FACETS

OPEN ACCESS | Review

X-ray modalities in the era of artificial intelligence: overview of self-supervised learning approach

Ivan Martinović[a,b], **Shitong Mao**[c], **Mehdy Dousty**[b,d], **Wuqi Li**[b], **Milena Đukanović**[a], **Errol Colak**[e,f], **and Ervin Sejdić**[b,g]

[a]Faculty of Electrical Engineering, University of Montenegro, Podgorica, Montenegro; [b]Edward S. Rogers Sr. Department of Electrical and Computer Engineering, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON, Canada; [c]Department of Head and Neck Surgery, University of Texas MD Anderson Cancer Center, Huston, TX, USA; [d]Vector Institute, Toronto, ON, Canada; [e]Department of Medical Imaging, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada; [f]Department of Medical Imaging, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada; [g]North York General Hospital, Toronto, ON, Canada

Corresponding author: **Ervin Sejdić** (email: ervin.sejdic@utoronto.ca)

Abstract

Self-supervised learning enables the creation of algorithms that outperform supervised pre-training methods in numerous computer vision tasks. This paper provides a comprehensive overview of self-supervised learning applications across various X-ray modalities, including conventional X-ray, computed tomography, mammography, and dental X-ray. Apart from the application of self-supervised learning in the interpretation phase of X-ray images, the paper also emphasizes the critical role of self-supervised learning integration in the preprocessing and archiving phase. Furthermore, the paper explores the application of self-supervised learning in multi-modal scenarios, which represents a key future direction in developing machine learning-based applications across the field of medicine. Lastly, the paper addresses the main challenges associated with the development of self-supervised learning applications tailored for X-ray modalities. The findings from the reviewed literature strongly suggest that the self-supervised learning approach has the potential to be a "*game-changer*", enabling the elimination of the current situation where many machine learning-based systems are developed but few are deployed in daily clinical practice.

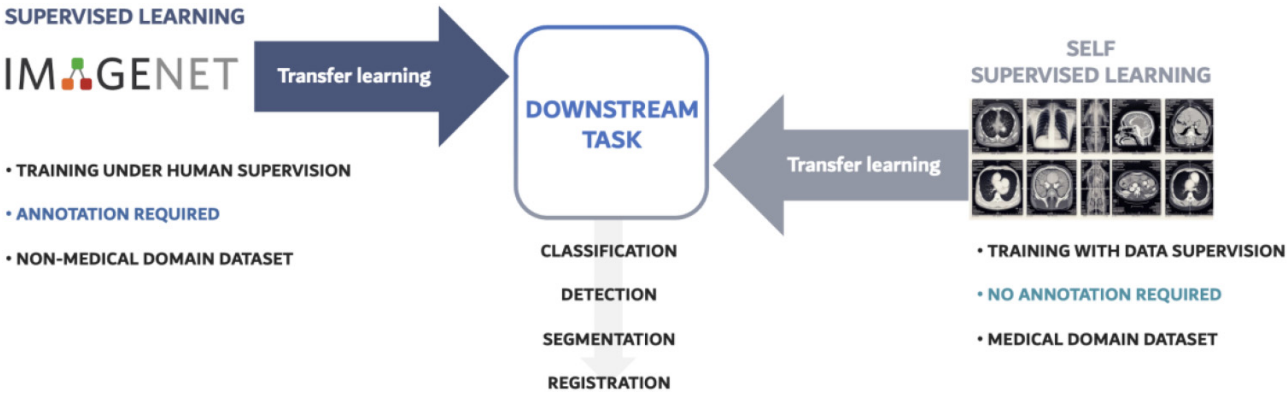
Key words: self-supervised learning, X-ray modalities, generative models, contrastive models, multi-modal scenarios

1. Introduction

Radiology has been established for over 125 years and has significantly reduced mortality rates from various diseases such as pneumonia, cancer, coronary heart disease, and nonfatal myocardial infarction (Howell 2011; The National Lung Screening Trial Research Team 2011; Wake et al. 2011; Crummy et al. 2018; Imai et al. 2018). Currently, radiologists are faced with the challenge of interpreting a vast amount of imaging data. The immense task of interpreting medical images leads to radiologists experiencing fatigue and burnout, consequently raising the probability of medical errors (Bercovich and Javitt 2018). To address these challenges, there is an urge to develop algorithms. Implementing automated medical image analysis offers several advantages, including enhanced sensitivity for subtle findings, prioritization of time-sensitive cases, automation of routine tasks, and alleviating the scarcity of radiologists in remote areas and developing countries (Çallı et al. 2021).

The deployment of machine learning (ML) in radiology holds promise for improving many aspects of the radiology workflow (Pierre et al. 2023), potentially resulting in increased diagnostic accuracy and efficiency, optimized treatment plans, enhanced quality of care, and a potential reduction in healthcare-related expenditures. The currently developed deep learning-based systems for automated medical image interpretation have reached levels comparable to practicing radiologists' performance in some tasks (Rajpurkar et al. 2017). Studies have also shown that using ML can assist physicians in identifying abnormalities in medical images more effectively (Leibig et al. 2022). Despite all these breakthroughs, several challenging aspects still need to be addressed. Deep learning relies on having access to a substantial amount of annotated data (Zhou et al. 2021). This issue is particularly problematic in medical applications because medical imaging datasets are significantly smaller (hundreds/thousands) compared to natural domain datasets (millions of samples). The annotation of medical imaging data adds another challenge as it requires a substantial amount of time, effort, and a team of experts (Mckinney et al. 2020). To tackle the issue of data scarcity, the transfer learning technique is a promising solution. This method involves pretraining systems on a large natural domain dataset and then employing them in the medical image domain. However, the domain gap between the two datasets is substantial because medical images have entirely different characteristics from natural images, such as

Fig. 1. Supervised versus self-supervised learning in medical imaging.



low image quality, latent feature distribution, lower resolution, 3D form, and similar image content across images (Zhou et al. 2019; Zhuang et al. 2019; Li et al. 2021*a*). Data imbalance is another issue for deep learning algorithms. This imbalance is particularly noticeable in the context of rare diseases, where most of the dataset comprises "normal" data, while instances of the disease are statistically scarce. Additionally, the ML systems frequently struggle with the designated tasks when data are collected from different acquisition centers and devices, introducing variation within the medical imaging domain, such as signal-to-noise ratio. All these practical challenges critically stagnate the robustness and generalizability of transfer learning.

Therefore, a logical question that arises is: *How is it possible to mitigate or overcome the aforementioned challenges?* To investigate possible answers to this question, investigators have developed systems for automated medical image analysis using a relatively new approach called self-supervised learning (SSL), which enables the extraction of meaningful representations from unlabeled data, thereby reducing the reliance on large, labeled datasets and potentially improving the efficiency and accuracy of medical image analysis. We find the SSL approach as a significant step forward in the application of ML in X-ray modalities. Therefore, in this paper, we summarize the contributions of SSL applications in X-ray modalities and provide a comprehensive overview of the current state of medical imaging. Moreover, we discuss the main challenges and future work in this field.

2. General overview of self-supervised learning

SSL allows learning representations without explicit supervision. The general pipeline of SSL involves two tasks: a pretext/proxy task and a downstream task (Shurrab and Duwairi 2022). The pretext task is performed in a supervised manner using unlabeled data, allowing the model to learn meaningful representations by creating pseudo labels. In the second step, the learned representations from the pretext task are transferred to the downstream task for fine-tuning (Liu et al. 2021*c*). One advantage of the SSL approach over supervised methods is that pre-training and fine-tuning can be performed using the same image distribution (Fig. 1) (Newell and Deng 2020).

SSL algorithms have demonstrated remarkable success in the field of computer vision, with some algorithms outperforming supervised pre-training methods in various computer vision tasks (Tomasev et al. 2022). It remains an open question whether these outstanding results can be replicated in the field of X-ray imaging, given the intrinsic disparities that distinguish it from natural image analysis.

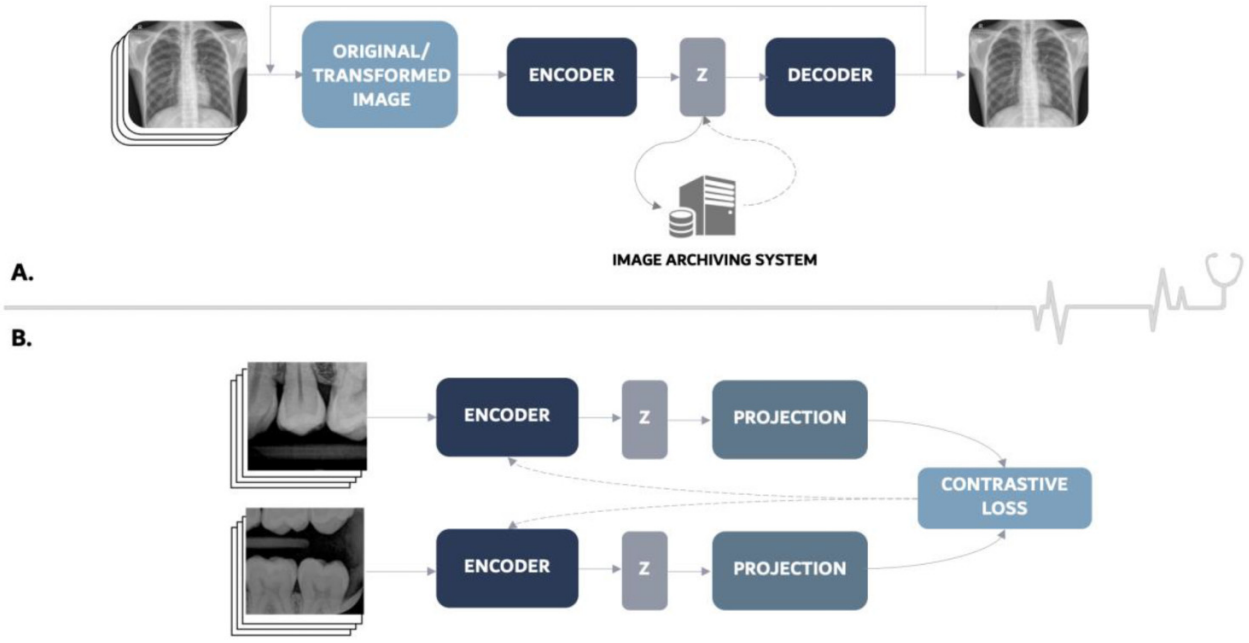
Currently, most SSL research focuses on developing pretext tasks since the pretext task is the key component where SSL takes place. SSL approaches are categorized in different ways based on model architectures and objectives (Liu et al. 2021*c*). Liu et al. (2021*c*) divided them into generative, contrastive, and generative–contrastive, while other authors classified them as predictive, generative, and contrastive (Shurrab and Duwairi 2022) or encoder, and encoder-decoder (Navarro et al. 2021). In this paper, our focus will be on generative and contrastive models as they are the most used approaches for X-ray modalities.

