 2018.
68. Tramèr, F.; Kurakin, A.; Papernot, N.; Goodfellow, I.; Boneh, D.; McDaniel, P. Ensemble Adversarial Training: Attacks and Defenses. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
69. Buchheim, C.; Kurtz, J. Min-max-min robustness: A new approach to combinatorial optimization under uncertainty based on multiple solutions. *Electron. Notes Discret. Math.* **2016**, *52*, 45–52. [[CrossRef](#)]
70. Esfahani, P.M.; Kuhn, D. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Program.* **2018**, *171*, 115–166. [[CrossRef](#)]
71. Ganin, Y.; Lempitsky, V. Unsupervised Domain Adaptation by Backpropagation. In Proceedings of the 32nd International Conference on Machine Learning, PMLR, Lille, France, 7–9 July 2015; pp. 1180–1189.
72. Al Khalil, Y.; Ayaz, A.; Lorenz, C.; Weese, J.; Pluim, J.; Breeuwer, M. Multi-modal brain tumor segmentation via conditional synthesis with Fourier domain adaptation. *Comput. Med. Imaging Graph.* **2024**, *112*, 102332. [[CrossRef](#)]
73. Gopinath, K.; Hoopes, A.; Alexander, D.C.; Arnold, S.E.; Balbastre, Y.; Billot, B.; Casamitjana, A.; Cheng, Y.; Chua, R.Y.Z.; Edlow, B.L.; et al. Synthetic data in generalizable, learning-based neuroimaging. *Imaging Neurosci.* **2024**, *2*, 1–22. [[CrossRef](#)]
74. Adragna, R.; Creager, E.; Madras, D.; Zemel, R. Fairness and Robustness in Invariant Learning: A Case Study in Toxicity Classification. *arXiv* **2020**. [[CrossRef](#)]
75. Yu, W.; Huang, Z.; Zhang, J.; Shan, H. SAN-Net: Learning generalization to unseen sites for stroke lesion segmentation with self-adaptive normalization. *Comput. Biol. Med.* **2023**, *156*, 106717. [[CrossRef](#)]
76. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In Proceedings of the NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QV, USA, 8–13 December 2014; pp. 3320–3328.
77. Long, M.; Cao, Y.; Wang, J.; Jordan, M. Learning transferable features with deep adaptation networks. In Proceedings of the ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 97–105.
78. Taşçı, B. Attention Deep Feature Extraction from Brain MRIs in Explainable Mode: DGXAINet. *Diagnostics* **2023**, *13*, 895. [[CrossRef](#)] [[PubMed](#)]
79. Krishnapriya, S.; Karuna, Y. Pre-trained deep learning models for brain MRI image classification. *Front. Hum. Neurosci.* **2023**, *17*, 1150120. [[CrossRef](#)] [[PubMed](#)]
80. Vimala, B.B.; Srinivasan, S.; Mathivanan, S.K.; Mahalakshmi; Jayagopal, P.; Dalu, G.T. Detection and classification of brain tumor using hybrid deep learning models. *Heliyon* **2023**, *13*, 23029. [[CrossRef](#)]
81. Seetha, J.; Raja, S.S. Brain Tumor Classification Using Convolutional Neural Networks. *Biomed. Pharmacol. J.* **2018**, *11*, 1457. [[CrossRef](#)]
82. Hu, S.Y.; Beers, A.; Chang, K.; Höbel, K.; Campbell, J.P.; Erdogumus, D.; Ioannidis, S.; Dy, J.; Chiang, M.F.; Kalpathy-Cramer, J.; et al. Deep feature transfer between localization and segmentation tasks. *arXiv* **2018**. [[CrossRef](#)]
83. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning Transferable Architectures for Scalable Image Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8697–8710.

84. Santoro, A.; Bartunov, S.; Botvinick, M.; Wierstra, D.; Lillicrap, T. Meta-learning with memory-augmented neural networks. In Proceedings of the ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1842–1850.
85. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR, Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
86. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated Machine Learning. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–19. [\[CrossRef\]](#)
87. Sadilek, A.; Liu, L.; Nguyen, D.; Kamruzzaman, M.; Serghiou, S.; Rader, B.; Ingerman, A.; Mellem, S.; Kairouz, P.; Nsoesie, E.O.; et al. Privacy-first health research with federated learning. *Npj Digit. Med.* **2021**, *4*, 132. [\[CrossRef\]](#)
88. Liu, Y.; Lian, L.; Zhang, E.; Xu, L.; Xiao, C.; Zhong, X.; Li, F.; Jiang, B.; Dong, Y.; Ma, L.; et al. Mixed-UNet: Refined class activation mapping for weakly-supervised semantic segmentation with multi-scale inference. *Front. Comput. Sci.* **2022**, *4*, 1036934. [\[CrossRef\]](#)
89. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [\[CrossRef\]](#)
90. Wang, Y.; Katsaggelos, A.K.; Wang, X.; Parrish, T.B. A deep symmetry convnet for stroke lesion segmentation. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 111–115.
91. Rockafellar, R.T.; Wets, R.J.B. *Variational Analysis*; Springer: Berlin/Heidelberg, Germany, 1998.
92. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit Lett.* **2006**, *27*, 861–874. [\[CrossRef\]](#)
93. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Int. J. Mach. Learn. Technol.* **2011**, *2*, 37–63. [\[CrossRef\]](#)
94. Hicks, S.A.; Strümke, I.; Thambawita, V.; Hammou, M.; Riegler, M.A.; Halvorsen, P.; Parasa, S. On evaluation metrics for medical applications of artificial intelligence. *Sci. Rep.* **2022**, *12*, 5979. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Demsar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
96. Wilcoxon, F. Individual Comparisons by Ranking Methods. In *Breakthroughs in Statistics*; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 1992.
97. Tibshirani, R.J.; Efron, B. *An Introduction to the Bootstrap*; Chapman & Hall/CRC: Boca Raton, FL, USA, 1993.
98. You, S.; Wiest, R.; Reyes, M. SaRF: Saliency regularized feature learning improves MRI sequence classification. *Comput. Methods Programs Biomed.* **2024**, *243*, 107867. [\[CrossRef\]](#)
99. Younis, E.M.; Mahmoud, M.N.; Albarrak, A.M.; Ibrahim, I.A. A Hybrid Deep Learning Model with Data Augmentation to Improve Tumor Classification Using MRI Images. *Diagnostics* **2024**, *14*, 2710. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Zhao, X.; Chen, K.; Wu, G.; Zhang, G.; Zhou, X.; Lv, C.; Wu, S.; Chen, Y.; Xie, G.; Yao, Z. Deep learning shows good reliability for automatic segmentation and volume measurement of brain hemorrhage, intraventricular extension, and peripheral edema. *Eur. Radiol.* **2020**, *31*, 5012–5020. [\[CrossRef\]](#)
101. Kok, Y.E.; Pszczolkowski, S.; Law, Z.K.; Ali, A.; Krishnan, K.; Bath, P.M.; Sprigg, N.; Dineen, R.A.; French, A.P. Semantic Segmentation of Spontaneous Intracerebral Hemorrhage, Intraventricular Hemorrhage, and Associated Edema on CT Images Using Deep Learning. *Radiol. Artif. Intell.* **2022**, *4*, e220096. [\[CrossRef\]](#)
102. Grøvik, E.; Yi, D.; Iv, M.; Tong, E.; Nilsen, L.B.; Latysheva, A.; Saxhaug, C.; Jacobsen, K.D.; Helland, Å.; Emblem, K.E.; et al. Handling missing MRI sequences in deep learning segmentation of brain metastases: A multicenter study. *NPJ Digit. Med.* **2021**, *4*, 33. [\[CrossRef\]](#)
