Medical Image Analysis 112 (2026) 104076

Contents lists available at ScienceDirect

Medical Image Analysis

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

A review of deep learning-based Unsupervised Anomaly Detection in brain MRI

Finn Behrendt [a,*], Debayan Bhattacharya [a], Lennart Maack [a], Julia Krüger [b], Roland Opfer [b], Alexander Schlaefer [a]

[a] Hamburg University of Technology, Am Schwarzenberg-Campus 1, Hamburg, 21073, Germany

[b] Jung Diagnostics GmbH, Röntgenstraße 24, Hamburg, 22335, Germany

ARTICLE INFO

Keywords:
Deep learning
Unsupervised Anomaly Detection
Brain MRI
Generative models
Survey

ABSTRACT

The manual assessment of brain Magnetic Resonance Imaging (MRI) scans can be labor-intensive and time-consuming for radiologists. Deep Learning methods have demonstrated the potential to aid this process. However, their effectiveness relies on the availability of large, annotated data sets. Unsupervised Anomaly Detection (UAD) presents a promising alternative, offering the potential to identify and localize anomalies without per-pixel annotations. Instead, a normative distribution is learned using healthy data, enabling the identification of abnormalities as deviations. This allows UAD methods to detect abnormalities that were unseen during training. This appealing feature has led to numerous studies proposing innovations and novel approaches. In this work, we provide a review of the literature and systematically collect and compare the proposed approaches. We observe that UAD has made significant advancements in brain MRI analysis. However, individual approaches are often evaluated in different contexts, i.e., changes in acquisition parameters, pre- and post-processing, and anomaly scoring. This variability makes it challenging to assess which models perform best, underscoring the need for comprehensive comparative studies concerning the specific context of MRI scans. Our collection, featuring public data sets, research studies, and open implementations, is available at our GitHub repository https://github.com/FinnBehrendt/Unsupervised-Anomaly-Detection-in-Brain-MRI.

Contents

1. Introduction 2
 1.1. Literature search 3
2. Anomaly detection 3
 2.1. Types of anomalies and detection levels 3
 2.2. Anomaly detection approaches 4
3. Approaches to UAD in brain MRI 4
 3.1. Feature modeling methods 4
 3.2. Reconstruction methods 4
 3.3. Synthetic anomaly/synthetic groundtruth-based methods 5
4. Data sets for UAD 5
5. Review of approaches to UAD in brain MRI 5
 5.1. Feature modeling methods 5
 5.2. Reconstruction methods 7
  5.2.1. Autoencoder 7
  5.2.2. Variational autoencoder 9
  5.2.3. GAN 12
  5.2.4. Diffusion models 13
  5.2.5. Other approaches and comparative studies 15
 5.3. Synthetic anomaly methods 15
6. Context, challenges and future directions of UAD in brain MRI 16

* Corresponding author.
E-mail address: finn.behrendt@tuhh.de (F. Behrendt).

https://doi.org/10.1016/j.media.2026.104076
Received 30 April 2024; Received in revised form 8 March 2026; Accepted 7 April 2026
Available online 15 April 2026

6.1. Data sets 16
6.2. Image acquisition parameters 17
6.3. Data representation 18
6.4. Pre-processing 18
6.5. Model type 18
6.6. Anomaly scoring 19
6.7. Post-processing 20
6.8. Summary 20
7. Conclusion 20
CRediT authorship contribution statement 20
Declaration of competing interest 21
References 21

1. Introduction

Magnetic resonance imaging (MRI) of the brain is a widely employed radiological examination for non-invasive diagnosis of neurological disorders. In recent decades, the continuous advancements in hardware and software technologies associated with MRI scanners have enabled the acquisition of high-resolution images, that provide valuable insights into the internal structures of the human body (Kabasawa, 2022; Börnert and Norris, 2020). The clinical relevance of modern MRIs has notably increased the annual volume of scans (McDonald et al., 2015). This increase in volume imposes additional burdens on radiologists, requiring labor-intensive evaluation of each MRI to avoid diagnostic errors (Bruno et al., 2015). However, interpreting MRIs can be challenging and human errors such as inattentional blindness are likely (Drew et al., 2013; van Hespen et al., 2021). This challenge is also evident in the high variability in interpretations among radiologists in clinical studies (Bauer et al., 2013; Porz et al., 2014). Hence, ongoing research aims to support radiologists by leveraging data-driven deep learning (DL) methods like convolutional neural networks (CNNs) trained in a supervised manner to detect specific pathologies (Porz et al., 2014; Menze et al., 2014; Lundervold and Lundervold, 2019; Soganciogl u et al., 2024). In this approach, extensive collections of annotated MRI scans are harnessed to learn correlations between the characteristics of pathologies and their manifestation in MRI images. This strategy can be highly effective when a large amount of data with high-quality annotations is available (Menze et al., 2014; Akkus et al., 2017; Crimi et al., 2019; Lundervold and Lundervold, 2019; Myronenko, 2019; Luu and Park, 2022). However, obtaining these annotations presents significant challenges, as it is a costly and time-consuming process. Ideally, the consensus annotation of multiple raters is collected to reduce the noise originating from inter-rater variability. Moreover, training a robust supervised CNN requires balanced, large-scale data sets, ensuring sufficient representation of each pathology class (Johnson and Khoshgoftaar, 2019; Fernando et al., 2020; Gooya et al., 2018; Shin et al., 2018). However, clinical data sets are often imbalanced in real-world scenarios. Another limitation of supervised methods is that even if trained with a perfectly annotated data set, only pathologies present in that specific data set can be detected, potentially overlooking other pathologies not accounted for in the training data distribution. An alternative to the supervised training paradigm is unsupervised anomaly detection (UAD). Instead of establishing a direct association between pathology features and ground truth annotations, the fundamental premise of UAD is to model the distribution of healthy anatomy and then identify any anomalies as deviations from this learned norm. UAD approaches have been applied to many domains, such as fraud detection (Bhattacharyya et al., 2011), industrial defect detection (Czimmermann et al., 2020), genetics (Tibshirani and Hastie, 2007) and medical diagnosis (Fernando et al., 2021).

While UAD is broadly applicable, brain MRI represents a specific application domain, as reflected by the particular clinical workflows and expertise required for neuroimaging. First, brain anatomy is broadly organized and exhibits approximate bilateral symmetry. Pathologies can manifest in diverse ways, ranging from pronounced structural changes, such as tumors with associated mass effect, to subtle intensity alterations such as small white matter lesions. Second, normal brain morphology depends on context. For example, brain structure changes with age, motivating models that account for age-related variability. Third, although the use of multiple MRI sequences (e.g., T1, T2, FLAIR) is not unique to brain MRI, it is commonly employed in neuroimaging, where anomalies can become more apparent through the joint interpretation of complementary sequences. These characteristics motivate modeling and evaluation strategies tailored to brain MRI. This is also reflected in the research interest in domain-specific UAD approaches (see Fig. 1).

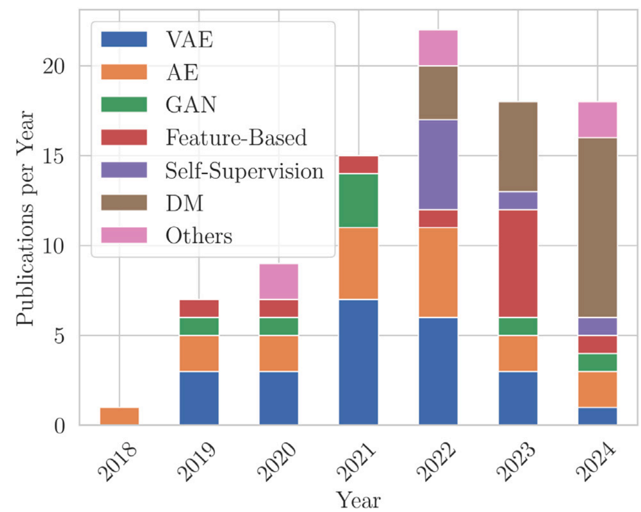

Fig. 1. Composition of the reviewed learning-based UAD approaches in brain MRI per year.

In brain MRI analysis, early approaches relied on comparing brain MRIs to a statistical brain atlas (Gering et al., 2002; Cuadra et al., 2002; Moon et al., 2002; Prastawa et al., 2003, 2004) or on asymmetries between the left and right brain hemispheres (Watkins et al., 2001; Woolard and Heckers, 2012). Furthermore, approaches incorporated pre-computed image-based features, which were subsequently input into machine learning algorithms such as Support Vector Machines (SVMs) or clustering methods to pinpoint outliers (van Leemput et al., 2001; Song and Wyrwicz, 2009; Wang et al., 2014; Andrade de Oliveira et al., 2015; Lapuyade-Lahorgue et al., 2017). In recent developments, DL methods such as CNNs are utilized due to their effective representation, particularly considering high-dimensional imaging data. Furthermore, reconstruction-based approaches as well as segmentation networks have been proposed for UAD, enabling voxel-wise anomaly scoring. However, even in unsupervised settings, training often relies on data labeled as healthy to define the reference distribution, making models sensitive to labeling inaccuracies. Although UAD can, in principle, be trained with partially labeled or contaminated data sets, current brain MRI studies often rely on curated healthy data to ensure a reliable reference distribution. Additionally, generalizability and robustness pose challenges. Variability in imaging protocols, scanner models, and patient populations introduces domain shifts that may be misclassified as pathologies. Addressing these domain shifts requires robust domain adaptation strategies and better integration of domain invariance mechanisms into UAD frameworks. Beyond these challenges, different classes of UAD approaches exhibit distinct methodological limitations. Methods based on feature representations often produce coarse anomaly maps, while approaches relying on synthetic training data may struggle to generalize to real pathologies. Reconstruction-based methods must balance accurate reconstructions without inadvertently reconstructing anomalies, often requiring additional post-processing and carefully designed anomaly scoring strategies. In addition, the computational complexity of many UAD methods limits their scalability and hinders applicability in clinical workflows, particularly when considering 3D processing. Finally, a lack of standardized evaluation and benchmarking frameworks further complicates comparisons between methods, making it difficult to assess their reliability or determine their readiness for clinical integration.

At the same time, the rapid development of UAD methods has led to a wide variety of methodological approaches, including novel architectures as well as variations and combinations of existing methods. This has led to several literature reviews providing general overviews of anomaly detection (Chalapathy and Chawla, 2019; Ruff et al., 2021) and surveys focusing on the medical domain (Fernando et al., 2021; Tschuchnig and Gadermayr, 2022). However, existing reviews typically address anomaly detection either in a general context or across diverse medical imaging domains. As discussed above, the specific characteristics of brain MRI motivate a dedicated review of modeling approaches, data handling, and evaluation strategies.

Our objective is to bridge this gap and to offer researchers a valuable resource by providing a thorough literature review of recent DL-based UAD methods for brain MRI, with three primary goals:

- We provide an introduction to commonly applied UAD methods and brain MRI data sets.
- We present an overview of research findings in the field of UAD for brain MRI, by systematically collecting and grouping recent work and highlighting the different approaches to UAD in brain MRI.
- We analyze existing approaches and explore the context and differences, challenges and future directions in this field.

The organization of our work is as follows:

1. **Systematic literature search:** We begin by describing the approach employed in our literature search, reporting used search queries and utilized resources.
2. **Exploration of data sets and methods for UAD:** We provide a collection of commonly used UAD data sets and provide an overview of common approaches to UAD in brain MRI.
3. **Discussion of the context, challenges and future directions:** We compare the reviewed approaches concerning variable aspects of the UAD pipeline that we identified as important. Furthermore, we discuss the progress made in the literature and highlight trends and challenges observed in the reviewed literature.

1.1. Literature search

Our literature search aimed to compile a collection of peer-reviewed studies that employed DL methods in an unsupervised manner to identify abnormalities in brain MRI scans. To initiate our literature search, we conducted an extensive search across relevant databases, namely PubMed, the SPIE digital library, Google Scholar and the IEEE Proceedings. The search was last performed in December 2024. We gathered various search terms, which were organized into the following query: *Brain AND (Unsupervised OR Self-Supervised OR Self Supervised OR Weakly-Supervised OR Weakly Supervised) AND (Anomaly Detection OR Anomaly Segmentation OR Outlier Detection)*. In addition to these database searches, we extended our scope by including research contributions from prominent conference proceedings such as IEEE Engineering in Medicine and Biology Society (EMBC), International Symposium on Biomedical Imaging (ISBI), Medical Image Computing and Computer Assisted Interventions (MICCAI) and Medical Imaging with Deep Learning (MIDL). Note that in contrast to databases like PubMed, for most conference proceedings, there are no advanced search queries that include keywords or general information other than the title of the publications. Therefore, we collected all publications within the aforementioned conference proceedings and refined this collection by implementing a post-filtering mechanism based on titles containing any of the following keywords: *Unsupervised OR Self-Supervised OR Self Supervised OR Weakly-Supervised OR Weakly Supervised OR Anomaly Detection OR Anomaly Segmentation OR Outlier Detection*. Finally, we added relevant work, suggested by other researchers. Our focus in this review was on solutions rooted in DL, which led us to initiate our search from the year 2013 onward, as we did not anticipate the existence of DL-based UAD methods before that period.

Our search yielded a compilation of 203 research items. Subsequently, we performed a manual review to exclude any false positives. Our exclusion criteria primarily targeted studies exploring modalities other than brain MRI or relying on supervised learning techniques. After manual filtering, we arrived at a refined selection of 89 research items that aligned with our criteria. This curated collection serves as the foundational resource for our comprehensive review.

2. Anomaly detection

Anomaly detection (AD) encompasses a class of machine learning tasks aimed at identifying data points that deviate significantly from a normative distribution. In medical imaging, such deviations can indicate the presence of pathological conditions. Formally, given a data space $\mathcal{X} \subseteq \mathbb{R}^D$, anomalies can be defined as elements $x \in \mathcal{X}$ that deviate from the normal data distribution $\mathbb{D}^+$ as stated by Ruff et al. (2021). In brain MRI applications, $\mathbb{D}^+$ is typically estimated from healthy patient data. The task of AD involves learning this normal distribution and identifying deviations that indicate anomalies. In practice, the sampled data distribution $\mathbb{D}$ may be a mixture of normal and anomalous data. This mixture can be expressed as:

$$\mathbb{D} = (1 - \eta) \cdot \mathbb{D}^+ + \eta \cdot \mathbb{D}^-$$

where $\mathbb{D}^+$ represents the normal data distribution, $\mathbb{D}^-$ represents the anomalous data distribution, and η is the contamination factor indicating the proportion of anomalies in the data set (Ruff et al., 2021). Many AD methods can be designed to account for small contamination factors ($\eta > 0$) without significant degradation in performance (Zhou and Paffenroth, 2017). However, in brain MRI applications, it is common practice to assume that the data distribution $\mathbb{D}$ is identical to $\mathbb{D}^+$, i.e., the contamination factor $\eta = 0$.

2.1. Types of anomalies and detection levels

Anomalies can be categorized into point anomalies, contextual anomalies, and collective anomalies (Ruff et al., 2021). Point anomalies are single data samples that deviate from the normal data distribution. Considering AD in brain MRIs, point anomalies can be single samples that indicate the presence of a tumor or, in the voxel-wise case, single voxels that deviate from the normal data distribution. Contextual anomalies are data samples that are considered anomalies in a specific context, but not in another context. For example, enlarged ventricles are a common finding in older patients due to normal aging processes, but they may be considered anomalous in younger patients. Similarly, vascular lesions, such as microbleeds or white matter hyperintensities, may be expected in older individuals but could signify pathological conditions if observed in younger patients. In a voxel-wise context, contextual anomalies can be single voxels that are considered anomalies in a specific region of the brain, but not in another region. Collective anomalies are groups of data samples that deviate from the normal data distribution. For example, they can correspond to groups of samples indicating the presence of a specific disease, such as tumors or multiple sclerosis. In the voxel-wise case, collective anomalies correspond to groups of voxels that form pathological patterns such as tumors or lesions. In practice, many clinically relevant abnormalities appear as spatial patterns rather than isolated voxel deviations. Consequently, many reconstruction- and feature-based methods primarily target deviations that form spatial patterns. Furthermore, some approaches explicitly incorporate contextual information to enhance the detection of contextual anomalies. While anomaly detection generally focuses on identifying irregularities within a known data distribution, Out-of-Distribution (OOD) detection identifies data points originating from entirely different domains. For example, a chest X-ray in a brain MRI data set constitutes OOD data, whereas a brain tumor is an anomaly within the MRI domain. The challenges of distinguishing OOD samples from anomalies are discussed in Section 6.2.

2.2. Anomaly detection approaches

Unsupervised Anomaly Detection (UAD): UAD relies solely on unlabeled data, typically derived from (mostly) healthy individuals, for training. In this setting, the goal is to model the underlying data distribution and identify deviations indicative of anomalies without requiring explicit anomaly labels.

Semi-Supervised Anomaly Detection: Semi-supervised anomaly detection incorporates partially labeled data sets, where a subset of anomalies is explicitly identified. This limited supervision aids in calibrating models and improves their ability to detect specific types of anomalies.

Weakly-Supervised Anomaly Detection: Weakly-supervised approaches leverage imprecise or indirect annotations, such as coarse labels or approximate regions of interest, to achieve more detailed anomaly detection and segmentation. These methods aim to enhance model performance while reducing dependence on precise, fine-grained labels. For instance, weakly supervised models can be used to segment anomalies while only having access to sample-level labels.

Supervised Anomaly Detection: Supervised anomaly detection requires fully labeled data sets for both normal and anomalous classes, transforming the task into a binary classification problem. In many cases, synthetic anomalies are generated using domain knowledge to simulate pathological conditions.

The supervised setting can be very effective when a large number of labeled data samples is available and the anomalies are well-defined. However, in practice AD aims to detect a wide range of anomalies that are sometimes unknown beforehand, making the unsupervised settings more common.

3. Approaches to UAD in brain MRI

In this section, we provide an overview of common approaches to the UAD task. We categorize the approaches based on their underlying anomaly scoring mechanism. First, we introduce feature-based approaches that typically score anomalies based on discrepancies or distances in the feature space of CNNs. Second, we introduce reconstruction-based approaches, which typically rely on encoder–decoder architectures. These include both deterministic models, such as autoencoders (AEs), and probabilistic or generative models (GM), such as variational AEs or diffusion models. Lastly, we introduce synthetic anomaly methods that are trained by using specific pretext tasks to allow the segmentation of pathologies at test time. A general overview of all approaches is shown in Fig. 2 and Table 1.

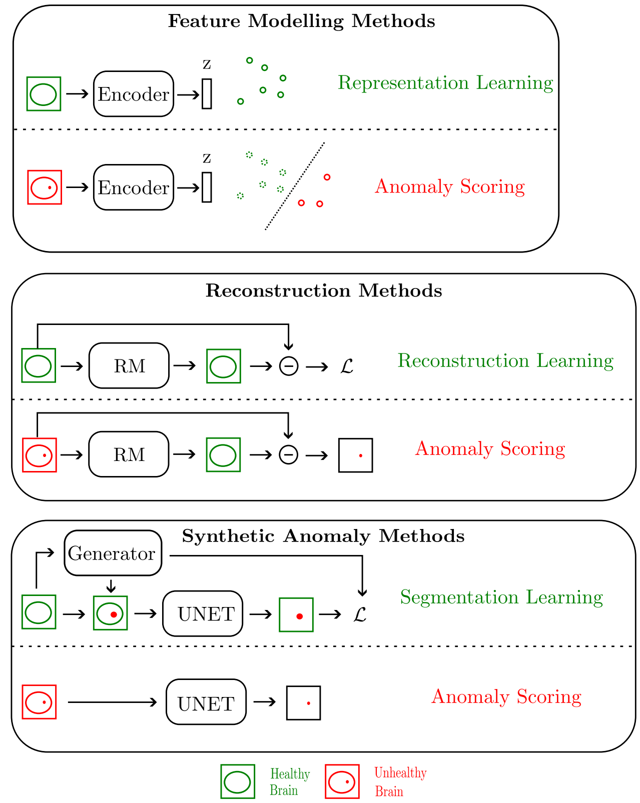

Fig. 2. Schematic drawing of the most common approaches for UAD in brain MRI. **Feature-Based:** A representation of healthy brains is learned and unhealthy brains are detected as outliers from the learned representation space. **Reconstruction-based:** A reconstruction model (RM), such as an autoencoder or a generative model, is trained to reconstruct healthy brains. Unhealthy structures are identified via reconstruction errors. **Synthetic anomaly methods:** An Unet (Ronneberger et al., 2015) is trained to detect synthetic anomalies, added to healthy brains by the generator. Unhealthy structures can directly be segmented by the trained Unet.

3.1. Feature modeling methods

Feature-based approaches are typically trained to encode normal or healthy images into a representation space, also called latent space. Assuming that the latent representation of the training data reflects the distribution of normal data, anomalies can then be identified as deviations from this learned representation of normality. Often, density estimation techniques are used to score outliers (Laurikkala et al., 2000), e.g., utilizing the Mahalanobis distance (Lee et al., 2018) or evaluating the log-likelihood of Gaussian Mixture Models (Roberts and Tarassenko, 1994; Lüth et al., 2023). Furthermore, machine learning models, like one-class SVMs (OC-SVM) (Schölkopf et al., 1999) are applied to directly learn a decision boundary, enabling a direct classification (Das et al., 2010; Alaverdyan et al., 2020). In the field of UAD in brain MRI, feature modeling techniques are often applied in global pathology detection as they do not directly provide a pixel-wise or voxel-wise localization of the unhealthy regions. Therefore, the feature discrepancies are often upsampled to the original size (Zimmerer et al., 2019a; Meissen et al., 2022c) or the discrimination is applied patch-wise (Martins et al., 2019c,b; Alaverdyan et al., 2020).

3.2. Reconstruction methods

Reconstruction-based approaches are the most common methods for UAD in brain MRI, using networks designed for image reconstruction. These approaches typically rely on encoder–decoder architectures that

Table 1
Overview of anomaly detection methods categorized by target task, minimal required supervision, and representative models. Classification tasks require anomaly scores across an entire image. Segmentation tasks, additionally require fine-grained anomaly localization at the pixel or voxel level. The table distinguishes minimal supervision levels, based on the degree of labeled data required for training.

Target task/Approach	Feature-based	Reconstruction-based	Synthetic groundtruth-based
Classification	Natively supported	Supported via aggregation	Natively supported
Segmentation	Supported via upsampling	Natively Supported	Natively supported
Typical architectures	GMM, OC-SVM, AE, VAE	AE, VAE, GAN, DDPM	Unet
Minimal supervision	Unsupervised	Unsupervised	Self-Supervised

learn to map input images into a latent representation and reconstruct them in the pixel space. Such models include deterministic reconstruction models, such as AEs, as well as probabilistic or generative models, such as variational AEs and diffusion models.

The core hypothesis of reconstruction-based methods is that the model is optimized to accurately reconstruct healthy structures observed during training but struggles to map or reconstruct structures and patterns that deviate from the training distribution. To identify unhealthy structures, these methods measure the discrepancy, typically quantified as the reconstruction error, between the input and its reconstruction. In regions corresponding to healthy structures, the reconstruction closely matches the input, resulting in minimal discrepancies. Conversely, in pathological regions, the reconstruction deviates significantly from the input, producing higher discrepancies. A voxel-wise anomaly score is then derived from these reconstruction errors, enabling the localization of anomalies.