2.1. Generative

The generative SSL approach allows the model to learn latent features from unlabeled data by modeling the underlying distribution of the input data (Shurrab and Duwairi 2022). A typical application is the photo restoration task (Huu et al. 2022). The basic architecture of generative models consists of two components: the encoder and decoder networks (Fig. 2*a*). The encoder network compresses the input data into a latent space (Z), while the decoder network reconstructs the compressed input from the latent space. Two commonly used generative models in X-ray modalities are auto-regressive models and auto-encoder models.

The auto-regressive methods allow learning meaningful representations from the image datasets by regenerating images pixel by pixel. Nagoor et al. (2022) stated that the auto-regressive model is one of the state-of-the-art models for estimating data distribution and pixel likelihood. Auto-regressive models are implemented across different architectures such as PixelRNN (van den Oord et al. 2016), PixelCNN

Fig. 2. Self-supervised learning approaches: (A) generative and (B) contrastive models-schematic diagrams.



(van den Oord et al. 2016), PixelCNN++ (Salimans et al. 2017), and Multiscale-PixelCNN (Reed et al. 2017). The main advantage of auto-regressive models in computer vision is their ability to effectively capture context dependency, allowing them to understand spatial relationships and dependencies in image data. However, a limitation of these models is their unidirectional accessibility to context, which hinders their ability to incorporate information from future positions and capture long-range dependencies and global structures in images (Liu et al. 2021*c*).

The auto-encoder models can discover structure within data to produce a compact representation of the input (Liu et al. 2021*c*). Although auto-encoders were first used 40 years ago, they remain the most popular generative models due to their flexibility and adaptability (Ehrhardt and Wilms 2022). Commonly used variants of the general auto-encoder architecture include denoising autoencoders, convolutional autoencoders, and variational autoencoders (Liu et al. 2021*c*). Denoising autoencoders learn representations invariant to noise by reconstructing a noise-free output from noisy input (e.g., Gaussian noise, Poisson noise, uniform noise, impulsive noise, etc.) (Liu et al. 2021*c*). However, denoising autoencoders overlook the 2D structure of images, resulting in redundancy and a global representation of features (Masci et al. 2011). On the other hand, convolutional autoencoders capture localized features by sharing weights across all input locations, enabling spatial locality, and reconstructing images using a linear combination of basic image patches based on the latent code (Masci et al. 2011). Variational autoencoders combine Bayesian variational inference with deep learning to learn the probability distribution of data through amortized variational inference (Singh and Ogunfunmi 2021) and the reparameterization trick.

2.2. Contrastive

Contrastive models learn by comparing an anchor with positive and negative instances generated through data augmentation (Fig. 2*b*). The methodology behind contrastive learning is remarkably intuitive and mirrors the learning mechanisms observed in children (Dehghan and Amasyali 2023). In a child's matching game (Ramani and Scalise 2020), the objective of contrastive learning is to group similar instances while keeping distant dissimilar instances. Contrastive learning frameworks could be divided into different ways (Jaiswal et al. 2020; Le-Khac et al. 2020; Kumar et al. 2022; Lu 2022), but the most common division is on context–instance and instance–instance contrasts (Liu et al. 2021*c*).

Context–instance contrast methods (global–local contrast) are centered around modeling the relationship between the instance's local features and its global context representation (Liu et al. 2021*c*). Two basic strategies of global–local contrast are predicting relative position and maximizing mutual information. The first category emphasizes the relative position of local components, which inherently suggests an understanding of the global context. For example, predicting the relative position of two organs within the human body requires comprehensive knowledge of the body's overall composition. Accordingly, scientists developed several models that could be used as the pretext task. These models solve tasks such as jigsaw, rotation prediction, and relative position tasks (Jaiswal et al. 2020; Liu et al. 2021*c*; Kumar et al. 2022). On the other hand, the second strategy ignores the relative position between local parts and instead focuses on learning the belonging relationship between local parts and the global context. This concept is derived from mutual information in statistics (Liu et al. 2021*c*). Essentially, mutual information assesses the association between two variables that are sampled simultaneously. In the global–local contrast

approach, the objective is to maximize mutual information. Use of the mutual information in contrastive learning tasks initiated with the Deep Infomax (Hjelm et al. 2019), and it has further evolved with the introduction of new models such as Augmented Multiscale Deep InfoMax-AMDIM (Bachman et al. 2019). Meanwhile, the use of mutual information is not only reserved for global–local contrast but is also applied in instance–instance contrast.

Instance–instance contrastive methods focus on modeling the relationships between instance-level local representations from different instances, and they could be further divided into two categories: cluster discrimination and instance discrimination.

The initial success of instance–instance contrastive methods was demonstrated by their ability to achieve performance comparable to the AlexNet supervised model (Liu et al. 2021*c*). This was achieved through the utilization of clustering-based techniques, such as Deep Cluster (Caron et al. 2019), which employs instance clustering to generate pseudo-labels. An advanced variant of DeepCluster, known as swapping assignments between multiple views (Caron et al. 2021), takes the concept further by incorporating online clustering principles and multi-view data augmentation strategies. The integration of data augmentation into the swapping assignments between multiple views method draws inspiration from instance discrimination-based approaches, which leverage these strategies as a substitute for the time-intensive clustering process. Recently, the development of instance–instance contrastive methods has predominantly centered around approaches based on instance discrimination. Well-known methods in this context include CMC (Tian et al. 2020*a*), MoCo (He et al. 2019), SimCLR (Chen et al. 2020), InfoMin (Tian et al. 2020*b*), BYOL (Grill et al. 2020), and SimSiam (Chen and He 2020). To achieve contrastive learning objectives, data augmentation techniques are used to generate similar or positive instances. However, the generation and utilization of dissimilar or negative instances vary across each method.

3. Applications in X-ray modalities

In the field of X-ray modalities, ML approaches have gained significant attention for automating the analysis of X-ray images. This has led to the development of systems that can perform various key tasks, such as segmenting organs, lesions, or tumors, classifying different diseases or conditions, detecting abnormalities, monitoring disease progression or treatment response, preprocessing X-ray images, and archiving X-ray images.

The manuscripts referenced in this review were sourced from extensive research databases including Springer Link, ScienceDirect, IEEE Explore, ArXiv, and PubMed. Keyword searches were performed across these databases using terms such as “medical imaging”, “deep learning”, “self-supervised learning”, “computed tomography”, and “X-ray imaging”. Given the narrow focus of our research, the most effective strategy was to formulate queries that combined two or more keywords. Consequently, we selected the query “self-supervised learning” AND “X-ray imaging” as the most relevant. A search conducted in March 2023 yielded a total of 5202 records (PubMed = 54; IEEE Explorer = 49; ScienceDirect = 4214; SpringerLink = 823; ArXiv = 62). After removing duplicates (primarily from PubMed and arXiv), 5162 unique articles remained. Following a title and abstract screening, 5107 articles were excluded based on irrelevance to the research objective. A full-text review was then conducted on the remaining 58 articles, resulting in the exclusion of eight additional studies, primarily due to the absence of downstream task evaluation or the use of non-human (animal) datasets. Figure 3 shows our screening strategy and screening process results.

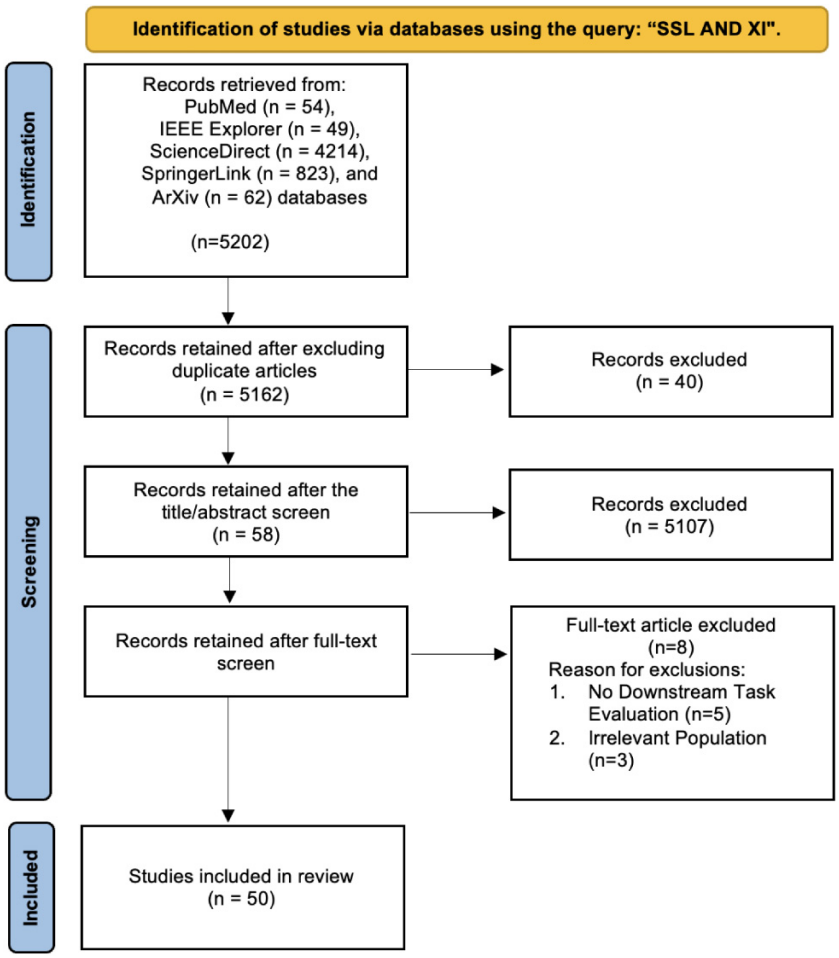
In the subsequent sections, we present a comprehensive summary of the 50 selected studies across various X-ray modalities. As a significantly larger number of publications about the application of SSL in X-ray modalities started to emerge in 2018, we took into consideration the literature starting from this point of time for this comprehensive review.

3.1. Conventional X-ray

Conventional X-ray holds historical significance in the application of ML in medical imaging as it was one of the first modalities in which ML was utilized to assist in the interpretation of medical images. The SSL approach has found different applications in this modality, with the main studies listed in Table 1.