103. Amin, J.; Sharif, M.; Anjum, M.A.; Raza, M.; Bukhari, S.A.C. Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI. *Cogn. Syst. Res.* **2020**, *59*, 304–311. [\[CrossRef\]](#)
104. Ali, R.R.; Yaacob, N.M.; Alqaryouti, M.H.; Sadeq, A.E.; Doheir, M.; Iqtait, M.; Rachmawanto, E.H.; Sari, C.A.; Yaacob, S.S. Learning Architecture for Brain Tumor Classification Based on Deep Convolutional Neural Network: Classic and ResNet50. *Diagnostics* **2025**, *15*, 624. [\[CrossRef\]](#)
105. Yurtsever, M.; Atay, Y.; Arslan, B.; Sagioglu, S. Development of brain tumor radiogenomic classification using GAN-based augmentation of MRI slices in the newly released gazi brains dataset. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 285. [\[CrossRef\]](#)
106. Celik, F.; Celik, K.; Celik, A. Enhancing brain tumor classification through ensemble attention mechanism. *Sci. Rep.* **2024**, *14*, 22260. [\[CrossRef\]](#)
107. Rajput, S.; Kapdi, R.; Roy, M.; Raval, M.S. A triplanar ensemble model for brain tumor segmentation with volumetric multiparametric magnetic resonance images. *Healthc. Anal.* **2024**, *5*, 100307. [\[CrossRef\]](#)
108. Saeed, T.; Khan, M.A.; Hamza, A.; Shabaz, M.; Khan, W.Z.; Alhayan, F.; Jamel, L.; Baili, J. Neuro-XAI: Explainable deep learning framework based on deeplabV3+ and bayesian optimization for segmentation and classification of brain tumor in MRI scans. *J. Neurosci. Methods* **2024**, *410*, 110247. [\[CrossRef\]](#)
109. Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H.R.; Xu, D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. *arXiv* **2022**. [\[CrossRef\]](#)

110. Van, M.H.; Carey, A.N.; Wu, X. Robust Influence-Based Training Methods for Noisy Brain MRI. In Proceedings of the Advances in Knowledge Discovery and Data Mining, PAKDD 2024, Taipei, Taiwan, 7–10 May 2024; pp. 246–257.
111. Tran, A.T.; Karam, G.A.; Zeevi, D.; Qureshi, A.I.; Malhotra, A.; Majidi, S.; Murthy, S.B.; Park, S.; Kontos, D.; Falcone, G.J.; et al. Improving the Robustness of Deep-Learning Models in Predicting Hematoma Expansion from Admission Head CT. *Am. J. Neuroradiol.* **2025**, ajnr.A8650. [\[CrossRef\]](#)
112. Zhou, S.; Cox, C.R.; Lu, H. Improving whole-brain neural decoding of fMRI with domain adaptation. In Proceedings of the International Workshop on Machine Learning in Medical Imaging (MLMI), Shenzhen, China, 13 October 2019; pp. 265–273.
113. Dong, D.; Fu, G.; Li, J.; Pei, Y.; Chen, Y. An unsupervised domain adaptation brain CT segmentation method across image modalities and diseases. *Expert Syst. Appl.* **2022**, 207, 118016. [\[CrossRef\]](#)
114. Awang, M.K.; Rashid, J.; Ali, G.; Hamid, M.; Mahmoud, S.F.; Saleh, D.I.; Ahmad, H.I. Classification of Alzheimer disease using DenseNet-201 based on deep transfer learning technique. *PLoS ONE* **2024**, 19, 0304995. [\[CrossRef\]](#)
115. Albalawi, E.; TR, M.; Thakur, A.; Kumar, V.V.; Gupta, M.; Khan, S.B.; Almusharraf, A. Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. *BMC Med. Imaging* **2024**, 24, 110. [\[CrossRef\]](#)
116. Nimeshika, G.N.; Subitha, D. Enhancing Alzheimer’s disease classification through split federated learning and GANs for imbalanced datasets. *PeerJ Comput. Sci.* **2024**, 10, e2459. [\[CrossRef\]](#)
