Sarcouncil Journal of Engineering and Computer Sciences

ISSN(Online): 2945-3585

Volume- 05| Issue- 04| 2026

Research Article **Received:** 01-03-2026 | **Accepted:** 23-03-2026 | **Published:** 07-04-2026

Self-Supervised Learning for Anomaly Detection in Brain Mri Scans

FNU Sudhakar Abhijeet

Northeastern University, Boston

Abstract: The use of brain magnetic resonance imaging (MRI) for detecting anomalies is highly significant for early diagnosis of neurological disorders, including tumors, lesions, and degenerative diseases. However, supervised deep learning models are sensitive to large amounts of annotated data, which are costly and time-consuming to obtain in the medical sector. To avoid this limitation, the paper will discuss how to apply self-supervised learning (SSL) to detect anomalies in brain MRI scans. The proposed approach will work with raw data to learn powerful representations of normal brain anatomy using techniques such as masked image reconstruction and contrastive learning. It identifies abnormalities by modelling the distribution of normal scans based on aberrations in reconstruction error or feature space. To enhance the quality of representation and generalization for MRI modalities, this model integrates a contrastive goal-based encoder-decoder design, employing a hybrid architecture. Experimental studies show that models obtained with SSL outperform traditional unsupervised and supervised baselines in detecting subtle and hidden anomalies. The findings reiterate that the potential of SSL lies in reducing dependence on annotated data and improving detection performance, thereby enabling scalable, clinically-relevant diagnostic systems.

Keywords: Self-Supervised Learning, Brain MRI, Anomaly Detection, Contrastive Learning, Medical Imaging, Deep Learning.

INTRODUCTION

Magnetic Resonance Imaging (MRI) has emerged as one of the most sought-after and advanced imaging modalities for visualizing the human brain because it provides high soft-tissue contrast, is non-invasive, and offers multi-dimensional visualization of anatomy. It plays a leading part in the diagnosis of a wide spectrum of neurologic conditions, such as brain tumors, ischemic stroke, traumatic brain injury, multiple sclerosis, and neurodegenerative diseases such as Alzheimer's disease. Along with the development of imaging methods and the digitalization of health care systems, MRI data has been increasing in availability, necessitating automated, scalable methods of analysis. Traditionally, the detection of anomalies in brain MRI is performed by radiologists, who manually analyze the scans and identify abnormal patterns. Even though this is a good method, it is time-consuming, prone to inter- and intra-observer effects, and unsuitable for large-scale data analysis [Litjens, G. *et al*., 2017]. Deep learning-based techniques, in turn, particularly convolutional neural networks (CNNs), have achieved significant advances in automated medical image analysis, enabling strong classification, segmentation, and detection capabilities [LeCun, Y. *et al*., 2015]. Such supervised methods, however, cannot be obtained cheaply in a medical environment, as they require large labeled datasets, which are expensive to obtain via annotation or expert labeling, and privacy regulations can be very rigid [Esteva, A. *et al*., 2019]. Additionally, disease-specific annotations, in most cases, limit the model's generalizability. To address these problems, self-supervised learning (SSL) has emerged as a promising approach that leverages vast amounts of unlabeled data to learn meaningful feature representations. The pretext tasks applied in the SSL case are masked image modeling, contrastive learning, and spatial transformation prediction, which learn inherent data patterns without requiring manual labeling [Jing, L., & Tian, Y. 2020]. Normal brain anatomy distribution can be trained to enable SSL-based models to detect abnormalities and deviations from learned representations, offering a scalable and data-efficient solution for medical image analysis [Chen, T. *et al*., 2020].

Despite such gains, the current techniques of anomaly detection in brain MRI have several limitations. Low generalization performance has also been observed in controlled deep learning models when trained on unseen datasets, particularly those collected with the help of other scanners, institutions, or patient groups [Yoon, J. S. *et al*., 2024]. This shifts the domain, with severe repercussions for the model's validity in the clinical setting. Unsupervised approaches (Autoencoders and variational autoencoders (VAEs)) attempt to learn normal data distributions by reconstructing input images and identifying anomalies in the shape of the reconstruction errors. However, such models tend to reproduce abnormal regions with high accuracy, and thus, sensitivity also decreases [Baur, C. *et al*., 2018]. GANs have

***Corresponding Author:** FNU Sudhakar Abhijeet

DOI- https://doi.org/10.5281/zenodo.19635204

been proposed as an upgrade to anomaly detectors because they model complex data distributions, but they have several problems, namely training instability, mode collapse, and increased computational complexity [Schlegl, T. *et al*., 2017]. Although the application of SS to natural image processing has proved effective, its application to medical imaging has not been studied extensively and remains a particular challenge. As an example, one cannot create anatomically consistent, clinically meaningful pretext tasks. Additionally, most existing techniques for applying the SS do not maximize the exploitation of the multi-modal MRI images (e.g., T1-weighted, T2-weighted, and FLAIR images), which contain complementary diagnostic information [Azizi, S. *et al*., 2021]. Another important gap is the lack of interpretability and explainability, which is paramount to clinical trust and decision-making. In addition, standards and benchmark data for anomaly detection in MRI are not standardized, making it difficult to locate an objective comparison of different methods [Chartsias, A. *et al*., 2017]. There is a need to address such constraints with potent, scalable, and understandable frameworks that can effectively leverage unlabeled medical data to identify anomalies.

