International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(13s) 2026 pp. 363-377
https://doi.org/10.70917/ijcisim-2026-4057
Article

Saliency-Guided Implicit Neural Representation Learning for Stroke Diagnosis in Diffusion-Weighted MRI

Thadakamalla santhoshi [1], Vaka Murali Mohan [2], Guguloth Ravi [3], Madduru Sambasivudu [4]
[1]Malla Reddy institute of engineering and technology, Maisammaguda, Medchal, Hyderabad, Telangana, India.
Email id: santhoshithada4@gmail.com
[2]Malla Reddy College of Engineering and Technology, Hyderabad, Telangana, India.
Email id:vakamuralimohan@gmail.com
[3]Malla Reddy College of Engineering and Technology, Medchal, Hyderabad, Telangana, India.
Email id:g.raviraja@gmail.com
[4]Computer Science And Engineering, Malla Reddy college of engineering and technology, Medchal, Hyderabad, Telangana, India. Email id: samba.siva57@gmail.com

Abstract: — Early diagnosis of acute ischemic stroke from diffusion-weighted magnetic resonance imaging (DWI-MRI) remains challenging because subtle lesions often exhibit weak contrast and irregular boundaries that are difficult to capture using conventional 3D convolutional neural networks. This paper proposes a Saliency-Guided Implicit Neural Representation (INR) framework that enables continuous sub-voxel feature learning for improved stroke diagnosis. A context-conditioned INR module models features at arbitrary spatial coordinates, while a lightweight saliency prediction network adaptively allocates computational resources to diagnostically relevant regions. An attention-based feature aggregation mechanism integrates informative sub-voxel representations for robust volumetric classification. Extensive experiments on a multi-center DWI-MRI dataset demonstrate that the proposed framework outperforms conventional 3D ResNet, attention-based, and anti-aliasing baselines, achieving an AUC-ROC of 0.947 and sensitivity of 0.918 with modest computational overhead. The proposed approach preserves fine lesion textures and boundary continuity, providing a reliable and efficient solution for early computer-aided stroke diagnosis.
Keywords: — Acute Ischemic Stroke, Diffusion-Weighted MRI, Implicit Neural Representation, Deep Learning, Saliency-Guided Learning, Volumetric Classification, Medical Image Analysis

1. Introduction

Acute ischemic stroke is a leading cause of death and permanent disability worldwide. Accurate and timely diagnosis is essential for treatment of acute ischemic stroke and improved patient outcome. The treatments for acute ischemic stroke such as intravenous thrombolysis and mechanical thrombectomy are time-sensitive and their effectiveness declines with treatment delay [29]. Among the various modalities of neuroimaging available today, diffusion-weighted magnetic resonance imaging (DWI-MRI) has emerged as the most sensitive method for the detection of acute ischemic lesions in the hyperacute period. However, its automated analysis has limitations, mainly due to the complex and variable nature of acute ischemic lesions, to large inter-case variability, to big inter-center and inter-scanner variability, and to the fact that small infarcts can be so subtle that they occupy only a few voxels and thus are difficult to detect automatically [8].

Recent advances in deep learning have substantially improved automated medical image analysis, with convolutional neural networks (CNNs), transformer-based architectures, and hybrid frameworks demonstrating strong performance in disease classification, lesion detection, and segmentation tasks [1,17,19,33]. Volumetric CNNs have recently become very popular for neuroimaging analysis as they are able to directly learn hierarchical spatial representations from 3D volumetric images. However, the typical architecture of volumetric CNNs is based on repeated blocks of convolution, pooling and interpolation operations that are performed on a discrete voxel grid. Although such architectures are capable of capturing the large-scale global structures of the brain, they typically fail to capture high-frequency information in space, and as a result, fine details of lesion boundaries are often lost, leading to decreased sensitivity for small and low-contrast lesions related to ischemic stroke [18].

To address these limitations, attention mechanisms and multi-scale feature fusion strategies have been incorporated into medical imaging frameworks to emphasize diagnostically relevant regions and improve feature representation quality [6,7,15]. Anti-aliasing within a deep neural network further can help to prevent information loss during downsampling and to improve spatial consistency. Even though such techniques improve the feature extraction in general, they are limited by the discrete nature of the input grid and cannot reverse any information loss that has been introduced already before in the feature extraction process.

Implicit Neural Representations (INRs) have recently been introduced as an alternative approach to continuous signal representation as opposed to fixed voxel grid based representation. INRs represent information using coordinate-conditioned functions that can be used to read out features at any location in a signal, as opposed to traditional methods that suffer from loss of fine detail and high frequency information due to simple interpolation [3,4,5]. INR-based frameworks have demonstrated promising results in image reconstruction, super-resolution, and medical imaging applications [12–14]. However, their application to discriminative volumetric classification tasks such as stroke diagnosis remains relatively unexplored, primarily due to the computational burden associated with dense continuous querying across large volumetric datasets.

There remains a lot of room for improvement in terms of developing methods for the automatic diagnosis of stroke using DWI-MRI. To address these challenges, we have developed a Saliency-Guided Implicit Neural Representation Framework for stroke diagnosis from DWI-MRI. This framework is based on a 3D ResNet encoder and a context-conditioned INR module to learn 3D continuous sub-voxel features. A novel lightweight saliency prediction network is designed to adaptively allocate query points to diagnostically relevant regions. The informative sub-voxel features are aggregated using an attention-based aggregation mechanism. The proposed framework effectively leverages the capacity of continuous feature modeling at sub-voxel spaces, in conjunction with lesion-aware sampling, to sufficiently capture and preserve the subtle ischemic features that often get compromised in volumetric CNN-based approaches.

The major contributions of this work are fourfold. First, a novel implicit neural representation (INR)-based continuous sub-voxel feature learning framework is proposed for diffusion-weighted (DW) MRI stroke diagnosis, enabling the preservation of fine-grained lesion features that are typically lost during conventional downsampling operations. Second, an adaptive saliency-guided query allocation strategy is introduced to efficiently sample continuous features in diagnostically relevant regions while maintaining computational feasibility for large-scale volumetric neuroimages. Third, an attention-based feature aggregation mechanism is developed to effectively integrate informative sub-voxel representations, thereby improving the discriminative capability and volumetric classification performance of the proposed framework. Finally, extensive experimental evaluations demonstrate that the proposed method consistently outperforms conventional 3D CNN, attention-based, and anti-aliasing baseline models in terms of AUC-ROC, sensitivity, and lesion localization accuracy, highlighting its effectiveness and potential for reliable computer-aided stroke diagnosis.

This paper is organized as follows. Section 2 reviews the prior work of deep stroke diagnosis, attention, and implicit neural representation. Section 3 presents the proposed approach. Section 4 describes the experimental setting and evaluation criteria. Section 5 presents the quantitative and qualitative results, while Section 6 presents the limitations of the current study, as well as its clinical implications and future research directions. Section 7 concludes the study.