117. Shi, C.; Wang, Y.; Wu, Y.; Chen, S.; Hu, R.; Zhang, M.; Qiu, B.; Wang, X. Self-supervised pretraining improves the performance of classification of task functional magnetic resonance imaging. *Front. Neurosci.* **2023**, 17, 1199312. [\[CrossRef\]](#)
118. Gryshchuk, V.; Singh, D.; Teipel, S.; Dyrba, M.; ADNI; AIBL; FTL DNI Study Groups. Contrastive Self-supervised Learning for Neurodegenerative Disorder Classification. *medRxiv* **2024**. [\[CrossRef\]](#)
119. Correia de Verdier, M.; Saluja, R.; Gagnon, L.; LaBella, D.; Baid, U.; Hoda Tahon, N.; Foltyn-Dumitru, M.; Zhang, J.; Alafif, M.; Baig, S.; et al. The 2024 Brain Tumor Segmentation (BraTS) Challenge: Glioma Segmentation on Post-treatment MRI. *arXiv* **2024**, arXiv:2405.18368. [\[CrossRef\]](#)
120. Jack, C.R., Jr.; Bernstein, M.A.; Fox, N.C.; Thompson, P.; Alexander, G.; Harvey, D.; Borowski, B.; Britson, P.J.; J, L.W.; Ward, C.; et al. The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* **2008**, 27, 685–691. [\[CrossRef\]](#)
121. Hooper, S.M.; Dunnmon, J.A.; Lungren, M.P.; Mastrodicasa, D.; Rubin, D.L.; Re, C.; Wang, A.; Patel, B.N. Impact of Upstream Medical Image Processing on Downstream Performance of a Head CT Triage Neural Network. *Radiol. Artif. Intell.* **2021**, 3, e200229. [\[CrossRef\]](#)
122. Hernandez Petzsche, M.R.; de la Rosa, E.; Hanning, U.; Wiest, R.; Valenzuela, W.; Reyes, M.; Meyer, M.; Liew, S.L.; Kofler, F.; Ezhov, I.; et al. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Sci. Data* **2022**, 9, 762. [\[CrossRef\]](#)
123. Nickparvar, M. Brain Tumor MRI Dataset. Available online: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset> (accessed on 6 April 2025).
124. Chilamkurthy, S.; Ghosh, R.; Tanamala, S.; Biviji, M.; Campeau, N.G.; Venugopal, V.K.; Mahajan, V.; Rao, P.; Warier, P. Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* **2018**, 392, 2388–2396. [\[CrossRef\]](#)
125. Sprigg, N.; Flaherty, K.; Appleton, J.P.; Al-Shahi Salman, R.; Bereczki, D.; Beridze, M.; Christensen, H.; Ciccone, A.; Collins, R.; Czlonkowska, A.; et al. Tranexamic acid for hyperacute primary IntraCerebral Haemorrhage (TICH-2): An international randomised, placebo-controlled, phase 3 superiority trial. *Lancet* **2018**, 391, 2107–2115. [\[CrossRef\]](#)
126. Qureshi, A.I.; Palesch, Y.Y.; Barsan, W.G.; Hanley, D.F.; Hsu, C.Y.; Martin, R.L.; Moy, C.S.; Silbergleit, R.; Steiner, T.; Suarez, J.I.; et al. Intensive Blood-Pressure Lowering in Patients with Acute Cerebral Hemorrhage. *N. Engl. J. Med.* **2016**, 375, 1033–1043. [\[CrossRef\]](#)
127. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. In Proceedings of the International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014.
128. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Montreal, QC, USA, 8–13 December 2014; pp. 2672–2680.
129. Bibi, N.; Wahid, F.; Ma, Y.; Ali, S.; Abbasi, I.A.; Alkhayyat, A. A Transfer Learning-Based Approach for Brain Tumor Classification. *IEEE Access* **2024**, 12, 111218–111238. [\[CrossRef\]](#)
130. Qin, C.; Li, B.; Han, B. Fast brain tumor detection using adaptive stochastic gradient descent on shared-memory parallel environment. *Eng. Appl. Artif. Intell.* **2023**, 120, 105816. [\[CrossRef\]](#)
131. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the IJCAI’95: Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; pp. 1137–1143.