The development of SSL algorithms for chest X-rays is especially complex due to the nature of disease/disorder detection, often requiring the identification of abnormalities within a limited number of pixels (Sowrirajan et al. 2020). In contrastive models, one main challenge is to generate positive image pairs as emphasized in the MoCo-CXR study (Sowrirajan et al. 2020). This study developed a specific data augmentation strategy for chest X-rays, as traditional augmentations used for natural images may not be appropriate due to the lack of meaningful representation in a chest X-ray. By using random rotation (10 degrees) and horizontal flipping, this approach created a pre-trained model that generated better representations and initializations for the detection of pathologies. MoCo-CXR demonstrated that SSL models outperformed ImageNet-pretrained models for higher-quality representations. Similar results were observed in the case of C2 L (Zhou et al. 2020) and MUSCLE (Liao et al. 2022) models. While C2 L exclusively utilized a chest X-ray dataset like MoCo-CXR, MUSCLE was developed using a broader multi-instance dataset encompassing the head, lungs, and bones. A different approach was proposed in a study (Truong et al. 2021), in which the authors developed SSL models that used the ImageNet dataset for a pretext task. A study (Azizi et al. 2021) attempted to combine the merits of C2L, MoCo-CXR, and (Truong et al. 2021), and further developed a two-stage pretext task. First, they used SimCLR on the ImageNet dataset. For the second step, they used a new model named multi-instance contrastive learning, designed for handling data from different projections (frontal and lateral) in chest X-ray datasets. Another study (Müller et al. 2022) showed that radiologist reports could enhance pretraining for down-

Fig. 3. Screening strategy, screening process, and result flow chart. SSL, self-supervised learning.



stream tasks on chest X-ray images. Moreover, a multi-modal RepsNet (Tanwani et al. 2022) model adapted pre-trained vision and language models to interpret medical images and automate report generation.

The COVID-19 pandemic also accelerated the progress of automated chest X-ray image analysis, as X-ray imaging is generally the first-line imaging test performed on symptomatic patients. A range of applications for this purpose have been proposed, such as (Gazda et al. 2021; Park et al. 2021; Sriram et al. 2021). The study (Park et al. 2021) places significant emphasis on the issue of overfitting. Overfitting is a common challenge when applying deep learning models in medical imaging, particularly in cases of rare or new diseases like COVID-19, where the available CXR dataset is limited compared to other conditions, impacting the generalization performance in real-world applications. Furthermore, the SALAD (Bozorgtabar et al. 2020) model is designed specifically to tackle the issue of overfitting in anomaly detection tasks.

3.2. Computed tomography (CT)

Computed tomography (CT) plays a significant role in diagnosing many diseases, resulting in approximately 300 million CT scans per year (Schöckel et al. 2020). Analyzing CT images is more demanding compared to X-ray images, due to the volume of imaging data to review, the multi-planar nature of CT, and improved soft tissue resolution (Zhuang et al. 2019; Pape et al. 2022; C.C.M. Professional). To enhance decision-making accuracy and efficiency, deep learning algorithms have been developed. However, due to limited annotated data and domain gaps, the SSL approach is preferred. Recent studies on the application of SSL in CT are listed in Table 2.

Various research studies have highlighted the need for a profound understanding of CT images when creating pretext tasks. For instance, applying 2D neural networks to 3D medical images resulted in the loss of crucial 3D anatomical information, leading to suboptimal performance (Zhou et al. 2019). To overcome the problem of the loss of crucial 3D anatomical information, some studies proposed 3D-based SSL algorithms for volumetric CT data, such as Rubik's cube recovery (Zhuang et al. 2019), models genesis (Zhou et al. 2019), semantic genesis (Haghighi et al. 2020), Rubik's cube++ (Tao et al. 2020), 3D context feature learning (Blendowski et al. 2019), and Vol2Flow (Bitarafan et al. 2022). Most of these pretext tasks have been validated on classification and segmentation downstream tasks, yielding promising results. Furthermore, certain pretext tasks specifically address other chal-

Table 1. The application of self-supervised learning to conventional X-ray

Model name/model architecture /*Code Link*	Downstream task	Performance	Dataset	Ref.
SALAD/auto-encoder/-	Anomaly detection	94.19% AUC (NIH), and 81.17% AUC (MURA)	NIH chest X-rays and MURA	Bozorgtabar et al. (2020)
C2l/momentum-based teacher-student architecture/https://github.com/funnyzhou/C2L_MICCAI2020	Classification	89.3% AUROC (CheXpert)	ChestX-ray14, MIMIC-CXR, CheXpert, and MURA	Zhou et al. (2020)
-/Modified models genesis: U-shaped convolutional neural network model combined with a convolution block attention module/-	Classification	98.6% averaged accuracy	CXR	Park et al. (2021)
MoCo-CXR/contrastive learning: MoCo/-	Classification	(CheXper) MoCo-CXR 0.813 AUC, ImageNet 0.775 AUC. (Shenzhen) MoCo-CXR 0.974 AUC	CheXper, Shenzhen	Sowrirajan et al. (2020)
-/Contrastive learning. Consists of three parts: a backbone neural network, a projection head, and a stochastic data augmentation module/-	Chest X-Ray classification	97.7% AUC	CheXper, ChestX-ray14, C19-Cohen, COVIDGR, Cell	Gazda et al. (2021)
MICLe/contrastive learning: adapted SimCLR/-	Chest X-ray classification	0.7689 $\pm$ 0.0010 AUC, improvement of 1.1% outperforming ImageNet pre-train	CheXpert	Azizi et al. (2021)
DVME/contrastive learning: SimCLR, SwAV, DINO; DVME- a model-agnostic technique to combine multiple self-supervised pretrained features for downstream tasks/-	Chest X-ray classification	DINO-0.6323 AUC (NIH Chest X-ray), SwAV - 0.5903 AUC. (APTOS), SwAV -0.6330 AUC, DVME 0.6566 AUC	NIH chest X-ray, Pneumonina chest X-ray; (APTOS, PatchCam)	Truong et al. (2021)
-/Contrastive learning-MoCo/https://github.com/facebookresearch/CovidPrognosis	Adverse event prediction from single images; Oxygen requirements prediction from single images, and adverse event prediction from multiple	0.786 AUC for predicting and 0.848 AUC of for predicting mortalities	MIMIC-CXR, CheXpert, NYU COVID	Sriram et al. (2021)
-/BYOL (Grill et al. 2020), SimCLR (Chen et al. 2020), PixelPro/-	Semantic segmentation and object detection tasks	The best methods on four tasks (BYOL on three and PixelPro on one task)	MIMIC-CXR 2	Müller et al. (2022)
MUSCLE/consists of two parts: Multi-Dataset Momentum Contrastive (Multi-Dataset MoCo) Learning, Multi-Task Continual Learning/-	Pneumonia classification, skeletal abnormality classification, lung segmentation, and tuberculosis detection	99.72% AUC (pneumonia classification), and 88.37% AUC (skeleton abnormality classification)	NIHCC, China-Set-CXR, Montgomery-Set-CXR, Indiana-CXR, RSNA Bone Age; Pneumonia, MURA, Chest Xray Masks, TBX	Liao et al. (2022)
RepsNet/encoder–decoder model. The encoder aligns the images with natural language descriptions via contrastive learning, while the decoder predicts answers from the encoded images/https://sites.google.com/view/repsnet	VQA-Rad (Medical visual question answering) and IU-Xray (report generation)	81.08% classification accuracy on VQA-Rad 2018 and 0.58 BLEU-1 score on IU-Xray	VQA-Rad, IU-XRay	Tanwani et al. (2022)

Note: Abbreviation:—model unnamed and/or code not available.

Table 2. The application of self-supervised learning to CT.

Model name/model architecture /*Code Link*	Downstream task	Performance	Dataset	Ref.
Rubik's cube recovery/involves two operators: cube rearrangement and cube rotation; and Siames—Octad network/-	(1) Brain hemorrhage classification and (2) brain tumor segmentation	(1) 83.8% accuracy (2) 76.2% mIoU using U-Net and 77.3% mIoU using 3D DUC	Brain Hemorrhage Dataset- internal dataset and BraTS-2018	Zhuang et al. (2019)
Models genesis/encoder–decoder/https://github.com/MrGiovanni/ModelsGenesis	Classification and segmentation (NCC, NCS, LCS, BMS)	NCC- 98.20 ± 0.5% (AUC) NCS- 77.62 ± 0.6% (IoU) LCS- 79.52 ± 4.7% (IoU) BMS- 90.60 ± 0.2% (IoU)	LIDC-IDRI, ChestX-ray8	Zhou et al. (2019)
-/REG2D and HEATMAP/-	Few-shot CT segmentation task	≈55% average Dice score (one-shot segmentation accuracy)	VISCERAL Anatomy3	Blendowski et al. (2019)
Semantic genesis/encoder–decoder, introduce two novel components: self-discovery and self-classification/https://github.com/JLiangLab/SemanticGenesis	Classification and segmentation (NCC, NCS, LCS, BMS)	NCC- 98.47 ± 0.2% (AUC) NCS- 77.24 ± 0.6% (IoU) LCS- 85.60 ± 1.9% (IoU) BMS- 68.80 ± 0.3% (IoU)	LUNA-2016, ChestX-ray14	Haghighi et al. (2020)
-/Alternative encoder–decoder CNN architectures: RS-AE, DE-UNET, RS-UNET/-	Reconstruct the skull defect removed during decompressive craniectomy performed after traumatic brain injury from post-operative CT images	The proposed direct estimation method based on the U-Net architecture (DE-UNet) outperforms all the other strategies	Dataset were provided by the University of Cambridge (Division of aesthesis, department of medicine)	Matzkin et al. (2020)
-/includes a spatial transformer stage that takes as input the NN-generated landmarks and compares the resulting aligned images/-	Automatically positioning and detecting landmarks	Phantom dataset-0.01% average registration loss 2D dataset-0.1% testing registration accuracy	Shepp-Logan (S-L) phantom (synthetic) dataset, 3D cranial CT scans of infants, 2D images of diatoms of four morphological classes	Bhalodia et al. (2020)
Rubik's cube++/Consists of two components: a generator and a discriminator/-	(1) Pancreas segmentation, (2) brain tissue segmentation	(1) 4-fold cross validation yielded 100%-84.08% DSC (2) 77.56% DSC	NIH Pancreas computed tomography, MRBrainS18	Tao et al. (2020)
-/Consists of 3D encoder, RE/SE module, augmentation module and 3D Siamese network/https://github.com/hongweilibran/imbalanced-SSL	Classification: discriminating high grade (H-grade) and low-grade tumor (L-grade), predicting lung cancer stages (i.e., I, II or III)	BraTS: Sensitivity/specificity 0.920/0.711 Lung cancer staging: Overall/Minor-class accuracy 0.538/0.372	a multi-center MRI dataset (BraTS), a lung CT dataset with 420 non-small cell lung cancer patients (NSCLC-radiomics)	Li et al. (2021*a*)
PCL/contrastive learning: U-Net encoder with projection head/https://github.com/dewenzeng/positional_cl	Multi organ segmentation	Achieved mean (standard deviation) on 5-fold cross-validation: CHD- 0.774(.03) and ACDC- 0.929(.00) for $M = 51$. M-number of patients used in the fine-tuning process	CHD, MMWHS, ACDC, and HVSMR	Zeng et al. (2021)
FCL/contrastive learning: MoCo (He et al. 2019)/-	Volumetric medical image segmentation	0.656 ± 0.052 DIC for $N = 1$, N-the number of annotated patients for fine-tuning on each client	ACDC MICCAI 2017 challenge dataset	Wu et al. (2021)
SAME/Pixel-level contrastive learning framework, breaks down image registration into three steps: affine transformation, coarse deformation, and deep deformable registration/-	(1) intra-phase registration and (2) cross-phase registration	(1) 54.42% Dice score, and (2) 50.96% Dice score	Internal dataset	Liu et al. (2021*b*)

Table 2. (*concluded*).