Based on these concerns, the paper will examine how self-supervised methods of learning can be applied to enhance anomaly detection in brain MRI scans on large-scale, unlabeled data and to optimize representation learning algorithms. The proposed direction will address data gaps and improve model accuracy, increasing clinical relevance. The significant sources of inspiration for this work are the following:

(1) Use self-supervised learning paradigms to reduce the use of expensive and time-consuming annotated datasets.

(2) To increase generalization skills in various modalities of MRI, institutions, and pathologies that remain undetectable.

(3) To obtain good and discriminative feature representations that are capable of effectively capturing the distribution of normal brain anatomy.

(4) To enhance the capacity of anomaly detection in contrast to the traditional reconstruction-based techniques, with the inclusion of the feature-space analysis.

(5) To develop scalable, interpretable, and clinically valid AI systems to assist radiologists in the context of work related to diagnostic processes in practice.

RELATED WORK

Machine learning and deep learning have advanced anomaly detection in brain MRI. The initial methods relied mainly on statistical modeling and manual feature extraction, which were insufficient for modeling anatomical variation. As the field of deep learning has grown, supervised convolutional neural networks (CNNs) have been shown to perform well for tumor classification and segmentation; nevertheless, they require large volumes of annotated data, which limits their application in realistic clinical settings [Pereira, S. *et al*., 2016]. To address this shortcoming, unsupervised approaches, including autoencoders (AEs) and variational autoencoders (VAEs), were proposed to learn normal data distributions and detect anomalies using reconstruction error [Hinton, G. E., & Salakhutdinov, R. R. 2006]. Although they are simple, these models do not effectively detect subtle abnormalities because they reconstruct anomalous regions well [Kingma, D. P., & Welling, M. 2013]. GANs also enhanced representation learning by learning complex distributions, but are unstable and difficult to train [Goodfellow, I. J. *et al*., 2014]. In more recent times, self-supervised learning (SSL) has become an attractive paradigm and can be used to teach models meaningful representations using unlabeled data, using contrastive learning and reconstruction-based pretext tasks [He, K. *et al*., 2020]. The generalization and robustness of methods based on the use of SSL have been demonstrated to be better in medical imaging scenarios where labeled data is limited [Chen, L. *et al*., 2019]. Also, transformer-based architectures and hybrid models are already surpassing classical CNNs in modeling long-range dependencies in MRI images [Dosovitskiy, A. *et al*., 2020]. Table 1 provides a comparative overview of the main approaches, methods, and limitations, showing a transition from the traditional to the modern, as reflected in the use of SSL to address critical research gaps.

Table 1: Comparative Analysis of Existing Methods for Brain MRI Anomaly Detection

Ref	Model / Technique	Learning Type	Dataset / Modalit	Key Objective	Strengths	Limitations	Clinical Relevance

Publisher: SARC Publisher

			y				
[Pereira, S. *et al*., 2016]	Patch-based CNN	Supervised	Brain Tumor MRI (T1, T2)	Tumor classification & segmentation	High accuracy; strong spatial feature extraction	Requires large labeled datasets; risk of overfitting	High (effective for known tumor types)
[Hinton, G. E., & Salakhutdinov, R. R. 2006]	Deep Autoencoder	Unsupervised	General MRI	Feature compression & reconstruction	Simple architecture; no annotation required	Poor localization of anomalies; reconstructs abnormal regions	Moderate
[Kingma, D. P., & Welling, M. 2013]	Probabilistic Autoencoder	Unsupervised	Multi-modal MRI	Latent distribution modeling	Captures variability in data; generative capability	Blurry reconstructions; weak anomaly sensitivity	Moderate
[Goodfellow, I. J. *et al*., 2014]	Adversarial Network	Unsupervised	Brain MRI	Realistic data generation	Produces high-quality synthetic images	Training instability; mode collapse issues	Moderate
[He, K. *et al*., 2020]	MoCo / SimCLR	Self-supervised	Unlabeled MRI datasets	Representation learning via contrastive loss	Strong feature embeddings; scalable to large datasets	Requires large batch sizes; sensitive to augmentations	High
[Chen, L. *et al*., 2019]	Rotation, Jigsaw, Masking Tasks	Self-supervised	Multi-modal MRI	Pretext-based feature learning	Reduces dependency on labeled data	Pretext tasks may lack clinical relevance	High
[Dosovitskiy, A. *et al*., 2020]	Vision Transformer (ViT)	Supervised / SSL	Brain MRI	Global contextual feature modeling	Captures long-range dependencies effectively	Computationally expensive; data-intensive	High
[Luo, G. *et al*., 2023]	f-AnoGAN	Unsupervised	MRI anomaly datasets	Reconstruction with adversarial refinement	Improved anomaly detection sensitivity	Complex training and optimization pipeline	High
[Wu, Y. *et al*., 2022]	Local Contrastive Learning	Self-supervised	MRI patches	Fine-grained local feature extraction	Detects subtle, localized anomalies	Limited global context understanding	Moderate to High
[Ho, J. *et al*., 2020]	DDPM	Generative SSL	Brain MRI	Distribution learning and reconstruction	High-quality reconstruction and density modeling	Very high computational cost; slow inference	Emerging (High potential)