132. Badža, M.M.; Barjaktarović, M. Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network. *Appl. Sci.* **2020**, *10*, 1999. [\[CrossRef\]](#)
133. Rastogi, D.; Johri, P.; Tiwari, V.; Elngar, A.A. Multi-class classification of brain tumour magnetic resonance images using multi-branch network with inception block and five-fold cross validation deep learning framework. *Biomed. Signal Process. Control* **2024**, *88*, 105602. [\[CrossRef\]](#)
134. Liu, J.; Deng, F.; Yuan, G.; Yang, C.; Song, H.; Luo, L. An Efficient CNN for Radiogenomic Classification of Low-Grade Gliomas on MRI in a Small Dataset. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 8856789. [\[CrossRef\]](#)
135. Taher, F.; Shoaib, M.R.; Emara, H.M.; Abdelwahab, K.M.; El-Samie, F.E.A.; Haweel, M.T. Efficient framework for brain tumor detection using different deep learning techniques. *Front. Public Health* **2022**, *10*, 959667. [\[CrossRef\]](#)
136. Usman, K.; Rajpoot, K. Brain tumor classification from multi-modality MRI using wavelets and machine learning. *Pattern Anal. Appl.* **2017**, *20*, 871–881. [\[CrossRef\]](#)
137. Allgaier, J.; Pryss, R. Cross-Validation Visualized: A Narrative Guide to Advanced Methods. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1378–1388. [\[CrossRef\]](#)
138. Pati, S.; Thakur, S.P.; Hamamcı, İ.E.; Baid, U.; Baheti, B.; Bhalerao, M.; Güley, O.; Mouchtaris, S.; Lang, D.; Thermos, S.; et al. GaNDLF: The generally nuanced deep learning framework for scalable end-to-end clinical workflows. *Commun. Eng.* **2023**, *2*, 23. [\[CrossRef\]](#)
139. Marklund, H.; Xie, S.M.; Zhang, M.; Balsubramani, A.; Hu, W.; Yasunaga, M.; Phillips, R.L.; Beery, S.; Leskovec, J.; Kundaje, A.; et al. WILDS: A Benchmark of in-the-Wild Distribution Shifts. *arXiv* **2020**. [\[CrossRef\]](#)
140. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6629–6640.
141. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training GANs. In Proceedings of the NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 2234–2242.
142. Barati, B.; Erfaninejad, M.; Khanbabaie, H. Evaluation of effect of optimizers and loss functions on prediction accuracy of brain tumor type using a Light neural network. *Biomed. Signal Process. Control* **2025**, *103*, 107409. [\[CrossRef\]](#)
143. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [\[CrossRef\]](#)
144. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [\[CrossRef\]](#)
145. Liu, S.; Liu, S.; Cai, W.; Che, H.; Pujol, S.; Kikinis, R.; Feng, D.; Fulham, M.J. ADNI. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1132–1140. [\[CrossRef\]](#)
146. Balaji, N.S.; Hemachandran, M.; Jansi, R. Precision Brain Tumor Detection Using Integrated Batch Normalization. In Proceedings of the 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 12–14 April 2024; pp. 438–444.
147. Alnowami, M.; Taha, E.; Alsebaei, S.; Anwar, S.M.; Alhawsawi, A. MR image normalization dilemma and the accuracy of brain tumor classification model. *J. Radiat. Res. Appl. Sci.* **2022**, *15*, 33–39. [\[CrossRef\]](#)
148. Mok, T.C.W.; Chung, A.C.S. Learning Data Augmentation for Brain Tumor Segmentation with Coarse-to-Fine Generative Adversarial Networks. In Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Granada, Spain, 16 September 2018; pp. 70–80.