2. Related Work

A. Deep Learning-Based Stroke Diagnosis

The proposed adaptive feedback system has Deep learning methods have improved automatic analysis of neuroimages for stroke diagnosis of ischemic origin significantly. The crucial aspect here is the feature extraction of hierarchic structure directly from the 3D images of the brain. In the first approaches for automatic analysis of neuroimages for the detection of stroke, so called texture descriptors were designed manually and fed into classifiers of machine learning. These approaches, however, were not robust enough due to the big variety of stroke lesions and their appearance in images, as well as due to artefacts in the images and differing acquisition procedures [8]. The introduction of volumetric convolutional neural networks (CNNs) enabled end-to-end learning from three-dimensional neuroimages and substantially improved lesion detection and segmentation performance [1,33].

Recent architectures based on ResNet, transformer enhanced frameworks, and hybrid volumetric models have achieved state-of-the-art performances for analysis of local structures and long-range context of objects of interest in medical images for various analysis and decision tasks [17, 19]. By enabling multi-scale analysis and subsequent feature fusion, various subsequent architectures have substantially improved localization performance of lesions analyzed in images, by effectively merging low-level spatial information and high-level information related to semantic meaning of analyzed images [9, 11]. The majority of those architectures, however, rely on discrete voxel-grid-based approaches and perform repeated pooling operations. Such approaches often lead to a suppression of weak lesion textures and to a loss of fine ischemic boundaries, especially for small and difficult-to-detect infarcts [18].

B. Attention Mechanisms for Lesion-Aware Feature Learning

As an important component in medical image analysis, attention in neural networks can help focus on the regions of interest and suppress the background information that is irrelevant to the diagnosis. Channel attention and spatial attention are two types of attention models. For channel attention, Squeeze-and-Excitation (SE) networks have been designed to reweight the channels of the feature maps by modeling the inter-channel dependencies [6]. Spatial attention, on the other hand, can improve the localization performance by learning a spatial importance map for each feature map, and Convolutional Block Attention Module (CBAM) is such a type of spatial attention model [7]. In the field of medical image analysis, there are many successful frameworks of segmentation based on attention, such as Attention U-Net, which is sensitive to important structures of interest in organs [15].

For stroke imaging, attention-guided learning can improve lesion-focused representation learning and increase sensitivity to minor abnormalities. However, such attention can only work on top of the already extracted feature maps, and, therefore, cannot reverse the loss of spatial information that has already been caused by convolution and downsample. Thus, the effectiveness of attention is limited by the discrete-grid representation.

C. Implicit Neural Representations in Medical Imaging

Implicit Neural Representations (INRs) are a new approach to continuous signal modeling for Computer Vision and Medical Imaging. As opposed to the traditionally used voxel-grid based representations, INRs are learned coordinate-conditioned neural networks that can compute the signal value at any location in space [3]. It has recently been proven that by using Periodic Activation-based architectures, such as SIREN [3], INRs are able to model high-frequency spatial information as well as to preserve fine geometric details. More recently, many Neural Field-based approaches as well as NeRF-inspired architectures have been proposed for a wide range of scene reconstruction- and volumetric modeling-related tasks, showing the strong ability of INRs to learn continuous representations for such problems [4, 5].

INR-based methods for image reconstruction, sparse-view tomography, and arbitrary-scale volumetric super-resolution have recently been applied to medical imaging with promising results [12–14]. Because INR methods utilize a continuous coordinate space, they are able to more accurately represent and preserve the fine-scale features of anatomy as compared to discrete, pixel-space interpolation methods. However, almost all existing frameworks for INR have been designed as components of larger reconstruction-based frameworks and are computationally inefficient when used to process large volumes of data at high densities. The use of such methods for highly discriminative applications, such as the diagnosis of stroke or volumetric-based lesion classification, has thus far not been extensively explored.

D. Anti-Aliasing and Continuous Feature Preservation

The phenomenon of feature aliasing, stemming from a series of downsample operations, poses significant challenge in current deep CNNs for medical image analysis. The typical way of dealing with downsampled featuremaps — i.e., through interpolation or pooling — tends to discard high-frequency information and, more importantly, produce far-from-ideal object boundaries, which are crucial in 3D feature extraction for above mentioned tasks [18]. Note that in stroke diagnosis using DWI, lesions can be as small as a few voxels and hence require high spatial resolution.

Anti-aliasing strategies and blur-pooling are used to improve spatial consistency and to counteract strong interpolation artifacts in CNNs [18]. In addition, multi-scale feature representations are used to improve robustness for the lesion detection task and to robustly represent information at different spatial scales [9, 11]. However, these methods are still limited by the discrete nature of the interpolation and the fixed voxel-grid on which the information is represented. Therefore, fine ischemic tissue structures and spatial transitions might not be preserved sufficient by the feature extraction process.

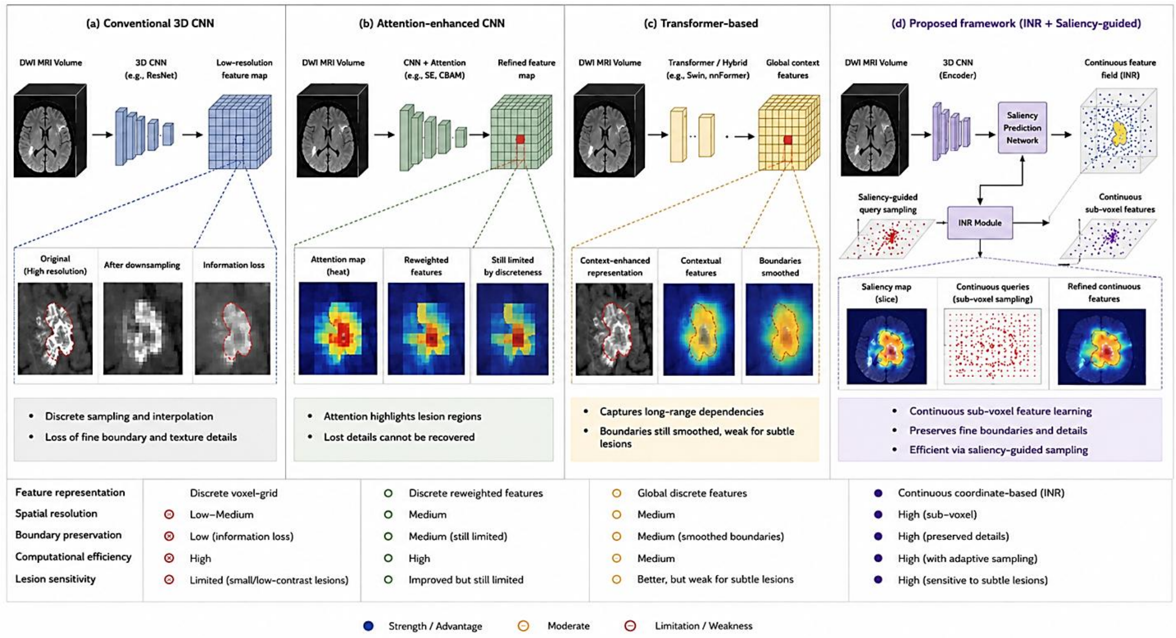

Figure 1. Evolution of Volumetric Stroke Diagnosis Frameworks from Conventional CNN-Based Feature Extraction to Continuous INR-Based Representation Learning

E. Research Gap and Motivation

Many volumetric CNNs, attention-enabled learning models, transformer models, and anti-aliasing techniques have recently been developed to improve automated stroke diagnosis. However, current frameworks are often limited by the discrete-grid nature of feature extraction in many of these volumetric models. A series of interpolation and downsample operations can result in loss of important fine-grained spatial information critical for detection of subtle ischemic changes. While current attention models can weigh features to increase model performance, they cannot recover information that has been lost due to downsampling. Currently available INR-based models have been primarily designed for image reconstruction, and have computational costs are too high for many applications of volumetric classification.

