

Review

# Consequential Advancements of Self-Supervised Learning (SSL) in Deep Learning Contexts

Mohammed Majid Abdulrazzaq<sup>1,2</sup>, Nehad T. A. Ramaha<sup>1</sup>, Alaa Ali Hameed<sup>3</sup>, Mohammad Salman<sup>4</sup>, Dong Keon Yon<sup>5</sup>, Norma Latif Fitriyani<sup>6,\*</sup>, Muhammad Syafrudin<sup>6,\*</sup> and Seung Won Lee<sup>7,\*</sup>

<sup>1</sup> Department of Computer Engineering, Demir Celik Campus, Karabuk University, 78050 Karabuk, Turkey; 2038166008@ogrenci.karabuk.edu.tr or moh.majid@uoanbar.edu.iq (M.M.A.); nehadramaha@karabuk.edu.tr (N.T.A.R.)

<sup>2</sup> Computer Center, University of Anbar, Anbar 31001, Iraq

<sup>3</sup> Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Istinye University, 34396 Istanbul, Turkey

<sup>4</sup> College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait

<sup>5</sup> Center for Digital Health, Medical Science Research Institute, Kyung Hee University Medical Center, Kyung Hee University, Seoul 02453, Republic of Korea

<sup>6</sup> Department of Artificial Intelligence and Data Science, Sejong University, Seoul 05006, Republic of Korea

<sup>7</sup> Department of Precision Medicine, School of Medicine, Sungkyunkwan University, Suwon 16419, Republic of Korea

\* Correspondence: norma@sejong.ac.kr (N.L.F.); udin@sejong.ac.kr (M.S.); swleemd@g.skku.edu (S.W.L.)

**Abstract:** Self-supervised learning (SSL) is a potential deep learning (DL) technique that uses massive volumes of unlabeled data to train neural networks. SSL techniques have evolved in response to the poor classification performance of conventional and even modern machine learning (ML) and DL models of enormous unlabeled data produced periodically in different disciplines. However, the literature does not fully address SSL's practicalities and workabilities necessary for industrial engineering and medicine. Accordingly, this thorough review is administered to identify these prominent possibilities for prediction, focusing on industrial and medical fields. This extensive survey, with its pivotal outcomes, could support industrial engineers and medical personnel in efficiently predicting machinery faults and patients' ailments without referring to traditional numerical models that require massive computational budgets, time, storage, and effort for data annotation. Additionally, the review's numerous addressed ideas could encourage industry and healthcare actors to take SSL principles into an agile application to achieve precise maintenance prognostics and illness diagnosis with remarkable levels of accuracy and feasibility, simulating functional human thinking and cognition without compromising prediction efficacy.

**Keywords:** deep learning (DL); self-supervised learning (SSL); machine learning (ML); cognition; classification; data annotation

**MSC:** 68T07; 68T05; 93E35



**Citation:** Abdulrazzaq, M.M.; Ramaha, N.T.A.; Hameed, A.A.; Salman, M.; Yon, D.K.; Fitriyani, N.L.; Syafrudin, M.; Lee, S.W. Consequential Advancements of Self-Supervised Learning (SSL) in Deep Learning Contexts. *Mathematics* **2024**, *12*, 758. <https://doi.org/10.3390/math12050758>

Academic Editor: Daniel-Ioan Curiac

Received: 11 January 2024

Revised: 23 February 2024

Accepted: 26 February 2024

Published: 3 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Concepts of AI, convolutional neural networks (CNNs), DL, and ML considered in the last few decades have contributed to multiple valuable impacts and core values to different scientific disciplines and real-life areas because of their amended potency in executing high-efficiency classification tasks of variant complex mathematical problems and difficult-to-handle subjects. However, some of them are more rigorous than others. More specifically, DL, CNN, and artificial neural networks (ANNs) have a more robust capability than conventional ML and AI models in making visual, voice, or textual data classifications [1].

The crucial rationale for these feasible models includes their significant classification potential in circumstances involving therapeutic diagnosis, maintenance, and production line prognostics. As these two processes formulate a prominent activity in medicine and engineering, the adoption of ML and DL models could contribute to numerous advantages and productivity records [2–6].

Unfortunately, documented historical information may not provide relevant recognition solutions for ML and DL, especially for new industry failure situations and recent manufacturing and production fault conditions since the characteristics and patterns of lately-reported problems do not approach past observed datasets. More classification complexity, in this respect, would increase [7].

A second concern pertaining to the practical classification tasks of ML and DL is their fundamental necessity for clear annotated data that can accelerate this procedure, offering escalated accuracy and performance scales [8].

Thirdly, in most data annotation actions, the latter may contribute to huge retardation, labor efforts, and expenses to be completely attained, particularly when real-life branches of science handle big data [9].

Resultantly, a few significant classification indices and metrics may be affected, namely accuracy, efficacy, feasibility, reliability, and robustness [10].

Relying on what had been explained, SSL was innovated with the help of extensive research and development (R&D), aiming to overcome these three obstacles at once. Thanks to researchers who addressed some beneficial principles in SSL to conduct flexible analysis of different data classification modes, such as categorizing nonlinear relationships, unstructured and structured data, sequential data, and missing data.

Technically speaking, SSL is considered a practical tactic for learning deep representations of features and crucial relationships from the existing data through efficient augmentations. Without time-consuming annotations of previous data, SSL models can generate a distinct training objective (considering pretext processes) that relies solely on unannotated data. To boost performance in additional classification activities, the features produced by SSL techniques should have a specific set of characteristics. The representations should be distinguished with respect to the downstream tasks while being sufficiently generic to be utilized on untrained actions [11].

The emergence of the SSL concept has resulted in core practicalities and workable profitabilities correlated with functional information prediction for diverse disciplines that have no prior annotated documented databases, contributing to preferable outcomes in favor of cost-effectiveness, time efficiency, computational effort flexibility, and satisfying precision [12].

Taking into account the hourly, daily, and weekly creation of massive data in approximately each life and science domain, this aspect could pose variant arduous challenges in carrying out proper identification of data, especially when more information is accumulated after a long period of time [13].

Within this framework of background, the motivation for exploring essential SSL practicalities arises from the increasing need to leverage vast amounts of unlabeled data to improve classification performance. Accordingly, the major goal of this article is to enable industrial engineering researchers and medical scientists to better understand major SSL significances and realize comprehensively their pivotal workabilities to allow active involvement of SSL in their work for conducting efficient predictions in diagnoses and prognostics.

To provide more beneficial insights on SSL incorporation into industry and medicine, a thorough review is carried out. It is hoped from this overview that its findings could actually clarify some of SSL's substantial rationale and innovatory influences to handle appropriate maintenance checks and periodic machine prognostics to make sure production progress and industrial processes are operating safely within accepted measures.

On the other hand, it is actually essential to emphasize the importance of this paper in elucidating a collection of multiple practicalities of SSL to support doctors and clinical

therapists in identifying the type of problem in visual data and, thus, following suitable treatment. This clinical action can sometimes be challenging, even for professionals. As a result, other approaches might be implemented, like costly consultations, which are not feasible.

Correspondingly, the organization of this paper is arranged based on the following sequence:

- Section 2 is prepared to outline the principal research method adopted to identify the dominant advantages of SSL algorithms in accomplishing efficient classification tasks without the identification of essential datasets crucial for training and testing procedures to maximize the model's classification effectiveness.
- Section 3 is structured to explain the extensive review's prominent findings and influential characteristics that allow SSL paradigms to accomplish different classification tasks, offering elevated scales of robustness and efficacy.
- Section 4 illustrates further breakthroughs and state of the art that have been lately implemented through several research investigations and numerical simulations to foster the SSL algorithms' categorization productivity and feasibility.
- Section 5 provides noteworthy illustrations and discussion pertaining to the evaluation of SSL serviceable applications and other crucial aspects for classifying and recognizing unlabeled data.
- Section 6 expresses the main research conclusions.
- Section 7 points out the imperative areas of future work that can be considered by other investigators to provide further modifications and enhancements to the current SSL models.
- Section 8 expresses the critical research limitations encountered in the review implementation until it is completed.

Totally, the paper's contribution is reflected in the following points:

1. Cutting much time, effort, and cost connected with essential data annotation for conventional DL and ML models adopted to support medical therapists in diagnosing the type of problem in visual databases,
2. Achieving the same relevance for industrial engineers, who wish to make machine prognostics as necessary periodic maintenance actions robustly,
3. Performing precise predictions of different problems in medicine, industry, or other important disciplines, where new behaviors of data do not follow previously noted trends, helps predict new data patterns flexibly and reliably in real-life situations.