It is important to note that this assumption does not extend to all approaches that leverage generative models and reconstruction errors for anomaly scoring. For example, methods based on density estimation can operate differently. Rather than purely relying on the assumption that the generative model replaces unhealthy structures with estimates of healthy ones, these methods utilize the estimated likelihood of the generative process to detect anomalies. Anomalous regions are identified as areas of low likelihood and can be corrected by replacing them with higher-likelihood representations during the generative process, effectively combining density estimation with the reconstruction-based approach (Marimont and Tarroni, 2021; Pinaya et al., 2022b; Marimont et al., 2023).

3.3. Synthetic anomaly/synthetic groundtruth-based methods

In the domain of UAD in brain MRI, another common strategy involves the use of synthetic training data or a synthetic groundtruth, which we refer to as "Synthetic Anomaly Methods". While these approaches can be categorized as self-supervised or even supervised, we distinguish them here to emphasize their specific focus on leveraging synthetic anomalies. This separation aims to reduce potential confusion, as the term "self-supervised" encompasses a broader range of techniques beyond synthetic anomaly generation, such as contrastive learning or masked image modeling. Synthetic Anomaly methods generate synthetic anomalies and integrate them into normal brain MRIs to create training data for anomaly detection. Subsequently, segmentation networks are trained to segment the added anomalies within the images. During testing, this segmentation network is directly employed to provide prediction masks for the detected pathologies. For the segmentation task, architectures akin to Unet are frequently deployed and trained with segmentation loss functions.

While this approach is shared among various methods, distinctions often arise in the nature and integration of synthetic anomalies into real images during training. A rudimentary approach coming from the generic image domain involves cutting and pasting the abnormal patterns into the real images (Li et al., 2021; Zavrtanik et al., 2021). Nevertheless, recognizing that the aim is to segment subtle pathologies that seamlessly blend with the normal brain anatomy, recent approaches adopt more sophisticated fusion strategies (Tan et al., 2021, 2022a). Additionally, a pivotal aspect of Synthetic Anomaly methods that rely on this "generate-and-detect" strategy is to generate pathologies with considerable variance. This is important as UAD methods aim to function as robust anomaly detectors capable of identifying a broad spectrum of anomalies (Zimmerer et al., 2022a).

4. Data sets for UAD

When developing DL models, the selection of appropriate data sets for training and evaluation is a critical consideration. While UAD methods can, in principle, be trained with contaminated data (i.e., $\eta > 0$) (Zhou and Paffenroth, 2017; Hendrycks et al., 2019; Ruff et al., 2021), only a minority of approaches in the domain of brain MRIs leverage such training strategies (Cai et al., 2023; Siddiquee et al., 2024). More commonly, training data sets consist of brain scans labeled as healthy, enabling models to learn a reference distribution of purely normal samples. Popular training data sets include both in-house data sets and publicly available data sets such as IXI, HCP, and Cam-CAN. Additionally, subsets of cognitively normal participants from data sets like ADNI, OASIS-3, and UK Biobank are frequently used.

Evaluation data sets, by contrast, are typically composed of scans with pathologies to assess the models' anomaly detection capabilities. Commonly used evaluation data sets include BraTS, which provides detailed segmentation masks for brain tumors, and data sets like ATLAS and ISLES, which focus on stroke lesion segmentation. Other specialized data sets, such as MSSEG, MSLUB, and WMH, are used for conditions like multiple sclerosis and white matter hyperintensities.

A comprehensive overview of the data sets, including their pathology types, number of subjects and imaging modalities, is provided in Table 2. Fig. 10 illustrates the distribution of these data sets across the studied approaches, highlighting the frequent use of in-house data sets, which are often tailored to specific research objectives but may lack generalizability. Further discussion on nuances such as the definition of "healthy" and the interpretation of abnormalities is provided in subsequent sections.

5. Review of approaches to UAD in brain MRI

In this chapter, we provide overview tables, grouped by the underlying UAD approaches. We describe the approaches briefly, focusing on the developed method and outcome of the respective studies.

5.1. Feature modeling methods

Feature extraction is commonly performed using either AE-based architectures (see Section 5.2.1 for a detailed introduction) or self-supervised learning frameworks such as SimCLR (T. Chen et al., 2020). These approaches learn representations in a lower-dimensional latent space that preserve relevant structural characteristics of the input data. In this latent space, statistical or machine learning methods, such as density estimation techniques like Gaussian Mixture Models or one-class classifiers, like OC-SVMs are employed. Anomalies are identified as deviations from this distribution, often quantified using measures like Mahalanobis distance, the negative log-likelihood, or the distance to the support vectors. This approach is illustrated in Fig. 2.

Table 2
Collection of publicly available MRI data sets used in UAD studies. DWI denotes Diffusion-weighted imaging and MRA denotes Magnetic Resonance Angiography. T1c denotes contrast-enhanced T1 weighting.

Data set	Pathology	Subjects	Weighting	Reference
IXI	None	560	T1, T2, PD, DWI, MRA	Biomedical Image Analysis Group (2026)
HCP	None	1200	T1, T2	van Essen et al. (2012)
CamCan (CC)	None	653	T1, T2, DWI, others	Taylor et al. (2017)
NFBS	None	125	T1	Puccio et al. (2016)
IBC	None	12	T1, T2, FLAIR	Pinho et al. (2020)
KIRBY	None	21	T1, T2, FLAIR, DWI	Landman et al. (2010)
CC359	None	359	T1	Souza et al. (2018)
Oasis (OA)	Mixed	1378	T1, T2, FLAIR, DWI, others	LaMontagne et al. (2018)
ADNI (AD)	Mixed	510	T1	Weiner et al. (2017)
UKB	Mixed	100 000	T1, T2, FLAIR, others	Sudlow et al. (2015)
FastMRI (fMR)	Mixed	6970	T1, T2, FLAIR	Zbontar et al. (2019)
MagNets (MG)	Mixed	78	T1, T2, FLAIR, others	Gullapalli (2011)
BR35H	Mixed	3000 (slices)	T1, T2, FLAIR	Hamada (2020)
PPMI	Mixed	600	T1, DWI	Parkinson Progression Marker Initiative (2011)
AGES Reykjavik Study (AGE)	Mixed	5764	T1, T2, FLAIR	Harris et al. (2007)
dHCP	Mixed	113	T1, T2	Hughes et al. (2017)
BraTS13 - BraTS21 (BR)	Tumor	30–2040	T1, T1c, T2, FLAIR	Baid et al. (2021)
CBS	Tumor	22	T1, T2, DWI	Pernet et al. (2016)
MSLUB (MB)	Multiple sclerosis	30	T1, T2, FLAIR	Lesjak et al. (2018)
MSSEG (MG)	Multiple sclerosis	38	FLAIR	Commowick et al. (2018)
MSISBI (MI)	Multiple sclerosis	82	T1, T2, FLAIR	Carass et al. (2017)
WMH	White matter lesions	60	T1, FLAIR	Kuijf et al. (2019)
ATLAS v1 (AT1)	Stroke	160	T1	Liew et al. (2022)
ATLAS v2 (AT2)	Stroke	655	T1	Liew et al. (2022)
ISLES (IL)	Stroke	250	FLAIR, DWI	Maier et al. (2017)

Table 3
Feature modeling methods for UAD in brain MRI. Tasks: Volume/Slice-wise Detection (Det), Segmentation (Seg). Representation: 2-Dimensional (2D), 3-Dimensional (3D), patched processing (-pp). Data set (Train/Eval): Inhouse (IH). Weighting: T1 contrast-enhanced (T1c), FLAIR (FL).

Citation	Method	Task	Dimension	Data set	Weighting
Martins et al. (2019a)	Symmetry based anomaly scoring	Det	3D-pp	IH/IH	T1
Alaverdyan et al. (2020)	One-Class SVM based on features of siamese AEs	Det, Seg	2D	IH/IH	T1, FL
Doorenbos et al. (2022)	OOD detection by the Mahalanobis distance based on SimCLR-features	Det	3D	MD/MD	T1
Kascenas et al. (2022b)	Context and local feature matching based on pretext tasks	Seg	2D-pp	BR/BR	T1, T1c, T2, FL
Meissen et al. (2022c)	Structural feature discrepancies as anomaly score	Det, Seg	2D	MD/MD, BR	T1
Pinon et al. (2023a)	One-Class SVM and Gaussian modeling based on AE-features	Seg	2D-pp	PPMI/PPMI	T1, DWI
Pinon et al. (2023b)	One-Class SVM based on AE-features	Seg	2D-pp	IH/WMH	T1, FL
Guo et al. (2023)	Encoder–decoder dissimilarity as anomaly score	Det	2D	BR35H/BR35H	T1, T2, FL
Frolova et al. (2023)	Feature-based anomaly score refinement	Det, Seg	2D	IXI/BR	T2
Lüth et al. (2023)	Gaussian modeling based on SimCLR-features	Det, Seg	2D	HPC/BR, IL	T2
Schwarz et al. (2024)	Student–teacher pyramid matching	Det, Seg	3D	IXI/BR	T2

Another direction within feature-based methods involves directly analyzing discrepancies between features extracted at different stages of a neural network. For example, anomaly detection can be performed by comparing features extracted from the encoder and decoder networks of an AE. Similarly, frameworks utilizing a student–teacher paradigm compare features derived from a pre-trained teacher model (often trained on general or mixed data) with those from a student model trained exclusively on normal samples. Often the cosine similarity in feature representations between the two models is used as an anomaly score. This approach is illustrated in 3. A general overview of feature-based approaches is provided in Table 3.

Alaverdyan et al. propose an approach utilizing a Siamese AE to extract features from patches of brain pairs (Alaverdyan et al., 2020). In addition to the reconstruction objective, the methods leverage the cosine similarity score between latent representations of the two encoders as a regularization term. Subsequently, the latent features are classified using an OC-SVM. According to the results, the OC-SVM identifies unhealthy patches given the extracted features and demonstrates the capability to discriminate epilepsy patients from healthy individuals.

Similar to Alaverdyan et al. Pinon et al. suggest using AE features to score anomalies in brain MRIs of *de novo* Parkinson patients (Pinon et al., 2023a). They explore using OC-SVMs to discriminate the feature vectors and measure the distance to a normative multivariate Gaussian distribution tailored to the features of healthy training data. The results indicate that reconstruction-based AEs are on par with the proposed method. Furthermore, they show competitive performance to supervised techniques, highlighting the challenge of the task.

Extending their work in a follow-up publication, Pinon et al. calibrate the OC-SVM with specific unhealthy samples of WMH, allowing for a margin of potential unhealthy patches (Pinon et al., 2023b). The authors state that this extension facilitates a patient-specific normal representation and decision, enhancing the method's adaptability to individual variances in brain anatomy.

Taking inspiration from radiological practices, Martins et al. employ symmetry-based anomaly detection by utilizing an AE to model typical asymmetries between the brain hemispheres (Martins et al., 2019a). This method extends previous approaches of the authors (Martins et al., 2019c,b) and utilizes AEs to capture normal variations in symmetry, identifying deviations from these patterns as anomalies during testing. The results show the potential of this approach in detecting epilepsy, highlighting the significance of symmetry analysis in brain MRI.

Doorenbos et al. employ the self-supervised technique SimCLR (T. Chen et al., 2020) to learn meaningful representations from different slices at varying plane views (Doorenbos et al., 2022). Anomaly scoring is done by fitting a multivariate Gaussian distribution to these representations and utilizing the Mahalanobis distance for scoring outlier slices. This approach achieved fourth place in the MOOD challenge's sample-wise detection task. Furthermore, the results suggest the potential for coarse localization of synthesized abnormalities.

Kascenas et al. propose a self-supervised context and feature matching approach for UAD (Kascenas et al., 2022b). In this approach, a

CNN is trained in a pretext task to differentiate between augmented patches of different locations, focusing on local and context information by employing different classification heads to the CNN backbone. At test time, each patch is processed by the network, and the classification head predicts match probabilities, which serve as anomaly scores. According to the results, this strategy demonstrates effectiveness in anomaly segmentation and outperforms several reconstruction-based baselines.

Frolova et al. propose to refine the anomaly score by relying on feature discrepancies (Frolova et al., 2023). They employ PatchCore (Roth et al., 2022) to obtain a coarse anomaly map that is multiplied with the detailed anomaly maps of different UAD methods. Furthermore, they extend PatchCore to 3D and train it on medical data. According to their results, all evaluated UAD methods show improved performance by the proposed refinement.

Lüth et al. investigate the use of contrastive pretext tasks using SimCLR to obtain feature representations (Lüth et al., 2023). Subsequently, a Gaussian Mixture Model is used to fit the healthy feature distribution using expectation maximization. The negative log-likelihood of the features is used for anomaly scoring. Furthermore, a pixel-wise score is derived by backpropagating the sample-wise log-likelihood to the input image. According to the results, their method outperforms reconstruction-based VAEs on most data sets.

Meissen et al. suggest leveraging the structural differences within the feature space of CNNs for anomaly scoring (Meissen et al., 2022c). Initially, they transform an input image into a predefined feature space of a ResNet CNN. Following this, they employ an AE to reconstruct the extracted features at multiple resolutions. Subsequently, they resize the features to match the original image size and compute the Structural Similarity Index Metric (SSIM) (Wang et al., 2004) between the input and the reconstructions. The results indicate that their proposed method is more effective in capturing anomalies that display small intensity gradients and primarily exhibit structural differences.

Similarly, Guo et al. introduce a method that utilizes the contrast between Encoder–Decoder features for UAD (Guo et al., 2023). Overall, cosine-similarity measures between encoder and decoder features are used as anomaly scores. Traditional feature-based methods often lean on pre-trained encoders from the natural image domain, which seldom align with the specific features needed for medical images. The authors address this by optimizing AE-based networks with contrastive learning to avoid the phenomenon of pattern collapse. The results indicate the methods' effectiveness, outperforming other state-of-the-art UAD approaches across various medical image types.

Schwarz et al. propose a 3D Student–Teacher Feature Pyramid Matching (3D-STFPM) approach for anomaly detection in brain MRIs (Schwarz et al., 2024). The framework leverages self-supervised patch learning and Axial-Coronal-Sagittal convolutions to adapt ImageNet-pretrained networks for 3D settings, following the approach in Wang et al. (2021). In this architecture, a teacher network is pre-trained on both healthy and unhealthy inputs to extract domain-specific features, while a student network is trained only on healthy inputs, with differences between their feature maps generating anomaly maps. The results show that the method outperforms a baseline f-AnoGAN in image- and pixel-wise anomaly detection on the BraTS data set.

5.2. Reconstruction methods

5.2.1. Autoencoder

AEs are widely used in UAD for brain MRI as shown in Table 4. As illustrated in Fig. 4, AEs consist of an encoder, $f(\boldsymbol{x})$, which maps the input $\boldsymbol{x}$ to a lower-dimensional latent representation $\boldsymbol{z}$, and a decoder, $g(\boldsymbol{z})$, which reconstructs the input as $\hat{\boldsymbol{x}} = g(f(\boldsymbol{x}))$. The model is optimized by minimizing the reconstruction error, typically defined as the pixel-wise mean squared error (MSE) or mean absolute error (MAE), to ensure accurate reconstruction of healthy, in-distribution data.

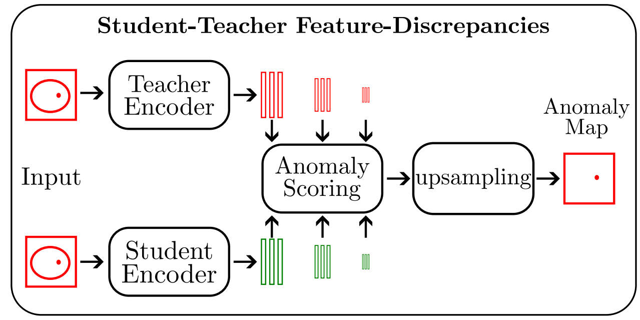

Fig. 3. Student–teacher anomaly detection: A pretrained Teacher network provides feature supervision for a Student network distilled on healthy data. At test time, feature discrepancies at different levels are upsampled to the input image resolution, generating a pixel-level anomaly map.

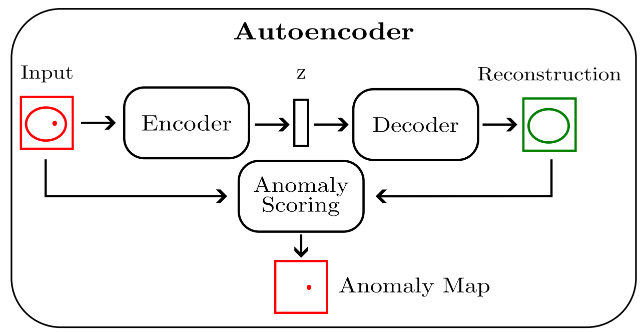

Fig. 4. Autoencoder-based anomaly detection: An autoencoder is trained to reconstruct healthy images. At test time, discrepancies between the input and its reconstruction highlight anomalies, forming an anomaly map.

While AEs can effectively capture the training data distribution of healthy brain MRIs, they often yield blurry reconstructions, making it challenging to discriminate smaller anomalies from noise in the residual map. Various strategies are proposed to enhance the quality of the reconstructed images.

Baur et al. explore using Laplacian pyramid representations to combine and reconstruct different frequency bands of MRI scans with AEs (Baur et al., 2020c). In this method, a sample is broken down using a 3-level Laplacian pyramid, each employing a separate AE to reconstruct the associated frequency component. According to the authors, this approach aids in creating high-quality reconstructions at different resolutions, which is crucial for small lesions. The results show that while the proposed method improves baseline AEs, it is outperformed by Gaussian-Mixture VAEs.

Integrating skip connections within the AE architecture, similar to those in Unets, emerges as another strategy to improve reconstructions. However, this can lead to content from the input image being copied, necessitating additional regularization methods.

Therefore, Baur et al. experiment with skip connections at different resolution levels and introduce dropout to the connections to prevent bypassing the latent space (Baur et al., 2020b). Furthermore, Bayesian AEs are built by deriving a predictive mean from 100 Monte Carlo samples given a certain dropout probability. According to the results, skip connections enhance segmentation slightly with no clear indication regarding the optimal placement of these connections. Furthermore, the Bayesian variant outperforms the classic AE baseline but falls short of the non-Bayesian version with skip connections.

In a subsequent exploration, Baur et al. further investigate the use of skip AEs for UAD, examining various pathologies (Baur et al., 2021b). They evaluate the skip AE against a straightforward thresholding method and a supervised U-Net CNN across several lesion types,

Table 4
Autoencoder-based approaches for UAD in brain MRI. Tasks: Volume/Slice-wise Detection (Det), Segmentation (Seg). Representation: 2-Dimensional (2D), 3-Dimensional (3D), patched processing (-pp). Data set (Train/eval): Inhouse (IH). Weighting: T1 contrast-enhanced (T1c), FLAIR (FL).

Citation	Method	Task	Dimension	Data set	Weighting
Atlason et al. (2019b,a)	AEs with linear unmixing	Seg	3D-pp	AGE/AGE	T1, T2, FL
Baur et al. (2020b)	AEs Laplacian pyramid decomposition	Seg	2D	IH/IH, MB	FL
Baur et al. (2020c)	Bayesian AEs with skip connections	Seg	2D	IH/IH	FL
Baur et al. (2021b)	Application Study of Bayesian skip AEs	Seg, Det	2D	IH/WMH, BR	FL
Aswani and Menaka (2021)	AEs with singular value decomposition	Seg	2D-pp	HCP, BR/BR	T1c
Naval Marimont and Tarroni (2021)	Auto-Decoders for implicit field learning	Seg	2D, 3D	HCP/BR	T2
Kascenas et al. (2022a)	Denoising AEs	Seg	2D	BR/BR	T1, T1c, T2, FL
Chen and Konukoglu (2022)	Constrained adversarial AEs	Seg	2D	HCP/BR	T2
Bercea et al. (2022)	AEs with disentangled latent space	Seg	2D	OA, AD/MB, MI, IH	FL
Behrendt et al. (2022b)	AEs with outlier cleaning	Det	3D	OA, IXI/BR	T1
Ma et al. (2022)	Application of AE reconstructions as a biomarker	Det	1D	UB/UB	T1, FL
Muñoz-Ramírez et al. (2022)	Application study of AEs	Seg	2D	IH/IH	DWI
Cai et al. (2023)	Dual AEs to leverage unlabeled data	Det	2D	BR35H/BR35H	T1, T2, FL
Luo et al. (2023)	Application study of 3D AEs	Det, Seg	3D	IXI/BR, IH	T2
Avci et al. (2024)	Dual Deformation AE	Seg	3D	AD/AD	T1
Rashmi et al. (2024)	Masked AE for anomaly detection	Seg	2D	IXI/BR, MB	T1, T2

including MS lesions, microangiopathy, and glioblastoma. The findings suggest that the unsupervised skip AE competes well against the supervised U-Net in many data sets, especially when adapting across data sets or when only a limited amount of training samples is available.

An interesting extension to AEs are denoising AEs (DAE). In contrast to the AE, the DAE is designed to learn the reconstruction of the original input $\boldsymbol{x}$ from a noise-corrupted version $\tilde{\boldsymbol{x}}$. Here, noise can be of various types, e.g. by cutting out entire areas of the image as is the case for CE-AEs (Pathak et al., 2016). Such denoising processes force the DAE to embed the structural characteristics of the underlying data distribution (Bengio et al., 2013; Alain and Bengio, 2014).

Following this principle, Kascenas et al. employ an Unet architecture as AE to aim for improved reconstructions (Kascenas et al., 2022a). To prevent the Unet from copying the image, they add a unique type of upscaled Gaussian noise into the image. The idea is to force the Unet to eliminate the noise from the image, preventing the copying task. During testing, the Unet reconstructs unhealthy brain scans without any added noise and demonstrates superior performance on the evaluated BraTS data set, outperforming standard AE approaches that rely on a more restricted, dense latent space.

Avci et al. introduce MORPHADE, a framework for analyzing Alzheimer's Disease using 3D deformable autoencoders (Avci et al., 2024). The method extends the method presented in Bercea et al. (2023c) to 3D, leveraging a dual deformation strategy. A constrained deformation field prediction reduces false positives by refining pseudo-healthy reconstructions, while an unconstrained deformation field prediction is used to identify atrophy through foldings in the deformation maps. The resulting anomaly maps are used to localize AD-related atrophy in clinically relevant brain regions, such as the hippocampus and amygdala. According to the results, the proposed method achieves an AUC of 80% for AD detection, outperforming both supervised and unsupervised baselines. Furthermore, the results indicate alignment with clinical atrophy assessments.

Another research direction focuses on structuring the latent space of AEs to improve the learned representations.

Chen et al. present an adversarial AE for the UAD task (Chen and Konukoglu, 2022). By regularizing the latent space structure through a loss function constraint, they aim to achieve better consistency within the latent representations. This constraint measures the distance between the latent vector of the original image and that of the reconstructed image. Their experiments indicate that this constraint enhances detection performance compared to VAEs. Moreover, the authors show that the weighting parameter of the constraint plays a crucial role in detection performance.