Model name/model architecture /*Code Link*	Downstream task	Performance	Dataset	Ref.
-/Variational autoencoders/-	Detect incorrect organ segmentations	0.92 AUC (kidney), 0.95 AUC (liver), 0.82 AUC (spleen)	Medical segmentation decathlon (liver and spleen), internal dataset for kidney segmentation, abdominal CT dataset from University of Wisconsin hospital and clinics	Sandfort et al. (2021)
MoCo-COVID-19/Contrastive learning-MoCo/-	Few-shot classification-automated diagnosis of COVID-19	0.931 ± 0.013 AUC	COVID-19 CT, dataset provided by the Italian Society of Medical and Interventional Radiology and preprocessed by MedSeg	Chen et al. (2021)
CMT-CNN/Multi-task contrastive learning	Classification COVID-19	CT (5.49%–6.45%) and X-ray (0.96%–2.42%)	Internal datasets (CT and X-ray)	Li et al. (2021*b*)
DeSD/Consists of online student network and a momentum teacher network/https://github.com/yeerwen/DeSD	Seven segmentation tasks (liver, kidney, hepaV, pancreas, colon, lung, spleen)	55.6% average Dice for 10% annotations and 75.8% average Dice for 100% annotations	DeepLesion; LiTS, KiTS, Hepatic Vessel (HepaV), Medical Segmentation Decathlon (MSD)	Ye et al. (2022)
Vol2Flow/Architecture is based on 3D-UNet structures (similar to VoxelMorph)/https://github.com/AdelehBitarafan/Vol2Flow	Multi-organ segmentation	82.20 ± 6.23 average DSC	C4KC-KiTS, CT-LN, CT-Pancrea, Sliver07, CHAOS, 3Dircadb-01, 3Dircadb-02	Bitarafan et al. (2022)
PrepNet/auto-encoder/-	COVID-19 classification	0.5343 cross-dataset average	SARS-CoV-2, UCSD COVID-CT, MosMed dataset	Amirian et al. (2022)
MAE/masked Autoencoders/-	CT abdomen multi-organ segmentation, magnetic resonance brain tumor segmentation, chest X-ray disease classification	83.5Avg DSC, 78.91 Avg DSC, 81.5% mAUC	BTCV, BRATS, CXR14	Truong Vu et al. (2021)
3DFPN-HS^2/consists of rotation module and rotation prediction network/-	Lung nodule detection	90.6% sensitivity at 1/8 false positive per scan on the LUNA16 dataset	LUNA16, SPIE-AAPM, LungTIME, and HMS	Liu et al. (2022)
-/Consists of a multi-modal keypoint detection module with attentive fusion for 2D patient joint localization, a self-supervised 3D mesh regression module/-	Automated isocentering with clinical CT scans	5.3/7.5/8.1 mm mean errors for abdomen/thorax/head respectively versus 13.2 mm median error of radiographers	SLP, internal RGBD data	Zheng et al. (2022)

Note: Abbreviation:—model unnamed and/or code not available; NCC- Nodule false positive reduction; NCS- Lung nodule segmentation; LCS- Liver segmentation; BMS- Brain tumor segmentation.

lenges in the medical image domain, such as Vol2Flow, which enhances generalizability by leveraging 3D medical images obtained from diverse acquisition conditions.

Recently, contrastive learning has gained prominence as a dominant SSL method due to its superior performance (Wang et al. 2023). To harness the potential of this method, a large number of algorithms have been developed specifically addressing the challenges of working with CT images, such as PCL (Zeng et al. 2021), FCL (Wu et al. 2021), SAME (Liu et al. 2021*b*), DeSD (Ye et al. 2022), MoCo-COVID-19 (Chen et al. 2021), and CMT-CNN (Li et al. 2021*b*). The challenge of creating contrastive learning algorithms for the medical domain is highlighted in the PCL publication. In this study, the authors focused on the issue of introducing numerous false negative pairs resulting from the utilization of state-of-the-art contrastive learning frameworks originally designed for the natural image domain. On the other side, the DeSD and FCL publications offer solutions for some of the challenges inherent to contrastive learning methods tailored specifically for the medical image domain. DeSD deals with the problem of weak supervision at the shallow layer, which has negative effects on downstream task performance. Conversely, FCL offers a new solution for a federated learning approach, which solves problems with limited amounts of medical data. It enables sharing image-level representation throughout different medical institutions while keeping source data private. Most automated CT image analysis systems use chest imaging datasets. Recently, many studies focused on using these systems to detect pneumonia or COVID-19, such as MoCo-COVID-19, CMT-CNN, and PrepNet (Amirian et al. 2022). Additionally, other SSL studies are being developed for specific applications such as decompressive craniectomy (Matzkin et al. 2020), statistical shape analysis (Bhalodia et al. 2020), radiomics (Li et al. 2021*a*), detection of incorrect organ segmentation (Sandfort et al. 2021), lung nodule detection (Liu et al. 2022), and 3D patient body modeling (Zheng et al. 2022).

3.3. Mammography

Mammography-based screening, alongside other medical imaging-based screening programs, plays a crucial role in preventive care systems. In the United States, mammography has significantly contributed to the consistent decline in breast cancer mortality rates over the past two decades (Ghesu et al. 2022). Furthermore, SSL algorithms have exhibited promising potential in the task of malignancy prediction, as evidenced by several notable studies listed in Table 3.

When developing SSL models for mammograms, the studies also need to consider the unique characteristics of mammograms. In a study focused on identifying cancer in mammography images (Truong Vu et al. 2021), the authors highlighted that traditional SSL tasks such as rotation prediction and colorization prediction may not be suitable for predicting the presence of cancer in mammography images (Truong Vu et al. 2021). This limitation arises from the fact that breast cancer tumors and other abnormalities in mammograms are typically small in scale, displayed in grayscale, and lack a defined orientation. As an alternative, the authors proposed solving a jigsaw puzzle task to generate features more suit-

Table 3. The application of self-supervised learning to mammography

Model name/model architecture/*Code Link*	Downstream task	Performance	Dataset	Ref.
MSVCL/Contrastive learning + CycleGAN/https://github.com/lizheren/MSVCL_MICCAI2021	Lesion detection	0.792 mAP (seen domain), 0.789 mAP (unseen domain)	Internal dataset of four vendors (GE, United Imaging Healthcare, Hologic, and Siemens); INbreast	Li et al. (2021*c*)
-/The jigsaw puzzle task/-	Classification	0.958 AUC (fine-tuned on a half or a quarter of the train set)	Internal datasets (the US, the UK)	Truong Vu et al. (2021)
-/Contrastive learning: lesion contrastive loss and normal contrastive loss/-	Classification	(Lesion type) 58.2 ±0.055% (Breast density) 78.1 ±0.029%	Internal dataset.	You et al. (2022)
-/SimCLR, BYOL, SWaV, ViT-MAEs/-	Classification	SWaV- 0.757 AUC (Linear Evaluation) 0.815 AUC (Finetuning)	CBIS-DDSM, CMMD	Miller et al. (2022)

Note: Abbreviation:—model unnamed and/or code not available.

Table 4. The application of self-supervised learning to dental X-rays

Model name/Model architecture	Downstream task	Performance	Dataset	Ref.
LCD-Net/Modified MoCoV2. LCD (the self-supervised pretrain stage and the two-branch network training stage	Segmentation and classification: diagnosis of tumors and cysts	LCD-Net 91.45% ACC (classification scores) LCD - 71.26 IoU (detection performance) LCD-Net 70.84 mIoU (segmentation performance)	Internal dataset	Hu et al. (2021)
-/SimCLR, BYOL, Barlow Twins	Dental caries classification	(Sensitivity) Barlow Twins -57.9%, SimCLR-57.2%, and BYOL-54.6%	Internal dataset	Taleb et al. (2022)

Note: Abbreviation:—model unnamed.

able for malignancy classification. Other studies attempted to use contrastive-based SSL algorithms for malignancy prediction (Li et al. 2021*c*; Miller et al. 2022; You et al. 2022). The study (Miller et al. 2022) identified the optimal image transformations for mammograms, including random crop with resizing to the original image size, gamma shift, contrast shift, and histogram equalization. The authors apply these augmentations to patches created by dividing the mammograms with a grid pattern, instead of the entire image, for better localization of the area of interest (e.g., lesion). Other studies (Li et al. 2021*c*; You et al. 2022) employ contrastive learning to address specific challenges in mammography datasets, such as domain gaps between datasets from different institutions, vendor domain gaps, false positive rates, and intra-class variations.

3.4. Dental X-ray

The phrase "*teeth are half of health*" portrays the significance of oral health to our overall health. Dentists commonly employ radiological procedures for diagnosing and treating dental problems. Deep learning has found extensive use in this field, while SSL holds the potential to offer even better performance. However, it is worth noting that SSL has not been widely applied in dental X-ray imaging, and key studies related to this are listed in Table 4.