## 2. Materials and Methods

### 2.1. Data Collection Approach

This study considers specific research steps, shown in Figure 1, to accomplish the primary research objective. The data collection process implemented in this article comprises secondary information collection, which relies on addressing beneficial ideas and constructive findings from numerous peer-reviewed papers and recent academic publications, examining variant benefits and many relevances of SSL in recognizing unspecified data, and bringing remarkable rates of workability, accuracy, reliability, and effectiveness.

### 2.2. The Database Selection Criteria

To upgrade the review outcomes' robustness, this work establishes a research foundation based on certain criteria, depicted in Figure 1, through which some aspects are taken into consideration, including the following:

- The multiple research publications analyzed and surveyed are more modern than in 2016. Thus, the latest results and state-of-the-art advantages can be extracted.
- The core focus of the inspected articles in this thorough overview is linked to SSL's significance in industry and medicine when involved in periodic machinery prognostics and clinical diagnosis, respectively.

- After completing the analysis of SSL’s relevant merits from the available literature, critical appraisal is applied, referring to some expert estimations and peer reviewer opinions to validate and verify the reliability and robustness of the paper’s overall findings.

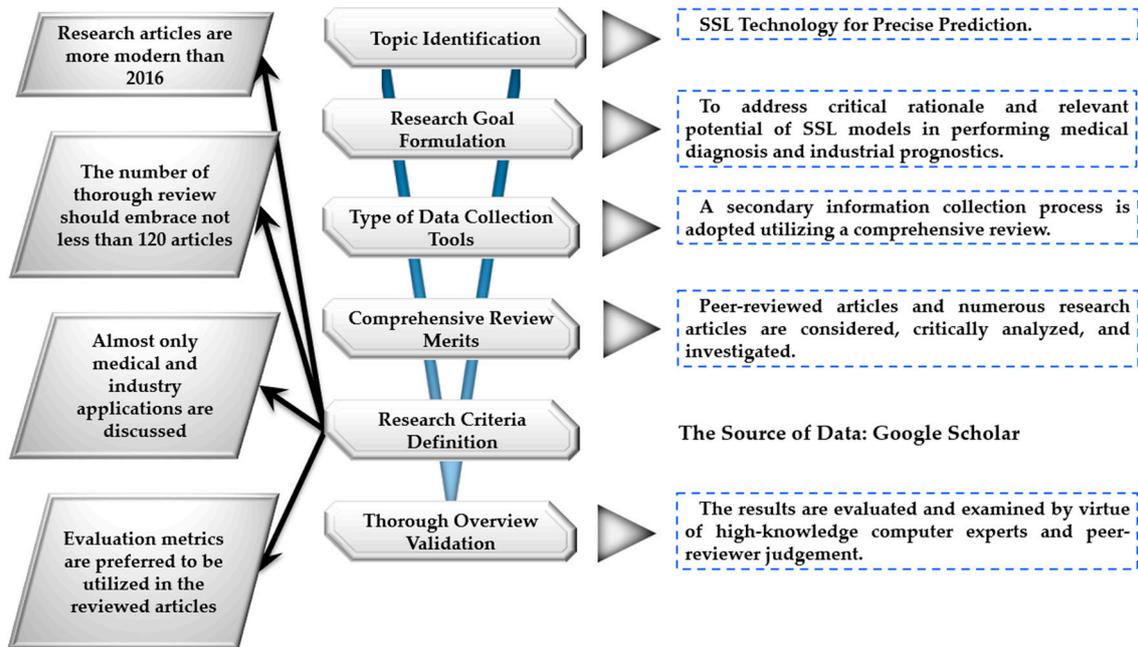


Figure 1. A flowchart of the comprehensive overview adopted in this paper.

### 3. Related Work

In this section, more explanation concerning the critical characteristics of SSL paradigms and their corresponding benefits and applications is provided, referring to existing databases from the global literature, which comprises recent academic publications and peer-reviewed papers.

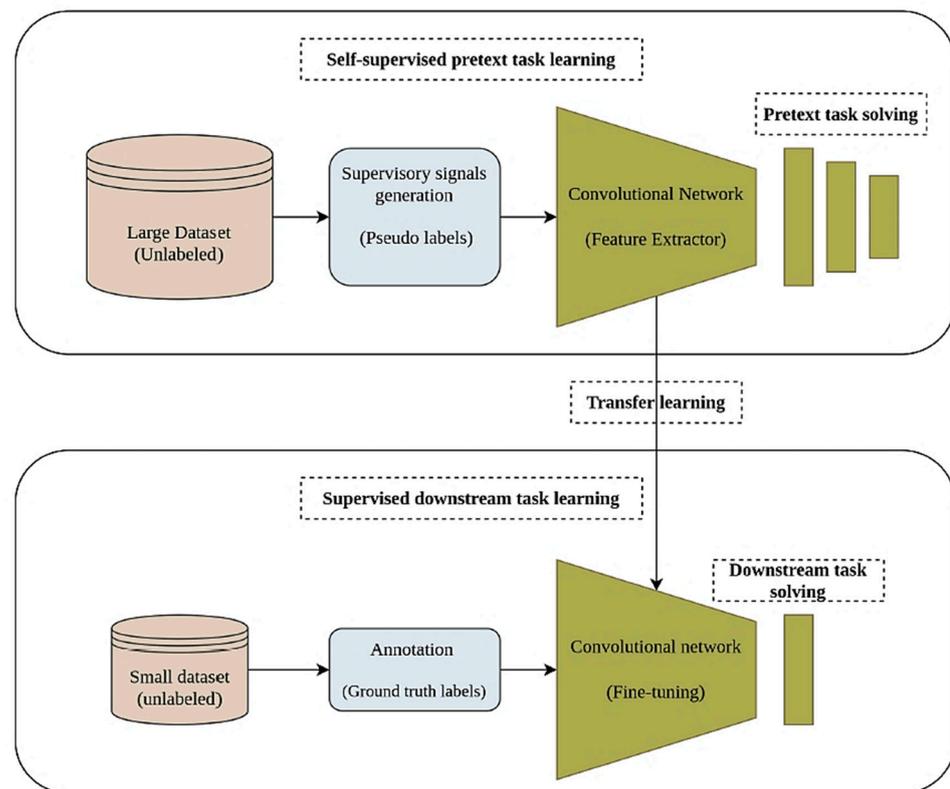
More illustration is offered on these aspects in the following sub-sections.

#### 3.1. Major Characteristics and Essential Workabilities of SSL

As illustrated above, supervised learning (SL) needs annotated data to train numerical ML models to enable an efficient classification process in various conventional ML models. On the contrary, unsupervised learning (USL) classification procedures do not require labeled data to accomplish a similar classification task. Rather, USL algorithms can rely solely on identifying meaningful patterns in existing unlabeled data without necessary training, testing, or preparation [14].

For the previously illustrated industrial and medical pragmatic practices, SSL can often be referred to as predictive learning (or pretext learning) (PxL). Labels can be generated automatically, transforming the unsupervised problem into a flexible, supervised one that can be solved viably.

Another favorable solution of SSL algorithms is their efficient categorization of data correlated with natural language processing (NLP). SSL can allow researchers to fill in blanks in databases when they are not fully complete or have a high-quality definition. As an illustration, with the application of ML and DL models, existing video data can be utilized to reconstruct previous and future videos. However, without relying on the annotation procedure, SSL takes advantage of patterns linked to the current video data to efficiently complete the categorization procedure of a massive video database [15,16]. Correspondingly, the critical working principles of the SSL approach can be illustrated in the workflow shown in Figure 2.



**Figure 2.** The major workflow related to SSL [17].

From Figure 2, during the pre-training stage (pretext task solving), feature extraction is carried out by pseudo-labels to enable an efficient prediction process. After that, transfer learning is implemented to initiate the SSL phase, in which a small dataset is considered to make data annotations (of ground truth labels). Then, fine-tuning is performed to achieve the necessary prediction task.