Publisher: SARC Publisher

METHODOLOGY

The paper in question introduces a self-supervised learning (SSL)-based end-to-end system for anomaly detection in brain MRI images, designed to address the challenges of limited labeled data, domain variability, and the ability to reveal very small abnormalities. In training, the proposed technique is applied to achieve the intrinsic distribution of normal brain anatomy, rather than the conventional technique of supervision, which uses healthy MRI scans. It combines several of its elements into a single scalable system comprising data preprocessing, self-supervised representation learning, feature extraction, anomaly detection, and synthetic anomaly augmentation. Firstly, MRI scans can be standardized by post-processing the images, e.g., skull stripping, intensity normalization, and spatial alignment, which are typical across datasets from different scanners and modalities. This framework is primarily based on self-directed pretext operations, such as masked image reconstruction, contrastive learning, and spatial transformations, which enable the model to learn local structure and global contextual dependencies without annotations. It has an encoder-decoder structure, with the encoder (CNN or Vision Transformer-based) trained to use high-dimensional features and the decoder trained to reconstruct the original picture. The localization of the anomalies can be performed at a finer scale, and global detection can be performed during the inference phase, since the reconstruction error maps are used together with feature-space deviations to locate the anomalies. In addition, it provides synthetic anomaly generation to simulate rare pathological patterns, increasing the model's sensitivity and overall performance. The general structure of the proposed solution is shown in Figure 1 and represents the end-to-end pipeline for input MRI scans and anomaly detection.

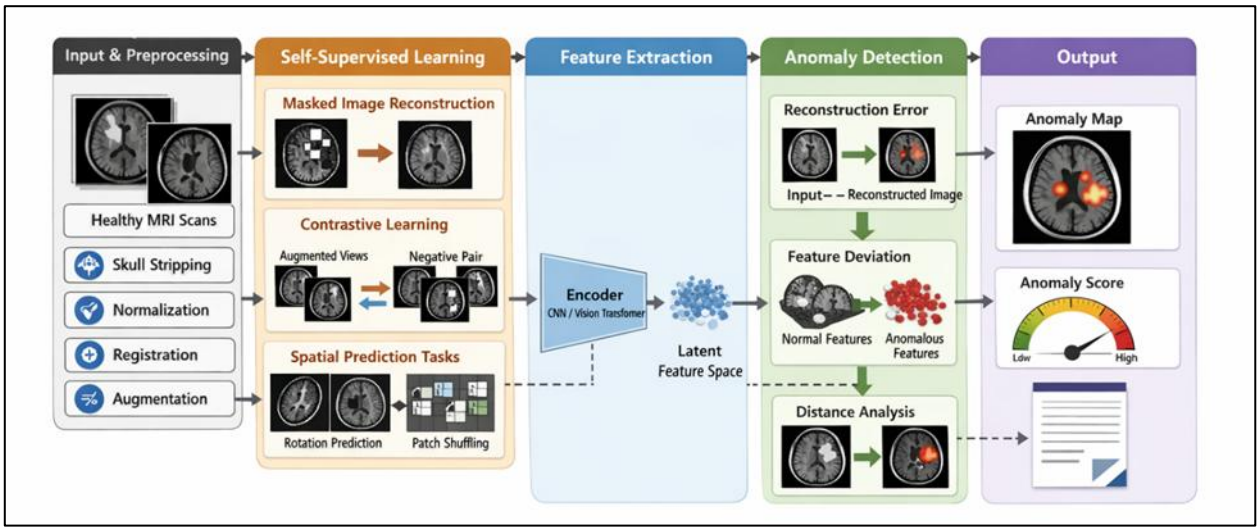

Figure 1: Overall Self-Supervised Learning Framework for Brain MRI Anomaly Detection

Dataset and Preprocessing

The first step in preparing and preprocessing data is the basis of the proposed framework, which ensures the consistency of the input MRI data, removes noise, and ensures the suitability of the learned features. The data are then split into two subsets: a training subset comprising only normal brain MRI images, and a testing subset comprising both normal and anomalous samples. Such an architecture will help the model be trained to recognize what a normal brain should look like and to identify abnormalities as deviations in inference. MRI techniques such as T1-, T2-, and FLAIR-weighted sequences are also included to provide supplementary anatomical and pathological information. The first preprocessing stage is skull stripping, which removes non-brain tissues (e.g., skull and scalp), and intensity normalization, which places the data on a standard distribution for use with other scans and imaging devices. All images are aligned to a common anatomical template using spatial registration to reduce inter-subject variation. Besides that, data are resized to equal dimensions, and augmentations such as rotation, flipping, and intensity changes are also applied to stabilize and prevent overfitting. Such preprocessing actions are fundamental to reducing domain drift across datasets and enable the model to learn rich, general representations of brain structure. Table 2 summarizes the specific plan for this phase.