The study (Taleb et al. 2022) proposed a contrastive learning-based approach (SimCLR, BYOL, and Barlow Twins) for the caries classification task. The notable advantage of this study lies in its label efficiency. Annotating the dataset for the SSL approach took approximately 71 working hours, whereas annotating the dataset for the supervised approach would have required over 7600 working hours (equivalent to 950 workdays). Additionally, scientists defined a suitable data augmentation strategy based on the nature of bitewing radiographs. This data augmentation strategy included: random resized cropping between 50% and 100% of input size, random horizontal flip with 50% probability, color adjustments (probabilities): brightness (20%) and contrast (10%), and saturation (10%), and random rotation angles between -20° and 20°. Contrastive learning methods have also been employed as pretraining models for the detection of jaw tumors and cysts. In a study (Hu et al. 2021), the authors proposed a method that addresses two main drawbacks of existing models: their heavy reliance on the number of lesion samples and the lack of reliability in diagnosis results. To overcome these drawbacks, they used many healthy samples for pretraining the model and designed a dual-branch network combined with the patch-covering data augmentation strategy with localization consistency loss.

3.5. Multi-modal applications

The multi-modal approach in medicine refers to the learning tasks that can process and integrate information from multiple types and sources of data, such as medical images and clinical reports, to provide a more comprehensive understanding of a patient's health condition. The literature suggests that implementing a multi-modal approach could improve the performance of automated medical image analysis (You et al. 2022). For these reasons, this holistic approach is widely applied in daily clinical practice nowadays. However, implementing a multi-modal approach in automated medical image analysis systems could be challenging due to privacy restrictions, high-class imbalance, and high annotation costs associated with medical data. SSL provides an opportunity to overcome these limitations, making the implementation of a multi-modal approach more feasible. Recent studies on multi-modal applications and SSL are listed in Table 5.

The selected multi-modal applications could be divided into two categories: the first category involves integrating medical image datasets from different medical imaging modalities, while the second category combines medical image datasets with text datasets. The most extensive integration of data from various medical imaging modalities is presented in the study (Ghesu et al. 2022). This study used a training dataset consisting of over 100 million medical images from X-ray, CT, magnetic resonance, and ultrasound modalities. The authors proposed an SSL method based on contrastive learning and online feature clustering. This method suggested an increase in accuracy and robustness to various image augmentations, as well as accelerated model convergence during training. Additionally, the results of the study demonstrated the effectiveness of the proposed method in improving the performance of different downstream tasks, such as classification/detection. In recent studies (Zhang et al. 2021; Zheng et al. 2021), it has been observed that while SSL has been extensively explored in classification and detection tasks, its application in medical segmentation tasks has been limited. To address these limitations, the authors proposed a new hierarchical SSL framework that learns task-agnostic knowledge from diverse medical image segmentation tasks using aggregated multi-domain datasets. Additionally, another study (Zhang et al. 2021) aimed to improve the accuracy of 3D tumor segmentation tasks using a SAR SSL method on CT and magnetic resonance datasets. This study

Table 5. The application of self-supervised learning to multi-modal applications

Model name/Model architecture/*Code Link*	Downstream task	Performance	Dataset	Ref.
-/Contrastive learning: the dual encoder framework/https://github.com/rwindsor1/biobank-self-supervised-alignment	Unsupervised rigid multi-modal scan registration, and cross-modal segmentation with opposite-modality annotations	0.927 Dice score	the UK Biobank (Sudlow et al. 2015)	Windsor et al. (2021)
HSSL/Contrastive learning. Consists of three hierarchical levels: image-level, task-level, group-level/-	Segmentations tasks: heart, prostate, spleen	heart-87.65% Dice score, prostate-68.58% Dice score, prostate-88.45% Dice score	Eight different data sources	Zheng et al. (2021)
SAR/Encoder–decoder architecture. Consists of: Multi-scale cube generator, transformation module, encoder, scale-aware module, decoder, and modality invariant adversarial learning module/-	Brain tumor segmentation, pancreas tumor segmentation	BraTS2018- 84.92% in average Dice. MSD- 75.68% average Dice (pancreas segmentation), 33.92% average Dice (tumor segmentation)	LUNA2016, LiTS2017, BraTS2018, MSD	Zhang et al. (2021)
CPRD/Consists of three specialized teacher models to focus on different body region respectively and then teach a student model to learn both intra- and inter-region features for Med-VQA/https://github.com/awenbocc/cprd	Med-VQA (Medical visual question answering)	SLAKE-EN accuracy 81.2% and 83.4% for "open-ended" and "closed-ended" questions respectively	SLAKE, SLAKE-EN, Medical Segmentation Decathlon.	Liu et al. (2021*a*)
M^{3}AE/Masked Autoencoders. Consists of different encoders and decoders for vision and language/https://github.com/zhjohnchan/M3AE	Medical visual question answering, medical image-text classification, medical image-caption retrieval	87.82% accuracy for closed-ended" questions on SLACK, 78.50% accuracy on MELINDA dataset, and the highest accuracy 66.65% on ROCO test set	ROCO, MedICaT, VQA-RAD, SLAKE, VQA-2019, MELINDA	Chen et al. (2022)
-/Online clustering—swapped prediction optimization, B. hybrid Self-Supervised—supervised lear/-	Abnormality detection in chest X-ray, brain metastases detection in magnetic resonance, brain hemorrhage detection in CT	Lesion detection- 0.94 AUC, brain metastasis detection-0.932 AUC. 85% acceleration of model convergence during training	Internal and public datasets	Ghesu et al. (2022)

Note: Abbreviation:—model unnamed and/or code not available.

is geared towards addressing challenges associated with tumor region segmentation, which arise from variations in scale, appearance, and geometric properties of the tumor regions.

In addition, an approach has been developed that utilizes SSL in medical imaging when two scan modalities are available for the same subject (Windsor et al. 2021). In this study, the authors proposed a multi-modal image-matching contrastive framework without any change, which utilizes two whole-body scans of magnetic resonance and dual-energy X-ray absorptiometry. This approach allows training a network to segment anatomical regions in magnetic resonance scans without the need for ground-truth magnetic resonance examples.

The second category of multi-modal applications involves medical vision-and-language models. These models aim to acquire generalized representations from extensive medical image-text datasets, which can be applied to various medical vision-and-language tasks such as medical visual question answering, medical image-text classification, and medical image-text retrieval. One example of such a model is presented in the study (Chen et al. 2022), which is a multi-modal masked autoencoder that learns cross-modal domain knowledge by reconstructing missing pixels and tokens from randomly masked images and texts.

3.6. Other application

Once X-ray images are acquired, they go through a series of steps, including preprocessing, interpretation, and archiving. SSL has found applications in all these phases, and recent studies have explored the potential of pretext tasks for improving preprocessing techniques and addressing the challenges of data archiving. The most notable studies are listed in Table 6.

The use of X-ray modalities in medical imaging comes with inherent risks due to the nature of X-rays, which could cause short-term and long-term negative effects on patients and clinical staff (Kim 2016). More recent contributions to reducing the negative effects of X-rays include the digitalization of image acquisition and the development of advanced preprocessing methods (Uffmann and Schaefer-Prokop 2009). One example is shown in the study by Zha et al. (2022), where the authors proposed a fast SSL solution for sparse-view cone beam computed tomography reconstruction. In this study, the reduction of radiation dose was achieved by decreasing the projection view in cone beam CT acquisition.

Additionally, the computational time required by the model is relatively short, ranging from 10 to 40 min, depending on the dataset used. Collectively, these results suggest making this model feasible for clinical CT applications. Another CT preprocessing application is proposed in the study (Roger et al. 2021) for performing interpolation of slices in CT. The PixelMiner model used a generative approach to accurately construct the texture and greatly improved the performance of downstream tasks. This SSL method has satisfactory robustness and generalizability because it performs well not only on the training dataset but also on externally validated datasets.

Reducing the radiation dose in medical imaging inevitably results in the creation of noise in medical images, thus necessitating the need for effective image-denoising methods. Biomedical image denoising is a challenging task because the noise corresponds directly to the signal strength and follows a Poisson distribution.

The study (Ta et al. 2022) proposed the Poisson2Sparse method for single-image denoising without ground-truth data. However, the need for image denoising is especially important for dynamic imaging methods because they use fast imaging techniques to improve temporal resolution, which can decrease the signal-to-noise ratio in each time frame. One possible application for dynamic medical imaging denoising was proposed in the study (Xu and Adalsteinsson 2021). This model, named Deformed2Self, combines single-image and multi-image denoising to improve image quality and uses a spatial transformer network to model motion between different slices.

In the previously described applications, an SSL approach was primarily considered as part of the preprocessing phase. One of the main challenges in the archiving phase is memory constraint, as medical images need to be archived for a specific number of years according to each country's regulations. For example, in the United States, medical images must be maintained for a period ranging from 5 to 10 years (Warren 2022). Traditional commercial compression algorithms may not be suitable for diagnostic radiology settings because they result in image quality losses that are unacceptable in a diagnostic radiology setting (Kwon et al. 2020). Among current imaging compression techniques, a deep learning-based autoencoder has become a promising option. Autoencoders consist of compression and decompression mechanisms. The autoencoders encode the medical image into a compressed representation and reconstruct the image by the decoder. Like other compression methods, this process causes information loss. Unfortunately, there is no clear guideline on an acceptable level of loss in a compression algorithm. Institutions such as the Food and Drug Administration only advise that the loss in compression "should be minimized as much as possible" (Warren 2022). Some studies, Warren (2022) and Barone identified the autoencoder-based methods, particularly those based on convolutional autoencoders, as being the most effective for medical image compression. These compression algorithms for medical images have shown promising results in reducing resource requirements for model training and image interpretation. For example, Joon et al. (Kwon et al. 2020) utilized a two-level vector quantized variational autoencoder framework for this purpose.