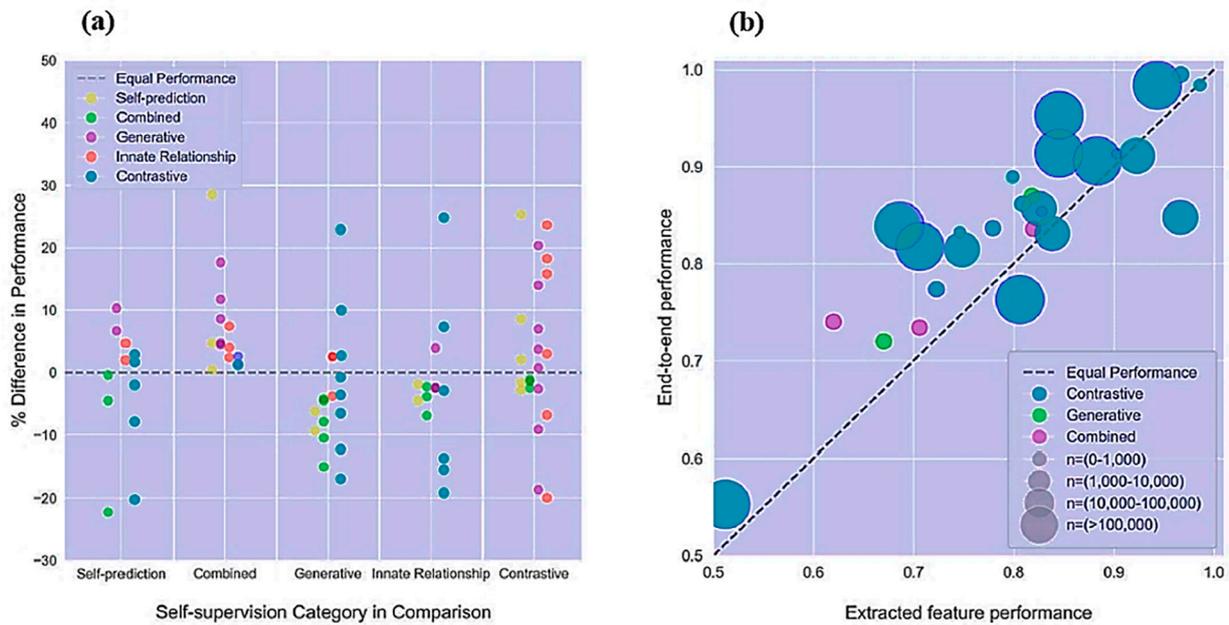
### 3.2. Main SSL Categories

Because it can be laborious to compile an extensively annotated dataset for a given prediction task, USL strategies have been proposed as a means of learning appropriate image identification without human guidance [18,19]. Simultaneously, SSL is an efficient approach through which a training objective can be produced from the data. Theoretically, a deep neural network (DNN) is trained on pretext tasks, in which labels are automatically produced without human annotation. The learned representations can be utilized to complete the pretext tasks. Familiar SSL sorts involve: (A) generative, (B) predictive, (C) contrastive, and (D) non-contrastive models. The multiple contrastive and noncontrastive tactics illustrated in this paper can be recognized as joint-embedded strategies.

However, more types of SSL are considered in some contexts. For example, a graphical illustration in [20] was created, explaining the performance rates that can be achieved when SSL is applied, focusing mainly on further SSL categories, as shown in Figure 3.

It can be realized from the graphical data expressed in Figure 3a that the variation in the performance between the self-prediction, combined, generative, innate, and contrastive SSL types fluctuates mostly between 10% and 10%. In Figure 3b, it can be noticed that end-to-end performance corresponding to contrastive, generative, and combined SSL algorithms varies between nearly 0.7 and 1.0, relating to an extracted feature performance that ranges approximately between 0.7 and 1.0.

In the following sections, more explanation is provided for some of these SSL categories.



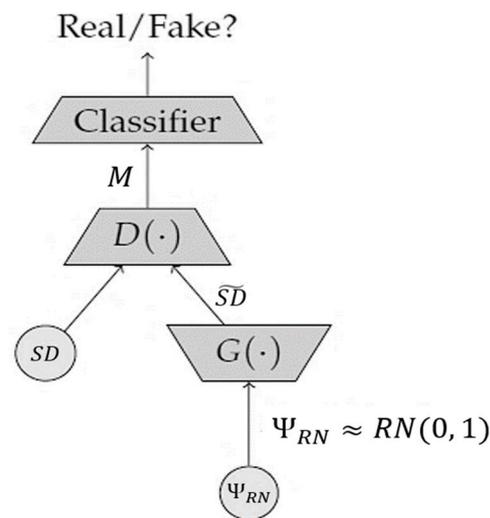
**Figure 3.** Two profiles of (a) some types of SSL classification with their performance level and (b) end-to-end performance of extracted features and the corresponding number of each type [20].

### 3.2.1. Generative SSL Models

Using an autoencoder to recreate an input image following compression is a common pretext operation. Relying on the first component of the network, called an encoder, the model should learn to compress all pertinent data from the image into a latent space with reduced dimensions to minimize the reconstruction loss. The image is then reconstructed by the latent space of a second network component called a decoder.

Researchers in [18,19,21–25] reported that denoising autoencoders could also provide reliable and stable identifications of images by learning to filter out noise. The network cannot learn the identity function owing to extra noise. By encoding the distribution parameters of a latent space, variational autoencoders (VAE) can advance the autoencoder model [26–29]. Both the reconstruction of error and extra factor, the Kull-Leibler divergence between an established latent distribution (often a unit-centered Gaussian distribution), and the encoder output are minimized during training. The samples from the resulting distribution can be obtained through this regularization of the latent space. To rebuild entire patches with only 25 percent of the visible patches, scholars in [30,31] have recently adopted vision transformers to create large masked autoencoders that work at the patch level rather than pixel-wise. Adding a class token to a sequence of patches or performing global mean pooling on all the patch tokens, as in this reconstruction challenge, yields reliable image representations.

A generative adversarial network (GAN) is another fundamental generative USL paradigm that has been extensively studied [32–34]. This architecture and its variants aim to mimic real data’s appearance and behavior by generating new data from random noise. To train a GAN, two networks compete in an adversarial minimax game, with one learning to turn the rate of random noise,  $\Psi_{RN} \approx RN(0, 1)$  into synthetic data,  $\tilde{SD}$ , which attempts to mimic the distribution of the original data. These aspects can be illustrated in Figure 4.



**Figure 4.** The architecture employed in GAN. Adopted from Ref. [35], used under Creative Commons CC-BY license.

In the adversarial method, a second network, which can be termed discriminator  $D(\cdot)$  was trained to distinguish between generated and authentic images from the original dataset. When the discriminator is certain that the input image is from the true data distribution, it reports a score of 1, whereas for the images produced by the generator, the score is zero. One possible estimation of this adversarial objective function,  $F_{AO}$ , can be accomplished by the following mathematical formula:

$$F_{AO} = \min_G \max_D \frac{1}{N} \sum_{i=1}^N \log(1 - D(G(\Psi_{RN_i}))) + \frac{1}{M} \sum_{i=1}^M \log(D(x_i)), \tag{1}$$

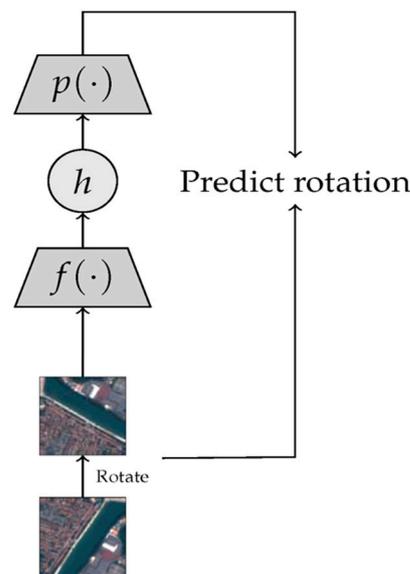
where:

$\Psi_{RN} \approx RN(0, 1)$ —A group of random noise vectors with an overall amount of  $N$   
 $SD \approx Q_{Data}$ —A dataset comprising a set of real images having a total number of  $M$ .

### 3.2.2. Predictive SSL Paradigms

Models trained to estimate the impact of artificial change on the input image express the second type of SSL technique. This strategy is inspired by understanding the semantic items and regions inside an image, which can be essential for accurately predicting the transformation. Scholars in [36] conducted analytical research to improve the performance of the model against random initialization and to approach the effectiveness obtained from the initialization with ImageNet pre-trained weights in benchmark computer vision datasets by pre-training a paradigm to predict the relative positions of two image patches.

Some researchers have confirmed the advantages of colored images [37]. In this method, the input image is first changed to grayscale. Next, a trained autoencoder converts the grayscale image back to its original color form by minimizing the average squared error between the reconstructed and original images. The encoder feature representations are considered in the subsequent downstream processes. The numerical RotNet approach [38] is another well-known predictive SSL approach, which represents a practical training process for mathematical schemes to help predict the rotation that is randomly implemented in the input image, as shown in Figure 5.