149. Alsaif, H.; Guesmi, R.; Alshammari, B.M.; Hamrouni, T.; Guesmi, T.; Alzamil, A.; Belguesmi, L. A Novel Data Augmentation-Based Brain Tumor Detection Using Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 3773. [\[CrossRef\]](#)
150. Aurna, N.F.; Abu Yousuf, M.; Abu Taher, K.; 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.* **2022**, *146*, 105539. [\[CrossRef\]](#)
151. Cheng, G.; Ji, H. Adversarial Perturbation on MRI Modalities in Brain Tumor Segmentation. *IEEE Access* **2020**, *8*, 206009–206015. [\[CrossRef\]](#)
152. Joel, M.Z.; Avesta, A.; Yang, D.X.; Zhou, J.-G.; Omuro, A.; Herbst, R.S.; Krumholz, H.M.; Aneja, S. Comparing Detection Schemes for Adversarial Images against Deep Learning Models for Cancer Imaging. *Cancers* **2023**, *15*, 1548. [\[CrossRef\]](#)
153. Han, Y.; Yoo, J.; Kim, H.H.; Shin, H.J.; Sung, K.; Ye, J.C. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magn. Reson. Med.* **2018**, *80*, 1189–1205. [\[CrossRef\]](#)
154. Dou, Q.; Ouyang, C.; Chen, C.; Chen, H.; Heng, P.-A. Unsupervised Cross-Modality Domain Adaptation of ConvNets for Biomedical Image Segmentations with Adversarial Loss. In Proceedings of the IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 691–697.
155. Deepak, S.; Ameer, P. Brain tumor classification using deep CNN features via transfer learning. *Comput. Biol. Med.* **2019**, *111*, 103345. [\[CrossRef\]](#)

156. Li, H.; Parikh, N.A.; He, L. A Novel Transfer Learning Approach to Enhance Deep Neural Network Classification of Brain Functional Connectomes. *Front. Neurosci.* **2018**, *12*, 491. [\[CrossRef\]](#)
157. Power, J.D.; Barnes, K.A.; Snyder, A.Z.; Schlaggar, B.L.; Petersen, S.E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **2012**, *59*, 2142–2154. [\[CrossRef\]](#)
158. Reuter, M.; Rosas, H.D.; Fischl, B. Highly accurate inverse consistent registration: A robust approach. *Neuroimage* **2010**, *53*, 1181–1196. [\[CrossRef\]](#)
159. Song, Y.-H.; Yi, J.-Y.; Noh, Y.; Jang, H.; Seo, S.W.; Na, D.L.; Seong, J.-K. On the reliability of deep learning-based classification for Alzheimer's disease: Multi-cohorts, multi-vendors, multi-protocols, and head-to-head validation. *Front. Neurosci.* **2022**, *16*, 851871. [\[CrossRef\]](#)
160. Varzandian, A.; Razo, M.A.S.; Sanders, M.R.; Atmakuru, A.; Di Fatta, G.; Biomarkers, T.A.I. Classification-Biased Apparent Brain Age for the Prediction of Alzheimer's Disease. *Front. Neurosci.* **2021**, *15*, 673120. [\[CrossRef\]](#)
161. Angkurawaranon, S.; Sanorsien, N.; Unrisong, K.; Inkeaw, P.; Sripan, P.; Khumrin, P.; Angkurawaranon, C.; Vaniyapong, T.; Chitapanarux, I. A comparison of performance between a deep learning model with residents for localization and classification of intracranial hemorrhage. *Sci. Rep.* **2023**, *12*, 9975. [\[CrossRef\]](#)
162. Do, L.-N.; Baek, B.H.; Kim, S.K.; Yang, H.-J.; Park, I.; Yoon, W. Automatic Assessment of ASPECTS Using Diffusion-Weighted Imaging in Acute Ischemic Stroke Using Recurrent Residual Convolutional Neural Network. *Diagnostics* **2020**, *10*, 803. [\[CrossRef\]](#)
163. Sharma, A.; Singh, P.K.; Chandra, R. SMOTified-GAN for Class Imbalanced Pattern Classification Problems. *IEEE Access* **2022**, *10*, 30655–30665. [\[CrossRef\]](#)