4. Main benefits and challenges

According to the reviewed literature, applying SSL in medical imaging offers a significant advantage by eliminating the need for large, labeled training datasets. This is particularly crucial when dealing with 3D volumetric medical data, as annotations for such data require considerable time from experienced clinicians (Tao et al. 2020). To address this challenge, previous studies have introduced the ImageNet pretraining strategy, which not only enhances accuracy but also expe-

Table 6. Self-supervised learning in applications for preprocessing and archive

Model name/Model architecture /*Code Link*	Task	Performance	Dataset	Ref.
-/Autoencoder/-	Compressing medical images	Convolutional autoencoder; the loss reached a plateau at $L = 0.0032$ convolutional autoencoder + PCA: the quality of the image, L ' 0.0036. Asymmetric autoencoder: if compared with CAE + PCA1024, the results are clearly worst even though the compression factor is the same	CRO Aviano: patch level dataset, full image dataset	Barone
Deformed2Self/Consists of three modules: a single-imaging denoising network; a spatial transformer network (STN); a multi-image denoising network/-	Dynamic imaging denoising	31.77% on PINCAT with 15% the standard deviation of Gaussian noise and 28.22% on ACDC with the same standard deviation of Gaussian noise	PINCAT, ACDC	Xu and Adalsteinsson (2021)
PixelMiner/Model based on PixelCNN++. PixelMiner includes many of the features of PixelCNN++, including vertically and horizontally stacked masked convolutions, gated convolutions, and the logistic mixture likelihood Loss/-	Slice interpolation of medical images	EPR 82% ($p < .01$), NRMSE of 0.11 ($p < .01$), (CCC) ≥ 0.85 ($p < .01$)	The radiological society of North America pulmonary embolism detection challenge ©	Roger et al. (2021)
NAF/Consists of four modules: ray sampling, position encoding, attenuation coefficient prediction, and projection synthesis/-	Sparse-view CBCT reconstruction (Cone Beam Computed Tomography)	PSNR/SSIM 33.05/.96 (chest) 34.14/.94 (jaw) 31.63/.94 (foot) 34.45/.95 (abdomen) 30.34/.88 (aorta)	LIDC-IDRI, Open scientific visualization, phantom internal dataset	Zha et al. (2022)
Poisson2Sparse/Sparsity and dictionary learning-based approach/https://github.com/tacalvin/Poisson2Sparse	Biomedical image denoising	Outperform existing state-of-the-art methods for SSL approaches under a variety of datasets and varying levels of Poisson noise, for example, by more ~2 dB PSNR in average performance	Florescent microscopy denoising dataset, PINCAT dataset	Ta et al. (2022)
FedFew/federated SSL/-	Multi-label classification	Accuracies 0.75 (edema), 0.73 (pneumonia), 0.77 (hernia)	ChestX-ray14	Dong et al. (2022)
-/Autoencoders/https://github.com/ciwarren/CCAE-MI	Image archiving	Compression ratio of 32:1, up to 96 percent SSIM, and an MSE of less than 0.003	Brain CT, brain magnetic resonance, COVID CT, chest X-ray	Warren (2022)

Note: Abbreviation:—model unnamed and/or code not available.

dites model convergence (Zhou et al. 2020). However, a notable limitation arises from the absence of an ImageNet medical image dataset for pretraining, resulting in a clear domain gap between natural images and medical images (Zhou et al. 2020). For instance, while ImageNet models are trained on relatively balanced datasets, they struggle to account for the inherent imbalanced nature commonly observed in medical data (Li et al. 2021*a*). This discrepancy presents a significant hurdle when attempting to transfer the learned representations from natural images to medical images, considering the substantial differences in their underlying distributions and characteristics. By introducing SSL approaches, this limitation is overcome as both the pretext task and downstream task are conducted on the same data domain, allowing for better alignment, and leveraging of the intrinsic properties of medical images. Additionally, studies have demonstrated that SSL learning can aid in the creation of a medical ImageNet by quickly generating initial rough annotations for unlabeled images, which can then be reviewed by experts (Zhou et al. 2019). This reduces the annotation efforts and accelerates the creation of a large, highly annotated medical ImageNet. Moreover, the benefits of SSL extend beyond the elimination of labeled datasets. They also contribute to the improvement of model robustness and generalizability, making them highly applicable in daily medical image analysis practice. For instance, in the study (Liu et al. 2022) the authors improved the robustness of the nodule detection across datasets collected by different vendors of CT scanners. Investigators also achieved the enhancement of model robustness and generalizability through the application of federated learning, in which SSL found its application.

In addition to the significant advantages over supervised learning approaches, it is worth noting that the creation of SSL methods presents a challenging task. In fact, SSL approaches rely on heuristics to design pretext tasks (Li et al. 2021*b*). This can be particularly challenging when applied to segmentation tasks (Zhuang et al. 2019). In segmentation, although the pre-trained weights can typically be adapted to the encoder part of the network, the decoder still requires random initialization. This can be problematic as the random initialization of the decoder may disrupt the pre-trained feature representation and counteract the improvements achieved through pre-training. Additionally, applying contrastive learning-based methods may not be practically feasible, especially in 3D datasets, as they require large batch sizes and/or negative pairs. This poses challenges for medical datasets because different images can have similar structures or organs, resulting in many false negative pairs (Zeng et al. 2021). For this reason, scientists often utilize Siamese networks to perform adaptation of existing models or propose novel contrastive learning networks. In the study (Li et al. 2021*b*), the authors even state that models in contrastive learning usually involve more parameters than their supervised counterparts. For instance, to achieve a comparable top-1 accuracy, the parameters of SimCLR are 16 times those of ResNet-50. Moreover, SSL faces similar challenges as other methods, such as the "black box" nature of CNN, and biases in datasets (Gazda et al. 2021).

5. Conclusion

This paper provided a comprehensive overview of SSL applications to X-ray modalities. The general pipeline of SSL involves a pretext/proxy task and a downstream task. The focus was given on pretext/proxy task, as it represents the actual point where SSL takes place. Various categorizations can be applied to SSL approaches, but we concentrated on generative and contrastive models. While contrastive models predominantly find their application in the interpretation phase of X-ray images, generative models take a lead role in the preprocessing and archiving phases. Studies demonstrated that formulating pretext tasks demands a profound understanding of the inherent nature of X-ray images. For instance, within contrastive models, a significant challenge lies in generating positive image pairs, leading to the development of specific data augmentation strategies tailored to the characteristics of X-ray images. Additionally, SSL has shown remarkable success in multi-modal scenarios, where integrating X-ray images with textual data or utilizing medical images from various modalities results in the formation of combined datasets. These datasets serve as a foundation for building SSL models that collectively aim to improve the performance of downstream tasks. Furthermore, SSL approaches could potentially mitigate limitations observed in supervised learning methods. The limitations include the requirement for access to extensive volumes of annotated data, issues related to data imbalances, considerations of robustness, and enhancing generalizability. From the reviewed literature, it is clear that the self-supervised learning approach has transformative potential by eliminating the current situation where many machine learning-based systems are developed but few are deployed in daily clinical practice.

Article information

Editor(s)
Parminder Raina, Samira Abbasgholizadeh Rahimi

History dates
Received: 7 October 2024
Accepted: 13 June 2025
Version of record online: 11 August 2025

Notes
This paper is part of a collection entitled "Digital health and machine learning integrated health care systems: why, when, and how?"

Data availability
All data supporting the findings of this review are included in this published article, as well as available in the cited pub-

lications and databases, such as PubMed, Scopus, and Web of Science. Specific details and access information are provided within the article.

Author information

Author ORCIDs

Ivan Martinović https://orcid.org/0000-0003-2723-0001

Author notes

Ervin Sejdić served as Guest Editor at the time of manuscript review and acceptance; peer review and editorial decisions regarding this manuscript were handled by Parminder Raina and Samira Rahimi.

Author contributions

Conceptualization: IM
Investigation: IM
Methodology: IM
Supervision: MĐ, ES
Writing – original draft: IM
Writing – review & editing: MS, MD, WL, EC, ES

Competing interests

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

Amirian, M., Montoya-Zegarra, J.A., Gruss, J., Stebler, Y.D., Bozkir, A.S., Calandri, M., et al. 2021. PrepNet: a convolutional auto-encoder to homogenize CT scans for cross-dataset medical image analysis. 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE. doi:10.1109/cisp-bmei53629.2021.9624344.

Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., et al. 2021. Big self-supervised models advance medical image classification [online]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE. doi:10.1109/ICCV48922.2021.00346.

Bachman, P., Hjelm, R.D., and Buchwalter, W. 2019. Learning representations by maximizing mutual information across views [online]. arXiv. Available from http://arxiv.org/abs/1906.00910 [accessed 27 April 2023].

Barone, F. Compressing medical images with minimal information loss. Master in High Performance Computing.

Bercovich, E., and Javitt, M.C. 2018. Medical Imaging: from Roentgen to the digital revolution, and beyond. Rambam Maimonides Medical Journal, **9**(4): e0034. doi:10.5041/RMMJ.10355. PMID: 30309440.

Bhalodia, R., Kavan, L., and Whitaker, R.T. 2020. Self-supervised discovery of anatomical shape landmarks. *In* Lecture notes in computer science. pp. 627–638. doi:10.1007/978-3-030-59719-1_61.

Bitarafan, A., Azampour, M.F., Bakhtari, K., Soleymani Baghshah, M., Keicher, M., and Navab, N. 2022. Vol2Flow: segment 3D volumes using a sequence of registration flows. *In* Computer science. pp. 609–618. doi:10.1007/978-3-031-16440-8_58.

Blendowski, M., Nickisch, H., and Heinrich, M.P. 2019. How to learn from unlabeled volume data: self-supervised 3D context feature learning. *In* Lecture notes in computer science. pp. 649–657. doi:10.1007/978-3-030-32226-7_72.

Bozorgtabar, B., Mahapatra, D., Vray, G., and Thiran, J.-P. 2020. SALAD: self-supervised aggregation learning for anomaly detection on X-rays. *In* Lecture notes in computer science. pp. 468–478. doi:10.1007/978-3-030-59710-8_46.

Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G., and Murphy, K. 2021. Deep learning for chest X-ray analysis: a survey. Medical Image Analysis, **72**: 102125. doi:10.1016/j.media.2021.102125.

Caron, M., Bojanowski, P., Joulin, A., and Douze, M. 2019. Deep clustering for unsupervised learning of visual features [online]. arXiv. Available from http://arxiv.org/abs/1807.05520 [accessed 27 April 2023].

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., and Joulin, A. 2021. Unsupervised learning of visual features by contrasting cluster assignments [online]. arXiv. Availablefrom http://arxiv.org/abs/2006.09882 [accessed 27 April 2023].

C.C.M. Professional. CT (Computed Tomography) scan. Cleveland Clinic. Available from https://my.clevelandclinic.org/health/diagnostics/4808-ct-computed-tomography-scan#:~:text=It%20usually%20takes%20about%2024,report%20that%20explains%20the%20findings [accessed 5 April 2023].

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. arXiv. doi:10.48550/arxic.2002.05709.

Chen, X., and He, K. 2020. Exploring simple siamese representation learning [online]. arXiv. Available from http://arxiv.org/abs/2011.10566 [accessed 27 April 2023].

Chen, X., Yao, L., Zhou, T., Dong, J., and Zhang, Y. 2021. Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. Pattern Recognition, **113**: 107826. doi:10.1016/j.patcog.2021.107826.

Chen, Z., Du, Y., Hu, J., Liu, Y., Li, G., Wan, X., and Chang, T.-H. 2022. Multi-modal masked autoencoders for medical vision-and-language pre-training. *In* Lecture notes in computer science. pp. 679–689. doi:10.1007/978-3-031-16443-9_65.