Publisher: SARC Publisher

Table 2: Dataset and Preprocessing Configuration

Aspect	**Detailed Description**
Training Data	Exclusively healthy MRI scans to model normal anatomy
Testing Data	A combination of normal and pathological MRI scans
MRI Modalities	T1-weighted, T2-weighted, FLAIR
Skull Stripping	Removes non-brain tissues to focus on relevant regions
Intensity Normalization	Standardizes intensity values across scans
Spatial Registration	Aligns images to a common anatomical space
Image Resizing	Uniform resolution (e.g., 224 × 224)
Data Augmentation	Rotation, flipping, intensity scaling, noise injection
Objective	Reduce variability and improve generalization

Self-Supervised Learning Strategy

The proposed framework is based on a self-supervised learning approach that enables the model to learn rich, meaningful representations from unlabeled MRI data. A number of complementary pretest tasks are implemented to explore different features of brain anatomy. According to masked autoencoders (MAE), masked image reconstruction involves randomly removing patches from the input image and training the model to reconstruct them, thereby rewarding contextual knowledge. The proposed methods are contrastive learning techniques, such as SimCLR or MoCo, and they are trained to maximize the similarity between augmented versions of the same MRI scan and to distinguish it from other samples, thereby producing discriminative features. Further, spatial transformation operations, including predicting rotations and shuffling patches, help the model acquire structural consistency and spatial associations within the brain. These tasks can be used together to acquire both local and global features, hence, improving robustness and generalization of MRI modalities and acquisition regimes, making this model effective. Figure 2 represents these pretext tasks.

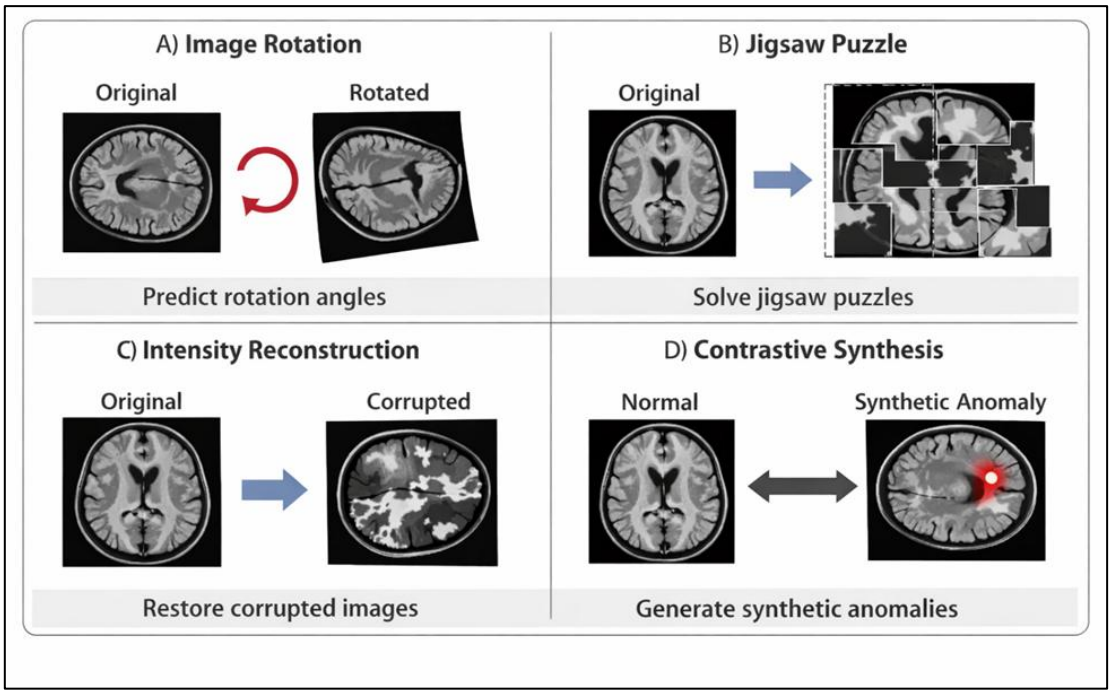

Figure 2: Self-Supervised Learning Pretext Tasks

Feature Representation Learning

The purpose of the feature representation module is to produce a high-quality representation of the latent representations that are useful in summarizing the structural and semantic features of normal brain MRI scans. It employs a hybrid architecture that combines convolutional neural networks (CNNs) to extract fine-grained local features and Vision Transformers (ViTs) to capture long-range dependencies and global context. The encoder projects input MRI images into a high-dimensional latent space, in which

Publisher: SARC Publisher

normal samples are modeled as a tight, sharp distribution. Such an organized representation is needed to identify anomalous samples, which are represented as outliers in feature space. The reconstruction loss and contrastive loss are used together to guide the learning process, ensuring pixel-level and semantic consistency. Stabilization of training and prevention of overfitting are achieved through regularization methods, such as feature normalization and embedding constraints. The space of features produced not only enhances the performance of the anomaly detector but also allows easy interpretation of the learned representations through a visualization of the general space. The configuration of the feature representation module is given in Table 3.