**Figure 5.** Flowchart configuration of the operating principles related to the intelligent RotNet algorithm relying on the SSL approach for accurate prediction results. From Ref. [35], used under Creative Commons CC-BY license.

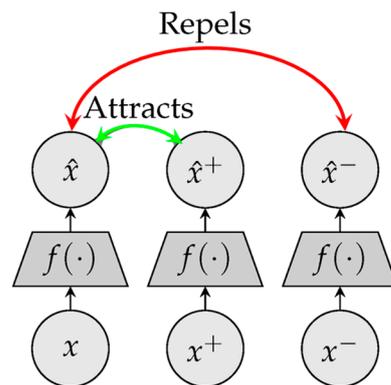
To improve the performance of the model in a dynamic rotation prediction task, the relevant characteristics that classify the semantic content of the image should first be extracted. Researchers in [39] considered a jigsaw puzzle to forecast the relative positions of the picture partitions using the shuffled SSL model. The Exemplar CNN was also addressed and trained in [40] to predict augmentations that can be applied to images by considering a wide variety of augmentation types. Cropping, rotation, color jittering, and contrast adjustment are examples linked to the enhancement classes gained by the Exemplar CNN model.

An SSL model can learn rich representations of the visual content by completing one of these tasks. However, the network may not be able to perform effectively on all subsequent tasks contingent on the pretext task and dataset. Because the orientation of objects is not as rigorously practical to handle in remote sensing datasets as in object-centric datasets, the prediction of random rotations of an image would not perform particularly well on such a dataset [41].

### 3.2.3. Contrastive SSL Paradigms

Forcing the features of various perspectives in a picture to be comparable is another strategy that can result in accurate representations. The resulting representations are independent of the particular enhancements needed to generate various image perspectives. However, the network can be converged to a stable representation that meets the invariance condition but is unrelated to the input image.

One typical approach to achieving this goal via the acquisition of various representations while avoiding the collapse problem is the contrastive loss. This type of loss function can be utilized to train the model to distinguish between views of the same image (positive) and views of distinct images (negative). Correspondingly, it seeks to obtain homogeneous feature representations for pairs with positive values while isolating features for negative pairs. The triplet loss investigated by researchers in [42] is the simplest form of this family. It requires a model to be trained such that the distance between the representations of a given anchor and its positive rates is smaller than the distance between the representations of the anchor and the random negative, as illustrated in Figure 6.



**Figure 6.** The architecture of the triplet loss function. From Ref. [35], used under Creative Commons CC-BY license.

In Figure 6, the triplet loss function is considered helpful in learning discriminative representations by learning an encoder that is able to detect the difference between negative and positive samples. Under this setting, the triplet Loss Function,  $\mathfrak{F}_{LossTriplet}$ , can be estimated using the following relationship:

$$\mathfrak{F}_{LossTriplet} = \max(\|f(x) - f(x^+)\| - \|f(x) - f(x^-)\| + m, 0), \tag{2}$$

where:

- $x^+$ —The positive vector value of the anchor  $x$
- $x^-$ —The negative vector value of the anchor  $x$
- $f(\cdot)$ —The embedding function
- $m$ —The value of the margin parameter.

In [43], the researchers examined the SimCLR method, which is one of the most well-known SSL strategies. It formulates a type of contrastive representational learning. Two versions of each training batch image can be generated using random-sampling augmentation. After these modified images are fed into the representational method, a prediction network can be utilized to map the representation onto a hypersphere of dimension,  $D$ .

The overall mathematical algorithm is trained to elevate the cosine similarity across the representation parameter,  $z$ , and its corresponding positive counterpart,  $z^+$  (belonging to the same original visual data) and to minimize the similarity between  $z$  and all other representations in the batch  $z^-$ , contributing to the following expression:

$$\mathbb{L}_F(z, z^+, z^-) = -\log \left( \frac{\exp(\langle z, z^+ \rangle / \tau)}{\sum_{z' \in z^- \cup \{z^+\}} \exp(\langle z, z' \rangle / \tau)} \right), \tag{3}$$

where:

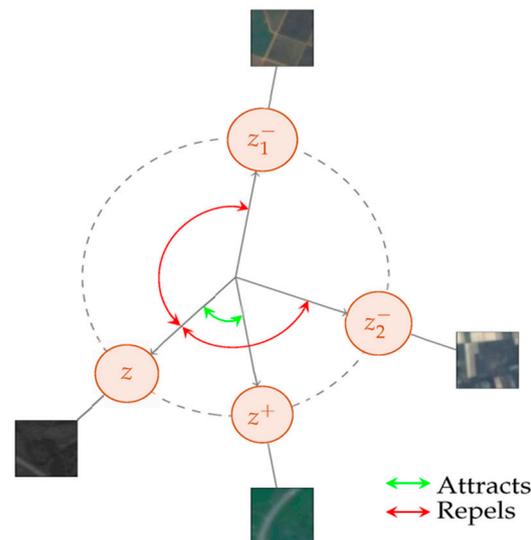
- $\langle z, z^+ \rangle$ —the dot product between  $z$  and  $z^+$ .
- $\tau$ —the temperature variable to scale the levels of similarity, distribution, and sharpness.
- $f(\cdot)$ — the embedding function.

At the same time, the algebraic correlation connected with the evaluation process of the complete loss function that assesses the cross-entropy of temperature, which can be dominated as normalized temperature cross-entropy, which is denoted by  $N\Theta - \mathfrak{X}S$ , is depicted in the following relation:

$$\mathbb{L}(N\Theta - \mathfrak{X}S) = \frac{1}{2N} \sum_{z, z^+, z^-} \mathbb{L}_F(z, z^+, z^-), \tag{4}$$

where  $N$  indicates the number of items in the dataset, such as images and textual characters.

Figure 7 shows that the NT-Xent loss [44] acts solely on the direction of the features confined to the  $D$ -dimensional hypersphere because the representations are normalized before calculating the function loss value.



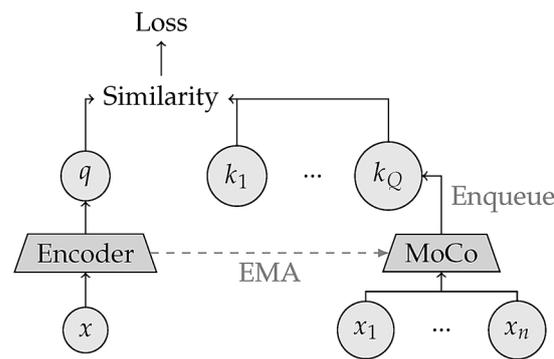
**Figure 7.** Description of contrastive loss linked to the 2-dimensional unit sphere with two negative parameters ( $z_1^-$  and  $z_2^-$ ) and one positive ( $z^+$ ) sample from the EuroSAT dataset. From Ref. [35], used under Creative Commons CC-BY license.

By maximizing the mutual data between the two perspectives, this loss ensures that the resulting representations are both style-neutral and content-specific.

In addition to SimCLR, they suggested the momentum contrast (MoCo) technique, which uses a reduced number of batches to calculate the contrastive loss while maintaining the same functional number of negative samples [45]. It employs an exponentially moving average (EMA)-updated momentum encoder whose values are updated by the main encoder's weights and a sample queue to increase the number of negative samples in each batch, as shown in Figure 8. To account for the newest positives, the oldest negatives from the previous batch were excluded. Other techniques, such as swapping assignments across views (SwAVs), correlate views to consistent clusters between positive pairs by clustering representations into a shared set of prototypes [44,46–48]. The entropy-regularized optimal transport strategy is also used in the same context to move representations between clusters in a manner that prevents them from collapsing into one another [46,49–53]. Finally, the cross-entropy between the optimal tasks in one branch and the anticipated distribution in the other is minimized by the loss. To feed sufficient negative samples to the loss function and avoid representations from collapsing, contrastive approaches often need large batch sizes.

As shown in Figure 8, at each step of the numerical analysis, only the major encoder amounts are updated based on the backpropagation process. The similarity aspects between the queue and encoded batch samples were then employed for contrastive loss.