While the adversarial training in Chen and Konukoglu (2022) impacts the whole structure of the latent manifold, Bercea et al. propose to boost the transfer learning capabilities of AEs by disentangling the latent space of different AEs in a federated learning setting (Bercea et al., 2022). This idea revolves around training global model parameters with diverse data sets, for example, from different institutions. They suggest a framework that disentangles model parameters and latent dimensions into shape and appearance, sharing only the shape parameters among the different data sets. By integrating this disentangled extension into the federated learning framework, they observe improved performance across all data sets, surpassing the AE baseline trained centrally with all data.

Taking a unique route, Atlason et al. harness a 3D AE coupled with linear unmixing for unsupervised detection of White Matter lesions (Atlason et al., 2019b). Instead of traditional reconstruction of input volumes, they propose reconstructing from five different channels corresponding to segmentation predictions of various brain tissue types, using linear unmixing for combination. Their results indicate the strong performance of this approach in segmenting white matter lesions, even exceeding a supervised baseline.

In a follow-up study, Atlason et al. suggest enhancements like a scale-invariant loss function, increased regularization, additional MR-sequences, inhomogeneity correction, and incorporating more data sets for evaluation (Atlason et al., 2019a). These steps result in an improved system with higher performance than the initial study.

Aswani et al. propose a dual AE with latent space optimization for brain tumor segmentation (Aswani and Menaka, 2021). Their work utilizes singular value decomposition and an auxiliary encoder that processes the AEs' reconstruction to obtain a rich latent representation. A decoder computes the final reconstruction from this latent space, which is then compared with the input scan for anomaly scoring. Vertical and horizontal crops are used as input within their method. The results indicate strong performance compared to other unsupervised approaches and similar performance to supervised segmentation networks.

Luo et al. investigate a residual 3D AE architecture and evaluate it based on brain tumors, MS lesions and cerebral infarction (Luo et al., 2023). Similar to Baur et al. (2021a), they show that a spatial latent space enables the AE to reconstruct unhealthy anatomy, contradicting the UAD principle. The results furthermore show that their method outperforms 2D versions of VAE and f-AnoGAN.

Marimont et al. diverge from traditional voxel-based representations, proposing to learn implicit field representations for 3D MRI scans (Naval Marimont and Tarroni, 2021). Through a continuous function mapping of spatial coordinates and latent variables to a probability distribution over intensity ranges, voxel intensities are classified into distinct clusters to generate images. Anomaly scoring is achieved through cross-entropy loss, comparing voxel-wise intensity clusters between test and restoration images. The results indicate this method's potential for accurate segmentation with moderate inference time. Rashmi et al. present a transformer-based masked autoencoder framework, leveraging computationally efficient Swin-Transformers (Rashmi et al., 2024). The approach leverages masked AEs to reconstruct images by cutting out regions of input images during training and tasking the transformer with inpainting. During inference, a sliding window approach is utilized. Residual maps between the reconstructed pseudo-healthy images and the original images are computed for anomaly scoring. Furthermore, the method incorporates extensive post-processing steps, including morphological filters and connected component analysis and a Gaussian mixture model-based segmentation to enhance the precision of the anomaly maps. The results indicate that the model, in combination with the extensive post-processing steps, demonstrates competitive performance on data sets like BraTS21 and MSLUB, compared to autoDDPMs (Bercea et al., 2023b).

Across the reviewed studies of brain MRIs, most approaches operate under the assumption that the training data consists solely of healthy or in-distribution samples. However, achieving this level of normality, especially in the medical domain, is challenging and human labeling errors can further lead to impure data sets.

Behrendt et al. investigate these effects of impure training data in the UAD setting (Behrendt et al., 2022b). They employ an AE architecture and simulate impure training data by injecting certain percentages of abnormal data into the training data set. The findings indicate that even a small fraction of unhealthy data within the training distribution can severely degrade the performance of AE-based UAD models. To address this, they suggest a straightforward outlier removal strategy during training and online data cleaning. According to the results, they manage to attenuate the negative impact of impure training data by periodically discarding outliers based on the average reconstruction error from the training set. However, this comes at the cost of removing samples labeled as healthy.

Cai et al. (2023) propose utilizing noisy or unlabeled data sets. The authors propose using ensembles of AEs for dual-distribution learning, incorporating a healthy data pool (normative AE) and a mixed data pool containing normal and abnormal images (mixed AE). At test time, the intra-discrepancy of the normative AEs' reconstructions and the inter-discrepancy of both AEs' reconstructions are used as anomaly scores. Furthermore, a self-supervised refinement network is used to improve the anomaly scores. The results indicate that strong sample-wise detection can be achieved even without the mixed data and adding the unknown and mixed data distribution can further increase the performance.

While the presented studies propose unique UAD approaches, some assess the efficacy of existing AE-based UAD approaches in different settings.

Ma et al. leverage the reconstruction errors of AEs as biomarkers for outlier detection within a large data set comprising 15,000 brain MRI scans (Ma et al., 2022). Alongside conventional biomarkers like ventricle volume or cortical thickness, they employ AEs in the UAD setting, utilizing the residual error as a biomarker for outlier detection. Their findings point to the presence of abnormal phenotypes within the data set, which could be identified using the AE-based biomarkers.

Munoz-Ramirez et al. explore and compare different AE architectures to detect subtle anomalies in brain scans of *de novo* Parkinson's patients (Muñoz-Ramírez et al., 2022). They compare spatial and dense AEs and VAEs, scoring anomalies based on the residual error between input and reconstruction. The results hint at the DL model's capability to spot potentially pathological areas in *de novo* Parkinson's patients without requiring expert delineation. This, they argue, could ease the extraction of neuroimaging biomarkers for Parkinson's disease in future research, although validation on larger cohorts is needed for confirmation.

5.2.2. Variational autoencoder

Variational Autoencoders (VAEs) (Kingma and Welling, 2014) are generative models that extend the principles of AEs by introducing a probabilistic latent space constrained by a prior distribution, typically modeled as a Gaussian distribution. A collection of VAE-based approaches is provided in Table 5.

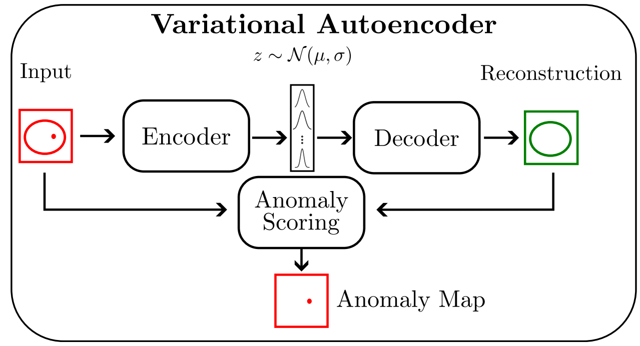

Fig. 5. Variational autoencoder-based anomaly detection: A variational autoencoder is trained to reconstruct healthy images. At test time, discrepancies between the input and its reconstruction highlight anomalies, forming an anomaly map. Additionally, some approaches leverage the KLD between the prior distribution enforced during training and the encoded distribution as an auxiliary anomaly score.

The VAE framework, illustrated in 5, approximates the posterior distribution of the latent variables using variational inference. The optimization objective minimizes the reconstruction error while regularizing the latent space by encouraging the approximate posterior to match the prior distribution. This is achieved through the Kullback–Leibler divergence (KLD) term, resulting in the VAE objective:

$$\mathcal{L}_{VAE} = \mathcal{L}_{Rec}(\boldsymbol{x}, \hat{\boldsymbol{x}}) + D_{KL}(q_\phi(z|\boldsymbol{x}) \parallel p_\theta(z)),$$

where $\mathcal{L}_{Rec}$ represents the reconstruction error, and D_{KL} enforces the structure of the latent space. The probabilistic nature of VAEs enables smooth interpolation within the latent space and facilitates sample generation. Additional flexibility can be achieved by using Gaussian Mixture Models instead of the standard Gaussian prior distribution, leading to Gaussian Mixture VAEs (GMVAE) (Dilokthanakul et al., 2017). However, similar to AEs, VAEs often produce blurry reconstructions (Bredell et al., 2022), which can challenge anomaly detection performance. Efforts to mitigate this issue include leveraging spatial context to improve reconstruction quality and enhance the detection of anomalies.

Zimmerer et al. propose a variational variant of the CE-AE (CE-VAE) for the UAD task (Zimmerer et al., 2019b). The objective is to utilize spatial context to learn discriminative features that remain robust against perturbations. Therefore, a two-path model with twin inputs and outputs is investigated. While one part mirrors a traditional VAE, the other is inspired by the CE-AE. The CE-VAE, acting as a denoising VAE, receives a distorted version of the input with a missing patch, contrasting the VAE branch, which receives a normal brain MRI. An additional term encompassing the CE reconstruction error is integrated into the VAE loss function. This approach aims to better harness the context of the input image, with results indicating enhanced performance in slice-wise and pixel-wise detection. The authors argue that the CE path in the CE-VAE facilitates better utilization of input context and latent variables for reconstruction and anomaly detection.

To further validate their proposed method (Zimmerer et al., 2019a), Zimmerer et al. extend their evaluation to real-world medical imaging applications, focusing on a large-scale population study involving 10,000 participants (Zimmerer et al., 2022b). Their method successfully identifies 31 previously unreported anomalies. However, a radiologist's evaluation underscores challenges in detecting subtle anomalies, casting doubts on the immediate clinical applicability of the method.

Shifting the focus to 3D context, Uzunova et al. (2019) exploit the volumetric information of brain MRI for UAD. They compare a 3D conditional VAE with a 2D counterpart, where the latter only uses small image patches as input, with the VAE conditioned by the patch location. The 3D VAE leverages 3D convolutional layers, contrasting

Table 5
Variational autoencoder-based approaches for UAD in brain MRI. Tasks: Volume/Slice-wise Detection (Det), Segmentation (Seg). Representation: 2-Dimensional (2D), 3-Dimensional (3D), patched processing (-pp). Data set (Train/eval): Inhouse (IH). Weighting: T1 contrast-enhanced (T1c), FLAIR (FL).

Citation	Method	Task	Dimension	Data set	Weighting
Akrami et al. (2020)	Robust VAEs for transfer learning	Seg	2D	MG/IL, BR	T1, T2, FL
Uzunova et al. (2019)	Position-conditioned 3D VAE	Seg	3D	BR/BR	T1, T1c, T2, FL
Zimmerer et al. (2019b)	Context denoising task for VAEs	Det, Seg	2D	HCP/BR, IL	T2
Zimmerer et al. (2019a)	KLD gradients as anomaly score	Det, Seg	2D	HCP/BR, IL	T2
Sato et al. (2019)	log-likelihood decomposition of VAEs	Seg	2D	IXI/BR, AT1	T1, T2
X. Chen et al. (2020)	Restoration by expectation maximizing of the VAE ELBO	Seg	2D	CC/BR, AT1	T1, T2
Albu et al. (2020)	Focusing on Dissimilarities in the Latent Space of a VAE	Det	2D	HCP/BR	T2
Bengs et al. (2021)	3D VAEs with spatial erasing	Seg	2D, 3D	IH/BR, AT1	T1, T1c
Heer et al. (2021)	Comparison of OOD and UAD tasks	Det, Seg	2D	CC/CC, BR	T1, T2
Lambert et al. (2021)	3D VAE	Seg	3D	AD IBC, KIRBY/MG, ML, WMH, BR	FL
Marimont and Tarroni (2021)	VQ-VAE with PixelSNAIL for density estimation	Det, Seg	2D	MD/MD	T1
Akrami et al. (2021, 2022)	Quantile regression with VAEs	Seg	2D	MG/IL	T1, T2
Silva-Rodríguez et al. (2022)	Activation maps to refine anomaly scores	Seg	2D	BR, ICH/BR ICH	FL
Chatterjee et al. (2022)	Compact context encoding VAE	Seg	2D	IXI, MD/MD BR	T1
Bengs et al. (2022)	3D VAEs with age incorporation	Det	3D	IH/BR	T1
Behrendt et al. (2022a)	Sequential 3D VAEs	Seg	2D, 3D	IXI/BR, AT1	T1
Meissen et al. (2022d)	Study on the reconstruction error as anomaly score	Seg	2D	MD/MD	T1
Pinaya et al. (2021, 2022b)	VQ-VAE with transformers for density estimation	Det, Seg	2D, 3D	UB/WMH, BR, MB	FL
Bercea et al. (2023a)	Soft-intro VAE	Det, Seg	2D	fMRI, IXI/fMRI	T1
Marimont et al. (2023)	VQ-VAE with masked image modeling	Seg	2D	HPC/BR	T2
Raad et al. (2023)	Application of VAEs in neonatal brain MRI	Det	3D	dHCP/dHCP	T2
Loizillon et al. (2024)	Application of β-VAEs for WMH detection	Det, Seg	3D	UKB, IH/MG, AD, IH	FLAIR

the 2D version based on fully connected layers. The results favor the multi-modal approach of the 3D model over solely utilizing T1c-weighting, although the 3D tasks' complexity prevents the 3D model from outperforming its 2D counterpart, according to the authors.

Similarly, Bengs et al. utilize the 3D-contextual information inherent to brain MRI scans for UAD, proposing 3D-VAEs as an extension to classical 2D VAEs (Bengs et al., 2021). They also explore a patch-based denoising task akin to Zimmerer et al. (2019b) to regularize training and enforce the use of 3D spatial context. In contrast to Uzunova et al. (2019), their findings indicate superior performance for 3D methods over 2D counterparts and that the spatial erasing strategy can improve segmentation performance.

Similarly, Lambert et al. compare spatial and dense AEs and VAEs to their 3D counterparts (Lambert et al., 2021). To prevent posterior collapse, the authors use an adapted loss function, following a cyclical annealing approach for the KLD. The results indicate that for all data sets, the segmentation performance is improved by the use of 3D architectures. Furthermore, it is shown that the spatial variants of both AE and VAE tend to reconstruct the lesions from the input MRI scans. Hence, their performance is low compared to their dense counterparts.

In a similar study design, the authors Raad et al. conduct an application study of 3D VAEs for UAD in neonatal brain MRIs (Raad et al., 2023). For dense and spatial architectures, 3D VAEs and 3D AEs are compared with varying latent space dimensions. The results show that VAEs can be used to identify abnormalities in neonatal brain MRI. Furthermore, large latent dimensions in a spatial shape show benefits regarding the Area under Receiver Operator curve (AUC) in the compared test cases.

Taking a different track, Bengs et al. propose utilizing brain age as an additional biomarker, contrasting with earlier approaches solely reliant on 2D or 3D structural context (Bengs et al., 2022). By incorporating age as additional information, the authors aim to mitigate the challenges of age-related ambiguities in brain MRIs. They examine various age-conditioned 3D VAEs that capture the age information of a given brain. Furthermore, a multi-task learning strategy is proposed where the VAE encoder derives features for age estimation. The mean absolute difference between the estimated and actual age serves as an additional anomaly score. The results suggest that appending age information to the VAEs' latent space enhances discriminative performance. Moreover, employing age estimation error as an additional biomarker significantly elevates detection performance.

In addition to modifications in architecture, various approaches emphasize enhancing the anomaly scoring mechanism, which is pivotal in obtaining the anomaly map from an unhealthy input and a pseudo-healthy reconstruction.

Zimmerer et al. focus on pixel-wise KLD- or ELBO-based scores instead of solely relying on the reconstruction error (Zimmerer et al., 2019a). These scores are derived using the derivative of either the ELBO term or the KLD with respect to the input image. The authors argue that these scores exhibit greater robustness to hyperparameter changes. Their results indicate that the proposed methods maintain similar performance across a wide spectrum of hyperparameters, aiding in avoiding tedious hyperparameter tuning and aligning towards a more assumption-free UAD approach with the proposed KLD-based Anomaly score.

Chen et al. integrate image restoration methods with VAEs and GMVAEs for UAD in brain MRI scans (X. Chen et al., 2020). The VAE or GMVAE assists in deriving the prior distribution for image restoration through expectation maximization. According to the results, their method outperforms its purely reconstruction-based counterparts, namely VAE, AAE, and AnoGAN. They also report a relationship between the Dice score and the size of the segmented lesion, with larger lesions yielding higher segmentation performance.

Silva-Rodriguez et al. leverage a VAE to derive network activation maps for improved anomaly delineation (Silva-Rodríguez et al., 2022). Training the VAE to reconstruct normal images, they introduce constraints on the attention maps generated from the encoder via GradCAM techniques (Selvaraju et al., 2017) to ensure comprehensive image coverage. The results indicate that their method, AMCons and its variations, outperforms prior approaches, elevating segmentation performance. The results highlight the superior efficacy of imposing constraints at the image level over pixel-wise constraints when using L2-penalty functions.

Bercea et al. present an approach employing soft-intro VAE (SI-VAE) as an alternative to standard VAEs (Bercea et al., 2023a). They propose an SI-VAE with a reversed multi-scale embedding loss to compare input and reconstruction features at multiple encoder stages. This aims to improve the generation of healthy anatomy while preserving input-reconstruction coherence. Furthermore, they aim to minimize false positives in the anomaly map by employing a blend of LPIPS and $l1$-loss as a pixel-wise anomaly score. Their results indicate robust performance in detecting global and local pathologies compared to recent baseline methods.

Table 6
GAN-based approaches for UAD in brain MRI. Tasks: Volume/Slice-wise Detection (Det), Segmentation (Seg). Representation: 2-Dimensional (2D), 3-Dimensional (3D), patched processing (-pp). Data set (Train/eval): Inhouse (IH). Weighting: T1 contrast-enhanced (T1c), FLAIR (FL).

Citation	Method	Task	Dimension	Data set	Weighting
Baur et al. (2019)	Combination of VAEs and GANs	Seg	2D	IH/IH	FL
Baur et al. (2020a)	CycleGAN for UAD	Seg	2D	IH/IH, WMH	FL, T1
Han et al. (2021)	Adjacent slice reconstruction with GANs	Det	2D	OA, IH/OA, IH	T1, T1c
Nguyen et al. (2021)	Image inpainting with GANs	Seg	2D	NFBS/CBS	T1
van Hespen et al. (2021)	GANomaly for UAD	Seg	2D-pp	IH/IH	T1, FL
Bercea et al. (2023d)	Combination of SI-VAEs and GANs	Seg	2D	fMRI, IXI/fMRI, AT2	T1
Siddiquee et al. (2024)	GAN-based image-to-image translation	Det, Seg	3D	AD, IXI/AD, IH	T1

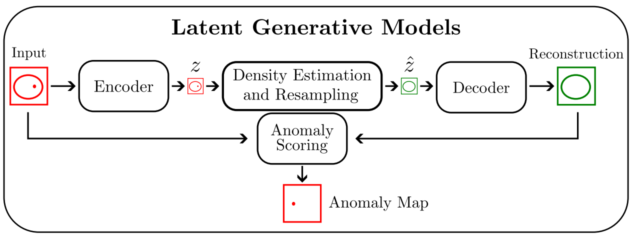

Fig. 6. Latent generative models for anomaly detection: An Encoder–Decoder architecture maps input images to a latent representation, where density estimation enables likelihood computation for each latent variable. Low-likelihood anomalies are replaced with samples from the learned normative distribution. Anomaly scoring is then performed by comparing the input to its resampled reconstruction.

Another avenue of research with VAEs focuses on their probabilistic properties.

Akrami et al. utilize an adapted version of the VAE to harness transfer learning, replacing the log-likelihood term of the VAE loss with the β-divergence (Akrami et al., 2020). This aims to mitigate the sensitivity of VAEs to outliers. According to the authors, this is a crucial step for successful transfer learning, especially when data set distributions diverge. The results showcase the robustness of their approach in capturing lesion locations and improved performance compared to a baseline VAE concerning the AUC values in re-training and transfer learning scenarios.

Sato et al. propose an anomaly measure for VAEs grounded on uncertainty estimation. They dissect the negative log-likelihood of a reconstructed pixel into predictable uncertainty and normalized error, assuming parametrized Gaussian distributions for each pixel (Sato et al., 2019). According to the results, their methodology consistently performs the baselines in T1-weighted MRI scans across the compared data sets and adds a layer of robustness against image variations.

Akrami et al. investigate deep quantile regression alongside VAEs, proposing a Quantile-Regression VAE (QR-VAE) (Akrami et al., 2021, 2022). Inspired by conformal predictions, this model estimates conditional quantiles for each input image, addressing the issue of underestimated variances. The QR-VAE, with its principled statistical approach for thresholding through computed p-values, outperforms the traditional VAE in detecting brain lesions, showcasing the potential of quantile regression in anomaly detection.

VAEs can be extended through hybrid approaches that incorporate advanced density estimation techniques for both anomaly detection and masking. These methods utilize VAEs or their quantized variants, such as Vector Quantized-VAEs (VQ-VAEs), to compress input data into a latent representation where anomalies are identified by estimating the probabilities of the latent variables. Latent values with low probabilities indicate outliers and can either be flagged directly as anomalies or replaced with more likely values based on the estimated probability distribution. While direct voxel-wise anomaly scoring is challenging to achieve with density estimation approaches, this replacement strategy enables the generation of a "healed" reconstruction image, which is then compared to the input, enabling fine-grained localization of anomalies. This framework is illustrated in Fig. 6.

For example, Marimont et al. utilize VQ-VAEs for compression and autoregressive modeling with PixelSNAIL to learn the probability density function of the latent representation of healthy brain MRI scans (Marimont and Tarroni, 2021). Their approach performs a form of inpainting within the latent space to replace latent variables with low likelihood with more likely ones. For pixel-wise anomaly detection, the pixel-wise difference between input and restoration is used as an anomaly map. For sample-level detection, the negative log-likelihood of the latents is used as anomaly score. The results show improved performance in comparison to plain VAEs on simulated pathologies.

Pinaya et al. take a similar approach and utilize VQ-VAEs in combination with an ensemble of transformer networks for density estimation instead of PixelSNAIL (Pinaya et al., 2021, 2022b). The results indicate a substantial performance improvement compared to the evaluated baseline methods across various real-world data sets, particularly the variant employing an ensemble of transformers.

In a subsequent work, Marimont et al. replace the autoregressive models used for estimating the density function and inpainting with two separate transformer-based networks (Marimont et al., 2023). One serves to identify abnormal tokens that have to be replaced. The second one is used to inpaint the abnormal tokens and is trained with masked image modeling. According to the results, their approach increases the segmentation performance while reducing inference time compared to earlier autoregressive approaches.

In an application study, Loizillon et al. assess the feasibility of using a β-VAE for anomaly detection of age-related WMHs in clinical routine brain MRIs (Loizillon et al., 2024). The model is pre-trained utilizing healthy MRIs from the UKB data set to learn the distribution of healthy brains and fine-tuned on 674 clinical FLAIR images from a heterogeneous in-house data set encompassing different hospitals with varying scanners and acquisition parameters. The approach achieves promising results, showing a correlation between Fazekas scores (a clinical WMH severity scale) and the detected lesion volumes, and demonstrates robustness to image quality variations. However, limitations include occasional failures to detect certain lesions due to poor reconstruction quality, particularly in heterogeneous or low-quality clinical data. The authors conclude that while UAD is encouraging, the invested VAE method requires further optimization before clinical deployment.