164. Wang, S.; Chen, Z.; You, S.; Wang, B.; Shen, Y.; Lei, B. Brain stroke lesion segmentation using consistent perception generative adversarial network. *Neural Comput. Appl.* **2022**, *34*, 8657–8669. [\[CrossRef\]](#)
165. Wang, C.; Li, Y.; Tsuboshita, Y.; Sakurai, T.; Goto, T.; Yamaguchi, H.; Yamashita, Y.; Sekiguchi, A.; Tachimori, H.; Hisateru Tachimori for the Alzheimer's Disease Neuroimaging Initiative. A high-generalizability machine learning framework for predicting the progression of Alzheimer's disease using limited data. *NPJ Digit. Med.* **2022**, *5*, 43. [\[CrossRef\]](#)
166. Lu, B.; Li, H.-X.; Chang, Z.-K.; Li, L.; Chen, N.-X.; Zhu, Z.-C.; Zhou, H.-X.; Li, X.-Y.; Wang, Y.-W.; Cui, S.-X.; et al. A practical Alzheimer's disease classifier via brain imaging-based deep learning on 85,721 samples. *J. Big Data* **2022**, *1*, 101. [\[CrossRef\]](#)
167. de la Rosa, E.; Reyes, M.; Liew, S.L.; Hutton, A.; Wiest, R.; Kaesmacher, J.; Hanning, U.; Hakim, A.; Zubal, R.; Valenzuela, W.; et al. A Robust Ensemble Algorithm for Ischemic Stroke Lesion Segmentation: Generalizability and Clinical Utility Beyond the ISLES Challenge. *arXiv* **2024**, arXiv:2403.19425. [\[CrossRef\]](#)
168. Sheller, M.J.; Edwards, B.; Reina, G.A.; Martin, J.; Pati, S.; Kotrotsou, A.; Milchenko, M.; Xu, W.; Marcus, D.; Colen, R.R.; et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Sci. Rep.* **2020**, *10*, 12598. [\[CrossRef\]](#)
169. Boudi, A.; He, J.; El Kader, I.A. Enhancing Alzheimer's Disease Classification with Transfer Learning: Finetuning a Pre-trained Algorithm. *Curr. Med. Imaging* **2024**, *20*, e15734056305633. [\[CrossRef\]](#)
170. Kim, H.J.; Roh, H.G. Imaging in Acute Anterior Circulation Ischemic Stroke: Current and Future. *Neurointervention* **2022**, *17*, 2–17. [\[CrossRef\]](#)
171. Qiu, S.; Joshi, P.S.; Miller, M.I.; Xue, C.; Zhou, X.; Karjadi, C.; Chang, G.H.; Joshi, A.S.; Dwyer, B.; Zhu, S.; et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain* **2020**, *143*, 1920–1933. [\[CrossRef\]](#)
172. Balzano, R.F.; Mannatrizio, D.; Castorani, G.; Perri, M.; Pennelli, A.M.; Izzo, R.; Papolizio, T.; Guglielmi, G. Imaging of Cerebral Microbleeds: Primary Patterns and Differential Diagnosis. *Curr. Radiol. Rep.* **2021**, *9*, 15. [\[CrossRef\]](#)
173. Sharrock, M.F.; Mould, W.A.; Hildreth, M.; Ryu, E.P.; Walborn, N.; Awad, I.A.; Hanley, D.F.; Muschelli, J. Bayesian Deep Learning Outperforms Clinical Trial Estimators of Intracerebral and Intraventricular Hemorrhage Volume. *J. Neuroimaging* **2023**, *32*, 968–976. [\[CrossRef\]](#)
174. Pan, D.; Zeng, A.; Jia, L.; Huang, Y.; Frizzell, T.; Song, X. Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning. *Front. Neurosci.* **2020**, *14*, 259. [\[CrossRef\]](#)
175. Yüce, M.; Öztürk, S.; Pamuk, G.G.; Varlık, C.; Cimilli, A.T. Automatic segmentation and volumetric analysis of intracranial hemorrhages in brain CT images. *Eur. J. Radiol.* **2025**, *184*, 111952. [\[CrossRef\]](#)
176. Piao, Z.; Gu, Y.H.; Jin, H.; Yoo, S.J. Intracerebral hemorrhage CT scan image segmentation with HarDNet based transformer. *Sci. Rep.* **2023**, *13*, 7208. [\[CrossRef\]](#) [\[PubMed\]](#)
177. Chang, C.S.; Chang, T.S.; Yan, J.L.; Ko, L. All Attention U-NET for Semantic Segmentation of Intracranial Hemorrhages In Head CT Images. In Proceedings of the IEEE Biomedical Circuits and Systems Conference (BioCAS), Taipei, Taiwan, 13–15 October 2022; pp. 600–604.