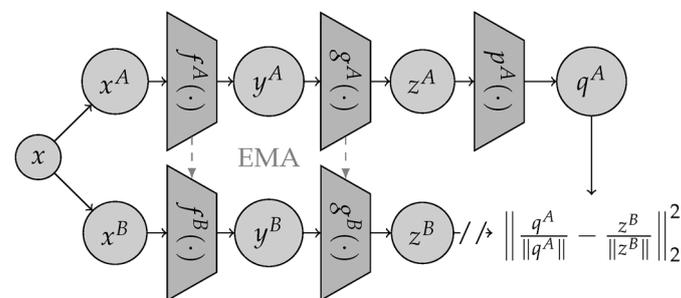
Compared with traditional prediction methods, joint-embedding approaches tend to generate broader representations. Nonetheless, their effectiveness in downstream activities may vary depending on the augmentation utilized. If a model consistently returns the same representations for differently cropped versions of the same image, it can effectively remove any spatial information about the image and will likely perform poorly in tasks such as semantic segmentation and object detection, which rely on this spatial information. Dense contrastive learning (DCL) has been proposed and considered by various researchers to address this issue [54–57]. Rather than utilizing contrastive loss on the entire image, it was applied to individual patches. This permitted the contrastive model to acquire representations that are less prone to spatial shifts.



**Figure 8.** An illustration of Q samples developed utilizing a momentum encoder whose amounts are modified. From Ref. [35], used under Creative Commons CC-BY license.

### 3.2.4. Non-Contrastive SSL Models

To train self-supervised models, alternative methods within joint-embedded learning frameworks can prevent the loss of contrastive elements. They classified these as approaches that do not rely on contrast. Bootstrap Your Own Latent (BYOL) is a system based on mentor-apprentice pairing [58–60]. The student network in a teacher-student setup is taught to mimic the teacher network’s output (or characteristics). This method is frequently utilized in knowledge distillation when the instructor and student models possess distinct architectures (e.g., when the student model is substantially smaller than the teacher model) [61]. The weights of the instructor network in BYOL are defined as the EMA of the student network weights. Two projector networks,  $g^A$  and  $g^B$ , are utilized after the encoders,  $f^A$  and  $f^B$ , to calculate the training loss. Subsequently, to extract representations at the image level, they retrain only the student encoder  $f^A$ . Additional asymmetry is introduced between the two branches by a predictor network superimposed on the student projector, as shown in Figure 9.



**Figure 9.** Architecture of the non-contrastive BYOL method, considering student A and lecturer B pathways to encode the dataset. From Ref. [35], utilized under Creative Commons CC-BY license.

In Figure 9, the teacher’s values are modified and updated by the EMA technique applied to the student amounts. The online branch is also supported by an additional network,  $p^A$ , which is known as the predictor [60].

SimSiam employs a pair of mirror-image networks and a predictor network at the end of a node [62–64]. The loss function employs an asymmetric stop gradient to optimize the pairwise alignments between positive pairs because the two branches have identical weights. Relying on a student-teacher transformer design known as self-distillation, DINO (self-distillation with no labels) defines the instructor as an EMA of the weights in the student network [65]. Next, the teacher network’s centered and sharpened outputs are utilized to train the student network to make exact predictions for a given positive pair.

Another non-contrastive learning model, known as the Barlow Twins, can be offered according to the information bottleneck theory, which eliminates the need for individual amounts for each branch of the teacher-student model considered in BYOL and

SimSiam [66,67]. This technique enhances the mutual information between two perspectives by boosting the cross-correlation of the matching characteristics provided by two identical networks and eliminating superfluous information in these representations. The Barlow twin loss function was evaluated by the following equation:

$$\mathcal{L}_{Barlow.Twins} = \sum_{i=1}^N (1 - C_{ii}^2) + \lambda \sum_{i=1}^N \sum_{j \neq i}^N C_{ij}^2 \quad (5)$$

where  $C$  is the cross-correlation matrix calculated by the following formula:

$$C_{ij} = \frac{\sum_b z_{bi}^A z_{bj}^B}{\sqrt{\sum_b (z_{bi}^A)^2} \sqrt{\sum_b (z_{bj}^B)^2}} \quad (6)$$

where  $z^A$  and  $z^B$  express the corresponding outcomes related to the two identical networks provided by the two views of a particular photograph.

Variance, invariance, and covariance regularization (VICReg) approaches have been recently proposed to enhance this framework [68–71]. In addition to invariance, which implicitly maximizes alignments between positive pairs, the loss terms are independent for every branch, unlike in low twins. Using distinct regularization for each pathway, this method allows for noncontrastive multimodal pre-training between text and photo pairs.

Most of these techniques train a linear classifier on the priority of representations as the primary performance metric. Researchers in [70] analyzed the beneficial impacts of ImageNet, whereas scholars in [69,72] examined CIFAR's advantages, which help accomplish an active analysis of object-centric visual datasets commonly addressed for the pre-training and linear probing phases of DL. Therefore, these techniques may not apply to image classification.

Scholars are invited to examine devoted review articles for further contributory information and essential fundamentals pertaining to SSL types [68,73].

### 3.3. Practical Applications of SSL Models

Before introducing the common applications and vital utilizations of SSL models to handle efficacious data classification and identification processes, their critical benefits should be identified as a whole. The commonly-addressed benefits and vital advantages of SSL techniques can be expressed as follows [74,75]:

- I. Minimizing the massive cost connected with data labeling phases is essential to facilitating a high-quality classification/prediction process.
- II. Alleviating the corresponding time needed to classify/recognize vital information in a dataset,
- III. Optimizing the data preparation lifecycle is typically a lengthy procedure in various ML models. It relies on filtering, cleaning, reviewing, annotating, and reconstructing processes through training phases.
- IV. Enhancing the effectiveness of AI models. SSL paradigms can be recognized as functional tools that allow flexible involvement in innovative human thinking and machine cognition.

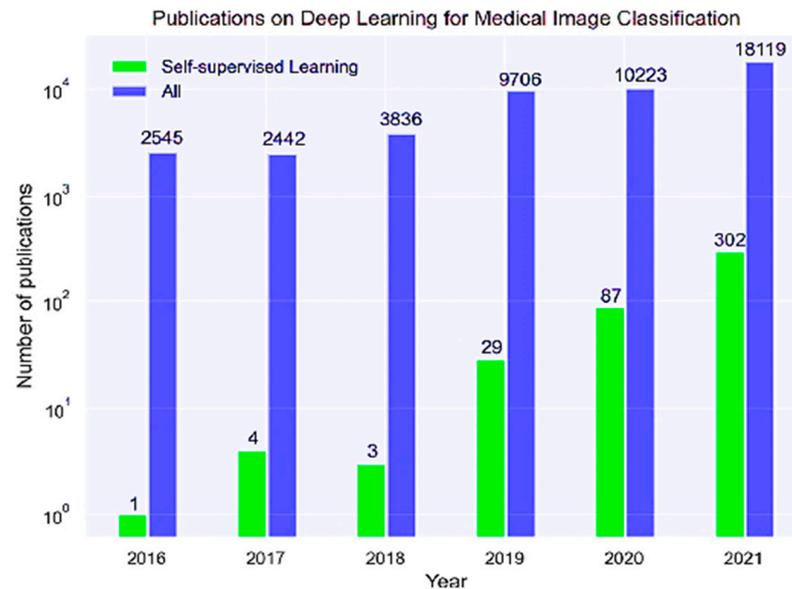
According to these practical benefits, further workable possibilities and effective prediction and recognition impacts can be explained in the following paragraphs, which focus mainly on medical and engineering contexts.

#### 3.3.1. SSL Models for Medical Predictions

Krishnan et al. (2022) [76] analyzed SSL models' application in medical data classification, highlighting the critical challenges of manual annotation of vast medical databases. They addressed SSL's potential for enhancing disease diagnosis, particularly in EHR and

some other visual clinical datasets. Huang et al. (2023) [20] conducted a systematic review affirming SSL's benefits in supporting medical professionals with precise classification and therapy identification from visual data, reducing the need for extensive manual labeling.

Figure 10 shows the number of DL, ML, and SSL research articles published between 2016 and 2021.



**Figure 10.** The number of articles on SSL, ML, and DL models utilized for medical data classification [20].

It can be concluded from the statistical data explained in Figure 10 that the number of research publications addressing ML and DL models' importance and relevance in the medical classification has been increasing per year. Similarly, the increasing trend was for the overall number of academic articles investigating the SSL, ML, and DL algorithms in conducting high-performance identification of problems in images of patients.