Meissen et al. highlight common pitfalls in using residual error for scoring anomalies, shedding light on the tendency of state-of-the-art UAD methods to perform "white-object-detection" rather than general anomaly detection (Meissen et al., 2022d). Their findings underscore the inability of AEs and VQ-VAEs to detect synthesized deviations in normal brain structure when the intensity remains unaltered.

Similarly, Heer et al. highlight the challenges and limitations of unsupervised lesion detection using VAEs, spotlighting the "OOD Blind Spot" (Heer et al., 2021). Their discourse emphasizes the difficulty in discerning anomalies stemming from actual lesions or domain shifts, urging further exploration to unravel the underlying causes and marking samples as OOD.

Table 7
Diffusion model-based approaches for UAD in brain MRI. Tasks: Volume/Slice-wise Detection (Det), Segmentation (Seg). Representation: 2-Dimensional (2D), 3-Dimensional (3D), patched processing (-pp). Data set (Train/eval): Inhouse (IH). Weighting: T1 contrast-enhanced (T1c), FLAIR (FL).

Citation	Method	Task	Dimension	Data set	Weighting
Pinaya et al. (2022a)	Diffusion models for sampling in a compressed space	Seg	2D	UB/BR MB WMH	FL
Wyatt et al. (2022)	Diffusion models with simplex noise	Seg	2D	NFBS/CBS	T1
Behrendt et al. (2023)	Patched Diffusion models	Seg	2D	IXI/BR, MB	T2
Behrendt et al. (2025)	context-conditioned Diffusion models	Seg	2D	IXI/BR, AT2, MB, WMH	T1, T2
Kascenas et al. (2023)	Investigation of the noise in Diffusion models	Seg	2D	BR	T1, T1c, T2, FL
Bercea et al. (2023b)	Self-masked sampling with Diffusion models	Seg	2D	IXI, fMRI/AT2	T1
Iqbal et al. (2023)	Frequency masked Diffusion models	Seg	2D	IXI/BR, MB	T2
Kumar et al. (2023)	Diffusion models with selective denoising	Seg	2D	BR/BR	T1, T1c, T2, FL
Liang et al. (2023)	Modality cycling with Diffusion models	Seg	2D	BR/BR	T1, T1c, T2, FL
Marimont et al. (2024)	Disyre v1: Cold diffusion with synthetic anomalies.	Seg	2D-pp	CC/BR, AT2	T1, T2
Naval Marimont et al. (2024)	Disyre v2 with sophisticated synthetic anomalies and test time ensembling	Seg	2D-pp	CC/BR, AT2	T1, T2
Behrendt et al. (2024b)	Adaptive ensemble of SSIM scores for robust anomaly scoring	Seg	2D	IXI/BR, AT2, MB, WMH	T1, T2
Behrendt et al. (2024c)	Leveraging the Mahalanobis Distance for Anomaly scoring	Seg	2D	IXI/BR, AT2, MB, WMH	T1, T2
Behrendt et al. (2024a)	Combining supervised and unsupervised anomaly detection	Seg	2D	IXI/BR, AT2	T1
Wolleb et al. (2024)	Masked Bernoulli Diffusion models	Seg	2D	BR	T1, T1c, T2, FL
Bercea et al. (2024)	Diffusion models with implicit guidance	Seg	2D	IXI/AT2	T1
Baugh et al. (2024)	Diffusion models with synthetic anomaly conditioning	Seg	2D	IXI/AT2	T1
Damudi and Kini (2024)	Single-step sampling with diffusion models	Seg	2D	NFBS/CBS	T1

5.2.3. GAN

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) are a class of generative models that use adversarial training to replicate the distribution of a given data set. A collection of GAN-based approaches is provided in Table 6. GANs consist of two neural networks: a generator, which creates samples from a random latent vector z, and a discriminator, which differentiates between real and generated data. The adversarial training process optimizes the generator to produce samples indistinguishable from real data, while the discriminator is optimized to detect samples that are generated by the generator network. This adversarial interplay is defined by the objective:

$$\min_G \max_D V(D,G) = \mathrm{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}\left[\log(D(\boldsymbol{x}))\right] + \mathrm{E}_{z\sim p_z(z)}\left[\log(1-D(G(z)))\right].$$

While GANs are known for producing sharp and high-quality samples compared to AEs and VAEs, they face significant challenges. Their training can be unstable, and they are prone to issues such as non-convergence and mode collapse (Kodali et al., 2018; Chu et al., 2020). Furthermore, since the vanilla GAN architecture generates images from a random noise vector rather than reconstructing images, their application in reconstruction-based UAD is not straightforward. To adapt GANs for UAD, methods such as AnoGAN (Schlegl et al., 2017) and its improved variant f-AnoGAN (Schlegl et al., 2019) derive a latent vector for a given input image, enabling anomaly detection, as illustrated in 7.

Another exploration is by Baur et al. where a CycleGAN Steganomaly (Chu et al., 2017) model is introduced, employing a style-transfer network to translate healthy data distributions into different styles corresponding to T1-, or FLAIR-weighted MRI scans (Baur et al., 2020a). Furthermore, filtering high-frequency bands from intermediate translations reveals significant advancements in Dice score and AUPRC, suggesting a superior performance of CycleGAN over AE-based methodologies, specifically with high-resolution data.

In another publication, Baur et al. fuse GANs and VAEs, creating a hybrid model called AnoVAEGAN to segment MS lesions (Baur et al., 2019). The adversarial training embedded in this setup demonstrates performance enhancements, indicating the potential synergy between GANs and VAEs.

Bao Nguyen et al. propose GAN-based inpainting (Nguyen et al., 2021). A target patch is shifted across the image, and the GAN is used to inpaint the patch region. Furthermore, superpixel segmentation is utilized as post-processing to obtain the segmentation. Their exploration of patch sizes suggests a performance peak for a 32 × 32 pixel window size, outperforming the baseline AnoGAN performance.

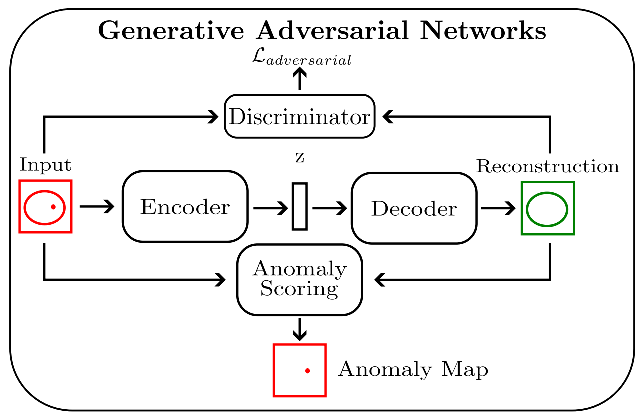

Fig. 7. Illustration of the f-AnoGAN framework: An encoder maps the input to a latent vector z, which a generator uses to reconstruct the image. During training (*), a discriminator is also trained, and an adversarial loss is minimized. At test time, the pixel-wise anomaly score, derived from input-reconstruction discrepancies, forms the anomaly map.

In another approach, Changhee Han et al. deploy a GAN-based reconstruction technique, employing an image-to-image GAN to predict adjacent slices (Han et al., 2021). Anomalies are identified by comparing the predicted slices to the real slice using the $l2$-error. According to the results, the proposed approach can detect accumulations of subtle anatomical anomalies and hyper-intense lesions, especially for late-stage Alzheimer's diseases. Furthermore, a notable enhancement in detection performance can be achieved by integrating the $l1$-loss into the Wasserstein GAN training objective.

Van Hespen et al. draw inspiration from the GANomaly framework, utilizing a GAN to compute anomaly scores based on the similarity between different latent manifolds (van Hespen et al., 2021). The results show that this methodology can yield a coarse segmentation of lesions while achieving a classification rate of 97.5% of the total brain infarct volume.

Bercea et al. propose a hybrid concept by integrating a GAN with a VAE-based GM for inpainting tasks (Bercea et al., 2023d). Through generating binary masks from a reversed SI-VAE's residual map, the subsequent GAN training facilitates inpainting, which, in turn, aids in reducing false positives and generating pseudo-healthy anatomy according to the results. Furthermore, the residual errors are combined with feature-based LPIPS scores to improve anomaly localization.

Table 8
Other approaches for UAD in brain MRI and comparative studies. Tasks: Volume/Slice-wise Detection (Det), Segmentation (Seg). Representation: 2-Dimensional (2D), 3-Dimensional (3D), patched processing (-pp). Data set (Train/eval): Inhouse (IH). Weighting: T1w (T1), T2w (T2), FLAIR (FL).

Citation	Method	Task	Dimension	Data set	Weighting
Saase et al. (2020)	Statistical methods for UAD	Det, Seg	3D	IH/IH	T1, T1c, T2, FLAIR, DWI
Baur et al. (2021a)	Comparative study for UAD	Seg	2D	IH/IH, MB, MG	FL
Meissen et al. (2022b)	Thresholding as a Baseline	Seg	2D	None/MB, BR, WMH	FL
Lagogiannis et al. (2023)	Comparative study for UAD	Det, Seg	2D	CC/AT1, BR	T1, T2
Zimmerer and Maier-Hein (2024)	Evaluation of UAD metrics	Det, Seg	2D	CC, MD/CC, MD	T1
Attyé et al. (2024)	Generative manifold learning	Det	3D	IH, AD/IH, AD	T1

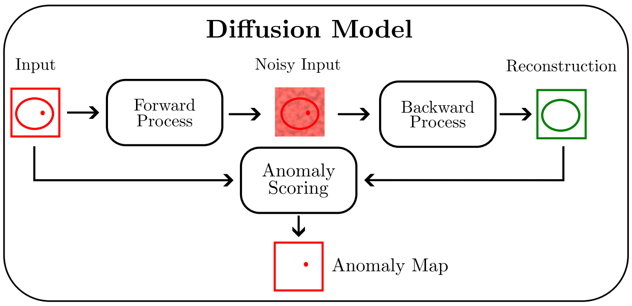

Fig. 8. Diffusion model-based anomaly detection: The forward process gradually adds noise to the input image up to a predefined noise level. In the backward process, noise is removed to generate a noise-free reconstruction. At test time, the anomaly score is computed by comparing the input with its reconstruction.

Siddiquee et al. propose a GAN-based image-to-image translation framework (Siddiquee et al., 2024). The model utilizes unannotated data sets containing a mix of healthy and unhealthy subjects to generate voxel-wise "additive maps" that transform input images into their corresponding healthy appearance. The difference between the input and the generated healthy image then serves as the basis for anomaly scoring. According to the results, the proposed approach outperforms existing methods in detecting Alzheimer's disease and headaches. While effective for patient-level detection, the generated localization maps are not highly precise, limiting their utility for segmentation. The method demonstrates strong potential for unsupervised detection in clinical scenarios with minimal labeled data.

5.2.4. Diffusion models

Denoising Diffusion Probabilistic Models (DDPM) represent a more recent class of GMs increasingly used for UAD in brain MRI. These models aim to learn the underlying data distribution through a two-step process: a forward process that gradually transforms input samples into noise and a backward process that generates samples from noise. In the forward process, a sample $\boldsymbol{x}_0$ is transformed into a noisy version $\boldsymbol{x}_t$ through a predefined noise schedule. This process progressively degrades the sample until it becomes pure noise at $t = T$. The backward process then reverses this transformation, recovering a reconstructed sample $\hat{\boldsymbol{x}}_0$ e.g., by using a trainable Unet. The optimization objective is derived via variational inference and simplified into the loss:

$$\mathcal{L}_{simple} = \|\epsilon - \epsilon_\theta(\boldsymbol{x}_t, t)\|^2,$$

as shown in Ho et al. (2020). Here ϵ represents the noise, added during the forward process and ϵ_θ is the predicted noise. Unlike traditional generative tasks that focus on creating entirely new samples, for the UAD tasks, DDPMs are often applied by starting the backward process from a partially noised version of the input rather than pure noise. A collection of DDPM-based approaches is provided in Table 7. The framework to use DDPMs in anomaly detection is illustrated in Fig. 8.

Pinaya et al. pioneer adapting latent DDPMs for UAD in brain MRI, combining a compression model with a DDPM for efficient anomaly segmentation (Pinaya et al., 2022a). Instead of solely relying on the model's capacities to reconstruct only healthy structures, the authors follow the combined approach of compressing the image to a latent representation and using the likelihood of the generative process to inpaint anomalies. Low-likelihood regions are identified and replaced with higher-likelihood representations during the reverse diffusion process. Through this "healing process", unlikely latent values are replaced with more probable ones via the DDPM, effectively filtering out anomalies in the reconstructed image. The results indicate robust performance with fast sampling comparison to autoregressive models that utilize transformers (Pinaya et al., 2022b).

Operating directly on the input image, Wyatt et al. introduce the AnoDDPM model, which limits the sampling steps at test time, thereby enhancing the efficiency of DDPMs for UAD in brain MRI (Wyatt et al., 2022). Furthermore, the authors propose to use a certain kind of simplex noise, instead of Gaussian noise. As shown in the results, the initialization with partially noised inputs, as opposed to pure noise, coupled with the adoption of structured simplex noise, can facilitate a more accurate tumor removal from unhealthy test images, thereby improving segmentation performance. This study underscores the pivotal role of the initial noise level and the noise structure in replicating unhealthy regions from the input image.

Behrendt et al. extend the application of DDPMs by employing a patch-wise approach for UAD (Behrendt et al., 2023). By adding simplex noise to individual image patches, reconstructing them, and stitching them back together, they aim to enhance image context understanding and anatomical coherence in brain image reconstruction. According to the results, the patch-wise methodology demonstrates superior performance to its non-patched counterpart, AnoDDPM, underscoring the potential of patched DDPMs in UAD.

Bercea et al. highlight the "noise-paradox" in DDPMs for UAD (Bercea et al., 2023b). They show the trade-off between the noise level chosen during evaluation and the detection of pathologies, particularly considering different pathology sizes. To overcome this issue, they propose to utilize a primary masking step based on an initial reconstruction, followed by a subsequent inpainting step, similar to Bercea et al. (2023d) but without the reliance on an external GAN-based inpainting model. Their results show that the approach outperforms AnoDDPM, particularly considering smaller lesion sizes.

The examination of noise types in DDPMs by Kascenas et al. provides an analysis of the noise type's impact on denoising tasks in UAD methods (Kascenas et al., 2023). Through comparative analysis between various noise types, including Gaussian, simplex, and coarse Gaussian noise, across their proposed DAEs (Kascenas et al., 2022a) and DDPMs, the study unveils performance enhancement achieved by the proposed coarse noise model. Although the authors show that DAEs outperform other methods in simplicity and inference speed, they accentuate the potential of DDPMs.

Liang et al. introduce a method called Masked Modality Cycles with Conditional Diffusion (Liang et al., 2023). This approach, akin to the CycleGan approach by Baur et al. (2020a), employs cyclic modality translation to learn the mapping from one sequence weighting to another for healthy brain MRIs. The authors accomplish the forward translation using a masked diffusion model, conditioned by the partially masked input image, to circumvent the reconstruction of abnormal structures. For the backward cycle, an Unet is utilized to obtain an

MRI in the original MRI weighting. The findings suggest that a simple Unet executing the cyclic transformation already demonstrates robust segmentation performance. Moreover, integrating the diffusion model can enhance the performance of the latent DDPM by Pinaya et al. (2022a) regarding the BraTS data set.

Iqbal et al. further develop the method of Behrendt et al. (2023) by applying the patching strategy in the Fourier space (Iqbal et al., 2023). Furthermore, they investigate a combination of cutmix (Yun et al., 2019) with the patched input image and the original. The results suggest improved performance compared to simple image-based patching.

Kumar et al. propose extending AnoDDPM with a selective denoising strategy (Kumar et al., 2023). They use masked attention modules inspired by Ristea et al. (2022) within the denoising Unet to alter the features for reconstruction. This aims to optimize the restoration process of potentially unhealthy structures. The results show improved segmentation performance of the proposed method compared to the baseline DDPM.

Behrendt et al. explore the use of SSIM as an anomaly score within DDPMs for unsupervised anomaly detection (Behrendt et al., 2024b). To improve robustness across varying pathology scales, they propose SSIM-ens, an adaptive method that computes a weighted average of SSIM scores across multiple kernel sizes, mitigating the limitations of relying on a fixed kernel. Their findings demonstrate that the kernel size significantly impacts segmentation performance when using SSIM. According to the results, the proposed approach mitigates this dependency, facilitating robust anomaly detection performance across a wide range of anomaly types.

Behrendt et al. (2025) propose utilizing conditioned DDPMs (cDDPMs) for UAD in brain MRI. This approach integrates a conditioning mechanism that supplies the denoising process with contextual information derived from the input image. A dedicated encoder network generates a context vector representing abstract input features, which adaptively conditions the feature maps at each level of the U-Net-based denoising network. This ensures that input image context is utilized without allowing direct copying, guiding the denoising process more effectively. The results indicate that the proposed method can enhance alignment between local intensity distributions of input and reconstructed images. Additionally, the adaptive conditioning mechanism shows promising results regarding generalization and domain adaptation capabilities.

Building upon the cDDPM architecture and SSIM, Behrendt et al. propose leveraging the Mahalanobis Distance (MHD) for anomaly scoring in UAD (Behrendt et al., 2024c). In their approach, the authors aim to address the differentiation between genuine anomalies and reconstruction artifacts. Instead of relying on a single reconstruction or aggregate statistics (e.g., mean or variance), the method utilizes multiple reconstructions from cDDPMs (Behrendt et al., 2025) to generate a pseudo-healthy distribution, incorporating both variance and inter-pixel covariance through the MHD. By considering the normal variation across reconstructions, this approach aims to reduce the influence of reconstruction artifacts, facilitating more accurate anomaly scoring and segmentation. The results suggest that the MHD enhances UAD performance by distinguishing true anomalies from common variations in pseudo-healthy reconstructions, providing a robust and precise method for anomaly detection. Furthermore, the authors state that the MHD can be used for any GM capable of generating multiple reconstructions of a given input image.

Marimont et al. introduce an approach that integrates synthetic anomalies into the diffusion process (Marimont et al., 2024), leveraging the concept of Cold Diffusion (Bansal et al., 2024). Instead of Gaussian noise, synthetic anomalies are progressively introduced during the forward process, and a U-Net is trained during the backward process to remove these anomalies while minimizing the reconstruction error between the anomaly-free input image and its reconstruction. Their method, referred to as DISYRE, demonstrates superior performance compared to certain baselines evaluated in Lagogiannis et al. (2023).

In a subsequent study, the authors present DISYRE v2 (Naval Marimont et al., 2024), which enhances the original approach by diversifying the synthetic anomalies and refining the anomaly scoring. The updated synthetic anomalies are based on a combination of intensity, texture, and shape alterations, randomly applied during each forward process. Additionally, an ensembling mechanism for different backward process start values is introduced. Results show that DISYRE v2 achieves improved performance over DISYRE v1.

Behrendt et al. propose a two-branch framework that combines the strengths of reconstruction-based UAD methods and synthetic anomalies to improve anomaly detection and segmentation accuracy (Behrendt et al., 2024a). The framework includes a GM which is trained to reconstruct healthy images, producing pseudo-healthy reconstructions for inputs with potential anomalies. A supervised segmentation model is then trained to segment deviations between the input image and its pseudo-healthy reconstruction. At inference, the anomaly maps generated by the reconstruction branch are combined with the segmentation predictions from the supervised branch to produce a unified anomaly score. The results indicate, that the framework facilitates enhanced performance and generalization across both synthetic and real pathologies. Furthermore, utilizing small subsets of real pathologies during training can enhance the segmentation performance for these pathologies while maintaining generalizability for previously unseen pathologies.

Bercea et al. present THOR, a diffusion-based model designed to guide the backward process in DDPMs by selectively focusing on abnormal regions (Bercea et al., 2024). The method iteratively replaces regions of presumably healthy anatomy with the original input image during the backward process at fixed intervals. These regions are identified by comparing intermediate deviations between the input and reconstruction. By concentrating the reconstruction process on abnormal areas, the proposed method aims to minimize false positives caused by reconstruction imperfections. The results indicate performance improvements in comparison to AnoDDPM (Wyatt et al., 2022), pDDPMs (Behrendt et al., 2023) and AutoDDPMs (Bercea et al., 2023b) considering the ATLAS v2 data set.

Wolleb et al. propose an approach to UAD utilizing latent Bernoulli diffusion models (Wolleb et al., 2024). The method compresses input images into a binary latent representation using an AE and applies a Bernoulli noise schedule for diffusion. A U-Net-based DDPM is trained to restore healthy representations, with flipping probabilities indicating potential anomalies. According to the results, the proposed approach achieves competitive performance with state-of-the-art baselines while reducing sampling time and memory consumption, making it particularly suitable for computationally constrained settings.

Baugh et al. propose to leverage image-conditioned diffusion models for anomaly detection in medical imaging. Their method introduces synthetic anomalies to healthy images, which are used as conditioning inputs for the reverse process in DDPMs (Baugh et al., 2024). The model's objective is to reconstruct the original image, with a noise-free conditioning input aiding the replication of healthy regions. This approach aims to remove synthetic anomalies while avoiding the reproduction of unhealthy structures. According to the results, the approach outperforms the baselines compared in Lagogiannis et al. (2023) demonstrating its efficacy in detecting anomalies across various data sets and modalities.

Damudi et al. compare single-step sampling with diffusion models to the full iterative sampling process that is traditionally used by DDPMs for image synthesis (Damudi and Kini, 2024). The authors directly use the output of the unet after a single denoising step, leveraging partial diffusion starting at different noise levels. Their results show that single-step sampling achieves comparable segmentation performance to the full-step AnoDDPM, while drastically reducing inference time.

Table 9
Synthetic anomaly methods for UAD in brain MRI. Tasks: Volume/Slice-wise Detection (Det), Segmentation (Seg). Representation: 2-Dimensional (2D), 3-Dimensional (3D), Channel Stacking (2.5D). Data set (Train/eval): Inhouse (IH), MOOD Challenge (MD).

Citation	Method	Task	Dimension	Data set	Weighting
Tan et al. (2022a)	Synthetic pathologies via patch interpolation.	Det, Seg	2D	MD/MD	T1
Meissen et al. (2022a)	Synthetic pathologies via polygon interpolation.	Det, Seg	2.5D	MD/MD	T1
Cho et al. (2022)	Synthetic pathologies via Copy paste	Det, Seg	3D	MD/MD	T1
Park et al. (2022)	Synthetic pathologies via 3D Cutout	Det, Seg	3D	MD/MD	T1
Tan et al. (2022b)	Segment synthetic pathologies with meta-learning	Det, Seg	2D	MD/MD	T1
Baugh et al. (2023)	Cross-validation for various synthetic pathologies	Det, Seg	3D	HPC/BR IL	T2
Xu et al. (2024)	synthetic brain tumors	Seg	3D	HCP/BR	T2

5.2.5. Other approaches and comparative studies

So far, we have provided an overview of different GMs for reconstruction-based UAD. In the following, we present alternative approaches and collect comparative studies. A collection of these alternative approaches is provided in Table 8.

Victor Saase et al. adopt simple statistical methods, like voxel-wise covariance models and linear projection with spatial patterns for UAD (Saase et al., 2020). The study finds these methods outperform a 3D-AE in voxel-wise performance while falling short in sample-wise evaluation. They speculate that spatial correlations learned by the 3D-AE could contribute to discrepancies between voxel- and sample-wise performances, shedding light on the capability of simple statistical methods in UAD.