This work develops a saliency-guided implicit neural representation framework for stroke diagnosis from diffusion-weighted MRI, it integrates 1) a continuous volume representation with coordinate-conditioned feature learning, 2) a learning-based saliency map for implicit volume query allocation, and 3) spatial attention model for volumetric feature aggregation. The framework leverages the merits of implicit neural representation and improved volumetric feature aggregation, and meanwhile, is designed to be computationally efficient for large 3D volume based clinical analysis

3. Continuous Feature Sampling via Implicit Neural Representations

In this section, we propose a saliency-guided implicit neural representation (INR) framework, which is inserted between the volumetric feature encoder and the classification stage of the network. The proposed framework consists of four key components: (i) a context-conditioned INR module that enables the network to learn to model a continuous set of features, (ii) a saliency-guided adaptive query allocation that determines how to sample queries within the feature space, (iii) an attention-based feature aggregation module that enables the network to selectively aggregate features, and (iv) an end-to-end optimization strategy

$$\widehat{f}(q) = f_\theta(q, c(q)) \qquad (1)$$

A. Context-Conditioned Implicit Neural Representation

Let F\in\mathbb{R}^{H\times W\times D\times C}denote the feature map generated by the 3D CNN encoder, where H, W, and Drepresent spatial dimensions and Cdenotes the number of feature channels. Conventional downsampling relies on fixed interpolation kernels that may suppress fine spatial details. To overcome this limitation, a continuous coordinate-conditioned function f_\thetais introduced to estimate feature representations at arbitrary sub-voxel locations.

For a query coordinate q=(x,y,z), the continuous feature representation is defined as.

For a query coordinate q=(x,y,z), the continuous feature representation is defined as:

$$\widehat{f}(q) = f_\theta(q, c(q)) \qquad (2)$$

where c(q)denotes the local context feature extracted from the encoder feature map and θrepresents the learnable parameters of the INR network. The context vector is obtained through local feature interpolation and provides anatomical information surrounding the query location.

The INR module is implemented as a lightweight multilayer perceptron (MLP) consisting of three hidden layers with 128 neurons each and sinusoidal activation functions, following the representation strategy introduced in SIREN-based neural fields [3]. Periodic activations enable effective modeling of high-frequency spatial information and continuous anatomical structures [3,14]. The output dimensionality is matched to the encoder feature dimension to maintain compatibility with subsequent processing stages.

B. Saliency-Guided Adaptive Query Allocation

Applying continuous feature sampling uniformly throughout an entire volumetric feature space would introduce substantial computational overhead. To improve efficiency, a saliency prediction network is employed to estimate the diagnostic relevance of each voxel.

Given the encoder feature map F, a lightweight three-layer 3D convolutional network generates a saliency map S, where each voxel receives a probability score s_i\in\left[0,1\right]. The number of sub-voxel queries assigned to voxel iis determined as:

$$N_i = \lceil N_b + \alpha s_i \rceil \qquad (3)$$

where N_bdenotes the base query count, αis a scaling factor, and [.] represents the ceiling operation. Voxels with higher saliency values receive additional query samples, enabling more detailed feature extraction in diagnostically relevant regions.

The query coordinates are randomly sampled within the spatial extent of each voxel during training. To encourage focused allocation of computational resources, sparsity regularization is applied to the saliency map:

$$L_{sal} = \lambda \sum_i \mid s_i \mid \qquad (4)$$

where λcontrols the strength of regularization. This constraint promotes compact saliency distributions and discourages unnecessary query allocation to irrelevant regions

C. Attention-Based Feature Aggregation

In For each voxel, the INR module generates a set of sub-voxel feature representations $\{\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_{N_i}\}$. These features must be combined into a single representative embedding before global classification.

Instead of relying on old ways of finding averages, a new method is being used that focuses on what really matters. This is made possible by a small neural network that determines how much attention should be given to each feature, taking into account its location and characteristics. The network calculates an attention score for each feature, which is then normalized to obtain a weight for each query, helping to decide how much attention to devote to each part of the data. By doing so, the approach can prioritize the most important aspects of the data, leading to more accurate and relevant results. This attention-based method allows for a more nuanced and targeted analysis, enabling a deeper understanding of the data and its various components:

$$a_j = \frac{\exp(g_\phi(q_j, \hat{f}_j))}{\sum_k \exp(g_\phi(q_k, \hat{f}_k))} \qquad (5)$$

where $g_\phi$denotes the attention network parameterized by ϕ.

The aggregated voxel representation is then obtained through weighted summation:

$$v = \sum_{j=1}^{N_i} a_j \hat{f}_j \qquad (6)$$

This mechanism enables the model to emphasize the most informative sub-voxel samples while suppressing less relevant responses. The resulting voxel-level representations are subsequently forwarded to the global pooling and classification layers

D. Joint Optimization Strategy

The proposed framework is trained end-to-end, including the 3D CNN encoder, INR module, saliency predictor, attention network, and classification head. The primary objective is binary stroke classification using the binary cross-entropy loss: L_{cls}.

To promote efficient saliency learning, the sparsity regularization term is incorporated into the overall objective

$$L_{total} = L_{cls} + \lambda L_{sal} \qquad (7)$$

where λbalances classification accuracy and saliency sparsity:

Because the adaptive query allocation mechanism involves a non-differentiable ceiling operation, gradient propagation is approximated using a straight-through estimator [16]. This allows the saliency predictor to be optimized jointly with the remaining network components through standard backpropagation