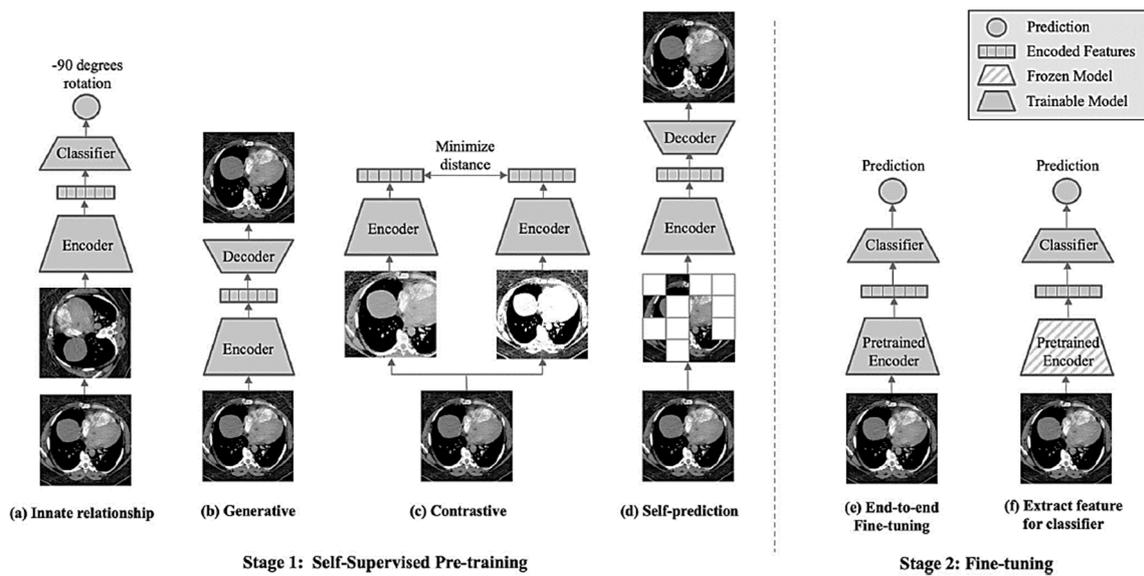
Besides these numeric figures, an explanation of the pre-training process of SSL and fine-tuning can be expressed in Figure 11.

It can be inferred from the data explained in Figure 11 that the pre-training SSL process takes into account four critical types to be accomplished, including (Figure 11a) innate relationship, (Figure 11b) generative, (Figure 11c) contrastive, and (Figure 11d) self-prediction. At the same time, there are two categories included in the fine-tuning process, which comprise end-to-end and feature extraction procedures.

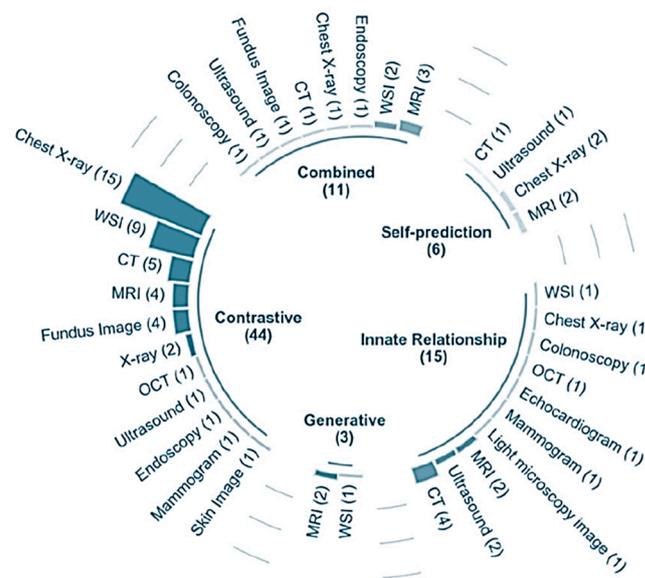
Before the classification process is done, SSL models are first trained. This step is followed by the encoding of image features. It follows the adoption of the classifier, which is important to enable precise prediction of the medical problem in the image.

In their overview [20], the scholars identified a collection of some medical disciplines in which SSL models can be advantageous in conducting the classification process flexibly, which can be illustrated in Figure 12.

From the data expressed in Figure 12, it can be inferred that the possible medical classification types and dataset categories are numerous in SSL models that can be applied reliably for efficient classification. As a result, this aspect makes SSL models more practical and feasible for carrying out robust predictions of problems in clinical datasets.



**Figure 11.** The two stages of pre-training and fine-tuning are considered in the classification of visual data [20].



**Figure 12.** The major categories correlated with medical classification can be done by SSL models [20].

Various studies have explored the application of SSL models in medical data classification, showcasing their efficacy in improving diagnostic accuracy and efficiency. Azizi et al. (2021) [77] demonstrated the effectiveness of SSL algorithms in classifying medical disorders within visual datasets, particularly highlighting advancements in dermatological and chest X-ray recognition. Zhang et al. (2022) [78] utilized numerical simulations to classify patient illnesses on X-rays, emphasizing the importance of understanding medical images for clinical knowledge. Bozorgtabar et al. (2020) [79] addressed the challenges of data annotation in medical databases by employing SSL methods for anomaly classification in X-ray images. Tian et al. (2021) [80] identified clinical anomalies in fundus and colonoscopy datasets using SSL models, emphasizing the benefits of unsupervised anomaly detection in large-scale health screening programs. Ouyang et al. (2021) [81] introduced longitudinal neighborhood embedding SSL models for classifying Alzheimer’s disease-related neurological problems, enhancing the understanding of brain disorders. Liu et al. (2021) [82] proposed an SSMT-SiSL hybrid model for chest X-ray data classification, highlighting the

potential of SSL techniques to expedite data annotation and improve model performance. Li et al. (2021) [83] addressed data imbalances in medical datasets with an SSL approach, enhancing lung cancer and brain tumor detection. Manna et al. (2021) [84] demonstrated the practicality of SSL pre-training in improving downstream operations in medical data classification. Zhao and Yang (2021) [85] utilized radiomics-based SSL approaches for precise cancer diagnosis, showcasing SSL’s vital role in medical classification tasks.

### 3.3.2. SSL Models for Engineering Contexts

In the field of engineering, SSL models may provide contributory practicalities, especially when prediction tasks in mechanical, industrial, electrical, or other engineering domains are necessary without the need for massive data annotations to train and test conventional models to accomplish this task accurately and flexibly.

In this context, Esrafilian and Haghghat (2022) [86] explored the critical workabilities of SSL models in providing sufficient control systems and intelligent monitoring frameworks for heating, ventilation, and air-conditioning (HVAC) systems. Typically, ML and DL models may not contribute to noteworthy advantages since complicated relationships, patterns, and energy consumption behaviors are not directly and clearly provided. The controller was created by employing a model-free reinforcement learning technique recognized with a double-deep Q-network (DDQN). Long et al. (2023) [87] proposed an SSL-based defect prognostics-trained DL model, SSDL, addressing the challenges of costly data annotation in industrial health prognostics. SSDL dynamically updates a sparse auto-encoder classifier with reliable pseudo-labels from unlabeled data, enhancing prediction accuracy compared with static SSL frameworks. Yang et al. (2023) [88] developed an SSL-based fault identification model for machine health prognostics, leveraging vibrational signals and one-class classifiers. Their SSL model, utilizing contrastive learning for intrinsic representation derivation, outperformed novel numerical models in fault prediction accuracy during simulations. Wei et al. (2021) [89] utilized SSL models for rotary machine failure diagnosis, employing 1-D SimCLR to efficiently encode patterns with a few unlabeled samples. Their DTC-SimCLR model combined data transformation combinations with a fixed feature encoder, demonstrating effectiveness in diagnosing cutting tooth and bearing faults with minimal labeled data. Overall, DTC-SimCLR had improved diagnosis accuracy and fewer samples. Figure 13 depicts a low-sample machine failure diagnosis approach.

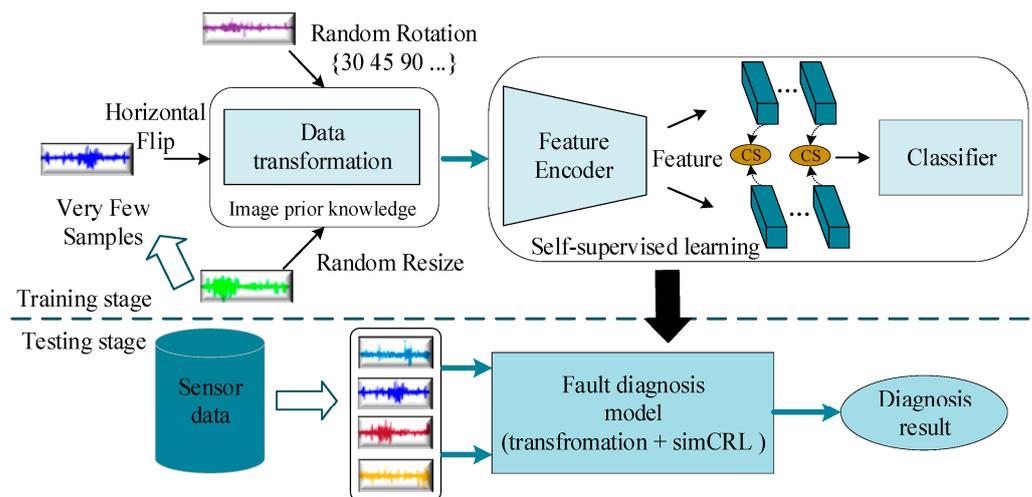


Figure 13. The formulated system for machine failure diagnosis needs very few samples [89].