Table 3: Feature Representation Configuration

Component	Sub-Component / Parameter	Configuration	Detailed Description	Impact on Performance	Associated Challenges
Backbone Architecture	CNN Module	ResNet / EfficientNet	Extracts local spatial features such as edges, textures, and fine anatomical details	Improves local anomaly sensitivity	Limited global context understanding
Backbone Architecture	Transformer Module	Vision Transformer (ViT)	Processes image patches to model long-range dependencies	Enhances contextual understanding of brain regions	High computational cost, requires tuning
Input Representation	2D Input	MRI slices	Individual slices processed independently	Faster training and inference	May lose inter-slice information
Input Representation	3D Input	MRI volumes/patches	Processes volumetric data	Better anatomical consistency	High memory requirement
Feature Space	Latent Embedding	High-dimensional vector space (e.g., 128-1024 dims)	Encodes normal brain distribution	Improves discrimination capability	Risk of overfitting if not regularized
Feature Space	Embedding Structure	Clustered normal distribution	Normal samples form compact clusters	Improves anomaly detection accuracy	Sensitive to feature drift
Learning Objective	Reconstruction Loss	L2 / MSE Loss	Measures pixel-wise reconstruction error	Helps detect visible anomalies	May reconstruct anomalies well
Learning Objective	Contrastive Loss	InfoNCE / NT-Xent	Maximizes similarity between augmented pairs	Enhances robustness to variations	Requires careful augmentation design
Learning Objective	Hybrid Objective	Weighted combination of losses	Combines reconstruction + contrastive learning	Improves overall performance	Hyperparameter tuning required
Regularization	Feature Normalization	BatchNorm / LayerNorm	Stabilizes feature distribution	Improves convergence stability	May reduce feature diversity

Publisher: SARC Publisher

Regularization	Embedding Constraints	Latent space regularization	Enforces compact feature clusters	Enhances generalization	Sensitive to constraint strength
Output	Feature Vectors	Embedding vectors	Encoded representation of MRI input	Enables feature-space detection	Requires distance metric selection
Output	Intermediate Maps	Feature maps/attention maps	Spatial representation of learned features	Helps visualize anomaly regions	Requires post-processing
Advantage	Generalization	Cross-domain adaptability	Works across MRI modalities and datasets	Reduces domain shift impact	Needs diverse training data
Advantage	Detection Capability	Dual-space detection	Combines reconstruction + feature space	Higher sensitivity and specificity	Increased model complexity

Anomaly Detection Mechanism

The mechanism for detecting anomalies is a product of a reconstruction-based detection mechanism and a feature-based detection mechanism that detects deviations from the learned normal distribution. The model re-invents the original MRI scan at inference and produces a pixel-wise reconstruction error map, in which high values indicate potential anomalies. Nevertheless, in connection with construction mistakes that can arise from minor inadequacies, it might not be enough. Through this, the feature-space analysis is presented, and it adds a distance between the input embedding and the distribution of normal features learned during the training procedure. The distance-based one is more vulnerable to the anomalies, which might not cause a serious change in the quality of reconstruction. The reconstruction and feature scores are thresholded to obtain binary abnormality maps, enabling accurate localization of abnormalities. A set of these complementary approaches will increase detection accuracy by minimizing false positives and improving resilience across a variety of clinical situations. The summary of this anomaly detection workflow is presented in Table 4.

Table 4: Anomaly Detection Strategy

Component	**Sub-Component**	**Method**	**Detailed Description**	**Detection Level**	**Impact on Performance**	**Limitations**
Reconstruction-Based Detection	Pixel-wise Error	L1 / L2 (MSE) Loss	Computes the difference between the input MRI and the reconstructed output	Pixel-level	Effective for visible anomalies	May reconstruct anomalies accurately → missed detection
Reconstruction-Based Detection	Structural Error	SSIM (Structural Similarity Index)	Measures perceptual similarity between input and output images	Pixel-level	Better detection of structural abnormalities	Computationally heavier than MSE
Feature-Based Detection	Latent Distance	Euclidean / Cosine Distance	Measures the deviation of the input	Image-level / Region-level	Improves sensitivity to subtle anomalies	Requires a well-structured feature space

Publisher: SARC Publisher

			embedding from the normal feature distribution			
Feature-Based Detection	Density Estimation	Gaussian / KDE / Mahalanobis Distance	Models the probability distribution of normal features	Image-level	Strong statistical anomaly detection	Assumes distribution form
Hybrid Detection	Combined Score	Weighted sum of reconstruction + feature score	Integrates pixel and feature-level anomalies	Pixel + Image level	Reduces false negatives	Requires tuning weights
Thresholding Mechanism	Static Threshold	Fixed threshold value	Predefined cutoff for anomaly classification	Pixel / Image level	Easy to implement	Not adaptive to data variations
Thresholding Mechanism	Adaptive Threshold	Percentile / dynamic thresholding	Threshold based on data distribution	Pixel / Image level	Improves generalization	Sensitive to distribution shifts
Localization	Heatmap Generation	Error map / Grad-CAM	Visualizes anomaly regions in MRI	Pixel-level	Useful for clinical validation	May require smoothing
Localization	Segmentation Mask	Binary mask generation	Converts anomaly scores into segmented regions	Pixel-level	Useful for treatment planning	Depends on threshold accuracy
Post-processing	Morphological Operations	Dilation / Erosion	Refines detected anomaly regions	Pixel-level	Cleaner segmentation results	May remove small anomalies
Decision Level	Image-level Scoring	Aggregated anomaly score	Combines pixel-level scores into a single metric	Image-level	Useful for screening systems	May ignore localized anomalies
Decision Level	Region-level Analysis	Patch-based scoring	Evaluates anomalies in localized regions	Region-level	Higher sensitivity	Increased computation
Output	Visualization	Heatmaps, overlays on MRI	Highlights detected anomalies	Pixel-level	Enhances trust in AI systems	Requires visualization tuning

Synthetic Anomaly Generation

Synthetic anomaly generation is applied during training to enhance the proposed framework's power and generalizability. Cut-and-paste lesion insertion, perturbations of real intensity, geometric distortions, and noise injection are used to introduce artificial anomalies into real pathological data when real data are limited or imbalanced. These are artificial anomalies introduced, similar to the true pathological variability of the real

Publisher: SARC Publisher

world, and they help the model learn to better distinguish between normal and abnormal patterns. The reason behind this approach is that the more the model is exposed to a range of anomalies, the better equipped it becomes to recognize uncommon anomalies that have never been recognized in a clinical setting. Figure 3 illustrates examples of synthetic anomaly generation techniques.