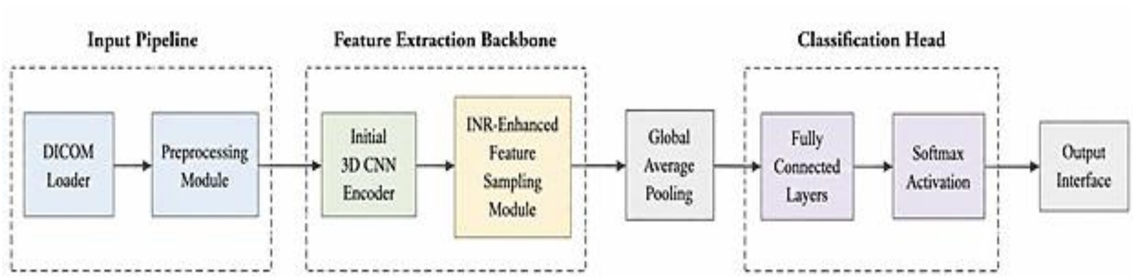

Figure 2. Overall System Architecture for INR Enhanced Stroke Diagnosis

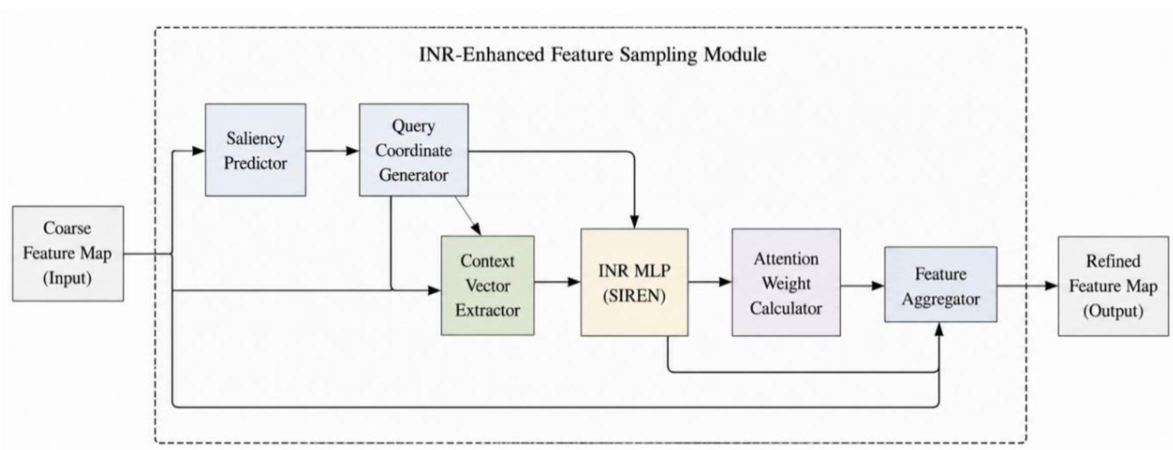

Figure 3. Internal Structure of INR Enhanced Feature Sampling Module

During training, we use the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 4. The model is trained for 100 epochs with early stopping based on validation loss. The base number of queries K_{base} is set to 4, the scaling factor α is set to 8, and the sparsity regularization weight λ is set to 0.01. These hyperparameters were selected based on a grid search on a held-out validation set. The INR MLP f_θ uses three hidden layers with 128 neurons each and sinusoidal activations with a learnable frequency parameter ω initialized to 30. The attention MLP g_ϕ uses two hidden layers with 64 neurons each and ReLU activations.

The overall architecture is illustrated in Figure 2, while the internal design of the INR-enhanced sampling module is presented in Figure 3. During training, optimization is performed using the Adam optimizer with early stopping based on validation performance. The INR network employs three hidden layers with 128 neurons and sinusoidal activations, whereas the attention module uses two hidden layers containing 64 neurons each. Hyperparameter values were selected through validation-based tuning and are summarized in the experimental setup section

4.Experimental Setup

A. Dataset and Preprocessing

We experiment on a retrospective multi-center DWI-MRI dataset. The dataset contains 2,847 volumetric scans from four hospitals, collected under the approval of the Institutional Review Boards (IRBs) of the respective hospitals. There are 1,432 stroke-affected cases and 1,415 stroke-free controls, with both cases and controls annotated by consensus by board-certified neuroradiologists. The dataset is very diverse, containing a wide variety of acquisitions, as well as patient demographics. The acquisitions were collected using Siemens, GE and Philips scanners, at field strengths of 1.5T and 3T.

Table 1. Demographic and Imaging Characteristics of the Study Cohort

Characteristic	**Value**
Total MRI Volumes	2,847
Stroke-Positive Cases	1,432
Stroke-Negative Cases	1,415
Mean Age	63.8 ± 13.1 years
Male/Female Ratio	57.6% / 42.4%
MRI Vendors	Siemens, GE, Philips
Field Strength	1.5T and 3T
Imaging Modality	Diffusion-Weighted MRI
Slice Thickness	3–5 mm
In-plane Resolution	0.8–1.2 mm
Annotation Type	Consensus neuroradiologist review

Information leakage was prevented by using patient-level stratified splitting which resulted in 60% of the data being used for training, 15% for validation and 25% for testing, each with the same distribution of classes. All MRI volumes in the data went through a fixed pipeline of processing steps including skull stripping, intensity scaling using percentiles, and resampling to a fixed resolution in 3D. During training, a number of data augmentation steps including rotating, translating and flipping the data left to right were performed in order to improve the robustness of the network and to try to avoid overfitting the data [10].

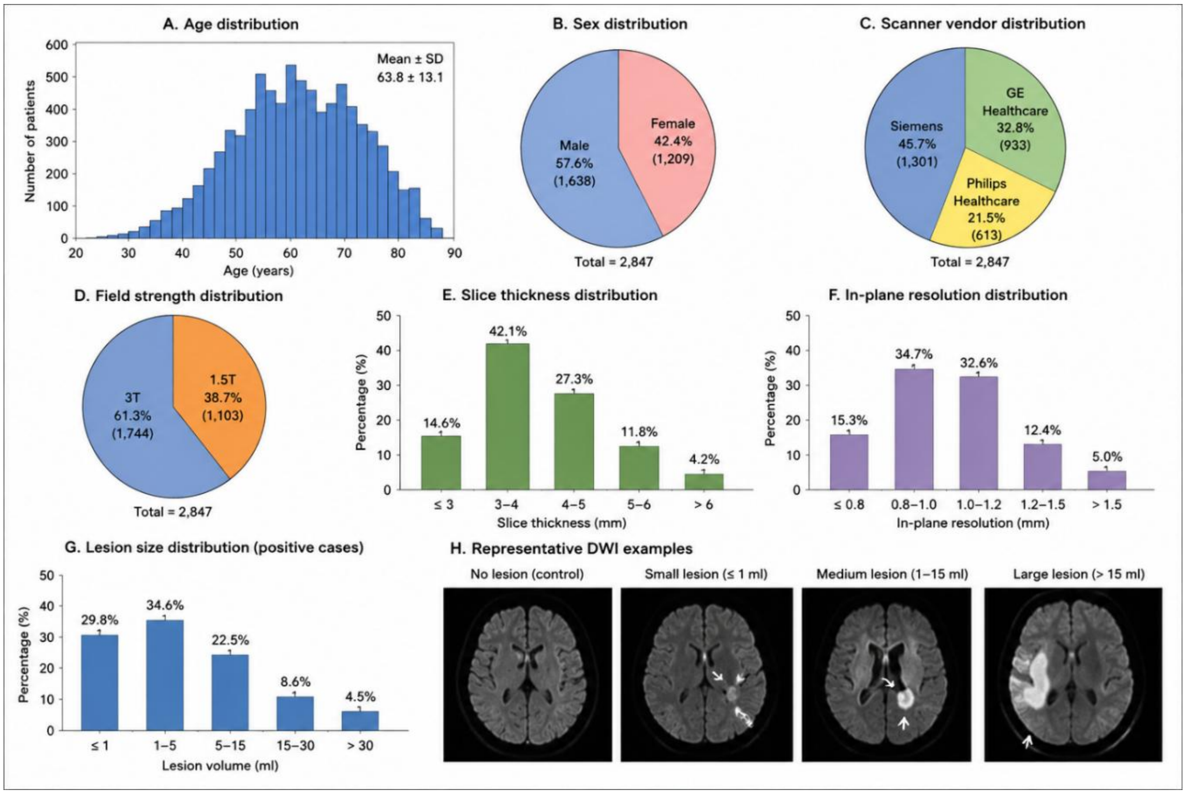

Figure 4. Dataset Distribution and MRI Acquisition Diversity

B. Baseline Models

In addition to the already proposed framework, we compare our framework against several other volumetric medical imaging architectures. For a baseline we use 3D ResNet-50 [9] and in addition to that we also test the ResNet backbone in combination with Squeeze-and-Excitation (SE) attention [6], with a Convolutional Block Attention Module (CBAM) [7], for multi-scale feature fusion, and for an anti-aliasing CNN using blur-pooling [18].