178. Nijati, M.; Tuersun, A.; Zhang, Y.; Yuan, Q.; Gong, P.; Abulizi, A.; Tuoheti, A.; Abulaiti, A.; Zou, X. A symmetric prior knowledge based deep learning model for intracerebral hemorrhage lesion segmentation. *Front. Physiol.* **2022**, *13*, 977427. [\[CrossRef\]](#)

179. Kiewitz, J.; Aydin, O.U.; Hilbert, A.; Gultom, M.; Nouri, A.; Khalil, A.A.; Vajkoczy, P.; Tanioka, S.; Ishida, F.; Dengler, N.F.; et al. Deep Learning-based Multiclass Segmentation in Aneurysmal Subarachnoid Hemorrhage. *Front. Neurol.* **2024**, *15*, 1490216. [[CrossRef](#)]
180. Wu, B.; Xie, Y.; Zhang, Z.; Ge, J.; Yaxley, K.; Bahadir, S.; Wu, Q.; Liu, Y.; To, M.S. BHSD: A 3D Multi-class Brain Hemorrhage Segmentation Dataset. In Proceedings of the Machine Learning in Medical Imaging: 14th International Workshop, MLMI, MICCAI, Vancouver, BC, Canada, 8 October 2023. Proceedings, Part I.
181. Asif, M.; Shah, M.A.; Khattak, H.A.; Mussadiq, S.; Ahmed, E.; Nasr, E.A.; Rauf, H.T. Intracranial Hemorrhage Detection Using Parallel Deep Convolutional Models and Boosting Mechanism. *Diagnostics* **2023**, *13*, 652. [[CrossRef](#)] [[PubMed](#)]
182. Umaphathy, S.; Murugappan, M.; Bharathi, D.; Thakur, M. Automated Computer-Aided Detection and Classification of Intracranial Hemorrhage Using Ensemble Deep Learning Techniques. *Diagnostics* **2023**, *13*, 2987. [[CrossRef](#)] [[PubMed](#)]
183. Nizarudeen, S.; Shanmughavel, G.R. Comparative analysis of ResNet, ResNet-SE, and attention-based RaNet for hemorrhage classification in CT images using deep learning. *Biomed. Signal Process. Control* **2024**, *8*, 105672. [[CrossRef](#)]
184. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.T.; Rocktäschel, T.; et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In Proceedings of the NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2020; pp. 9459–9474.
185. Ayubcha, C.; Sajed, S.; Omara, C.; Veldman, A.B.; Singh, S.B.; Lokesh, Y.U.; Liu, A.; Aziz-Sultan, M.A.; Smith, T.R.; Beam, A. Improved Generalizability in Medical Computer Vision: Hyperbolic Deep Learning in Multi-Modality Neuroimaging. *J. Imaging* **2024**, *10*, 319. [[CrossRef](#)]
186. Bao, Q.; Mi, S.; Gang, B.; Yang, W.; Chen, J.; Liao, Q. MDAN: Mirror Difference Aware Network for Brain Stroke Lesion Segmentation. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 1628–1639. [[CrossRef](#)]
187. Wu, H.; Chen, X.; Li, P.; Wen, Z. Automatic Symmetry Detection from Brain MRI Based on a 2-Channel Convolutional Neural Network. *IEEE Trans. Cybern.* **2021**, *51*, 4464–4475. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.