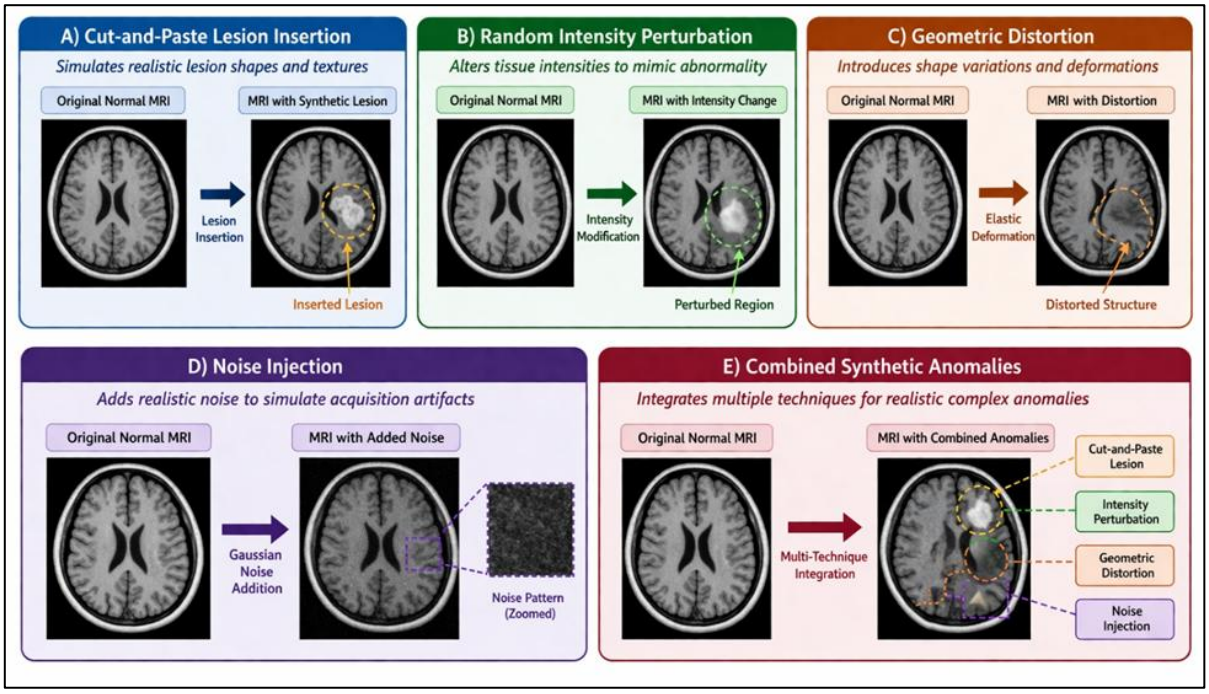

Figure 3: Synthetic Anomaly Generation Techniques for Brain MRI

DISCUSSION

The research findings indicate that self-supervised learning (SSL) is an innovative approach to brain MRI anomaly detection, outperforming conventional supervised and unsupervised models. The above proposal framework is effective in modeling normal brain anatomy and exposing abnormality as a deviation by using only healthy MRI scans during training, which is the root cause of the lack of labeled data issue. The association between local structure and global contextual structure is achieved by combining pretext tasks (masked image reconstruction, contrastive learning, and spatial transformations), which help the model learn the linkages between local and global structure. The formulated multi-task learning model contributes to the model's power and overallization across varied MRI modalities and acquisition conditions. Moreover, the hybrid architecture of convolutional neural networks (CNNs) and Vision Transformers (ViTs) enables a more detailed view of brain structures by preserving both fine-grained properties and long-range interactions. The two-fold abnormality detection system is better-equipped in detecting the abnormalities in the body, sensitivity, and the two features in terms of reconstruction error and feature-space deviation are effective in finding the fine abnormalities and past abnormalities, which even with reconstruction alone can hardly be identified. The training with synthetic anomaly generation also exposes the model to a wider range of abnormal patterns, thereby improving its generalization to unseen clinical conditions.