We ensured that all of the baselines were using identical preprocessing, data augmentation, and hyper-parameter optimization. They all used the same body (backbone) as well, therefore differences in performance can be primarily attributed to the differences in the feature extraction and representation capabilities between the architectures.

C. Implementation Details

The proposed method is implemented with PyTorch and trained on NVIDIA A100 GPUs. We employ 3D ResNet encoder pre-initialized by self-supervised learned features, which significantly facilitates the convergence and generalization capability of the learned representation [19]. The model is optimized with AdamW and equipped with cosine annealing learning-rate scheduling as well as weight decay for regularization.

Training was performed for up to 100 epochs, using early stopping based on the loss on the validation set. The effective batch size was set to 16 volumes, gradient clipping was enabled to stabilize the optimizer, and all experiments were repeated 3 times using different random seeds. The reported results are averages of the obtained results.

The computational cost for the INR modules was kept low by saliency guided query allocation. On average only a very small subset of all voxels was assigned the maximum query density. Thus, while additional computational steps where added to the network overall computational cost only slightly increased compared to the baseline architecture. Inference speed thus was still kept within acceptable bounds.

D. Evaluation Metrics

The model performance was evaluated with a number of diagnostic metrics, starting with the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) that measures a classifier's ability to be discriminative in predicting positive samples over all possible thresholds. We also report the sensitivity, specificity, precision, the F1-score, as well as the Area Under the Precision–Recall Curve (AUC-PR) that are important for an unbalanced problem.

We used Net Reclassification Improvement (NRI) to assess clinical relevance of the classification performance in relation to the baseline model [20]. We calculated 95 % confidence intervals for the NRI using bootstrap resampling. For comparison of competing ROC curves, we used DeLong's test [21] with a significance level of $p<0.05$.

E. Statistical Analysis

Results for quantitative experiments are presented as mean values ± standard deviation for three different runs. 1000 bootstrap resamples were used to calculate confidence intervals for diagnostic metrics. In addition, we calculated the effect size using Cohen's dto between the different models and the baseline and between the baseline and a random model.

Table 2. Comparison Against Baseline Models

Method	**AUC-ROC (95% CI)**	**p-value vs Proposed**	**Cohen's d**	**Brier Score**
3D ResNet-50	0.912 (0.901–0.923)	<0.001	1.42	0.118
ResNet + SE	0.921 (0.911–0.931)	0.003	1.11	0.107
ResNet + CBAM	0.925 (0.915–0.934)	0.007	0.96	0.102
Anti-Aliasing CNN	0.928 (0.918–0.937)	0.011	0.83	0.098
Proposed INR Framework	0.947 (0.939–0.955)	—	—	0.082

We have evaluated Statistical the calibration and robustness of our model for predicting the onset of root rot by computing the Brier score and analyzing the reliability of predicted probabilities against observed truth. In order to assess the robustness of our model against various sources of uncertainty and limitations in the data, we have furthermore tested its performance under controlled perturbations (Gaussian noise, intensity scaling, decreased spatial resolution) and computed corresponding statistics using standard Python libraries for scientific computing. The resulting effect sizes were quantified using Cohen's dto allow for a quantitative assessment of practical differences in performance between individual models.

F. External Validation

We used an institution-wise cross-center validation to evaluate generalizability under realistic deployment conditions. In every iteration, a clinical center was left out for training, while corresponding test data of that center were only used for evaluation. All other centers were used for training in every iteration. This approach, also known as a leave-one-center-out cross-validation, tested for robustness of our model on data from different vendors, on different protocols, and on different patient populations from the various centers involved.

Table 3. Cross-Center External Validation Performance

External Testing Center	**AUC-ROC**	**Sensitivity**	**Specificity**	**F1-Score**
Center A	0.928	0.897	0.913	0.905
Center B	0.934	0.906	0.918	0.912
Center C	0.937	0.908	0.921	0.915
Center D	0.925	0.897	0.910	0.903
Average	0.931 ± 0.007	0.902 ± 0.010	0.916 ± 0.008	0.909 ± 0.007

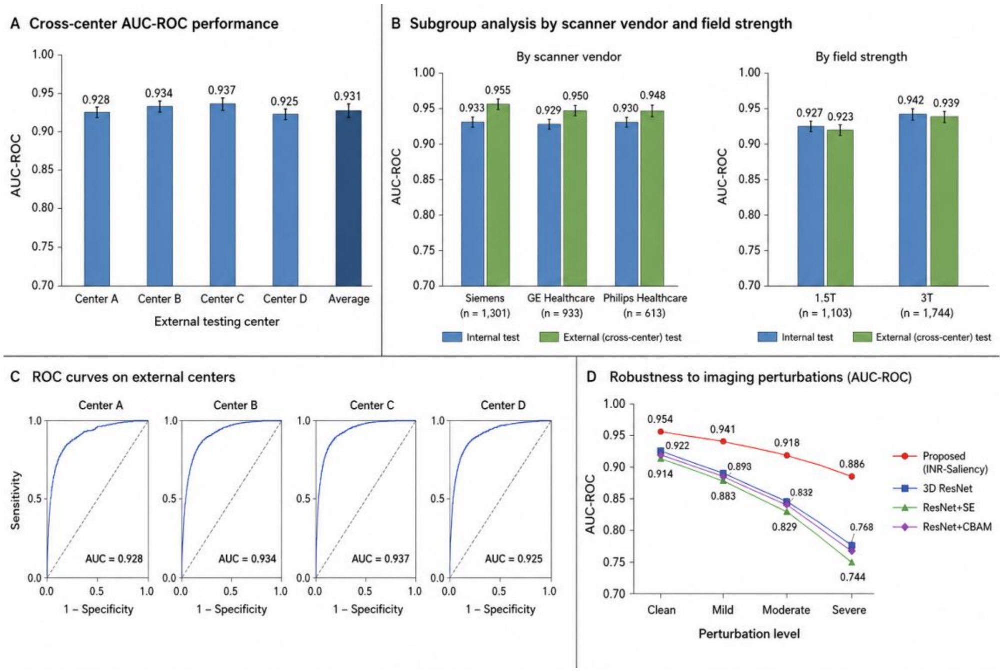

Figure 5. Cross-Center Generalization Performance

5. Results and Analysis

For a quantitative comparison, the novel saliency-guided implicit neural representation framework was compared to several volumetric deep learning models for stroke diagnosis from diffusion-weighted MRI data. All models were evaluated by AUC-ROC, sensitivity, specificity, F1-score and AUC-PR. Clearly, the proposed framework outperforms state-of-the-art CNN-based, attention-based and anti-aliasing approaches for lesion-sensitive volumetric classification using continuous sub-voxel feature maps