Furthermore, the procedure related to the SSL in SimCLR can be expressed in Figure 14.

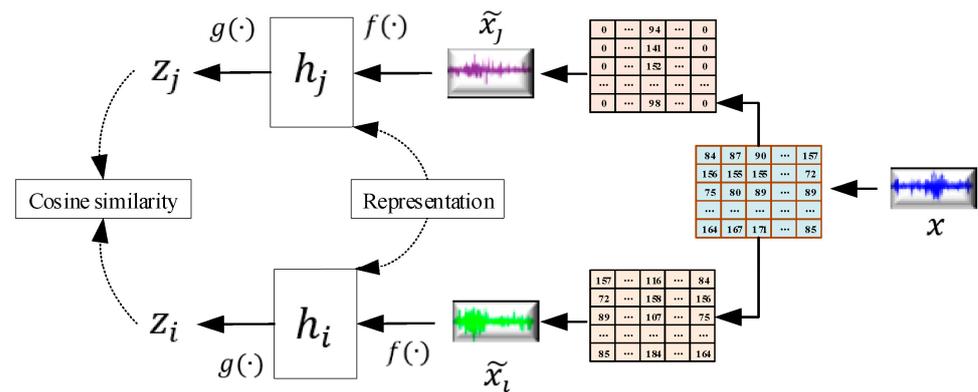


Figure 14. The procedure related to the SSL in SimCLR [89].

Simultaneously, Table 1 indicates the critical variables correlated with the 1D SimCLR.

Table 1. The major variables linked to the 1D SimCLR [89].

No.	Variable Category	Magnitude
1	Input Data	A Length of 1024 Data Points
2	Temperature	10
3	Feature Encoder	Sixteen Convolutional Layers
4	Output Size	128
5	Training Epoch	200

Above these examples, Lei et al. (2022) [90] addressed SSL models in predicting the temperature of aluminum correlated with industrial engineering applications. Through their numerical analysis, they examined how changing the temperature of the pot or electrolyte could affect the overall yield of aluminum during the reduction process through their proposed deep long short-term memory (D-LSTM).

On the other side, Xu et al. (2022) [91] identified the contributory rationale of functional SSL models to offer alternative solutions to conventional human defect detection methods that became insufficient. Bharti et al. (2023) [92] remarked that deep SSL (DSSL) contributed to significant relevance in the industry owing to its potency in reducing the time and effort required by humans for data annotation by manipulating operational procedures carried out by robotic systems, taking into account the CIFAR-10 dataset. Hannan et al. (2021) [93] implemented SSL prediction to estimate the state of charge (SOC) correlated with lithium-ion (Li-ion) batteries precisely in electric vehicles (EVs) to ensure their maximum cell lifespan.

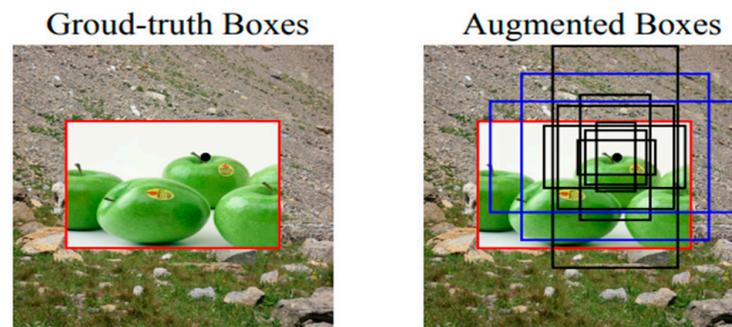
### 3.3.3. Patch Localization

Regarding the critical advantages and positive gains of SSL models in conducting active processes of patch localization, several authors confirmed the significant effectiveness and valuable merits of innovative SSL schemes in accomplishing optimum activities of recognition and detection related to a defined dataset of patches. For instance, Li et al. (2021) [94] estimated the substantial contributions of SSL in identifying visual defects or irregularities in an image without relying on abnormal training data. The patch localization of visual defects involves grid classes, wood, screws, metal nuts, hazelnuts, and bottles.

Although SSL has made great strides in the field of image classification, there is moderate effectiveness in making precise object recognition. Through their analysis, Yang et al. (2021) [95] aimed to improve self-supervised, pre-trained models for object detection. They proposed a novel self-supervised pretext algorithm called instance localization, proposing an augmentation strategy for the image-bounding boxes. Their results confirmed that their pre-trained algorithm for object detection was improved, but it became less effective in ImageNet semantic classification and more so in image patch localization. Object detection

considering the PASCAL VOC and MSCOCO datasets revealed that their method achieved state-of-the-art transfer learning outcomes.

The red box in their result, expressed in Figure 15, indicates the base truth bounding box linked to the foreground image. However, the right-hand photo shows a group of anchor boxes positioned in the central area related to a singular spatial location. By improving the multiple anchors using variant scales, positions, and aspect ratios, the base truth pertaining to the blue boxes can be augmented, offering an intersection over union (IoU) level greater than 0.5.



**Figure 15.** Bounding boxes addressed for spatial modeling [95].

To train an end-to-end model for anomaly identification and localization using only normal training data, Schlüter et al. (2022) [96] created a flexible self-supervision patch categorization model called natural synthetic anomalies (NSA). Their NSA harnessed Poisson photo manipulation to combine scaled patches of varying sizes from multiple photographs into a single coherent entity. Compared with other data augmentation methods for unsupervised anomaly identification, this aspect helped generate a wider variety of synthetic anomalies that were more akin to natural sub-image inconsistencies. Natural and medical images were employed to test their proposed technique, including the MVTec AD dataset, indicating the efficient capability of identifying various unknown manufacturing defects in real-world scenarios.

### 3.3.4. Context-Aware Pixel Prediction

Learning visual representations from unlabeled photographs has recently witnessed a rapid evolution owing to self-supervised instance discrimination techniques. Nevertheless, the success of instance-based objectives in medical imaging is unknown because of the large variations in new patients' cases compared with previous medical data. Context-aware pixel prediction focuses on understanding the most discriminative global elements in an image (such as the wheels of a bicycle). According to the research investigation conducted by Taher et al. (2022) [97], instance discrimination algorithms have poor effectiveness in downstream medical applications because the global anatomical similarity of medical images is excessively high, resulting in complicated identification tasks. To address this shortcoming, scholars have innovated context-aware instance discrimination (CAiD), a lightweight but powerful self-supervised system, considering: (a) generalizability and transferability; (b) separability in embedding space; and (c) reusability and systematic reusability. The authors addressed the dice similarity coefficient (DSC) as a measure related to the similarity between two datasets that are often indicated as binary arrays. Similarly, authors in [98] proposed a teacher-student strategy for representation learning, wherein a perturbed version of an image serves as an input for training a neural net to reconstruct a bag-of-visual-words (BoW) representation referring to the original image. The BoW targets are generated by the teacher network, and the student network learns representations while simultaneously receiving online training and an updated visual word vocabulary.

Liu et al. (2018) [57] distinguished some beneficial yields of SSL models in identifying information from defined datasets of context-aware pixel databases. To train the CNN models necessary for depth evaluation from monocular endoscopic data without a

priori modeling of the anatomy or coloring, the authors implemented the SSL technique, considering a multiview stereo reconstruction technique.

### 3.3.5. Natural Language Processing

Fang et al. (2020) [15] considered SSL to classify essential information in certain defined datasets related to natural language processing. Scholars explained that pre-trained linguistic models, such as bidirectional encoder representations from transformers (BERTs) and generative pre-trained transformers (GPTs), have proved their considerable effectiveness in executing active linguistic classification tasks. Existing pretraining techniques rely on auxiliary prediction tasks based on tokens, which may not be effective for capturing sentence-level semantics. Thus, they proposed a new approach that recognizes contrastive self-supervised encoder representations using transformers (CERTs). Baevski et al. (2023) [99] highlighted critical SSL models' relevance to high-performance data identification correlated with NLP. They explained that currently available techniques of unsupervised learning tend to rely on resource-intensive and modal-specific aspects. They added that the Data2vec model expresses a practical learning paradigm that can be generalized and broadened across several modalities. Their study aimed to improve the training efficiency of this model to help handle the precise classification of NLP problems. Park and Ahn (2019) [100] inspected the vital gains of SSL to lead to efficient detection of NLP. Researchers proposed a new approach dedicated to data augmentation that considers the intended context of the data. They suggested a label-asked language model (LMLM), which can effectively employ the masked language model (MLM) in data with label information by including label data for the mask tokens adopted in the MLM. Several text classification benchmark datasets were examined in their work, including the Stanford sentiment treebank-2 (SST2), multi-perspective question answering (MPQA), text retrieval conference (TREC), Stanford sentiment treebank-5 (SST5), subjectivity (Subj), and movie reviews (MRs).