Crummy, A.B., Strother, C.M., and Mistretta, C.A. 2018. The history of digital subtraction angiography. Journal of Vascular and Interventional Radiology, **29**(8): 1138–1141. doi:10.1016/j.jvir.2018.03.030. PMID: 30055783.

Dehghan, S., and Amasyali, M.F. 2023. SelfCCL: curriculum contrastive learning by transferring self-taught knowledge for fine-tuning BERT. Applied Sciences, **13**(3): 1913. doi:10.3390/app13031913.

Dong, N., Kampffmeyer, M., and Voiculescu, I. 2022. Learning underrepresented classes from decentralized partially labeled medical images. *In* Lecture notes in computer science. pp. 67–76. doi:10.1007/978-3-031-16452-1_7.

Ehrhardt, J., and Wilms, M. 2022. Autoencoders and variational autoencoders in medical image analysis. *In* Biomedical image synthesis and aimulation. Elsevier. pp. 129–162. doi:10.1016/B978-0-12-824349-7.00015-3.

Gazda, M., Plavka, J., Gazda, J., and Drotar, P. 2021. Self-supervised deep convolutional neural network for chest X-ray classification. IEEE Access, **9**: 151972–151982. doi:10.1109/ACCESS.2021.3125324.

Ghesu, F.C., Georgescu, B., Mansoor, A., Yoo, Y., Neumann, D., Patel, P., et al. 2022. Self-supervised learning from 100 million medical images [online]. arXiv. Available from http://arxiv.org/abs/2201.01283 [accessed 16 July 2022].

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P.H., and Buchatskaya, E. 2020. Bootstrap your own latent: a new approach to self-supervised learning. arXiv. Available from https://arxiv.org/abs/2006.07733.

Haghighi, F., Hosseinzadeh Taher, M.R., Zhou, Z., Gotway, M.B., and Liang, J. 2020. Learning semantics-enriched representation via self-discovery, self-classification, and self-restoration. *In* Lecture notes in computer science. pp. 137–147. doi:10.1007/978-3-030-59710-8_14.

He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. 2019. Momentum contrast for unsupervised visual representation learning. arXiv (Cornell University). doi:10.48550/arxiv.1911.05722.

Hjelm, R.D., Fedorov, A, Lavoie-Marchildon, S, Grewal, K, Bachman, P, Trischler, A, and Bengio, Y 2019. Learning deep representations by mutual information estimation and maximization [online]. arXiv. Available from http://arxiv.org/abs/1808.06670 [accessed 27 April 2023].

Howell, J.D. 2011. Coronary heart disease and heart attacks, 1912–2010. Medical History, **55**(3): 307–312. doi:10.1017/S0025727300005317. PMID: 21792252.

Hu, J., Feng, Z., Mao, Y., Lei, J., Yu, D., and Song, M. 2021. A location constrained dual-branch network for reliable diagnosis of jaw tu-

mors and cysts. *In* Lecture notes in computer science. pp. 723–732. doi:10.1007/978-3-030-87234-2_68.

Huu, M.-K.N., Ngo, V.Q., Nguyen, T.-D., Nguyen, V.-T., and Ngo, T.D. 2022. Antique photo restoration and colorization via generative model. *In* 2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), IEEE, Phu Quoc, Vietnam. pp. 1–6. doi:10.1109/MAPR56351.2022.9924704.

Imai, S., Akahane, M., Konishi, Y., and Imamura, T. 2018. Benefits of computed tomography in reducing mortality in emergency medicine. Open Medicine, **13**(1): 394–401. doi:10.1515/med-2018-0058. PMID: 30234160.

Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., and Makedon, F. 2020. A survey on contrastive self-supervised learning. Technologies, **9**(1): 2. doi:10.3390/technologies9010002.

Kim, D.I. 2016. PREFACE: how dangerous are X-ray studies that we undertake every day?. Journal of Korean Medical Science, **31**: S2. doi:10.3346/jkms.2016.31.S1.S2.

Kumar, P., Rawat, P., and Chauhan, S. 2022. Contrastive self-supervised learning: review, progress, challenges and future research directions. International Journal of Multimedia Information Retrieval, **11**(4): 461–488. doi:10.1007/s13735-022-00245-6.

Kwon, Y.J. (Fred), Toussie, D., Reina, G.A., Tang, P.T.P., Doshi, A.H., and Oermann, E.K. 2020. Autoencoder image compression algorithm for reduction of resource requirements [online]. *In* Presented at the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada. p. 5. Available from http://www.cse.cuhk.edu.hk/~qdou/public/medneurips2020/27_27_camera_ready-compressed.pdf [accessed 27 April 2023].

Leibig, C., Brehmer, M., Bunk, S., Byng, D., Pinker, K., and Umutlu, L. 2022. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. The Lancet Digital Health, **4**(7): e507–e519. doi:10.1016/S2589-7500(22)00070-X.

Le-Khac, P.H., Healy, G., and Smeaton, A.F. 2020. Contrastive representation learning: a framework and review. IEEE Access, **8**: 193907–193934. doi:10.1109/ACCESS.2020.3031549.

Liao, W., Xiong, H., Wang, Q., Mo, Y., Li, X., Liu, Y. et al., 2022. MUSCLE: multi-task self-supervised continual learning to pre-train deep models for X-ray images of multiple body parts. *In* Lecture notes in computer science. pp. 151–161. doi:10.1007/978-3-031-16452-1_15.

Li, H., Xue, F.F., Chaitanya, K., Luo, S., Ezhov, I., Zhang, J. et al., 2021a. Imbalance-aware self-supervised learning for 3D radiomic representations. *In* Lecture notes in computer science. pp. 36–46. doi:10.1007/978-3-030-87196-3_4.

Li, J., Zhao, G., Tao, Y., Zhai, P., Chen, H., He, H., and Cai, T. 2021b. Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19. Pattern Recognition, **114**: 107848. doi:10.1016/j.patcog.2021.107848.

Li, Z., Cui, Z., Wang, S., Qi, Y., Ouyang, X., Chen, Q. et al., 2021c. Domain generalization for mammography detection via multi-style and multi-view contrastive learning. *In* Lecture notes in computer science. pp. 98–108. doi:10.1007/978-3-030-87234-2_10.

Liu, B., Zhan, L.-M., and Wu, X.-M. 2021a. Contrastive pre-training and representation distillation for medical visual question answering based on radiology images. *In* Lecture notes in computer science. pp. 210–220. doi:10.1007/978-3-030-87196-3_20.

Liu, F., Yan, K., Harrison, A.P., Guo, D., Lu, L., Yuille, A.L. et al., 2021b. SAME: deformable image registration based on self-supervised anatomical embeddings. *In* Lecture notes in computer science. pp. 87–97. doi:10.1007/978-3-030-87202-1_9.

Liu, J., Cao, L., Akin, O., and Tian, Y. 2022. Robust and accurate pulmonary nodule detection with self-supervised feature learning on domain adaptation. Frontiers in Radiology, **2**. doi:10.3389/fradi.2022.1041518.

Liu, X., Zhang, F., Hou, Z., Mian, Li, Wang, Z., Zhang, J., and Tang, J. 2021c. Self-supervised learning: generative or contrastive. IEEE Transactions on Knowledge and Data Engineering, 1. doi:10.1109/TKDE.2021.3090866.

Lu, Z. 2022. Brief introduction to contrastive learning pretext tasks for visual representation [online]. arXiv. Availablefrom http://arxiv.org/abs/2210.03163 [accessed 27 April 2023].

Masci, J., Meier, U., Cireşan, D., and Schmidhuber, J. 2011. Stacked convolutional auto-encoders for hierarchical feature extraction. *In* Lecture notes in computer science. pp. 52–59. doi:10.1007/978-3-642-21735-7_7.

Matzkin, F., Newcombe, V., Stevenson, S., Khetani, A., Newman, T., Digby, R. et al., 2020. Self-supervised skull reconstruction in brain CT images with decompressive craniectomy. *In* Lecture notes in computer science. pp. 390–399. doi:10.1007/978-3-030-59713-9_38.

Mckinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., et al. 2020. International evaluation of an AI system for breast cancer screening. Nature, **577**(7788): 89–94. doi:10.1038/s41586-019-1799-6.

Miller, J.D., Arasu, V.A., Pu, A.X., Margolies, L.R., Sieh, W., and Shen, L. 2022. Self-supervised deep learning to enhance breast cancer detection on screening mammography [online]. arXiv. Available from http://arxiv.org/abs/2203.08812 [accessed 6 April 2023].

Müller, P., Kaissis, G., Zou, C., and Rueckert, D. 2022. Radiological reports improve pre-training for localized imaging tasks on chest X-rays. *In* Lecture notes in computer science. pp. 647–657. doi:10.1007/978-3-031-16443-9_62.

Nagoor, O.H., Whittle, J., Deng, J., Mora, B., and Jones, M.W. 2022. Sampling strategies for learning-based 3D medical image compression. Machine Learning with Applications, **8**: 100273. doi:10.1016/j.mlwa.2022.100273.

Navarro, F., Watanabe, C., Shit, S., Sekuboyina, A., Peeken, J.C., Combs, S.E., and Menze, B.H. 2021. Evaluating the robustness of self-supervised learning in medical imaging [online]. arXiv. Available from http://arxiv.org/abs/2105.06986 [accessed 26 April 2023].

Newell, A., and Deng, J. 2020. How useful is self-supervised pretraining for visual tasks? *In* 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Seattle, WA, USA. pp. 7343–7352. doi:10.1109/CVPR42600.2020.00737.

Pape, P., Jensen, A.H., Bergdal, O., Munch, T.N., Rudolph, S.S., and Rasmussen, L.S. 2022. Time to CT scan for patients with acute severe neurological symptoms: a quality assurance study. Scientific Reports, **12**(1): 15269. doi:10.1038/s41598-022-19512-x.

Park, J., Kwak, I.-Y., and Lim, C. 2021. A deep learning model with self-supervised learning and attention mechanism for COVID-19 diagnosis using chest X-ray images. Electronics, **10**(16): 1996. doi:10.3390/electronics10161996.

Pierre, K., Haneberg, A.G., Kwak, S., Peters, K.R., Hochhegger, B., Sananmuang, T., et al. 2023. Applications of artificial intelligence in the radiology roundtrip: process streamlining, workflow optimization, and beyond. Seminars in Roentgenology, **58**(2): 158–169. doi:10.1053/j.ro.2023.02.003.