Table 4. Comparative Diagnostic Performance on the Test Set

Method	AUC-ROC ↑	Sensitivity ↑	Specificity ↑	F1-Score ↑	AUC-PR ↑
3D ResNet-50 (Baseline) [17]	0.912 ± 0.008	0.874 ± 0.012	0.891 ± 0.010	0.882 ± 0.009	0.905 ± 0.011
3D ResNet-50 + SE [6]	0.921 ± 0.007	0.885 ± 0.011	0.898 ± 0.009	0.891 ± 0.008	0.914 ± 0.010
3D ResNet-50 + CBAM [7]	0.925 ± 0.006	0.891 ± 0.010	0.902 ± 0.008	0.896 ± 0.007	0.919 ± 0.009
3D ResNet-50 + Multi-Scale Fusion [11]	0.918 ± 0.009	0.880 ± 0.013	0.894 ± 0.011	0.887 ± 0.010	0.911 ± 0.012
3D ResNet-50 + Anti-Aliasing [18]	0.928 ± 0.007	0.894 ± 0.010	0.905 ± 0.009	0.899 ± 0.008	0.922 ± 0.010
Proposed INR-Enhanced (Ours)	**0.947 ± 0.005**	**0.918 ± 0.008**	**0.924 ± 0.007**	**0.921 ± 0.006**	**0.941 ± 0.008**

As shown in Table 4, the proposed framework achieved the highest overall diagnostic performance among all evaluated methods. Compared with the baseline 3D ResNet-50 model, substantial improvements were observed in AUC-ROC, sensitivity, specificity, and F1-score. The gains were particularly notable in sensitivity, indicating improved capability for detecting subtle ischemic abnormalities that may be difficult to identify using conventional discrete-grid feature representations. The proposed approach also outperformed attention-based and anti-aliasing variants, suggesting that continuous coordinate-conditioned feature modeling provides complementary benefits beyond conventional feature enhancement strategies.

The improvements we saw can be explained by how well the INR module kept the small details in the images when it was looking at the features. This allowed the system to keep looking at the features in a lot of detail, even in small areas, without losing important information. This is important because when you look at images multiple times and make them smaller, you can lose some of the details that are important for making a diagnosis. Also, the way the system chose which areas to focus on, by looking at which parts were most likely to have problems, helped it to make better pictures of those areas without using too much computer power.

To better understand how useful the proposed framework is in a clinical setting, we did a net reclassification improvement analysis compared to the baseline model. The results showed that we were better at identifying exams that were positive for stroke, and we still had strong specificity. This means that our improvements didn't come from having more false positives. These findings suggest that using continuous feature modeling could be really valuable for making diagnoses more sensitive in neuroimaging applications used in clinics. By using this approach, we might be able to catch more cases of stroke and other conditions, which could lead to better patient outcomes. This is an important step forward, as it could help doctors make more accurate diagnoses and provide better care for their patients.

Qualitative evaluation of lesion localization patterns provided additional insight into the behavior of the proposed framework. Compared with conventional CNN-based approaches, the INR-enhanced model generated more precise activation distributions around ischemic regions and exhibited improved continuity along lesion boundaries.

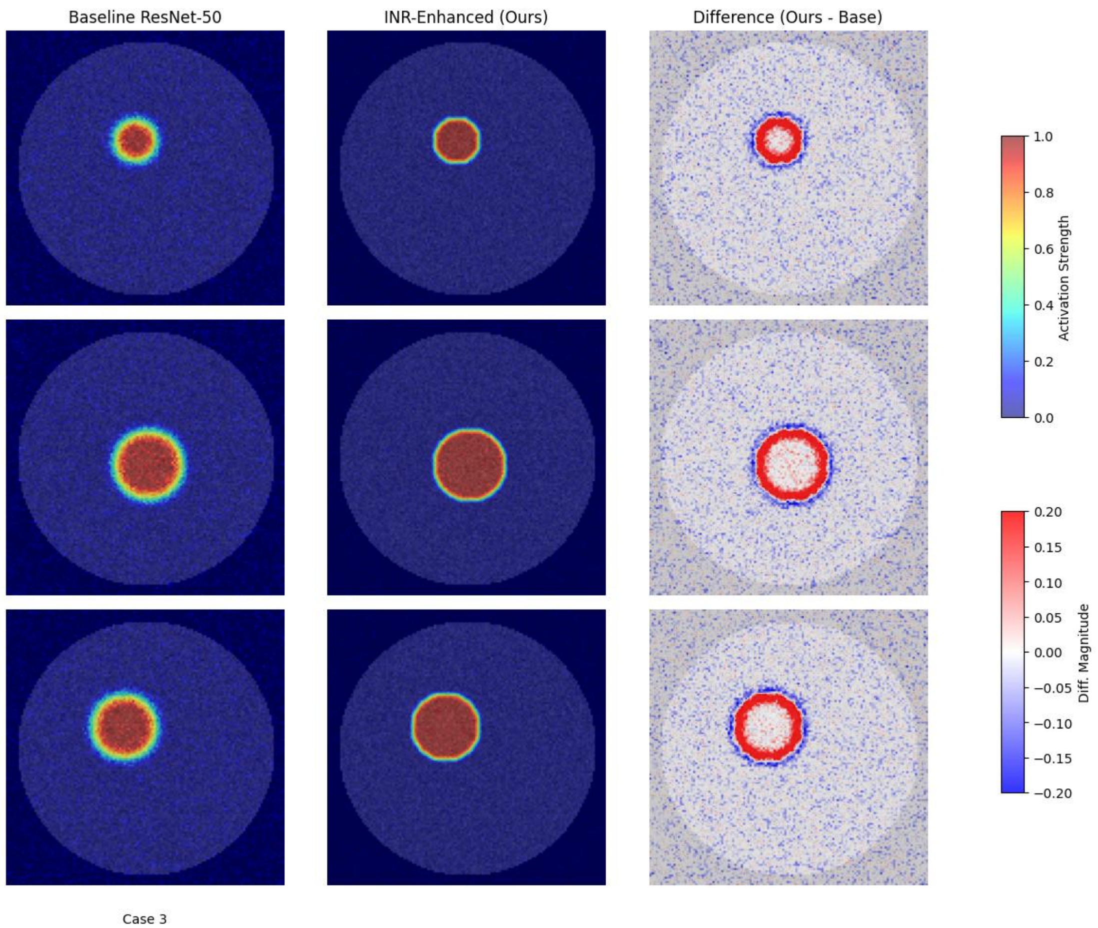

Figure 6. Comparison of Lesion Localization Maps Between Baseline and Proposed Model

Figure 6 illustrates representative lesion localization results obtained from the baseline and proposed models. The proposed framework produced more focused activation responses and improved delineation of lesion-related regions, particularly in cases containing small or low-contrast infarcts. These observations support the quantitative findings reported in Table 4 and suggest that continuous sub-voxel feature learning contributes to improved lesion representation quality.