Meissen et al. challenge the reconstruction-based UAD principle by comparing it to simple thresholding with histogram equalization (Meissen et al., 2022b). Their findings demonstrate that many UAD methods can be outperformed by the thresholding strategy for hyper-intense pathologies, underlining the limitations of current UAD approaches.

Baur et al. conduct a comparative study across different AE-based models for UAD, striving for a fair evaluation of numerous proposed methodologies (Baur et al., 2021a). They compare several models, including AE, VAE, AAE and GANs. The study emphasizes the superior performance of restoration-based methods, especially with the restoration-based VAE of X. Chen et al. (2020), in lesion segmentation tasks. They recommend the VAE for UAD due to their reduced hyperparameters, stable training behavior, and competitive performance. Similarly, Lagogiannis et al. evaluate UAD in medical images, benchmarking new methods within a fixed framework for evaluation (Lagogiannis et al., 2023). They compare feature-based methods, reconstruction-based and Synthetic Anomaly Methods methods, indicating promising performance for recent feature-modeling methods. They also speculate that these methods could further improve by leveraging recently developed self-supervised pre-training algorithms.

Attyé et al. investigate generative manifold learning for quantitative MRI analysis (Attyé et al., 2024). The model employs UMAP (Uniform Manifold Approximation and Projection) to construct a latent manifold of healthy brain structures, identifying a "digital twin" for each patient from the closest healthy controls in this space. The objective is to detect anomalies such as cortical hypertrophy or atrophy by comparing the patient's brain metrics to the personalized normative values derived from their digital twin. The results indicate superior sensitivity compared to traditional lifespan models in detecting abnormalities, such as epilepsy-related cortical hypertrophy and hippocampal atrophy in Alzheimer's disease.

Zimmerer et al. critically examine evaluation metrics for anomaly localization in medical imaging, emphasizing the limitations of current methods derived from either image segmentation or out-of-distribution detection (Zimmerer and Maier-Hein, 2024). They introduce the Soft-InstanceIoU metric, which combines instance segmentation principles with continuous anomaly scores. The study highlights the importance of tailored metrics for evaluating heatmaps, object detection, and localization tasks, demonstrating through controlled experiments and benchmarks that metric choice significantly influences model assessment. Their findings underscore the need for nuanced and task-specific evaluation in anomaly localization.

5.3. Synthetic anomaly methods

Exploring synthetic anomaly methods is another trajectory for UAD in brain MRI. A common strategy involves training a segmentation network to detect synthetic anomalies that are randomly added during training. A collection of GAN-based approaches is provided in Table 9. A notable catalyst for developing such "generate and detect" strategies is the MOOD challenge, introduced by Zimmerer et al. In a summarizing paper (Zimmerer et al., 2022a), the authors compare different submitted solutions, unveiling a strong correlation between the algorithms' performance and the perceived difficulty in anomaly detection. The results underscore the higher performance for global anomalies compared to local anomalies with semantic variances. The authors question the universal applicability of the proposed algorithms across varying pathology types due to their frequent reliance on prior knowledge of pathologies' shapes and appearances.

Therefore, ongoing effort is put into developing and designing synthetic anomalies that generalize well to a wide range of pathologies.

Tan et al. introduce Foreign Patch Interpolation (FPI) to generate subtle anomalies. In their approach, patch regions are extracted from two independent samples at the same location and replaced with an interpolation between both patches. An Unet-like architecture is trained to predict the location and interpolation factor of the patch. This approach achieves first place in global and local detection tasks of the MOOD 2020 challenge (Tan et al., 2022a).

Meissen et al. explore artificially synthesized pathologies for UAD (Meissen et al., 2022a). Subsequently, they employ an Unet to segment generated pathologies during training and utilize the Unet to segment real pathologies during testing. Although their method performs strongly on synthetic data, it fails to generalize to specific pathologies in the MOOD 2021 challenge.

Park et al. experiment with cutting out shapes and replacing them with patches of varying intensity to simulate pathologies, leveraging a transformer-based 3D Unet for segmentation, and securing third and fifth places in the global and local detection tasks of the MOOD 2021 challenge, respectively (Park et al., 2022).

Cho et al. utilize an Unet for classification and segmentation by employing a training method for anomaly generation with a copy-paste strategy. During training, they contrast the normal image with the generated abnormal image, processing the volumes in 3D patches. They evaluate their approach in the MOOD 2021 challenge, achieving first place in global and local detection (Cho et al., 2022).

Independent of the MOOD challenge, Tan et al. investigate meta-learning strategies to enhance the adaptability of synthetic anomaly methods like FPI (Tan et al., 2022b). They employ an optimization strategy involving inner and outer training loops. The results indicate improvement in the performance and the adaptability of synthetic anomaly methods.

Baugh et al. propose using a wide variability of synthetic anomalies together with a cross-validation strategy to avoid overfitting certain anomaly types (Baugh et al., 2023). Anomalies include patch-blending, image deformation and intensity modulations. Furthermore, the authors extend the Poisson image interpolation proposed in Tan et al. (2021) to 3D and propose continuous anomaly labels. The results indicate superior or similar performance compared to reconstruction-based VAEs and the CRADL method.

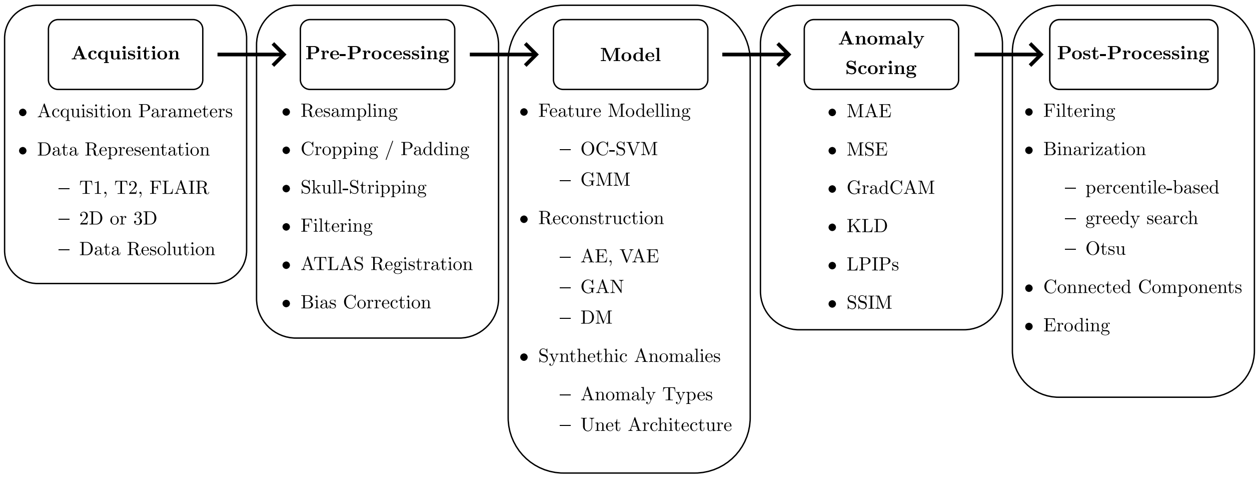

Fig. 9. General pipeline for UAD in brain MRI from data acquisition to the post-processing of anomaly maps.

Xu et al. propose a framework for brain tumor synthesis and segmentation (Xu et al., 2024). The method aims to synthesize tumors with realistic heterogeneity in shape, size, and intensity by leveraging prior knowledge of brain tissues. These synthetic tumors are integrated into a multi-target segmentation network trained to segment both brain tissues and tumors. According to the results, the proposed approach achieves performance close to supervised methods, outperforming unsupervised baselines. Notably, while showing robust segmentation performance, the method is tailored for brain tumor segmentation and generalization beyond this task is not evaluated.

6. Context, challenges and future directions of UAD in brain MRI

Approaching UAD in brain MRI is not just about selecting DL models. As illustrated in Fig. 9, this process encompasses data acquisition, pre-processing, model selection, anomaly scoring, and post-processing steps. The combination of these individual steps forms the context of UAD in brain MRI and can significantly influence performance outcomes. In this chapter, we spotlight the key differences in the context of the approaches reviewed and discuss the impact of individual settings. Furthermore, we identify existing challenges and point to potential future directions in this field.

6.1. Data sets

Data is the core of machine learning, influencing performance even in unsupervised approaches. Several widely used data sets, listed in Table 2, come with their respective pathologies and image acquisition parameters. The utilization of this data for training and evaluation is depicted in Fig. 10, which shows that the reviewed studies frequently employ IXI, MOOD, and BraTS for training and BraTS, MOOD and ISLES for evaluation. In-house data is also widely used, often motivated by specific clinical applications and data requirements. Overall, the size of data sets used for evaluation is increasing. For instance, the number of annotated samples in the updated versions of the ATLAS and BraTS2021 data sets increased roughly by a factor of two and three, respectively, from 2017 to 2021. Furthermore, initiatives like the public MOOD challenge (Zimmerer et al., 2022a) aim for comprehensive benchmarking of UAD approaches. This shift towards large data sets facilitates a more robust evaluation of UAD methods in brain MRI. However, while publicly available data sets adequately represent stroke lesions and tumors, the amount of annotated collections of other pathologies, such as MS lesions or WMH, is much lower. This imbalance is a critical factor as the type and size of the pathologies represented in the data can significantly influence UAD performance. Studies by Bengs et al. (2021) and Lagogiannis et al. (2023) demonstrate this correlation between pathology size and segmentation performance, reporting an increase in performance with increasing pathology size. Specifically, Bercea et al. (2023b) report a Dice score of 36.77% for lesions in the largest 25th percentile, compared to a Dice score of only 7.46% for lesions in the smallest 25th percentile in the ATLAS data set. However, it should be noted that the Dice score is inherently sensitive to object size. Therefore, lower Dice scores for small pathologies may partly reflect this metric-related effect rather than solely indicating inferior segmentation performance. Overall, these results indicate that the performance of individual models depends on the task and data at hand, as also demonstrated in Baur et al. (2021a). The results from the MOOD challenge highlight the utility of synthetic anomalies for systematically evaluating UAD methods in brain MRI. However, studies by Cai et al. (2023) and Lagogiannis et al. (2023) demonstrate that approaches that perform well in detecting the synthetic anomalies used for evaluation in the MOOD challenge often fail to generalize to data sets containing real pathologies. These findings indicate that evaluating novel approaches should incorporate a broad spectrum of data sets containing various pathologies with different morphologies. While synthetic anomalies provide controlled benchmarks, they cannot fully replicate the complexity and variability of real-world data. Ideally, evaluation should rely on sufficiently diverse real-world data sets that capture a broad range of pathologies. However, such data sets are often limited in practice. Therefore, synthetic anomalies can be a useful complementary tool for controlled experimentation and benchmarking if combined with real pathological data sets to ensure a comprehensive and clinically meaningful assessment of UAD performance.

A challenge that is particularly important in the field of UAD is that existing evaluation data sets may have incomplete annotations. For instance, Bercea et al. (2023b) demonstrate that in the commonly used ATLAS v2 data set, several scans show substantial hypo-intense imaging artifacts that are not annotated. This lack of comprehensive labeling skews the evaluation of UAD models, which aim to reconstruct healthy anatomy and flag deviations. Furthermore, most data sets only annotate overt pathologies, ignoring anatomical changes caused by phenomena such as space-occupying lesions. These structural changes, though not annotated by radiologists, would be detected as anomalies by successful UAD systems, potentially leading to false positive segmentation results.

Biases in data acquisition further compound these challenges. Publicly available data sets often reflect biases stemming from demographic or institutional factors, such as overrepresentation of specific scanner models, field strengths, or acquisition protocols. These biases can limit the generalizability of UAD models to unseen domains or underrepresented populations. For example, data collected primarily from high-field 3T scanners may not generalize well to lower-field 1.5T scans, potentially leading to domain shifts that might be detected as

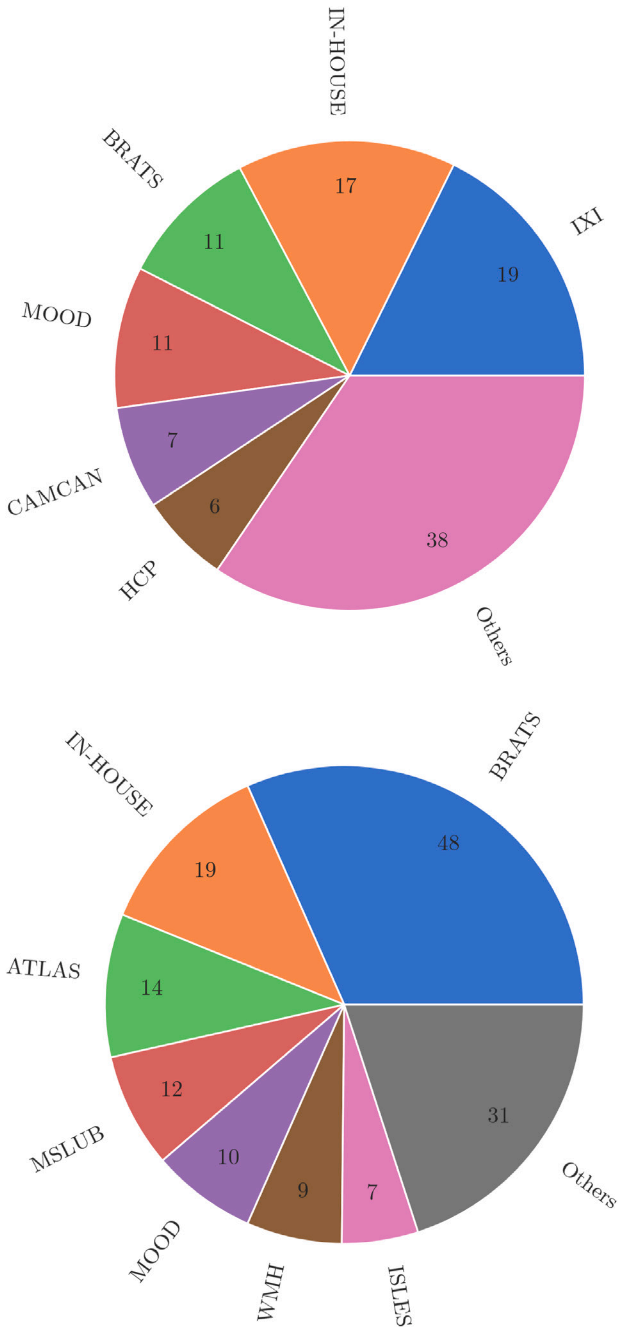

Fig. 10. Distribution of used data sets across all reviewed approaches. Top: Training data sets, Bottom: Evaluation data sets.

pathological anomalies. Furthermore, variations in demographic representation, such as age, gender, or ethnicity, could result in models that perform suboptimally for underrepresented cohorts, emphasizing the need for diverse and representative data sets. Ambiguities in the clinical significance of certain abnormalities also complicate evaluation. For instance, WMH associated with vascular aging may be considered normal in older populations, whereas visually similar features caused by subtle strokes are pathologically significant (Wood et al., 2021). Such ambiguities necessitate careful curation and interpretation of evaluation data sets to ensure that models are tested in clinically meaningful ways. These challenges underscore the importance of developing evaluation protocols designed for UAD tasks that account for the complexity and variability inherent in real-world clinical scenarios.

Future efforts should focus on creating data sets with more granular and comprehensive annotations, capturing a wide spectrum of anomalies, including subtle structural changes, imaging artifacts, and clinically ambiguous features. Additionally, integrating demographic and scanner metadata into data sets will facilitate the evaluation of model generalizability across diverse populations and imaging conditions. This approach not only addresses biases resulting from the overrepresentation of specific scanner types, field strengths, or demographic groups but also enables models to be conditioned on contextual factors, such as patient age or scanner characteristics, thereby reducing ambiguity by providing context. Additionally, the development of standardized benchmarking frameworks is imperative to ensure consistent and fair comparisons across UAD approaches. These frameworks should incorporate both synthetic and real-world pathologies and include different anomaly categories, allowing for a thorough assessment of models' ability to detect diverse types of anomalies.

In addition to evaluation data, acquiring sufficient and high-quality training data is a challenge for UAD methods in brain MRI. While UAD approaches, in principle, can be trained with unlabeled or contaminated data (contamination factor $\eta > 0$) (Zhou and Paffenroth, 2017; Hendrycks et al., 2019; Ruff et al., 2021), brain MRI studies often rely on training data sets labeled as healthy ($\eta = 0$) to ensure the models learn a reference distribution of normal anatomy. These healthy labels are crucial for the performance of UAD systems, as they enable the detection of deviations that indicate anomalies. Studies show that even minor labeling errors can significantly reduce model effectiveness (Behrendt et al., 2022b; Pinaya et al., 2022b). Recent studies (Cai et al., 2023; Siddiquee et al., 2024) suggest that incorporating partially or fully unlabeled data sets can improve detection performance while reducing dependency on curated labels. Leveraging contaminated data sets, where anomalies are present during training, offers a way to enhance model robustness and better reflect real-world data variability. However, this approach remains largely underexplored in the field of UAD for brain MRI, highlighting a significant opportunity for further research and development.

6.2. Image acquisition parameters

The MRI acquisition parameters are a critical consideration when studying brain MRIs, as they directly affect image contrasts and the intensity distribution of anatomical structures, including pathologies. Among the reviewed approaches, T1, T2, and FLAIR weightings are used in 64%, 38%, and 38% of studies, respectively. These differences in weighting significantly influence segmentation performance, as demonstrated by Sato et al. (2019), who observe higher segmentation performance with T2-weighted scans compared to T1-weighted scans in the BraTS data set. Similarly, Uzunova et al. (2019) report that contrast-enhanced T1 scans yield superior Dice scores compared to T2 and FLAIR scans. Combining multiple weightings further improves performance, achieving a Dice score of 50%, highlighting the complementary information provided by different MRI sequences.

While there is no clear trend favoring a specific weighting across the reviewed approaches, some weightings, such as FLAIR, are particularly effective for specific pathologies. For instance, hyper-intense tumors in combination with the hypo-intense cerebrospinal fluid in FLAIR-weighted scans allow straightforward thresholding to achieve performance comparable to state-of-the-art UAD approaches concerning BraTS data (Meissen et al., 2022b). However, such methods are ineffective for pathologies that manifest as subtle structural changes rather than intensity differences. Studies by Meissen et al. (2022d) and Lagogiannis et al. (2023) reveal that reconstruction-based UAD methods often fail to detect structural anomalies, underscoring the need to improve GMs and to develop anomaly scoring mechanisms that incorporate structural information.

Beyond these methodological insights, a critical challenge in deploying UAD systems lies in achieving domain invariance across diverse acquisition parameters. Variability in MRI acquisition, such as differences in scanner models, field strengths, or imaging protocols, can introduce domain shifts. As highlighted by Heer et al. (2021), these domain shifts can mimic pathological anomalies, leading to false positive detections. For example, changes in magnetic field strength may alter image intensity distributions, causing a model trained on one domain to misinterpret healthy regions in another domain as anomalous. Approaches like cDDPMs (Behrendt et al., 2025), which are adaptively conditioned on the original image to preserve intensity information during the reconstruction process, show promise in enhancing reconstruction coherence and improving domain adaptation. However, it is required to evaluate domain shifts and pathologies as separate phenomena to disentangle these effects fully. Heer et al. propose specific strategies to separate domain-shifted samples from those containing true lesions (Heer et al., 2021). These methods rely on capturing global distribution differences versus localized reconstruction errors, offering promising avenues to enhance robustness. Hence, in addition to the aforementioned need for comprehensive evaluation data sets, further work should focus on integrating these insights into scalable and reliable UAD frameworks and benchmarks.

6.3. Data representation

Brain MRIs, typically acquired slice-wise, usually result in volumetric images. Several studies indicate that using this inherent 3D information can be beneficial. For instance, Bengs et al. (2021) report that using 3D VAEs instead of 2D VAEs improves the Dice score by 6.4% and 28.4% for tumors and stroke lesions, respectively. Similarly, Naval Marimont and Tarroni (2021) report that a 3D CNN outperforms its 2D counterpart by 5.1% on the BraTS data set. Also Behrendt et al. (2022b), who feed the different latent vectors of individual slices to a transformer, report improvements in the segmentation performance by 10.0% and 10.5% for the BraTS and ATLAS data sets, respectively. In the study of Pinaya et al. (2022b), 3D processing led to a relative performance improvement of 14.9% and 58.6%, for the BraTS and UKB data sets, respectively. However, in the same study, 3D processing substantially decreased the performance, with Dice scores dropping from 37.8% to 13.3% (MSLUB) and from 42.9% to 13.3% (WMH). This demonstrates that extending existing architectures to 3D does not necessarily lead to consistently improved performance. Consequently, the majority of the considered approaches perform slice-wise processing, ignoring the information of spatial neighborhoods across slices.

Another key factor contributing to this imbalance is the computational demand for 3D models. Generally, 3D models have a higher parameter count than 2D models, and processing 3D volumes necessitates significant computational memory (Pinaya et al., 2022b; Behrendt et al., 2022b). Moreover, the high parameter count in 3D models is coupled with data scarcity, as the number of training samples effectively decreases to the number of volumes rather than the number of slices as in slice-wise processing. This combination of increased parameters and reduced training samples can lead to overfitting, potentially impacting the performance of 3D models (Bengs et al., 2021).

There is a noticeable gap in recent literature regarding a comprehensive study that contrasts 2D and 3D methodologies, considering the changing number of parameters and different data set sizes. Investigating these effects is essential to understand the advantages and limitations of 3D methods for UAD in brain MRI. A valuable research direction is seen in the exploration of efficient 3D processing strategies. For instance, parameter-efficient approaches such as latent diffusion models (Pinaya et al., 2022a) or wavelet diffusion models (Friedrich et al., 2024) could significantly reduce computational costs and, thus, make 3D processing more applicable to the field of UAD in brain MRI.

6.4. Pre-processing

The reviewed approaches apply several pre-processing strategies specifically designed for brain MRIs. Given the general size of the image data and variations in the actual size and image resolution, focusing on relevant regions and resizing them to some fixed values is common practice. The size in the reviewed papers ranges from 32 to 256 voxels per axis, with several studies indicating that larger images often result in improved performance. For instance, Zimmerer et al. (2019a) demonstrate that changing the input size from 32×32 pixels to 192×192 pixels increases the AUC from approximately 70% to 80% for the baseline VAE. Similarly, Baur et al. (2020c) report that an input size of 256×256 pixels results in a relative increase of the Dice score of 145% compared to an input size of 64×64 pixels when studying MS lesions. However, Baur et al. (2020c) and X. Chen et al. (2020) show that for the BraTS data set, an input size of 128×128 pixels outperforms larger input sizes of 256×256. Cropping is another frequent adaptation, particularly trimming a certain number of top and bottom slices where no significant information is assumed. In more extreme cases, evaluations focus on one to four central slices only (Pinaya et al., 2021, 2022b; Bercea et al., 2023a,d,b). While this approach reduces computational demands, Meissen et al. (2022b) demonstrated that evaluating only the central slices can substantially boost performance metrics compared to using all slices in an MRI volume. However, this selective evaluation may lead to overestimated performance, raising concerns about whether models generalize well to other slices or different imaging protocols. In addition to these strategies, many methods incorporate skull stripping, which involves the removal of the skull from MRI scans using CNN-based tools like ROBEX (Iglesias et al., 2011) or HD-Bet (Isensee et al., 2019). Another frequent pre-processing step uses registration techniques to align the scans with a standardized MRI space, such as MNI space (Mazziotta et al., 2001) or SRI space (Rohlfing et al., 2010). Often, bias correction filters are applied to eliminate artifacts. Some approaches also utilize histogram equalization (Akrami et al., 2020; Martins et al., 2019a; Akrami et al., 2021, 2022; Meissen et al., 2022a).