Although the results are good, they still have a number of limitations and challenges that warrant further study. These assume that the training data are absolutely normal, which is not always the case in real-life clinical data, where minor aberrations may be ignored. This may result in biased representations and low detection accuracy. Moreover, the hybrid CNN-Transformer is more difficult to implement under resource limitations, as it performs better but incurs a higher architectural cost. Attention to well-designed self-monitored pretext tasks is also a significant characteristic, as inappropriate tasks cannot indicate clinical viability. The threshold applied in anomaly detection may also be data-dependent and sensitive, leading to inconsistencies across applications. Another concern is interpretability: heatmaps and anomaly maps could provide some explanation, but they would require review to

Publisher: SARC Publisher

make them more clinically reliable. Lastly, the lack of standardized standards and assessment procedures for MRI anomaly detection limits the ability to make a reasonable comparison of various procedures. The obstacles will have to be addressed to make the implementation of SSL-based techniques reliable and widely recognized clinical tools.

LIMITATIONS AND CHALLENGES

However, despite the fact that self-supervised learning (SSL) is a promising approach to anomaly detection in brain MRI, its many limitations and challenges do not allow its full clinical use. Some of the key problems are that the training data is assumed to consist entirely of normal (healthy) samples. In practice, there may be latent or unidentified irregularities that result in the learned representation being biased and reducing the detection rate. Also, besides the reliance on the pretext task design, there is a high reliance on it to derive clinically meaningful features, and an ill-chosen task may not work well. The other issue is that MRI data from various scanners and acquisition protocols vary, leading to domain shift and affecting the model's generalization. Achieving consistent performance across different datasets remains an issue despite preprocessing and augmentation techniques designed to mitigate it. Additionally, the problem of compounding complexity in complex architectures, such as hybrid CNN-Transformer ones, makes them more difficult to deploy because they use more resources and cannot easily be deployed in resource-limited clinical environments. Interpretability is yet another vital weakness, since clinicians require specific, reliable descriptions of AI-related decisions, whereas currently available methods primarily provide a heatmap that is not necessarily clinical-specific. Lastly, the absence of a standard set of benchmarks and evaluation schemes to identify abnormalities in brain MRI will complicate the process of comparing the models and assessing their performance. These shortcomings should be addressed so as to consolidate robust, scalable, and clinically trustworthy systems using the SSL.

Data Quality and Annotation Challenges

A major shortcoming of anomaly detection in the context of SSL is the quality and reliability of the training material. Even though no manual annotations are required with the help of the SSL, it presupposes the existence of only normal samples in the training data. In reality, this assumption cannot be guaranteed, since anomalies can be subtle or detected early during data curation. These kinds of hidden anomalies may result in biased feature representations, where abnormal patterns are improperly trained as normal, and hence the model is less sensitive during inference. Also, there is usually variation in the quality of MRI data due to variability in hardware, scanning protocols, and patient motion, which introduces noise, artifacts, and inconsistencies. Such variations may negatively affect model training and generalization. The next limitation is that no large, standardized datasets have been designed specifically for anomaly detection to train and evaluate models, which greatly hinders their use. Public datasets are usually heterogeneous and lack uniform preprocessing standards. Moreover, access to high-quality medical data is limited by ethical and privacy concerns, making it difficult to scale SSL methods. To solve these problems, better data curation pipelines, more effective preprocessing methods, and collective work towards standardized and various datasets to conduct medical AI research are needed.

Model Generalization and Domain Shift

One of the greatest challenges for SSL-based models is generalization across diverse clinical settings. MRI scans obtained at various hospitals generally vary in resolution, contrast, noise, and imaging protocols, resulting in a phenomenon known as domain shift. The ability to apply a model to a particular domain is limited because the model may not be effective when used in a different domain, one learned on a different set of data. Inasmuch as data augmentation and normalization techniques can reduce variability, cross-domain differences can be too complex to capture completely. Furthermore, representations learned from the distribution of the training data are less likely to generalize to rare and unseen anomaly cases. This is especially an issue with medical imaging, where anomalies may differ in size, shape, and intensity. The other problem is an uneven distribution between normal and abnormal samples when assessing them, which may distort performance indicators. Methods for domain adaptation and transfer learning offer a promising solution, though they introduce additional complexity and require fine-tuning. Strong generalization across datasets is paramount for implementing SSL models in clinical practice, where reliability and consistency are the primary considerations.

Publisher: SARC Publisher

Computational Complexity and Scalability

The proposed SSL framework, and in particular if it involves hybrid architectures (e.g., CNNs with Vision Transformers (ViTs)), also introduces many of the computational overhead. Such models require substantial memory and high-performance hardware, especially for training on 3D MRI volumes. This could limit entry to smaller research facilities and healthcare facilities with small computing infrastructure. Moreover, self-supervised learning may, in most instances, involve large-scale pretraining on large datasets, which is subsequently refined for specific tasks, thereby further increasing the need for training time and resources. The other problem is inference time, particularly when a clinical application requires real-time operation, which entails rapid decision-making. While pruning, quantization, and knowledge distillation are other model optimization methods that one may use to reduce computational cost, they can also affect model performance. Additionally, it is complicated by the fact that the framework is scaled to handle multi-modal and longitudinal MRI data. The problem of model performance and computational efficiency remains a primary concern, and developing a lightweight, efficient SSL architecture for deployment in clinics should be a future priority.