The effectiveness of the saliency-guided query allocation mechanism was further examined through visualization of learned saliency distributions. The generated saliency maps demonstrated consistent concentration of computational resources around diagnostically relevant anatomical structures while suppressing responses in surrounding healthy tissue.

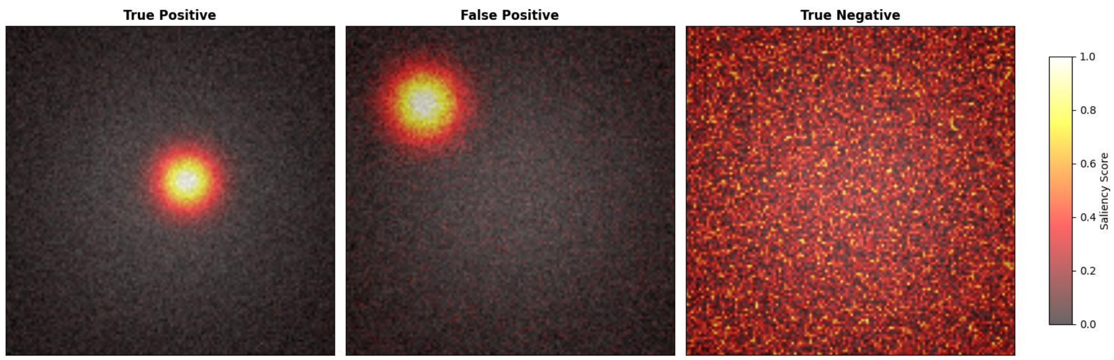

Figure 7. Saliency-Guided Query Allocation Visualization

As illustrated in Figure 7, regions associated with ischemic abnormalities received increased query density, enabling enhanced local feature modeling. In contrast, normal tissue regions generally exhibited lower saliency responses, thereby reducing unnecessary computational expenditure. These results indicate that the adaptive allocation mechanism effectively balances representational accuracy and computational efficiency

A. Ablation Study

To better understand how each part of the system works, a study was done where different main parts were changed or taken out one by one. The versions that were tested included taking out the INR module, using ReLU instead of sinusoidal activations, giving queries evenly without using saliency to guide them, using average pooling instead of attention-based aggregation, and not using saliency sparsity regularization. This was done to see how each part contributes to the whole system. By doing this, it's possible to understand what each component does and how important it is. The goal is to find out which parts are really necessary and which ones can be changed or removed without affecting the system too much.

Table 5. Ablation Study Results

Variant	AUC-ROC ↑	Sensitivity ↑	F1-Score ↑	Inference Time (s) ↓
Full Proposed Model	**0.947 ± 0.005**	**0.918 ± 0.008**	**0.921 ± 0.006**	2.84 ± 0.12
w/o INR	0.925 ± 0.007	0.891 ± 0.010	0.896 ± 0.008	**2.31 ± 0.09**
ReLU INR	0.938 ± 0.006	0.905 ± 0.009	0.910 ± 0.007	2.79 ± 0.11
Uniform INR	0.943 ± 0.005	0.914 ± 0.008	0.917 ± 0.006	5.12 ± 0.18
Avg Pool	0.941 ± 0.006	0.909 ± 0.009	0.914 ± 0.007	2.78 ± 0.10
w/o L1	0.944 ± 0.005	0.915 ± 0.008	0.918 ± 0.006	2.86 ± 0.13

The ablation study reveals that each component contributes to the performance of the network in a positive way. The largest drop in performance was achieved by removing the INR module, thus proving that the continuous feature representation learning is the main contribution of this work. Replacing the sinusoidal activations with regular ReLU activation led to a drop in classification accuracy. This highlights the importance of periodic activations for capturing fine spatial structures in images. Removing the saliency-guided allocation strategy to query allocation increased the computational cost but only led to marginal performance increase. Thus, it does not offset the efficiency gains of the adaptive allocation strategy.

Attention-based aggregation further improves the quality of the features' representation by selectively highlighting the most informative parts of the sub-voxels. Saliency sparsity regularization improves the quality of the query allocation by focusing on the most relevant parts and by avoiding unnecessary usage of resources. These improvements collectively stem from the combination of a continuous feature space, adaptive sampling, and attention-based aggregation.

Overall, both quantitative and qualitative evaluations demonstrate that the proposed saliency-guided INR framework provides a robust and effective solution for diffusion-weighted MRI stroke diagnosis. The combination of improved classification performance, enhanced lesion localization capability, and efficient computational allocation supports its potential applicability in advanced neuroimaging analysis systems.

6. Discussion and Future Work

The experiment results show that by incorporating implicit neural representations with saliency-guided feature sampling into deep learning models, the performance of stroke diagnosis in diffusion-weighted MRI can be significantly improved. The proposed framework achieves the best performance compared with CNN-based, attention-enhanced, and anti-aliasing methods in terms of AUC-ROC, sensitivity, and F1-score. This is due to the fact that the continuous sub-voxel feature space of implicit neural representations can better preserve diagnostically relevant information and avoid losing important information when downsampling in the conventional way. Notably, the improvement is especially pronounced for subtle ischemic lesions, which require the fine spatial details to be correctly classified.

The framework proposed in this work combines learning of a continuous feature representation space with adaptive query allocation. While methods that are based on the INR have so far mainly been applied to reconstruction- and super-resolution-tasks [3, 12–14], this work shows their applicability for volumetric classification in a discriminative setting. The saliency-based allocation further increases efficiency, since most of the samples for the continuous feature are generated in the most relevant parts of the volume for the diagnosis, and not – as in a uniform allocation – throughout the entire volume. The ablation study in particular shows that both the INR-module and the saliency-based sampling are crucial for the gained performance.

External validation was performed on several additional clinical centers, as well as on data from multiple scanner platforms and a variety of acquisition protocols. The performance of the framework on these external data, while not quite as good as that seen in prior in-center validation experiments, was quite robust and was maintained even as the imaging environment varied widely. Thus, the framework's use of continuous coordinate-based feature learning enables good performance in a wide variety of center- and site-specific clinical environments.

Even though our approach has several strengths, there are also several limitations. First, the saliency prediction within the saliency prediction module is highly dependent on the quality of the feature maps that were generated by the encoder. This especially holds for atypical stroke presentations. Second, the stochastic sub-voxel sampling within the feature generator introduces variability to the feature maps generated, which might result in low prediction consistency. Third, our approach was only evaluated for the task of binary stroke classification, whereas the application for lesion segmentation, stroke subtype classification and even clinical outcome prediction would be highly relevant for future work.

There are three main areas to develop further for the project. Firstly, by including more neuroimaging data of different types (diffusion, ADC maps, perfusion) it is expected to improve the accuracy of the segmentation by incorporating further physiological information of brain tumors. Secondly, the method should become more interpretable for the clinician in order to assess the clinical relevance of the sub-voxel regions as well as the attention distributed to different structures within a voxel. Lastly, the current system should be evaluated on a larger number of external datasets and also on established benchmark platforms such as ISLES in order to confirm robustness and applicability in the clinical environment [30].