While these pre-processing steps can reduce memory requirements and facilitate model training, they can inadvertently introduce biases. For example, excessive cropping may exclude peripheral regions of the brain where anomalies could occur, potentially biasing models towards detecting central anomalies only. Similarly, normalization techniques, such as histogram equalization, may obscure subtle intensity-based anomalies, limiting the ability of models to capture fine-grained pathological changes. Registration, while useful for standardization, can introduce variability depending on the tool or algorithm used, further complicating generalizability. Furthermore, skull stripping can lead to the removal of extracranial abnormalities, making the detection of these conditions impossible (Wood et al., 2021). In summary, selecting pre-processing steps can substantially affect the UAD performance, particularly regarding the size of the processed volumes or slices. The lack of systematic evaluation of these pre-processing strategies across studies adds to these challenges. Despite the evidence that pre-processing choices can significantly impact UAD performance, the reviewed literature does not establish a standard pipeline or quantitatively assess the trade-offs of different strategies. Consequently, the selection of pre-processing steps is often empirical, relying on specific task requirements and available data. Future work should focus on systematically evaluating pre-processing strategies to better understand their impact on model performance and reduce biases introduced during data preparation. This effort could help establish standardized processing frameworks tailored to specific scenarios.

6.5. Model type

UAD methods in brain MRI can be categorized into reconstruction-based, synthetic anomaly, and feature-based approaches, each with specific strengths and limitations. As Fig. 1 indicates, reconstruction-based approaches dominate the field, likely reflecting their broad applicability and their operation directly in image space, making them well-suited for segmentation tasks. These methods rely on GMs to reconstruct healthy anatomy, with anomalies identified as deviations from the reconstructed image. However, high reconstruction fidelity often results in the replication of anomalous structures, leading to false negatives, while overly regularized models may produce generic or blurred reconstructions, introducing false positives. Recent advancements, such as incorporating skip connections and spatial latent dimensions, have enhanced the information flow within GMs, improving reconstruction accuracy (Baur et al., 2021b). Regularization objectives like denoising further prevent models from directly replicating anomalous inputs (Kascenas et al., 2022a). DDPMs exemplify this approach, presenting a promising GM for reconstruction-based UAD (Wyatt et al., 2022; Pinaya et al., 2022b; Behrendt et al., 2023). However, the characteristics and magnitude of the added noise in DDPMs can affect the detection performance on specific pathologies, posing a challenge for a general UAD solution (Kascenas et al., 2023; Bercea et al., 2023b). Furthermore, as reconstruction-based approaches utilize the anomaly maps derived from discrepancies between the input and reconstruction for segmentation, these methods are highly sensitive to reconstruction artifacts due to imperfect reconstructions. This necessitates robust post-processing techniques and anomaly scoring functions, as detailed in Sections 6.6 and 6.7.

Promising future directions for reconstruction-based approaches include adaptive masked regularization strategies that apply varying levels of regularization to specific regions of the input, as well as the integration of metadata, such as patient age or scanner information, to condition GMs and reduce ambiguities associated with certain pathologies (Wood et al., 2021; Bengs et al., 2022).

Synthetic anomaly methods introduce artificial anomalies during training, allowing direct segmentation of predefined anomaly types, trained in a supervised fashion. This strategy has shown promise in controlled scenarios, such as the MOOD challenge, where simulated pathologies are part of the evaluation. However, as recent studies show, using simulated pathologies adds a strong bias towards specific anomalies seen during training. The study (Cai et al., 2023) demonstrates that synthetic anomaly methods only perform well on a few particular data sets and are often outperformed by reconstruction-based AEs. Similarly, Lagogiannis et al. (2023) report that synthetic anomaly methods perform poorly across all evaluated data sets. Therefore, recent studies aim to generate a wide range of diverse anomalies to cover the morphology of potential pathologies and employ specific cross-validation strategies to reduce the risk of overfitting to the synthetic anomalies (Baugh et al., 2023).

Feature-based approaches have shown strong performance in non-medical anomaly detection tasks, such as industrial defect detection (e.g., MVTEC Bergmann et al., 2021), where large-scale pre-training on natural images enables robust feature extraction. Although industrial images can differ from natural image data sets such as ImageNet, they often share low-level statistics, allowing pretrained representations to transfer effectively. In contrast, the transfer to medical imaging data, including brain MRI, can be less straightforward because of differences in appearance, signal characteristics, and underlying semantics. In addition, anomalies in brain MRI are often subtle and context-dependent, which may limit the effectiveness of approaches based solely on feature-space discrepancies. These factors may contribute to the comparatively lower adoption of feature-based methods in brain MRI reported in the current literature, although this relationship is not yet fully understood and warrants further systematic investigation. Domain-adapted pre-training on large-scale medical data sets, such as UKB, therefore appears to be a promising direction for improving feature-based methods in brain MRI. Furthermore, recent advancements in self-supervised learning, such as masked image modeling approaches (e.g., MAEs He et al., 2022 or their CNN-based counterparts such as Spark (Tian et al., 2023)) and improved contrastive and self-distillation-based methods (e.g., BYOL Grill et al., 2020, MoCo He et al., 2020, or DINO Caron et al., 2021; Oquab et al., 2023), could enhance feature-based methods. Conducting a comprehensive comparison of these approaches in the context of UAD could provide valuable insights into their relative strengths and limitations, thereby guiding the development of optimized feature extraction strategies.

Overall, each approach has specific use cases and limitations. Reconstruction-based methods are well-suited for tasks requiring fine-grained anomaly localization and detecting a wide spectrum of pathologies. Synthetic anomaly methods excel in controlled environments where the anomaly types are well-defined, but they require careful design to avoid overfitting to specific anomalies. Feature-based methods can offer computational efficiency and may be particularly useful for tasks where coarse or sample-level anomaly detection is sufficient.

Hybrid approaches that combine these methodologies are promising for advancing UAD in brain MRI. For example, feature-based methods can be integrated with reconstruction-based frameworks by applying density estimation in the latent space of GMs (Marimont et al., 2023; Pinaya et al., 2022a). Modeling the latent vector distribution of normal brain MRIs allows for efficient density estimation in a lower-dimensional space, while anomaly detection remains in the image space, enabling precise, pixel-wise comparisons. Furthermore, coarse feature discrepancies from feature-based methods such as LPIPs can refine reconstruction-based anomaly scores, reducing the impact of background artifacts induced by imperfect reconstructions.

Synthetic anomaly methods can also complement reconstruction-based approaches by providing precise segmentation maps for known pathologies, which can be combined with the general anomaly maps of reconstruction-based methods. This integration enables enhanced segmentation performance for specific pathologies while maintaining sensitivity to unseen anomalies. Following such a strategy, as demonstrated in Behrendt et al. (2024a), offers a pathway towards achieving supervised-level performance in segmentation while preserving the flexibility of unsupervised methods.

Furthermore, the concept of leveraging semi-supervised learning further enriches this integration. Semi-supervised anomaly detection methods, such as Deep SAD (Ruff et al., 2020), utilize a limited amount of labeled normal and anomalous data to guide the model in distinguishing normal from anomalous patterns. This principle can be particularly valuable in UAD for brain MRIs, where labeled anomalies (e.g., certain pathologies) can guide the detection process while retaining flexibility for identifying novel anomalies.

6.6. Anomaly scoring

Anomaly scoring is a critical aspect in the pipeline of UAD. As discussed above, reconstruction-based approaches derive the anomaly score as the deviation between the potentially abnormal input image and its reconstruction. The most common metrics for this measurement are intensity-based metrics, such as pixel-wise l_1- or l_2-errors. However, these metrics face challenges when detecting anomalies that exhibit structural changes rather than intensity differences (Meissen et al., 2022b). Studies propose alternative anomaly scores to improve segmentation performance. For example, Zimmerer et al. (2019a) suggest a KLD-based anomaly score, reporting performance improvements compared to the l_1-error. The authors Silva-Rodríguez et al. (2022) report that using CNN activation maps for anomaly scoring can improve the Dice score by 85%, compared to the l_1-error. However, Lagogiannis et al. (2023) contradicts these findings, reporting that these approaches lead to poor generalization across different data sets. As an alternative, anomaly scoring functions that assess structural similarities have been shown to outperform intensity-based scoring functions (Bergmann et al., 2020; Lagogiannis et al., 2023; Behrendt et al., 2024b). Another approach involves measuring structural similarities of input and reconstruction in the feature space of pre-trained CNNs rather than the image space (Meissen et al., 2022c). Learned perceptual similarity scores have also demonstrated their value in refining the anomaly scores derived by reconstruction-based approaches. These methods leverage the complementary information from the coarse but robust feature-based anomaly map and the highly sensitive reconstruction-based residual map to enhance segmentation performance (Bercea et al., 2023a,d). Future research should focus on developing hybrid anomaly scoring strategies that integrate and combine the strengths of these methods to achieve robust segmentation across diverse pathologies. For instance, ensemble-based structural similarities (Behrendt et al., 2024b) could be extended to multiple scales, potentially improving generalization across different pathology sizes. Furthermore, integrating reconstruction-based and synthetic anomaly methods to directly learn the difference between inputs and reconstructions (Zavrtanik et al., 2021), holds potential for clinically relevant UAD in brain MRIs. This approach is particularly valuable in weakly supervised settings, where a small, labeled subset of data can be leveraged effectively to enhance model performance and generalizability (Behrendt et al., 2024a).

6.7. Post-processing

Typically, the raw error metrics of UAD methods lead to small local artifacts and noise in the anomaly maps, necessitating further post-processing steps. Typical steps include median filtering and removing connected components with a volume below a certain threshold. However, only a few studies explore how these post-processing steps can impact performance. For instance, Kascenas et al. (2022a) found that using median filtering can improve the performance of DAEs and VAEs by 2% and 39%, respectively, when considering the Dice score for the BraTS data set. This finding suggests that specific models benefit more from additional post-processing steps than others. The same study also reports that removing connected components with a volume no larger than 20 voxels after binarization leads to only moderate performance improvements. Similar trends were observed for DDPMs in Behrendt et al. (2025), where median filtering consistently showed the most substantial impact across various post-processing strategies. Furthermore, the results of Behrendt et al. (2025) demonstrate that the effectiveness of different post-processing strategies varies across different data sets. This highlights that while post-processing can significantly enhance segmentation performance, it often introduces a degree of supervision. For example, the kernel size of the commonly used median filter can lead to specialization towards specific pathology sizes, as larger kernel sizes would erase small pathologies. Consequently, kernel sizes vary across the reviewed approaches. Most approaches chose a kernel size of $5 \times 5 \times 5$, although some approaches use $7 \times 7 \times 7$ (Akrami et al., 2021, 2022), or $3 \times 3 \times 3$ (Bercea et al., 2022) kernel sizes.

For UAD approaches, finding an appropriate threshold for binarization is crucial. Many studies utilize a data set of unhealthy MRI scans for a greedy search of the threshold that maximizes segmentation performance. Other approaches consider the marginal percentiles of reconstruction errors on healthy data or use OTSU thresholding. In their study, Silva-Rodríguez et al. (2022) demonstrate that the threshold search can substantially affect the segmentation performance. Specifically, they report Dice scores of 69.3% and 58.7% for thresholds found by using unhealthy validation data, compared to thresholds based on the 98% percentile of the healthy validation data. These results suggest that calibrating the threshold on unhealthy validation data improves the segmentation performance. However, this strategy introduces a level of supervision and contradicts the principle of UAD as it may limit generalizability to other pathologies not included in the calibration set.

A thorough comparative analysis of post-processing strategies across various UAD approaches is essential to fully understand their impact on model performance. Future studies should focus on systematically evaluating different post-processing techniques to identify generalizable methods that perform well across diverse pathologies and imaging conditions. This evaluation should prioritize transparency and fairness to ensure a consistent assessment of novel approaches to UAD in brain MRI. To reduce reliance on potentially bias-inducing post-processing, future research could explore adaptive post-processing strategies that dynamically adjust based on the characteristics of individual scans and thereby enhance robustness while minimizing the risk of overfitting to specific pathologies or data sets. Combining the probabilistic properties of DDPMs with uncertainty-based measures such as the Mahalanobis distance (Behrendt et al., 2024c) represents a promising direction for robust anomaly score refinement. Furthermore, such an approach could provide calibrated confidence levels or prediction sets (Akrami et al., 2022), providing robust thresholds for the final anomaly segmentation.

6.8. Summary

This chapter underscores that beyond selecting specific DL models, the entire pipeline design can have a substantial effect on anomaly detection performance and should be tailored specifically for the context of UAD in brain MRI. This pipeline includes data selection, acquisition, pre-processing, post-processing, and anomaly scoring. Importantly, these elements are interconnected. For instance, the quality and type of generative models, along with the available data, can dictate the necessary pre- and post-processing steps and the choice of suitable anomaly scores. This interconnectivity underscores the necessity of a complete understanding of the impact of different elements in the UAD pipeline when designing and evaluating novel approaches. Therefore, in addition to addressing the individual challenges of generative models, synthetic anomaly methods, and feature modeling approaches, there is a crucial need to systematically compare and evaluate different aspects of the entire UAD pipeline along with the DL approach.

7. Conclusion

UAD in brain MRI represents a promising research direction, as it identifies abnormal structures as deviations from a learned norm, thereby enabling the detection of pathologies that are unseen during training. In this paper, we provide a systematic literature review and categorize the collected DL-based UAD approaches. We present their key differences within the overall UAD pipeline and highlight challenges faced by these approaches, while also suggesting potential avenues for future research.

Our review shows that reconstruction methods are most commonly applied across the literature, followed by feature modeling and synthetic anomalies. However, substantial differences exist across and within these categories in terms of data handling and processing, model architectures, and anomaly scoring strategies. This complexity makes it challenging to identify which elements of individual approaches contribute to robust performance across different scenarios. Therefore, comparative studies that consider this context of UAD in brain MRI while comparing different state-of-the-art approaches are important. Furthermore, an important direction is to address the issue of potentially impure or unlabeled training data, which could enhance the applicability of UAD in real-world scenarios. Moreover, clinical application studies that compare existing UAD methods in various real-world settings could shed light on the strengths, weaknesses and clinical significance of UAD in brain MRI.

CRediT authorship contribution statement

Finn Behrendt: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Debayan Bhattacharya:** Writing – review & editing, Conceptualization. **Lennart Maack:** Writing – review & editing, Conceptualization. **Julia Krüger:** Writing – review & editing, Funding acquisition. **Roland Opfer:** Writing – review & editing, Funding acquisition. **Alexander Schlaefer:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Funding

This work was partially funded by grant number KK5208101KS0 and ZF4026303TS9 and by the Free and Hanseatic City of Hamburg (Interdisciplinary Graduate School), Germany.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D.L., Erickson, B.J., 2017. Deep learning for brain MRI segmentation: State of the art and future directions. J. Digit. Imaging 30 (4), 449–459. http://dx.doi.org/10.1007/s10278-017-9983-4, URL: https://link.springer.com/article/10.1007%2Fs10278-017-9983-4.

Akrami, H., Joshi, A., Aydore, S., Leahy, R., 2021. Quantile regression for uncertainty estimation in VAEs with applications to brain lesion detection. Inf. Process. Med. Imaging 12729, 689–700. http://dx.doi.org/10.1007/978-3-030-78191-0_53.

Akrami, H., Joshi, A., Aydore, S., Leahy, R., 2022. Deep quantile regression for uncertainty estimation in unsupervised and supervised lesion detection. Mach. Learn. Biomed. Imaging 1 (IPMI 2021), 1–23. http://dx.doi.org/10.59275/j.melba.2022-6751.

Akrami, H., Joshi, A.A., Li, J., Aydore, S., Leahy, R.M., 2020. Brain lesion detection using a robust variational autoencoder and transfer learning. In: IEEE 17th International Symposium on Biomedical Imaging. pp. 786–790. http://dx.doi.org/10.1109/isbi45749.2020.9098405.

Alain, G., Bengio, Y., 2014. What regularized auto-encoders learn from the data-generating distribution. J. Mach. Learn. Res. 15 (1), 3563–3593.

Alaverdyan, Z., Jung, J., Bouet, R., Lartizien, C., 2020. Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging: Application to epilepsy lesion screening. Med. Image Anal. 60, 101618. http://dx.doi.org/10.1016/j.media.2019.101618, URL: https://www.sciencedirect.com/science/article/pii/S1361841519301562.

Albu, A., Enescu, A., Malagò, L., 2020. Tumor detection in brain MRIs by computing dissimilarities in the latent space of a variational AutoEncoder. Proc. North. Light. Deep. Learn. Work. 1, 6. http://dx.doi.org/10.7557/18.5172.

Andrade de Oliveira, A., Carthery-Goulart, M.T., Oliveira Júnior, P.P.d., Carrettiero, D.C., Sato, J.R., 2015. Defining multivariate normative rules for healthy aging using neuroimaging and machine learning: an application to Alzheimer's disease. J. Alzheimer's Dis.: JAD 43 (1), 201–212. http://dx.doi.org/10.3233/JAD-140189, URL: https://pubmed.ncbi.nlm.nih.gov/25079801/.

Aswani, K., Menaka, D., 2021. A dual autoencoder and singular value decomposition based feature optimization for the segmentation of brain tumor from MRI images. BMC Med. Imaging 21 (1), 82. http://dx.doi.org/10.1186/s12880-021-00614-3, URL: https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-021-00614-3.

Atlason, H., Askell Love, M.D., Sigurdur Sigurdsson, Vilmundur Gudnason, M.D., Ellingsen, L.M., 2019a. Unsupervised brain lesion segmentation from MRI using a convolutional autoencoder. pp. 372–378. http://dx.doi.org/10.1117/12.2512953.

Atlason, H., Love, A., Sigurdsson, S., Gudnason, V., Ellingsen, L.M., 2019b. SegAE: Unsupervised white matter lesion segmentation from brain MRIs using a CNN autoencoder. NeuroImage: Clin. 24, 102085. http://dx.doi.org/10.1016/j.nicl.2019.102085.

Attyé, A., Renard, F., Anglade, V., Krainik, A., Kahane, P., Mansencal, B., Coupé, P., Calamante, F., 2024. Data-driven normative values based on generative manifold learning for quantitative MRI. Sci. Rep. 14 (1), 7563.

Avci, M.Y., Chan, E., Zimmer, V.A., Rueckert, D., Wiestler, B., Schnabel, J.A., Bercea, C.I., 2024. Unsupervised analysis of Alzheimer's disease signatures using 3D deformable autoencoders. In: MICCAI Student Board EMERGE Workshop: Empowering Medical Image Computing & Research Through Early-Career Expertise.

Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al., 2021. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314.

Bansal, A., Borgnia, E., Chu, H.-M., Li, J., Kazemi, H., Huang, F., Goldblum, M., Geiping, J., Goldstein, T., 2024. Cold diffusion: Inverting arbitrary image transforms without noise. Adv. Neural Inf. Process. Syst. 36.

Bauer, S., Wiest, R., Nolte, L.-P., Reyes, M., 2013. A survey of MRI-based medical image analysis for brain tumor studies. Phys. Med. Biol. 58 (13), R97–129. http://dx.doi.org/10.1088/0031-9155/58/13/R97.

Baugh, M., Reynaud, H., Marimont, S.N., Cechnicka, S., Müller, J.P., Tarroni, G., Kainz, B., 2024. Image-conditioned diffusion models for medical anomaly detection. In: International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging. Springer, pp. 117–127.

Baugh, M., Tan, J., Müller, J.P., Dombrowski, M., Batten, J., Kainz, B., 2023. Many tasks make light work: Learning to localise medical anomalies from multiple synthetic tasks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 162–172.

Baur, C., Graf, R., Wiestler, B., Albarqouni, S., Navab, N., 2020a. SteGANomaly: Inhibiting cyclegan steganography for unsupervised anomaly detection in brain MRI. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. In: Lecture Notes in Computer Science, vol. 12262, Springer International Publishing, pp. 718–727.

Baur, C., Stefan Denner, Benedikt Wiestler, Nassir Navab, Shadi Albarqouni, 2021a. Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study. Med. Image Anal. 69, 101952. http://dx.doi.org/10.1016/j.media.2020.101952, URL: https://www.sciencedirect.com/science/article/pii/S1361841520303169.

Baur, C., Wiestler, B., Albarqouni, S., Navab, N., 2019. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. In: Image Processing, Computer Vision, Pattern Recognition, and Graphics, vol. 11383, Springer International Publishing, Cham, pp. 161–169. http://dx.doi.org/10.1007/978-3-030-11723-8_16.

Baur, C., Wiestler, B., Albarqouni, S., Navab, N., 2020b. Bayesian skip-autoencoders for unsupervised hyperintense anomaly detection in high resolution brain mri. In: 2020 IEEE 17th International Symposium on Biomedical Imaging. ISBI, pp. 1905–1909. http://dx.doi.org/10.1109/ISBI45749.2020.9098686.

Baur, C., Wiestler, B., Albarqouni, S., Navab, N., 2020c. Scale-space autoencoders for unsupervised anomaly segmentation in brain MRI. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. In: Lecture Notes in Computer Science, vol. 12264, Springer International Publishing, pp. 552–561. http://dx.doi.org/10.1007/978-3-030-59719-1_54.

Baur, C., Wiestler, B., Muehlau, M., Zimmer, C., Navab, N., Albarqouni, S., 2021b. Modeling healthy anatomy with artificial intelligence for unsupervised anomaly detection in brain MRI. Radiol.: Artif. Intell. 3 (3), e190169. http://dx.doi.org/10.1148/ryai.2021190169.

Behrendt, F., Bengs, M., Bhattacharya, D., Krüger, J., Opfer, R., Schlaefer, A., 2022a. Capturing inter-slice dependencies of 3D brain MRI-scans for unsupervised anomaly detection. In: Medical Imaging with Deep Learning. URL: https://openreview.net/forum?id=db8wDgKH4p4.

Behrendt, F., Bengs, M., Rogge, F., Kruger, J., Opfer, R., Schlaefer, A., 2022b. Unsupervised anomaly detection in 3D brain MRI using deep learning with impured training data. In: 2022 IEEE 19th International Symposium on Biomedical Imaging. ISBI, pp. 1–4. http://dx.doi.org/10.1109/ISBI52829.2022.9761443.

Behrendt, F., Bhattacharya, D., Krüger, J., Opfer, R., Schlaefer, A., 2023. Patched diffusion models for unsupervised anomaly detection in brain MRI. In: Medical Imaging with Deep Learning. URL: https://openreview.net/forum?id=O-uZr5S1tJE.