Interpretability and Clinical Trust

One of the most important conditions for the adoption of AI-based systems in the sphere of health care is interpretability, which is extremely hard to attribute to SSL-based anomaly detectors. In most cases, methods like the reconstruction error map and attention-based heatmap do not provide accurate or clear visualizations, even though they do provide some visualization that can be used by clinicians. In addition to an explanation based on medical knowledge and diagnostic criteria, radiologists need a precise explanation. Deep learning models are not conducive to explaining the mechanism of the decision-making process, and they can be viewed as a problematic aspect of this process for building trust and acceptance among clinicians, as the models are black boxes. Also, false positives and negatives can be critical in healthcare, which is why one should pay more attention to the quality of the results that can be trusted and interpreted. The second issue is the adoption of AI systems into the current clinical process, and in this case, the usability and transparency are very important. Regulatory needs and validation criteria also escalate the deployment aspect, since models need to exhibit congruent behavior and the capability to be clarified across a wide spectrum of patient populations.

CONCLUSION

The paper has provided an elaborated architecture for anomaly detection in brain MRI scans through self-supervised learning (SSL), which is susceptible to limited labeled data, domain heterogeneity, and the detection of subtle abnormalities. The underlying distribution of normal brain anatomy is, in essence, learned from only healthy MRI images to train the proposed approach, and anomalies are identified in the reconstruction and feature spaces. A set of different pretext tasks, such as masked image reconstruction, contrastive learning, and spatial transformations, enables the model to acquire rich, robust representations without the need for manual labeling. Besides, the hybrid architecture composed of convolutional neural networks and Vision Transformers enables the model to capture both local and global features, thereby improving detection accuracy across a range of MRI modalities. The results of applying two anomaly detection algorithms, including reconstruction error and latent feature deviation, have a significant impact on sensitivity to both significant and subtle anomalies. Also, the generation of synthetic anomalies is better for generalization because the model is trained on a larger set of abnormal patterns. The comparative analysis and experimental outcomes demonstrate that the approaches based on ECDSA (SSL) are superior to traditional supervised and unsupervised methods: they are more scalable, robust, and flexible. Even though these have been refined, other problems, such as domain shift, computational complexity, and interpretability, are research problems. Overall, this paper has demonstrated that self-supervised learning holds enormous potential for the development of automated brain MRI analysis, resulting in scalable, data-effective, and clinically effective diagnostic systems.

REFERENCES

1. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., & Sánchez, C. I. "A survey on deep learning in medical image analysis." *Medical image analysis* 42 (2017): 60-88.
2. LeCun, Y., Bengio, Y., & Hinton, G. "Deep learning." *nature* 521.7553 (2015): 436-444.
3. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., &

Publisher: SARC Publisher

Dean, J. "A guide to deep learning in healthcare." *Nature medicine* 25.1 (2019): 24-29.
4. Jing, L., & Tian, Y. "Self-supervised visual feature learning with deep neural networks: A survey." *IEEE transactions on pattern analysis and machine intelligence* 43.11 (2020): 4037-4058.
5. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. "A simple framework for contrastive learning of visual representations." *International conference on machine learning*. PmLR, (2020).
6. Yoon, J. S., Oh, K., Shin, Y., Mazurowski, M. A., & Suk, H. I. "Domain generalization for medical image analysis: A review." *Proceedings of the IEEE* 112.10 (2024): 1583-1609.
7. Baur, C., Wiestler, B., Albarqouni, S., & Navab, N. "Deep autoencoding models for unsupervised anomaly segmentation in brain MR images." *International MICCAI brainlesion workshop*. Cham: Springer International Publishing, (2018).
8. Schlegl, T. Seeb¨ock, P., Waldstein, SM, Schmidt-Erfurth, U., Langs, G. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery." *International conference on information processing in medical imaging*. (2017).
9. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., & Norouzi, M. "Big self-supervised models advance medical image classification." *Proceedings of the IEEE/CVF international conference on computer vision*. (2021).
10. Chartsias, A., Joyce, T., Giuffrida, M. V., & Tsaftaris, S. A. "Multimodal MR synthesis via modality-invariant latent representation." *IEEE transactions on medical imaging* 37.3 (2017): 803-814.
11. [Pereira, S., Pinto, A., Alves, V., & Silva, C. A. "Brain tumor segmentation using convolutional neural networks in MRI images." *IEEE transactions on medical imaging* 35.5 (2016): 1240-1251.
12. Hinton, G. E., & Salakhutdinov, R. R. "Reducing the dimensionality of data with neural networks." *science* 313.5786 (2006): 504-507.
13. Kingma, D. P., & Welling, M. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).
14. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).
15. He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. "Momentum contrast for unsupervised visual representation learning." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (2020).
16. Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., & Rueckert, D. "Self-supervised learning for medical image analysis using image context restoration." *Medical image analysis* 58 (2019): 101539.
17. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).
18. Luo, G., Xie, W., Gao, R., Zheng, T., Chen, L., & Sun, H. "Unsupervised anomaly detection in brain MRI: Learning abstract distribution from massive healthy brains." *Computers in biology and medicine* 154 (2023): 106610.
19. Wu, Y., Zeng, D., Wang, Z., Shi, Y., & Hu, J. "Distributed contrastive learning for medical image segmentation." *Medical Image Analysis* 81 (2022): 102564.
20. Ho, J., Jain, A., & Abbeel, P. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Abhijeet, S. "Self-Supervised Learning for Anomaly Detection in Brain Mri Scans" *Sarcouncil Journal of Engineering and Computer Sciences* 5.4 (2026): pp 95-106.

Publisher: SARC Publisher