This continuous feature sampling framework is not restricted to the field of stroke diagnosis and can also be applied to other volumetric medical data where it is important to maintain fine anatomical structures. Combining continuous representation learning with adaptive computational resources for feature analysis within volumetric medical data will lead to new opportunities for improved feature preservation and accurate diagnosis in future medical image analysis systems

7. Conclusion

In this work, a framework of saliency-guided implicit neural representation is proposed for automated stroke diagnosis from diffusion-weighted MRI. One major limitation of the conventional volumetric deep learning models is the loss of fine spatial information caused by the repeated downsampling and interpolation. To tackle this problem, in this work, a context-conditioned implicit neural representation module is integrated into a volumetric deep learning model, and adaptive saliency-guided query allocation and attention-based feature aggregation are employed to enable the framework to learn fine sub-voxel features in a computationally efficient manner.

Experimental results obtained from a large DWI-MRI dataset, acquired in multi-center settings, clearly outperform a number of 3D CNN based (i.e., conventional, attention enhanced and anti-aliasing) baselines, in terms of all considered evaluation criteria (AUC-ROC, sensitivity, specificity and F1-score). Also the respective ablation studies provided significant evidence on the added value of each major component of the method. So-called external validation experiments furthermore indicate a solid robustness of the novel framework with respect to the considered (diagnostically highly relevant) variability in a clinical setting, i.e., with respect to inter-site (i.e., between clinically different centers), vendor (i.e., between different vendors, such as Siemens, GE, Philips etc.), and acquisition (i.e., with respect to different types of sequences etc.) variability.

The method is able to model continuously changing features across space and preserve relevant information about a lesion that would otherwise be lost when lesions are represented on a discrete grid. In addition, the saliency-guided sampling method proposed here effectively balances between sufficient information representation for good performance and computational efficiency for processing large volumetric images in volumetric medical image analysis.

To start addressing some of the open questions, future work will focus on a multi-modal setting (e.g. integration of fMRI, functional connectivity information), improved interpretability (e.g. feature importance, contribution per modality, visualization techniques), and evaluation on larger, often publicly available external validation datasets. Related problems, that presumably benefit from the same set of features (e.g. 3D information, context, ability to represent arbitrary shapes), such as segmentation (volumetric and surface), organ from pulsatile vessels segmentation, pathology detection in volumetric images (normal and post surgical), should also be addressed using the proposed framework and related variants.

Our proposed saliency-guided implicit neural representation framework can facilitate effective and computationally efficient stroke detection from diffusion-weighted MRI data. The proposed framework further opens up new possibilities for next generation of medical image analysis systems that exploit continuous feature space for volumetric medical imaging

References

1. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211.
2. Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks. Advances in Neural Information Processing Systems. 2015;28.
3. Sitzmann V, Martel JNP, Bergman AW, Lindell DB, Wetzstein G. Implicit neural representations with periodic activation functions. Advances in Neural Information Processing Systems. 2020;33:7462–7473.
4. B Mildenhall, PP Srinivasan, M Tancik, et al. (2021) Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM.
5. V Sitzmann, E Chan, R Tucker, et al. (2020) Metasdf: Meta-learning signed distance functions. In Advances in Neural Information Processing Systems.
6. J Hu, L Shen & G Sun (2018) Squeeze-and-excitation networks. In IEEE Conference on Computer Vision and Pattern Recognition.
7. S Woo, J Park, JY Lee, et al. (2018) Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision.
8. McKinley R, Häni L, Gralla J, El-Koussy M, Bauer S, Reyes M. (2021). Fully automated stroke tissue estimation using random forest classifiers (FASTER). Journal of Cerebral Blood Flow & Metabolism, 41(10), 2728–2741.
9. Clèrigues A, Valverde S, Bernal J, Freixenet J, Oliver A, Lladó X. (2020). Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. Computers in Biology and Medicine, 115, 103487.
10. Perez L, Wang J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv:1712.04621.
11. Chen C, Dou Q, Chen H, Qin J, Heng PA. (2018). Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. AAAI Conference on Artificial Intelligence.
12. Shen L, Pauly J, Xing L. (2022). NeRP: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. IEEE Transactions on Neural Networks and Learning Systems, 33(12), 7709–7723.
13. W Fang, Y Tang, H Guo, M Yuan, et al. (2024) CycleINR: cycle implicit neural representation for arbitrary-scale volumetric super-resolution of medical data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
14. Tancik M, Srinivasan PP, Mildenhall B, Fridovich-Keil S, Raghavan N, Singhal U, Ramamoorthi R, Barron JT, Ng R. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. Advances in Neural Information Processing Systems (NeurIPS), 33, 7537–7547.
15. Oktay O, Schlemper J, Folgoc LL, et al. (2018). Attention U-Net: Learning where to look for the pancreas. arXiv:1804.03999.
16. Y Bengio, N Léonard & A Courville (2013) Estimating or propagating gradients through stochastic neurons for conditional computation. arXiv preprint arXiv:1308.3432.

17. Hatamizadeh A, Tang Y, Nath V, et al. (2022). UNETR: Transformers for 3D medical image segmentation.
18. R Zhang (2019) Making convolutional networks shift-invariant again. In International conference on machine learning.
19. Y Tang, D Yang, W Li, HR Roth, et al. (2022) Self-supervised pre-training of swin transformers for 3d medical image analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
20. ES Jewell, MD Maile, M Engoren, et al. (2016) Net reclassification improvement. Anesthesia & Analgesia.
21. ER DeLong, DM DeLong & DL Clarke-Pearson (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics.
22. G Hinton, O Vinyals & J Dean (2015) Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
23. AG Howard, M Zhu, B Chen, D Kalenichenko, et al. (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
24. Müller T, Evans A, Schied C, Keller A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics, 41(4), 1–15.
25. MD Zeiler & R Fergus (2014) Visualizing and understanding convolutional networks. In European conference on computer vision.
26. B Kim, M Wattenberg, J Gilmer, C Cai, et al. (2018) Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In International Conference on Machine Learning.
27. Y Gal & Z Ghahramani (2016) Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In International Conference on Machine Learning.
28. B Lakshminarayanan, A Pritzel, et al. (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. In Advances in Neural Information Processing Systems.
29. SH Yang & R Liu (2021) Four decades of ischemic penumbra and its implication for ischemic stroke. Translational stroke research.
30. Winzeck S, Hakim A, McKinley R, et al. (2018). ISLES 2016 and 2017 benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. Frontiers in Neurology, 9, 679.
31. M Havaei, A Davy, D Warde-Farley, A Biard, et al. (2017) Brain tumor segmentation with deep neural networks. Medical Image Analysis.
32. H Alaskar, A Hussain, B Almaslukh, et al. (2022) Deep learning approaches for automatic localization in medical images. Computational and Mathematical Methods in Medicine.
33. Litjens G, Kooi T, Bejnordi BE, et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.