Behrendt, F., Bhattacharya, D., Maack, L., Krüger, J., Opfer, R., Mieling, R., Schlaefer, A., 2024b. Diffusion models with ensembled structure-based anomaly scoring for unsupervised anomaly detection. In: 2024 IEEE International Symposium on Biomedical Imaging. ISBI, pp. 1–4. http://dx.doi.org/10.1109/ISBI56570.2024.10635828.

Behrendt, F., Bhattacharya, D., Maack, L., Krüger, J., Opfer, R., Schlaefer, A., 2024a. Combining reconstruction-based unsupervised anomaly detection with supervised segmentation for brain MRIs. In: Medical Imaging with Deep Learning. URL: https://openreview.net/forum?id=iWfUcg4FrD.

Behrendt, F., Bhattacharya, D., Mieling, R., Maack, L., Krüger, J., Opfer, R., Schlaefer, A., 2024c. Leveraging the mahalanobis distance to enhance unsupervised brain MRI anomaly detection. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Springer Nature Switzerland, Cham, pp. 394–404.

Behrendt, F., Bhattacharya, D., Mieling, R., Maack, L., Krüger, J., Opfer, R., Schlaefer, A., 2025. Guided reconstruction with conditioned diffusion models for unsupervised anomaly detection in brain MRIs. Comput. Biol. Med. 186, 109660. http://dx.doi.org/10.1016/j.compbiomed.2025.109660, URL: https://www.sciencedirect.com/science/article/pii/S0010482525000101.

Bengio, Y., Yao, L., Alain, G., Vincent, P., 2013. Generalized denoising auto-encoders as generative models. In: Proceedings of the 26th International Conference on Neural Information Processing Systems. NIPS13, Curran Associates Inc, Red Hook, NY, USA, pp. 899–907.

Bengs, M., Behrendt, F., Krüger, J., Opfer, R., Schlaefer, A., 2021. Three-dimensional deep learning with spatial erasing for unsupervised anomaly segmentation in brain MRI. Int. J. Comput. Assist. Radiol. Surg. 16 (9), 1413–1423. http://dx.doi.org/10.1007/s11548-021-02451-9.

Bengs, M., Behrendt, F., Max-Heinrich Laves, Krüger, J., Opfer, R., Schlaefer, A., 2022. Unsupervised anomaly detection in 3D brain MRI using deep learning with multi-task brain age prediction. In: Karen Drukker, Iftekharuddin, K.M. (Eds.), Medical Imaging 2022: Computer-Aided Diagnosis. vol. 12033, SPIE, 1203314. http://dx.doi.org/10.1117/12.2608120.

Bercea, C., Benedikt Wiestler, Daniel Rueckert, Julia A Schnabel, 2023a. Generalizing unsupervised anomaly detection: Towards unbiased pathology screening. Med. Imaging Deep. Learn. URL: https://openreview.net/forum?id=8ojx-Ld3yjR.

Bercea, C., Michael Neumayr, Daniel Rueckert, Julia A Schnabel, 2023b. Mask, stitch, and re-sample: Enhancing robustness and generalizability in anomaly detection through automatic diffusion models. In: ICML 3rd Workshop on Interpretable Machine Learning in Healthcare. IMLH, URL: https://openreview.net/forum?id=kTpafpXrqa.

Bercea, C.I., Rueckert, D., Schnabel, J.A., 2023c. What do aes learn? challenging common assumptions in unsupervised anomaly detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 304–314.

Bercea, C.I., Wiestler, B., Rueckert, D., Albarqouni, S., 2022. Federated disentangled representation learning for unsupervised brain anomaly detection. Nat. Mach. Intell. 4 (8), 685–695. http://dx.doi.org/10.1038/s42256-022-00515-2.

Bercea, C.I., Wiestler, B., Rueckert, D., Schnabel, J.A., 2023d. Reversing the abnormal: Pseudo-healthy generative networks for anomaly detection. 14224, pp. 293–303. http://dx.doi.org/10.1007/978-3-031-43904-9_29,

Bercea, C.I., Wiestler, B., Rueckert, D., Schnabel, J.A., 2024. Diffusion models with implicit guidance for medical anomaly detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 211–220.

Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C., 2021. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. Int. J. Comput. Vis. 129 (4), 1038–1059.

Bergmann, P., Fauser, M., Sattlegger, D., Steger, C., 2020. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 4182–4191.

Bhattacharyya, S., Jha, S., Tharakunnel, K., Westland, J.C., 2011. Data mining for credit card fraud: A comparative study. Decis. Support Syst. 50 (3), 602–613. http://dx.doi.org/10.1016/j.dss.2010.08.008, URL: https://www.sciencedirect.com/science/article/pii/S0167923610001326. On quantitative methods for detection of financial fraud.

Biomedical Image Analysis Group, 2026. IXI Dataset – brain development. Available at: https://brain-development.org/ixi-dataset/ (accessed April 2026).

Börnert, P., Norris, D.G., 2020. A half-century of innovation in technology-preparing MRI for the 21st century. Br. J. Radiol. 93 (1111), 20200113. http://dx.doi.org/10.1259/bjr.20200113.

Bredell, G., Flouris, K., Chaitanya, K., Erdil, E., Konukoglu, E., 2022. Explicitly minimizing the blur error of variational autoencoders. In: The Eleventh International Conference on Learning Representations.

Bruno, M.A., Walker, E.A., Abujudeh, H.H., 2015. Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction. Radiogr.: Rev. Publ. Radiol. Soc. North Am. Inc 35 (6), 1668–1676. http://dx.doi.org/10.1148/rg.2015150023.

Cai, Y., Chen, H., Yang, X., Zhou, Y., Cheng, K.-T., 2023. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. Med. Image Anal. 86, 102794. http://dx.doi.org/10.1016/j.media.2023.102794.

Carass, A., Roy, S., Jog, A., Cuzzocreo, J.L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C.H., Jorge Cardoso, M., Cawley, N., Ciccarelli, O., Wheeler-Kingshott, C.A.M., Ourselin, S., Catanese, L., Deshpande, H., Maurel, P., Commowick, O., Barillot, C., Tomas-Fernandez, X., Warfield, S.K., Vaidya, S., Chunduru, A., Muthuganapathy, R., Krishnamurthi, G., Jesson, A., Arbel, T., Maier, O., Handels, H., Iheme, L.O., Unay, D., Jain, S., Sima, D.M., Smeets, D., Ghafoorian, M., Platel, B., Birenbaum, A., Greenspan, H., Bazin, P.-L., Calabresi, P.A., Crainiceanu, C.M., Ellingsen, L.M., Reich, D.S., Prince, J.L., Pham, D.L., 2017. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. NeuroImage 148, 77–102. http://dx.doi.org/10.1016/j.neuroimage.2016.12.064.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging properties in self-supervised vision transformers. In: Proceedings of the International Conference on Computer Vision. ICCV.

Chalapathy, R., Chawla, S., 2019. Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

Chatterjee, S., Sciarra, A., Dünnwald, M., Tummala, P., Agrawal, S.K., Jauhari, A., Kalra, A., Oeltze-Jafra, S., Speck, O., Nürnberger, A., 2022. StRegA: Unsupervised anomaly detection in brain MRIs using a compact context-encoding variational autoencoder. Comput. Biol. Med. 149, 106093. http://dx.doi.org/10.1016/j.compbiomed.2022.106093, URL: https://www.sciencedirect.com/science/article/pii/S0010482522008010.

Chen, X., Konukoglu, E., 2022. Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders. In: Medical Imaging with Deep Learning.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: Proceedings of the 37th International Conference on Machine Learning. ICML '20, JMLR.org.

Chen, X., You, S., Tezcan, K.C., Konukoglu, E., 2020. Unsupervised lesion detection via image restoration with a normative prior. Med. Image Anal. 64, 101713. http://dx.doi.org/10.1016/j.media.2020.101713.

Cho, J., Kang, I., Park, J., 2022. Self-Supervised 3D Out-of-Distribution Detection via Pseudoanomaly Generation. Springer, Cham, pp. 95–103. http://dx.doi.org/10.1007/978-3-030-97281-3_15, URL: https://link.springer.com/chapter/10.1007/978-3-030-97281-3_15.

Chu, C., Minami, K., Fukumizu, K., 2020. Smoothness and stability in GANs. In: 8th International Conference on Learning Representations. ICLR, URL: https://arxiv.org/pdf/2002.04185.

Chu, C., Zhmoginov, A., Sandler, M., 2017. CycleGAN, a master of steganography. In: von Luxburg, U., Guyon, I. (Eds.), Advances in Neural Information Processing Systems 31. NIPS 2017, Curran Associates Inc.

Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S.C., Girard, P., Améli, R., Ferré, J.-C., Kerbrat, A., Tourdias, T., Cervenansky, F., Glatard, T., Beaumont, J., Doyle, S., Forbes, F., Knight, J., Khademi, A., Mahbod, A., Wang, C., McKinley, R., Wagner, F., Muschelli, J., Sweeney, E., Roura, E., Lladó, X., Santos, M.M., Santos, W.P., Silva-Filho, A.G., Tomas-Fernandez, X., Urien, H., Bloch, I., Valverde, S., Cabezas, M., Vera-Olmos, F.J., Malpica, N., Guttmann, C., Vukusic, S., Edan, G., Dojat, M., Styner, M., Warfield, S.K., Cotton, F., Barillot, C., 2018. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Sci. Rep. 8 (1), 13650.

Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (Eds.), 2019. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II. In: Lecture Notes in Computer Science, vol. 11384, Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-030-11726-9.

Cuadra, M.B., Gomez, J., Hagmann, P., Pollo, C., Villemure, J.-G., Dawant, B.M., Thiran, J.-P., 2002. Atlas-based segmentation of pathological brains using a model of tumor growth. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002: 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part I 5. Springer, pp. 380–387.

Czimmermann, T., Ciuti, G., Milazzo, M., Chiurazzi, M., Roccella, S., Oddo, C.M., Dario, P., 2020. Visual-based defect detection and classification approaches for industrial applications—a survey. Sens. 20 (5), 1459.

Damudi, M.Z., Kini, A.S., 2024. Single-step sampling approach for unsupervised anomaly detection of brain MRI using denoising diffusion models. Int. J. Biomed. Imaging 2024 (1), 2352602. http://dx.doi.org/10.1155/ijbi/2352602, URL: https://onlinelibrary.wiley.com/doi/abs/10.1155/ijbi/2352602.

Das, S., Matthews, B.L., Srivastava, A.N., Oza, N.C., 2010. Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 47–56.

Dilokthanakul, N., Mediano, P.A.M., Garnelo, M., Lee, M.C.H., Salimbeni, H., Arulkumaran, K., Shanahan, M., 2017. Deep unsupervised clustering with Gaussian mixture variational autoencoders. In: 5th International Conference on Learning Representations. ICLR 2017, Toulon, France, URL: https://arxiv.org/pdf/1611.02648.

Doorenbos, L., Sznitman, R., Márquez-Neila, P., 2022. SS3D: Unsupervised Out-of-Distribution Detection and Localization for Medical Volumes. Springer, Cham, pp. 111–118. http://dx.doi.org/10.1007/978-3-030-97281-3_17, URL: https://link.springer.com/chapter/10.1007/978-3-030-97281-3_17.

Drew, T., Võ, M.L.-H., Wolfe, J.M., 2013. The invisible gorilla strikes again: sustained inattentional blindness in expert observers. Psychol. Sci. 24 (9), 1848–1853. http://dx.doi.org/10.1177/0956797613479386.

Fernando, T., Gammulle, H., Denman, S., Sridharan, S., Fookes, C., 2020. Deep learning for medical anomaly detection – A survey. arXiv preprint 2012.02364v1. URL: http://arxiv.org/pdf/2012.02364v1.

Fernando, T., Gammulle, H., Denman, S., Sridharan, S., Fookes, C., 2021. Deep learning for medical anomaly detection–a survey. ACM Comput. Surv. 54 (7), 1–37.

Friedrich, P., Wolleb, J., Bieder, F., Durrer, A., Cattin, P.C., 2024. Wdm: 3d wavelet diffusion models for high-resolution medical image synthesis. In: MICCAI Workshop on Deep Generative Models. Springer, pp. 11–21.

Frolova, D., Katrutsa, A., Oseledets, I., 2023. Feature-based pipeline for improving unsupervised anomaly segmentation on medical images. In: International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging. Springer, pp. 115–125.

Gering, D.T., Grimson, W.E.L., Kikinis, R., 2002. Recognizing deviations from normalcy for brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 388–395.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Proceedings of the 27th International Conference on Neural Information Processing Systems. NIPS14, pp. 2672–2680, URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Gooya, A., Goksel, O., Oguz, I., Burgos, N. (Eds.), 2018. Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al., 2020. Bootstrap your own latent-a new approach to self-supervised learning. Adv. Neural Inf. Process. Syst. 33, 21271–21284.

Gullapalli, R.P., 2011. Investigation of Prognostic Ability of Novel Imaging Markers for Traumatic Brain Injury (TBI). Defense Technical Information Center, Fort Belvoir, VA.

Guo, J., Lu, S., Jia, L., Zhang, W., Li, H., 2023. Encoder-decoder contrast for unsupervised anomaly detection in medical images. IEEE Trans. Med. Imaging PP, http://dx.doi.org/10.1109/TMI.2023.3327720.

Hamada, A., 2020. Brain tumor detection dataset. Kaggle dataset. https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection.

Han, C., Rundo, L., Murao, K., Noguchi, T., Shimahara, Y., Milacski, Z.Á., Koshino, S., Sala, E., Nakayama, H., Satoh, S., 2021. MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. BMC Bioinformatics 22 (Suppl 2), 31. http://dx.doi.org/10.1186/s12859-020-03936-1.

Harris, T.B., Launer, L.J., Eiriksdottir, G., Kjartansson, O., Jonsson, P.V., Sigurdsson, G., Thorgeirsson, G., Aspelund, T., Garcia, M.E., Cotch, M.F., Hoffman, H.J., Gudnason, V., for the Age, G.S.-R.S.I., 2007. Age, gene/environment susceptibility–reykjavik study: Multidisciplinary applied phenomics. Am. J. Epidemiol. 165 (9), 1076–1087. http://dx.doi.org/10.1093/aje/kwk115.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16000–16009.

He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9729–9738.

Heer, M., Postels, J., Chen, X., Konukoglu, E., Albarqouni, S., 2021. The OOD blind spot of unsupervised anomaly detection. Med. Imaging Deep. Learn. 286–300, URL: https://proceedings.mlr.press/v143/heer21a.html.

Hendrycks, D., Mazeika, M., Dietterich, T., 2019. Deep anomaly detection with outlier exposure. In: International Conference on Learning Representations. URL: https://openreview.net/forum?id=HyxCxhRcY7.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Hughes, E.J., Winchman, T., Padormo, F., Teixeira, R., Wurie, J., Sharma, M., Fox, M., Hutter, J., Cordero-Grande, L., Price, A.N., Allsop, J., Bueno-Conde, J., Tusor, N., Arichi, T., Edwards, A.D., Rutherford, M.A., Counsell, S.J., Hajnal, J.V., 2017. A dedicated neonatal brain imaging system. Magn. Reson. Med. 78 (2), 794–804. http://dx.doi.org/10.1002/mrm.26462.

Iglesias, J.E., Liu, C.-Y., Thompson, P.M., Tu, Z., 2011. Robust brain extraction across datasets and comparison with publicly available methods. IEEE Trans. Med. Imaging 30 (9), 1617–1634. http://dx.doi.org/10.1109/TMI.2011.2138152.

Iqbal, H., Khalid, U., Chen, C., Hua, J., 2023. Unsupervised anomaly detection in medical images using masked diffusion model. In: Cao, X., Xu, X., Rekik, I., Cui, Z., Ouyang, X. (Eds.), Machine Learning in Medical Imaging. In: Lecture Notes in Computer Science, vol. 14348, Springer Nature Switzerland, Cham, pp. 372–381. http://dx.doi.org/10.1007/978-3-031-45673-2_37.

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.-P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K.H., Kickingereder, P., 2019. Automated brain extraction of multisequence MRI using artificial neural networks. Hum. Brain Mapp. 40 (17), 4952–4964. http://dx.doi.org/10.1002/hbm.24750.

Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. J. Big Data 6 (1), 1–54. http://dx.doi.org/10.1186/s40537-019-0192-5.

Kabasawa, H., 2022. MR imaging in the 21st century: Technical innovation over the first two decades. Magn. Reson. Med. Sci. : MRMS : Off. J. Jpn. Soc. Magn. Reson. Med. 21 (1), 71–82. http://dx.doi.org/10.2463/mrms.rev.2021-0011.

Kascenas, A., Pugeault, N., O'Neil, A.Q., 2022a. Denoising autoencoders for unsupervised anomaly detection in brain MRI. In: Konukoglu, E., Menze, B., Venkataraman, A., Baumgartner, C., Dou, Q., Albarqouni, S. (Eds.), Proceedings of the 5th International Conference on Medical Imaging with Deep Learning. In: Proceedings of Machine Learning Research, vol. 172, PMLR, pp. 653–664, URL: https://proceedings.mlr.press/v172/kascenas22a.html.

Kascenas, A., Sanchez, P., Schrempf, P., Wang, C., Clackett, W., Mikhael, S.S., Voisey, J.P., Goatman, K., Weir, A., Pugeault, N., Tsaftaris, S.A., O'Neil, A.Q., 2023. The role of noise in denoising models for anomaly detection in medical images. Med. Image Anal. 90, 102963. http://dx.doi.org/10.1016/j.media.2023.102963.

Kascenas, A., Young, R., Jensen, B.S., Pugeault, N., O'Neil, A.Q., 2022b. Anomaly detection via context and local feature matching. In: 2022 IEEE 19th International Symposium on Biomedical Imaging. ISBI, pp. 1–5. http://dx.doi.org/10.1109/ISBI52829.2022.9761524.

Kingma, D., Welling, M., 2014. Auto-encoding variational Bayes. In: Bengio, Y., LeCun, Y. (Eds.), 2nd International Conference on Learning Representations. In: ICLR 2014, Banff, AB, Canada, URL: https://arxiv.org/pdf/1312.6114.pdf.

Kodali, N., Abernethy, J., Hays, J., Kira, Z., 2018. On convergence and stability of GANs. In: 6th International Conference on Learning Representations. ICLR, URL: https://arxiv.org/pdf/1705.07215.

Kuijf, H.J., Biesbroek, J.M., de Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., et al., 2019. Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge. IEEE Trans. Med. Imaging 38 (11), 2556–2568.

Kumar, K., Chakraborty, S., Roy, S., 2023. Self-supervised diffusion model for anomaly segmentation in medical imaging. In: Maji, P., Huang, T., Pal, N.R., Chaudhury, S., De, R.K. (Eds.), Pattern Recognition and Machine Intelligence. Springer Nature Switzerland, Cham, pp. 359–368.

Lagogiannis, I., Meissen, F., Kaissis, G., Rueckert, D., 2023. Unsupervised pathology detection: A deep dive into the state of the art. IEEE Trans. Med. Imaging PP, http://dx.doi.org/10.1109/TMI.2023.3298093.

Lambert, B., Louis, M., Doyle, S., Forbes, F., Dojat, M., Tucholka, A., 2021. Leveraging 3d information in unsupervised brain mri segmentation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, pp. 187–190. http://dx.doi.org/10.1109/ISBI48211.2021.9433894.

LaMontagne, P.J., Keefe, S., Lauren, W., Xiong, C., Grant, E.A., Moulder, K.L., Morris, J.C., Benzinger, T.L., Marcus, D.S., 2018. OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer's disease. Alzheimer's Dement. 14, P1097–P1097.

Landman, B.A., Huang, A.J., Gifford, A., Vikram, D.S., Lim, I.A.L., Farrell, J.A.D., Bogovic, J.A., Hua, J., Chen, M., Jarso, S., Smith, S.A., Joel, S., Mori, S., Pekar, J.J., Barker, P.B., Prince, J.L., van Zijl, P.C.M., 2010. Multi-parametric neuroimaging reproducibility: A 3T resource study. NeuroImage 54 (4), 2854–2866. http://dx.doi.org/10.1016/j.neuroimage.2010.11.047.

Lapuyade-Lahorgue, J., Xue, J.-H., Ruan, S., 2017. Segmenting multi-source images using hidden Markov fields with copula-based multivariate statistical distributions. IEEE Trans. Image Process. 26 (7), 3187–3195. http://dx.doi.org/10.1109/TIP.2017.2685345.

Laurikkala, J., Juhola, M., Kentala, E., Lavrac, N., Miksch, S., Kavsek, B., 2000. Informal identification of outliers in medical data. In: Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology. vol. 1, pp. 20–24, 1.

Lee, K., Lee, K., Lee, H., Shin, J., 2018. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. Adv. Neural Inf. Process. Syst. 31.

Lesjak, Ž., Galimzianova, A., Koren, A., Lukin, M., Pernuš, F., Likar, B., Špiclin, Ž., 2018. A novel public MR image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus. Neuroinformatics 16 (1), 51–63.

Li, C.-L., Sohn, K., Yoon, J., Pfister, T., 2021. Cutpaste: Self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9664–9674.

Liang, Z., Anthony, H., Wagner, F., Kamnitsas, K., 2023. Modality cycles with masked conditional diffusion for unsupervised anomaly segmentation in MRI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 168–181.

Liew, S.-L., Lo, B.P., Donnelly, M.R., Zavaliangos-Petropulu, A., Jeong, J.N., Barisano, G., Hutton, A., Simon, J.P., Juliano, J.M., Suri, A., et al., 2022. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. Sci. Data 9 (1), 320.

Loizillon, S., Jacob, Y., Aurélien, M., Dormont, D., Colliot, O., Burgos, N., Group, A.S., et al., 2024. Detecting brain anomalies in clinical routine with the β-VAE: Feasibility study on age-related white matter hyperintensities. In: Medical Imaging with Deep Learning.

Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Z. Med. Phys. 29 (2), 102–127. http://dx.doi.org/10.1016/j.zemedi.2018.11.002.

Luo, G., Xie, W., Gao, R., Zheng, T., Chen, L., Sun, H., 2023. Unsupervised anomaly detection in brain MRI: Learning abstract distribution from massive healthy brains. Comput. Biol. Med. 154, 106610. http://dx.doi.org/10.1016/j.compbiomed.2023.106610, URL: https://www.sciencedirect.com/science/article/pii/S0010482523000756.

Lüth, C.T., Zimmerer, D., Koehler, G., Jaeger, P.F., Isenensee, F., Maier-Hein, K.H., 2023. Contrastive representations for unsupervised anomaly detection and localization. In: BVM Workshop. Springer, pp. 246–252.

Luu, H.M., Park, S.-H., 2022. Extending nn-unet for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp. 173–186.

Ma, Z., Reich, D.S., Dembling, S., Duyn, J.H., Koretsky, A.P., 2022. Outlier detection in multimodal MRI identifies rare individual phenotypes among more than 15,000 brains. Hum. Brain Mapp. 43 (5), 1766–1782. http://dx.doi.org/10.1002/hbm.25756, URL: https://pubmed.ncbi.nlm.nih.gov/34957633/.