Rajpurkar, P., Irvin, J, Zhu, K, Yang, B, Mehta, H, Duan, T, et al. 2017. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning [online]. arXiv. Available from http://arxiv.org/abs/1711.05225 [accessed 26 April 2023].

Ramani, G.B., and Scalise, N.R. 2020. It's more than just fun and games: play-based mathematics activities for head start families. Early Childhood Research Quarterly, **50**: 78–89. doi:10.1016/j.ecresq.2018.07.011.

Reed, S., van den Oord, A., Kalchbrenner, N., Colmenarejo, S.G., Wang, Z., Belov, D., and de Freitas, N. 2017. Parallel multiscale autoregressive density estimation [online]. arXiv. Availablefrom http://arxiv.org/abs/1703.03664 [accessed 26 April 2023].

Roger, W., Lambin, P., Keek, S., Beuque, M., Primakov, S., Lavrova, E., et al. 2021. Towards texture accurate slice interpolation of medical images using PixelMiner. Research Square (Research Square), doi:10.21203/rs.3.rs-586453/v1.

Salimans, T., Karpathy, A., Chen, X., and Kingma, D.P. 2017. PixelCNN++: improving the PixelCNN with discretized logistic mixture likelihood and other modifications [online]. arXiv. Availablefrom http://arxiv.org/abs/1701.05517 [accessed 26 April 2023].

Sandfort, V., Yan, K., Graffy, P.M., Pickhardt, P.J., and Summers, R.M. 2021. Use of variational autoencoders with unsupervised learning to detect incorrect organ segmentations at CT. Radiology: Artificial Intelligence, **3**(4): e200218. doi:10.1148/ryai.2021200218.

Schöckel, L., Jost, G., Seidensticker, P., Lengsfeld, P., Palkowitsch, P., and Pietsch, H. 2020. Developments in X-ray contrast media and the potential impact on computed tomography. Investigative Radiology, **55**(9): 592–597. doi:10.1097/RLI.0000000000000696.

Singh, A., and Ogunfunmi, T. 2021. An overview of variational autoencoders for source separation, finance, and bio-signal applications. Entropy, **24**(1): 55. doi:10.3390/e24010055.

Shurrab, S., and Duwairi, R. 2022. Self-supervised learning methods and applications in medical imaging analysis: a survey. PeerJ Computer Science, **8**: e1045. doi:10.7717/peerj-cs.1045.

Sowrirajan, H., Yang, J., Ng, A.Y., and Rajpurkar, P. 2020. MoCo-CXR: MoCo pretraining improves representation and transferability of chest X-ray models. arXiv. doi:10.48550/arxiv.2010.05352.

Sriram, A., Muckley, M., Sinha, K., Shamout, F., Pineau, J., Geras, K.J., et al. 2021. COVID-19 prognosis via self-supervised representation learning and multi-image prediction [online]. arXiv. Available from http://arxiv.org/abs/2101.04909 [accessed 27 April 2023].

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. 2015. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Medicine, **12**(3): e1001779. doi:10.1371/journal.pmed.1001779.

Ta, C.-K., Aich, A., Gupta, A., and Roy-Chowdhury, A.K. 2022. Poisson2Sparse: self-supervised poisson denoising from a single image. *In* Lecture Notes in Computer Science. pp. 557–567. doi:10.1007/978-3-031-16452-1_53.

Taleb, A., Rohrer, C., Bergner, B., De Leon, G., Rodrigues, J.A., Schwendicke, F., et al. 2022. Self-supervised learning methods for label-efficient dental caries classification. Diagnostics, **12**(5): 1237. doi:10.3390/diagnostics12051237.

Tanwani, A.K., Barral, J., and Freedman, D., 2022. RepsNet: combining vision with language for automated medical reports. *In* Lecture notes in computer science. pp. 714–724. doi:10.1007/978-3-031-16443-9_68.

Tao, X., Li, Y., Zhou, W., Ma, K., and Zheng, Y. 2020. Revisiting Rubik's Cube: self-supervised learning with volume-wise transformation for 3D medical image segmentation. *In* Lecture notes in computer science. pp. 238–248. doi:10.1007/978-3-030-59719-1_24.

The National Lung Screening Trial Research Team. 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine, **365**(5): 395–409. doi:10.1056/NEJMoa1102873.

Tian, Y., Krishnan, D., and Isola, P. 2020a. Contrastive multiview coding [online]. arXiv. Available from http://arxiv.org/abs/1906.05849 [accessed 27 April 2023].

Tian, Y., Sun, C., Poole, B., Krishnan, D., Schmid, C., and Isola, P. 2020b. What makes for good views for contrastive learning? [online]. arXiv. Available from http://arxiv.org/abs/2005.10243 [accessed 27 April 2023].

Tomasev, N., Bica, I., McWilliams, B., Buesing, L.H., Pascanu, R., Blundell, C., and Mitrovic, J. 2022. Pushing the limits of self-supervised ResNets: can we outperform supervised learning without labels on ImageNet?[online]. arXiv. Available from http://arxiv.org/abs/2201.05119 [accessed 26 April 2023].

Truong, T., Mohammadi, S., and Lenga, M. 2021. How transferable are self-supervised features in medical image classification tasks? [online]. arXiv. Available from http://arxiv.org/abs/2108.10048 [accessed 5 April 2023].

Truong Vu, Y.N., Tsue, T., Su, J., and Singh, S. 2021. An improved mammography malignancy model with self-supervised learning. *In* Medical imaging 2021: computer-aided diagnosis. p. 29. doi:10.1117/12.2582318.

Uffmann, M., and Schaefer-Prokop, C. 2009. Digital radiography: the balance between image quality and required radiation dose. European Journal of Radiology, **72**(2): 202–208. doi:10.1016/j.ejrad.2009.05.060.

van den Oord, A., Kalchbrenner, N., and Kavukcuoglu, K. 2016. Pixel recurrent neural networks [online]. arXiv. Available from http://arxiv.org/abs/1601.06759 [accessed 26 April 2023].

Wake, R., Yoshiyama, M., Iida, H., Takeshita, H., Kusuyama, T., Kanamitsu, H., et al. 2011. History of coronary angiography. *In* Advances in the diagnosis of coronary atherosclerosis. *Edited by* S. Kirac. InTech. doi:10.5772/22578.

Wang, W.-C., Ahn, E., Feng, D., and Kim, J. 2023. A review of predictive and contrastive self-supervised learning for medical images [online]. arXiv. Available from http://arxiv.org/abs/2302.05043 [accessed 11 July 2023].

Warren, C. 2022. Analysis of compressive convolutional autoencoders for image archiving in medical informatics [online]. Master's theses. Michigan Technological University, United States. Available from https://www.mtu.edu/cs/graduate/dissertations/pdfs/an-analysis-of-compressive-convolutional-autoencoders-for-image.pdf.

Windsor, R., Jamaludin, A., Kadir, T., and Zisserman, A. 2021. Self-supervised multi-modal alignment for whole body medical imaging. *In* Lecture notes in computer science. pp. 90–101. doi:10.1007/978-3-030-87196-3_9.

Wu, Y., Zeng, D., Wang, Z., Shi, Y., and Hu, J., 2021. Federated contrastive learning for volumetric medical image segmentation. *In* Lecture notes in computer science. pp. 367–377. doi:10.1007/978-3-030-87199-4_35.

Xu, J., and Adalsteinsson, E. 2021. Deformed2Self: self-supervised denoising for dynamic medical imaging. *In* Lecture notes in computer science. pp. 25–35. doi:10.1007/978-3-030-87196-3_3.

Ye, Y., Zhang, J., Chen, Z., and Xia, Y. 2022. DeSD: self-supervised learning with deep self-distillation for 3D medical image segmentation. *In* Lecture notes in computer science. pp. 545–555. doi:10.1007/978-3-031-16440-8_52.

You, K., Lee, S., Jo, K., Park, E., Kooi, T., and Nam, H. 2022. Intra-class contrastive learning improves computer aided diagnosis of breast cancer in mammography. *In* Lecture notes in computer science. pp. 55–64. doi:10.1007/978-3-031-16437-8_6.

Zeng, D., Wu, Y., Hu, X., Xu, X., Yuan, H., Huang, M. et al., 2021. Positional contrastive learning for volumetric medical image segmentation. *In* Lecture notes in computer science. pp. 221–230. doi:10.1007/978-3-030-87196-3_21.

Zha, R., Zhang, Y., and Li, H. 2022. NAF: neural attenuation fields for sparse-view CBCT reconstruction. *In* Lecture notes in computer science. pp. 442–452. doi:10.1007/978-3-031-16446-0_42.

Zhang, X., Feng, S., Zhou, Y., Zhang, Y., and Wang, Y. 2021. SAR: scale-aware restoration learning for 3D tumor segmentation. *In* Lecture notes in computer science. pp. 124–133. doi:10.1007/978-3-030-87196-3_12.

Zhou, H.-Y., Yu, S., Bian, C., Hu, Y., Ma, K., and Zheng, Y. 2020. Comparing to learn: surpassing ImageNet pretraining on radiographs by comparing image representations. *In* Lecture notes in computer science. pp. 398–407. doi:10.1007/978-3-030-59710-8_39.

Zhou, S.K, Greenspan, H., Davatzikos, C., Duncan, J.S., Van Ginneken, B., Madabhushi, A., et al. 2021. A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, **109**(5): 820–838. doi:10.1109/JPROC.2021.3054390.

Zhou, Z., Sodha, V., Siddiquee, M.M.R., Feng, R., Tajbakhsh, N., Gotway, M.B., and Liang, J. 2019. Models Genesis: generic autodidactic models for 3D medical image analysis. *In* Lecture notes in computer science. pp. 384–393. doi:10.1007/978-3-030-32251-9_42.

Zhuang, X., Li, Y., Hu, Y., Ma, K., Yang, Y., and Zheng, Y. 2019. Self-supervised feature learning for 3D medical images by playing a Rubik's Cube. *In* Lecture notes in computer science. pp. 420–428. doi:10.1007/978-3-030-32251-9_46.

Zheng, H., Han, J., Wang, H., Yang, L., Zhao, Z., Wang, C., Chen, D.Z. et al., 2021. Hierarchical self-supervised learning for medical image segmentation based on multi-domain data aggregation. *In* Lecture notes in computer science. pp. 622–632. doi:10.1007/978-3-030-87193-2_59.

Zheng, M., Planche, B., Gong, X., Yang, F., Chen, T., and Wu, Z. 2022. Self-supervised 3D patient modeling with multi-modal attentive fusion. *In* Lecture notes in computer science. pp. 115–125. doi:10.1007/978-3-031-16449-1_12.