Maier, O., Menze, B.H., von der Gablentz, J., Hani, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., Christiaens, D., Dutil, F., Egger, K., Feng, C., Glocker, B., Goetz, M., Haeck, T., Halme, H.-L., Havaei, M., Iftekharuddin, K.M., Jodoin, P.-M., Kamnitsas, K., Kellner, E., Korvenoja, A., Larochelle, H., Ledig, C., Lee, J.-H., Maes, F., Mahmood, Q., Maier-Hein, K.H., McKinley, R., Muschelli, J., Pal, C., Pei, L., Rangarajan, J.R., Reza, S.M.S., Robben, D., Rueckert, D., Salli, E., Suetens, P., Wang, C.-W., Wilms, M., Kirschke, J.S., Kr Amer, U.M., Münte, T.F., Schramm, P., Wiest, R., Handels, H., Reyes, M., 2017. ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. 35, 250–269.

Marimont, S.N., Baugh, M., Siomos, V., Tzelepis, C., Kainz, B., Tarroni, G., 2024. DISYRE: Diffusion-inspired synthetic restoration for unsupervised anomaly detection. In: 2024 IEEE International Symposium on Biomedical Imaging. ISBI, pp. 1–5. http://dx.doi.org/10.1109/ISBI56570.2024.10635161.

Marimont, S.N., Siomos, V., Tarroni, G., 2023. MIM-OOD: Generative masked image modelling for out-of-distribution detection in medical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 35–44.

Marimont, S.N., Tarroni, G., 2021. Anomaly detection through latent space restoration using vector quantized variational autoencoders. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, pp. 1764–1767. http://dx.doi.org/10.1109/ISBI48211.2021.9433778.

Martins, S., Barbara Caroline Benato, Bruna Ferreira Silva, Clarissa Lyn Yasuda, Alexandre Xavier Falcão, 2019a. Modeling Normal Brain Asymmetry in MR Images Applied to Anomaly Detection Without Segmentation and Data Annotation. SPIE, pp. 71–80. http://dx.doi.org/10.1117/12.2512873.

Martins, S., Ruppert, G., Reis, F., Yasuda, C.L., Falcão, A.X., 2019b. A supervoxel-based approach for unsupervised abnormal asymmetry detection in mr images of the brain. In: 2019 IEEE 16th International Symposium on Biomedical Imaging. ISBI 2019, pp. 882–885. http://dx.doi.org/10.1109/ISBI.2019.8759166.

Martins, S., Telea, A.C., Falcao, A.X., 2019c. Extending supervoxel-based abnormal brain asymmetry detection to the native image space. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference 2019. vol. 2019, pp. 450–453. http://dx.doi.org/10.1109/EMBC.2019.8857447, URL: https://pubmed.ncbi.nlm.nih.gov/31945935/.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B., 2001. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). Philos. Trans. R. Soc. Lond. Ser. B 356 (1412), 1293–1322. http://dx.doi.org/10.1098/rstb.2001.0915.

McDonald, R.J., Schwartz, K.M., Eckel, L.J., Diehn, F.E., Hunt, C.H., Bartholmai, B.J., Erickson, B.J., Kallmes, D.F., 2015. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. Academic Radiol. 22 (9), 1191–1198. http://dx.doi.org/10.1016/j.acra.2015.05.007.

Meissen, F., Kaissis, G., Rueckert, D., 2022a. AutoSeg - Steering the Inductive Biases for Automatic Pathology Segmentation. Springer, Cham, pp. 127–135. http://dx.doi.org/10.1007/978-3-030-97281-3_19, URL: https://link.springer.com/chapter/10.1007/978-3-030-97281-3_19.

Meissen, F., Kaissis, G., Rueckert, D., 2022b. Challenging current semi-supervised anomaly segmentation methods for brain MRI. In: Crimi, A., Bakas, S. (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp. 63–74.

Meissen, F., Paetzold, J., Kaissis, G., Rueckert, D., 2022c. Unsupervised anomaly localization with structural feature-autoencoders. In: International MICCAI Brainlesion Workshop. Springer, pp. 14–24.

Meissen, F., Wiestler, B., Kaissis, G., Rueckert, D., 2022d. On the pitfalls of using the residual error as anomaly score. URL: https://arxiv.org/pdf/2202.03826.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., 2014. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging 34 (10), 1993–2024.

Moon, N., Bullitt, E., Van Leemput, K., Gerig, G., 2002. Automatic brain and tumor segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002: 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part I 5. Springer, pp. 372–379.

Muñoz-Ramírez, V., Kmetzsch, V., Forbes, F., Meoni, S., Moro, E., Dojat, M., 2022. Subtle anomaly detection: Application to brain MRI analysis of de novo Parkinsonian patients. Artif. Intell. Med. 125, 102251. http://dx.doi.org/10.1016/j.artmed.2022.102251, URL: https://www.sciencedirect.com/science/article/pii/S0933365722000161.

Myronenko, A., 2019. 3D MRI brain tumor segmentation using autoencoder regularization. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. In: Lecture Notes in Computer Science, vol. 11384, Springer International Publishing, Cham, pp. 311–320. http://dx.doi.org/10.1007/978-3-030-11726-9_28.

Naval Marimont, S., Siomos, V., Baugh, M., Tzelepis, C., Kainz, B., Tarroni, G., 2024. Ensembled cold-diffusion restorations for unsupervised anomaly detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 243–253.

Naval Marimont, S., Tarroni, G., 2021. Implicit Field Learning for Unsupervised Anomaly Detection in Medical Images. Springer, Cham, pp. 189–198. http://dx.doi.org/10.1007/978-3-030-87196-3_18, URL: https://link.springer.com/chapter/10.1007/978-3-030-87196-3_18.

Nguyen, B., Feldman, A., Bethapudi, S., Jennings, A., Willcocks, C.G., 2021. Unsupervised region-based anomaly detection in brain MRI with adversarial image inpainting. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, pp. 1127–1131. http://dx.doi.org/10.1109/ISBI48211.2021.9434115.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y.B., Li, S.-W., Misra, I., Rabbat, M.G., Sharma, V., Synnaeve, G., Xu, H., Jégou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P., 2023. DINOv2: Learning robust visual features without supervision. arXiv abs/2304.07193. URL: https://api.semanticscholar.org/CorpusID:258170077.

Park, S., Balint, A., Hwang, H., 2022. Self-supervised Medical Out-of-Distribution Using U-Net Vision Transformers. Springer, Cham, pp. 104–110. http://dx.doi.org/10.1007/978-3-030-97281-3_16, URL: https://link.springer.com/chapter/10.1007/978-3-030-97281-3_16.

Parkinson Progression Marker Initiative, 2011. The Parkinson progression marker initiative (PPMI). Prog. Neurobiol. 95 (4), 629–635. http://dx.doi.org/10.1016/j.pneurobio.2011.09.005.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A., 2016. Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, URL: https://arxiv.org/pdf/1604.07379.

Pernet, C.R., Gorgolewski, K.J., Job, D., Rodriguez, D., Whittle, I., Wardlaw, J., 2016. A structural and functional magnetic resonance imaging dataset of brain tumour patients. Sci. Data 3, 160003. http://dx.doi.org/10.1038/sdata.2016.3.

Pinaya, W.H.L., Graham, M.S., Gray, R., Da Costa, P.F., Tudosiu, P.-D., Wright, P., Mah, Y.H., MacKinnon, A.D., Teo, J.T., Jager, R., et al., 2022a. Fast unsupervised brain anomaly detection and segmentation with diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 705–714.

Pinaya, W.H.L., Tudosiu, P.-D., Gray, R., Rees, G., Nachev, P., Ourselin, S., Cardoso, M.J., 2021. Unsupervised brain anomaly detection and segmentation with transformers. In: Heinrich, M., Dou, Q., de Bruijne, M., Lellmann, J., Schläfer, A., Ernst, F. (Eds.), Proceedings of the Fourth Conference on Medical Imaging with Deep Learning. In: Proceedings of Machine Learning Research, vol. 143, PMLR, pp. 596–617, URL: https://proceedings.mlr.press/v143/pinaya21a.html.

Pinaya, W.H.L., Tudosiu, P.-D., Gray, R., Rees, G., Nachev, P., Ourselin, S., Cardoso, M.J., 2022b. Unsupervised brain imaging 3D anomaly detection and segmentation with transformers. Med. Image Anal. 79, 102475. http://dx.doi.org/10.1016/j.media.2022.102475.

Pinho, A.L., Amadon, A., Gauthier, B., Clairis, N., Knops, A., Genon, S., Dohmatob, E., Torre, J.J., Ginisty, C., Becuwe-Desmidt, S., Roger, S., Lecomte, Y., Berland, V., Laurier, L., Joly-Testault, V., Médiouni-Cloarec, G., Doublé, C., Martins, B., Salmon, E., Piazza, M., Melcher, D., Pessiglione, M., van Wassenhove, V., Eger, E., Varoquaux, G., Dehaene, S., Hertz-Pannier, L., Thirion, B., 2020. Individual brain charting dataset extension, second release of high-resolution fMRI data for cognitive mapping. Sci. Data 7 (1), 353. http://dx.doi.org/10.1038/s41597-020-00670-4.

Pinon, N., Oudoumanessah, G., Trombetta, R., Dojat, M., Forbes, F., Lartizien, C., 2023a. Brain subtle anomaly detection based on auto-encoders latent space analysis: Application to de novo Parkinson patients. In: 2023 IEEE 20th International Symposium on Biomedical Imaging. ISBI, pp. 1–5. http://dx.doi.org/10.1109/ISBI53787.2023.10230351.

Pinon, N., Trombetta, R., Lartizien, C., 2023b. One-class SVM on siamese neural network latent space for unsupervised anomaly detection on brain MRI white matter hyperintensities. Med. Imaging Deep. Learn. URL: https://openreview.net/forum?id=_c9r6-HCEaN.

Porz, N., Bauer, S., Pica, A., Schucht, P., Beck, J., Verma, R.K., Slotboom, J., Reyes, M., Wiest, R., 2014. Multi-modal glioblastoma segmentation: man versus machine. PLoS One 9 (5), e96873. http://dx.doi.org/10.1371/journal.pone.0096873.

Prastawa, M., Bullitt, E., Ho, S., Gerig, G., 2004. A brain tumor segmentation framework based on outlier detection. Med. Image Anal. 8 (3), 275–283. http://dx.doi.org/10.1016/j.media.2004.06.007, URL: https://www.sciencedirect.com/science/article/pii/S1361841504000295. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003.

Prastawa, M., Bullitt, E., Moon, N., Van Leemput, K., Gerig, G., 2003. Automatic brain tumor segmentation by subject specific modification of atlas priors1. Academic Radiol. 10 (12), 1341–1348.

Puccio, B., Pooley, J.P., Pellman, J.S., Taverna, E.C., Craddock, R.C., 2016. The preprocessed connectomes project repository of manually corrected skull-stripped T1-weighted anatomical MRI data. GigaScience 5 (1), 45. http://dx.doi.org/10.1186/s13742-016-0150-5.

Raad, J.D., Chinnam, R.B., Arslanturk, S., Tan, S., Jeong, J.-W., Mody, S., 2023. Unsupervised abnormality detection in neonatal MRI brain scans using deep learning. Sci. Rep. 13 (1), 11489. http://dx.doi.org/10.1038/s41598-023-38430-0.

Rashmi, K., Das, A., Matcha, N., Ram, K., Sivaprakasam, M., 2024. Ano-swinMAE: Unsupervised anomaly detection in brain MRI using swin transformer based masked auto encoder. In: Medical Imaging with Deep Learning.

Ristea, N.-C., Madan, N., Ionescu, R.T., Nasrollahi, K., Khan, F.S., Moeslund, T.B., Shah, M., 2022. Self-supervised predictive convolutional attentive block for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Roberts, S., Tarassenko, L., 1994. A probabilistic resource allocating network for novelty detection. Neural Comput. 6 (2), 270–284.

Rohlfing, T., Zahr, N.M., Sullivan, E.V., Pfefferbaum, A., 2010. The SRI24 multichannel atlas of normal adult human brain structure. Hum. Brain Mapp. 31 (5), 798–819. http://dx.doi.org/10.1002/hbm.20906.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer Assisted Intervention. pp. 234–241.

Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P., 2022. Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14318–14328.

Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., Dietterich, T.G., Muller, K.-R., 2021. A unifying review of deep and shallow anomaly detection. Proc. IEEE 109 (5), 756–795. http://dx.doi.org/10.1109/JPROC.2021.3052449.

Ruff, L., Vandermeulen, R.A., Görnitz, N., Binder, A., Müller, E., Müller, K.-R., Kloft, M., 2020. Deep semi-supervised anomaly detection. In: International Conference on Learning Representations. https://openreview.net/forum?id=HkgH0TEYwH.

Saase, V., Wenz, H., Ganslandt, T., Groden, C., Maros, M.E., 2020. Simple statistical methods for unsupervised brain anomaly detection on MRI are competitive to deep learning methods. arXiv preprint arXiv:2011.12735.

Sato, K., Hama, K., Matsubara, T., Uehara, K., 2019. Predictable uncertainty-aware unsupervised deep anomaly segmentation. In: 2019 International Joint Conference on Neural Networks. IJCNN 2019, IEEE, Piscataway, NJ, pp. 1–7. http://dx.doi.org/10.1109/IJCNN.2019.8852144.

Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U., 2019. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Med. Image Anal. 54, 30–44.

Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G., 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: The 25th Biennial International Conference on Information Processing in Medical Imaging. IPMI, Springer International Publishing, pp. 146–157, URL: https://arxiv.org/pdf/1703.05921.

Schölkopf, B., Williamson, R.C., Smola, A., Shawe-Taylor, J., Platt, J., 1999. Support vector method for novelty detection. Adv. Neural Inf. Process. Syst. 12.

Schwarz, J., Will, L., Wellmer, J., Mosig, A., 2024. A patch-based student-teacher pyramid matching approach to anomaly detection in 3D magnetic resonance imaging. In: Medical Imaging with Deep Learning.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 618–626. http://dx.doi.org/10.1109/ICCV.2017.74.

Shin, H.-C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M., 2018. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Gooya, A., Goksel, O., Oguz, I., Burgos, N. (Eds.), Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham, pp. 1–11.

Siddiquee, M.M.R., Shah, J., Wu, T., Chong, C., Schwedt, T.J., Dumkrieger, G., Nikolova, S., Li, B., 2024. Brainomaly: Unsupervised neurologic disease detection utilizing unannotated t1-weighted brain mr images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7573–7582.

Silva-Rodríguez, J., Naranjo, V., Dolz, J., 2022. Constrained unsupervised anomaly segmentation. Med. Image Anal. 80, 102526.

Sogancioglu, E., Ginneken, B.v., Behrendt, F., Bengs, M., Schlaefer, A., Radu, M., Xu, D., Sheng, K., Scalzo, F., Marcus, E., Papa, S., Teuwen, J., Scholten, E.T., Schalekamp, S., Hendrix, N., Jacobs, C., Hendrix, W., Sánchez, C.I., Murphy, K., 2024. Nodule detection and generation on chest X-Rays: NODE21 challenge. IEEE Trans. Med. Imaging 43 (8), 2839–2853. http://dx.doi.org/10.1109/TMI.2024.3382042.

Song, X., Wyrwicz, A.M., 2009. Unsupervised spatiotemporal fMRI data analysis using support vector machines. NeuroImage 47 (1), 204–212. http://dx.doi.org/10.1016/j.neuroimage.2009.03.054, URL: https://www.sciencedirect.com/science/article/pii/S1053811909002924.

Souza, R., Lucena, O., Garrafa, J., Gobbi, D., Saluzzi, M., Appenzeller, S., Rittner, L., Frayne, R., Lotufo, R., 2018. An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. NeuroImage 170, 482–494. http://dx.doi.org/10.1016/j.neuroimage.2017.08.021.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., Collins, R., 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12 (3), e1001779.

Tan, J., Hou, B., Batten, J., Qiu, H., Kainz, B., 2022a. Detecting outliers with foreign patch interpolation. Mach. Learn. Biomed. Imaging 1 (April 2022 issue), 1–27, URL: https://www.melba-journal.org/papers/2022:013.html.

Tan, J., Hou, B., Day, T., Simpson, J., Rueckert, D., Kainz, B., 2021. Detecting Outliers with Poisson Image Interpolation. Springer, Cham, pp. 581–591. http://dx.doi.org/10.1007/978-3-030-87240-3_56, URL: https://link.springer.com/chapter/10.1007/978-3-030-87240-3_56.

Tan, J., Kart, T., Hou, B., Batten, J., Kainz, B., 2022b. MetaDetector: Detecting Outliers by Learning to Learn from Self-supervision. Springer, Cham, pp. 119–126. http://dx.doi.org/10.1007/978-3-030-97281-3_18, URL: https://link.springer.com/chapter/10.1007/978-3-030-97281-3_18.

Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., Tyler, L.K., Cam-Can, Henson, R.N., 2017. The cambridge centre for ageing and neuroscience (cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. NeuroImage 144 (Pt B), 262–269.

Tian, K., Jiang, Y., Lin, C., Wang, L., Yuan, Z., et al., 2023. Designing BERT for convolutional networks: Sparse and hierarchical masked modeling. In: The Eleventh International Conference on Learning Representations.

Tibshirani, R., Hastie, T., 2007. Outlier sums for differential gene expression analysis. Biostatistics 8 (1), 2–8.

Tschuchnig, M.E., Gadermayr, M., 2022. Anomaly detection in medical imaging-a mini review. In: Data Science–Analytics and Applications: Proceedings of the 4th International Data Science Conference–IDSC2021. Springer, pp. 33–38.

Uzunova, H., Schultz, S., Handels, H., Ehrhardt, J., 2019. Unsupervised pathology detection in medical images using conditional variational autoencoders. Int. J. Comput. Assist. Radiol. Surg. 14 (3), 451–461. http://dx.doi.org/10.1007/s11548-018-1898-0.

van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., 2012. The human connectome project: a data acquisition perspective. NeuroImage 62 (4), 2222–2231. http://dx.doi.org/10.1016/j.neuroimage.2012.02.018.

van Hespen, K.M., Zwanenburg, J.J.M., Dankbaar, J.W., Geerlings, M.I., Hendrikse, J., Kuijf, H.J., 2021. An anomaly detection approach to identify chronic brain infarcts on MRI. Sci. Rep. 11 (1), 7714. http://dx.doi.org/10.1038/s41598-021-87013-4, URL: https://www.nature.com/articles/s41598-021-87013-4.

van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans. Med. Imaging 20 (8), 677–688. http://dx.doi.org/10.1109/42.938237, URL: https://pubmed.ncbi.nlm.nih.gov/11513020/.

Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612. http://dx.doi.org/10.1109/TIP.2003.819861.

Wang, G., Han, S., Ding, E., Huang, D., 2021. Student-teacher feature pyramid matching for anomaly detection. arXiv preprint arXiv:2103.04257.

Wang, R., Li, C., Wang, J., Wei, X., Li, Y., Hui, C., Zhu, Y., Zhang, S., 2014. Automatic segmentation of white matter lesions on magnetic resonance images of the brain by using an outlier detection strategy. Magn. Reson. Imaging 32 (10), 1321–1329. http://dx.doi.org/10.1016/j.mri.2014.08.010, URL: https://www.sciencedirect.com/science/article/pii/S0730725X14002409.

Watkins, K.E., Paus, T., Lerch, J.P., Zijdenbos, A., Collins, D.L., Neelin, P., Taylor, J., Worsley, K.J., Evans, A.C., 2001. Structural asymmetries in the human brain: a voxel-based statistical analysis of 142 MRI scans. Cereb. Cortex (New York, NY: 1991) 11 (9), 868–877. http://dx.doi.org/10.1093/cercor/11.9.868.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Morris, J.C., Petersen, R.C., Salazar, J., Saykin, A.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q., 2017. The Alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. Alzheimer's Dement. : J. Alzheimer's Assoc. 13 (5), 561–571.

Wolleb, J., Bieder, F., Friedrich, P., Zhang, P., Durrer, A., Cattin, P.C., 2024. Binary noise for binary tasks: Masked bernoulli diffusion for unsupervised anomaly detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 135–145.

Wood, D.A., Kafiabadi, S., Al Busaidi, A., Guilhem, E., Montvila, A., Agarwal, S., Lynch, J., Townend, M., Barker, G., Ourselin, S., et al., 2021. Automated triaging of head MRI examinations using convolutional neural networks. In: Medical Imaging with Deep Learning. PMLR, pp. 813–841.

Woolard, A.A., Heckers, S., 2012. Anatomical and functional correlates of human hippocampal volume asymmetry. Psychiatry Res. 201 (1), 48–53. http://dx.doi.org/10.1016/j.pscychresns.2011.07.016.

Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G., 2022. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 650–656.

Xu, P., Lyu, J., Lin, L., Cheng, P., Tang, X., 2024. LF-SynthSeg: Label-free brain tissue-assisted tumor synthesis and segmentation. IEEE J. Biomed. Health Inform..

Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y., 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6023–6032.

Zavrtanik, V., Kristan, M., Skočaj, D., 2021. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8330–8339.

Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., Parente, M., Geras, K.J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzal, M., Romero, A., Rabbat, M., Vincent, P., Yakubova, N., Pinkerton, J., Wang, D., Owens, E., Zitnick, C.L., Recht, M.P., Sodickson, D.K., Lui, Y.W., 2019. fastMRI: An open dataset and benchmarks for accelerated MRI.

Zhou, C., Paffenroth, R.C., 2017. Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '17, Association for Computing Machinery, New York, NY, USA, pp. 665–674. http://dx.doi.org/10.1145/3097983.3098052.

Zimmerer, D., Full, P.M., Isensee, F., Jager, P., Adler, T., Petersen, J., Kohler, G., Ross, T., Reinke, A., Kascenas, A., Jensen, B.S., O'Neil, A.Q., Tan, J., Hou, B., Batten, J., Qiu, H., Kainz, B., Shvetsova, N., Fedulova, I., Dylov, D.V., Yu, B., Zhai, J., Hu, J., Si, R., Zhou, S., Wang, S., Li, X., Chen, X., Zhao, Y., Marimont, S.N., Tarroni, G., Saase, V., Maier-Hein, L., Maier-Hein, K., 2022a. MOOD 2020: A public benchmark for out-of-distribution detection and localization on medical images. IEEE Trans. Med. Imaging 41 (10), 2728–2738. http://dx.doi.org/10.1109/TMI.2022.3170077.

Zimmerer, D., Isensee, F., Petersen, J., Kohl, S., Maier-Hein, K., 2019a. Unsupervised anomaly localization using variational auto-encoders. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer International Publishing, Cham, pp. 289–297.

Zimmerer, D., Kohl, S., Petersen, J., Isensee, F., Maier-Hein, K., 2019b. Context-encoding variational autoencoder for unsupervised anomaly detection. In: International Conference on Medical Imaging with Deep Learning. MIDL, In: Proceedings of Machine Learning Research, vol. 102, PMLR, URL: https://arxiv.org/pdf/1907.12258.

Zimmerer, D., Maier-Hein, K., 2024. Beyond heatmaps: A comparative analysis of metrics for anomaly localization in medical images. In: International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging. Springer, pp. 138–148.

Zimmerer, D., Paech, D., Lüth, C., Petersen, J., Köhler, G., Maier-Hein, K., 2022b. Unsupervised anomaly detection in the wild. In: Bildverarbeitung FÜr Die Medizin 2022: Proceedings, German Workshop on Medical Image Computing, Heidelberg, June 26-28, 2022. Springer, pp. 26–31.