### 3.3.6. Auto-Regressive Language Modeling

Elnaggar et al. (2022) [101] published a paper shedding light on valuable SSL roles in handling the active classification of datasets connected to the modeling of auto-regressive language. The scholars trained six models, four auto-encoders (BERT, Albert, Electra, and T5), and two auto-regressive prototypes (Transformer-XL and XLNet) on up to 393 billion amino acids from UniRef and BFD. The Summit supercomputer was utilized to train the protein LMs (pLMs), which required 5616 GPUs and a TPU Pod with up to 1024 cores. Lin et al. (2021) [102] performed numerical simulations, exploring the added value of three SSL models, notably (I) autoregressive predictive coding (APC), (II) contrastive predictive coding (CPC), and (III) wav2vec 2.0, in performing flexible classification and reliable recognition of datasets engaged in auto-regressive language modeling. Several any-to-any voice conversion (VC) methods have been proposed, like AUTOVC, AdaINVC, and FragmentVC. To separate the feature material from the speaker information, AUTOVC and AdaINVC utilize source and target encoders. They proposed a new model, known as S2VC, which harnesses SSL by considering multiple features of the source and target linked to the VC model. Chung et al. (2019) [103] proposed an unsupervised auto-regressive neural model to help students learn generalized representations of speech. Their speech representation learning approach was developed to maintain information for various downstream applications to remove noise or speaker variability.

### 3.4. Commonly-Utilized Feature Indicators of SSL Models' Performance

Specific formulas in [104,105] were investigated to examine different SSL paradigms in carrying out their classification task, particularly the prediction and identification of faults and errors in machines, which can support maintenance specialists in selecting the most appropriate repair approach. These formulas formulate practical feature indicators to monitor signals that can be prevalently utilized by maintenance engineers to identify the

health state of machines. Twenty-four typical feature indicators were addressed, referring to Zhang et al. (2022) [106]. These indices can enable maintenance practitioners to locate optimum maintenance strategies to apply to industrial machinery, helping to handle current failure issues flexibly. Those twenty-four feature indicators are shown in Table 2.

**Table 2.** Prevalent signal feature indicators utilized to examine and diagnose machine and industrial equipment health.

Feature Indicator Type	Major Formula	Eq#
Mean Value	$MV = \frac{1}{N} \sum_{n=1}^N x(n)$	(7)
Standard Deviation	$SD = \sqrt{\left(\frac{1}{N-1}\right) \left(\sum_{n=1}^N \left[x(n) - \frac{1}{N} \sum_{n=1}^N x(n)\right]^2\right)}$	(8)
Square Root Amplitude	$SRA = \left(\frac{1}{N} \sum_{n=1}^N \sqrt{ x(n) ^2}\right)^2$	(9)
Absolute Mean Value	$AMV = \frac{1}{N} \sum_{n=1}^N  x(n) $	(10)
Skewness	$S = \frac{1}{N} \sum_{n=1}^N (x(n))^3$	(11)
Kurtosis	$K = \frac{1}{N} \sum_{n=1}^N (x(n))^4$	(12)
Variance	$V = \frac{1}{N} \sum_{n=1}^N (x(n))^2$	(13)
Kurtosis Index	$KI = \frac{1}{N} \sum_{n=1}^N (x(n))^4 / \sqrt{\left(\frac{1}{N} \sum_{n=1}^N (x(n))^2\right)^2}$	(14)
Peak Index	$PI = \max x(n)  / \left(\frac{1}{N-1}\right) \left(\sum_{n=1}^N \left[x(n) - \frac{1}{N} \sum_{n=1}^N x(n)\right]^2\right)$	(15)
Waveform Index	$WI = \sqrt{\left(\frac{1}{N-1}\right) \left(\sum_{n=1}^N \left[x(n) - \frac{1}{N} \sum_{n=1}^N x(n)\right]^2\right)} / \frac{1}{N} \sum_{n=1}^N  x(n) $	(16)
Pulse Index	$PuI = \max x(n)  / \frac{1}{N} \sum_{n=1}^N  x(n) $	(17)
Skewness Index	$SI = \frac{1}{N} \sum_{n=1}^N (x(n))^3 / \left(\sqrt{\frac{1}{N} \sum_{n=1}^N (x(n))^2}\right)^3$	(18)
Frequency Mean Value	$FMV = \frac{1}{N} \sum_{n=1}^N s(n)$	(19)
Frequency Variance	$FV = \frac{1}{N-1} \sum_{n=1}^N \left[s(n) - \frac{1}{N} \sum_{n=1}^N s(n)\right]^2$	(20)
Frequency Skewness	$FS = \frac{1}{N \left(\frac{1}{N-1} \sum_{n=1}^N \left[s(n) - \frac{1}{N} \sum_{n=1}^N s(n)\right]^2\right)^{(3/2)}} \sum_{n=1}^N \left[s(n) - \frac{1}{N} \sum_{n=1}^N s(n)\right]^3$	(21)
Frequency Steepness	$FSt = \frac{1}{N \left(\frac{1}{N-1} \sum_{n=1}^N \left[s(n) - \frac{1}{N} \sum_{n=1}^N s(n)\right]^2\right)^2} \sum_{n=1}^N \left[s(n) - \frac{1}{N} \sum_{n=1}^N s(n)\right]^4$	(22)
Gravity Frequency	$GF = \frac{\sum_{n=1}^N f_i s(n)}{\sum_{n=1}^N s(n)}$	(23)
Frequency Standard Deviation	$FSD = \sqrt{\left[\sum_{n=1}^N \left(f_i - \frac{\sum_{n=1}^N f_i s(n)}{\sum_{n=1}^N s(n)}\right)^2 s(n) / N \sum_{n=1}^N s(n)\right]}$	(24)
Frequency Root Mean Square	$FRMS = \sqrt{\left[\frac{\sum_{n=1}^N f_i^2 s(n)}{\sum_{n=1}^N s(n)}\right]}$	(25)
Average Frequency	$AF = \sqrt{\left[\frac{\sum_{n=1}^N f_i^4 s(n)}{\sum_{n=1}^N f_i^2 s(n)}\right]}$	(26)
Regularity Degree	$RD = \sum_{n=1}^N \left[ f_i^2 s(n) / \sqrt{\left(\frac{\sum_{n=1}^N s(n)}{\sum_{n=1}^N f_i^4 s(n)}\right)} \right]$	(27)

Table 2. Cont.

Feature Indicator Type	Major Formula	Eq#
Variation Parameter	$VP = \left( \sqrt{\left[ \sum_{n=1}^N \left( f_i - \sum_{n=1}^N \left[ \frac{f_i s(n)}{\sum_{n=1}^N s(n)} \right] \right)^2 \right) \frac{s(n)}{N \sum_{n=1}^N S(n)} \right) / \sum_{n=1}^N \left[ \frac{f_i s(n)}{\sum_{n=1}^N s(n)} \right]} \right) \quad (28)$	(28)
Eighth-Order Moment	$8^{th} OM = \left( \sum_{n=1}^N f_i - \sum_{n=1}^N \left[ \frac{f_i s(n)}{\sum_{n=1}^N s(n)} \right] \right)^3 \frac{s(n)}{N} \left( \sqrt{\left[ \sum_{n=1}^N \left( f_i - \sum_{n=1}^N \left[ \frac{f_i s(n)}{\sum_{n=1}^N s(n)} \right] \right)^2 \right) \frac{s(n)}{N \sum_{n=1}^N S(n)} \right)^3 \quad (29)$	(29)
Sixteenth-Order Moment	$16^{th} OM = \left[ \sum_{n=1}^N \left( f_i - \sum_{n=1}^N \left[ \frac{f_i s(n)}{\sum_{n=1}^N s(n)} \right] \right)^4 \right] \frac{s(n)}{N} \left[ \sqrt{\left[ \sum_{n=1}^N \left( f_i - \sum_{n=1}^N \left[ \frac{f_i s(n)}{\sum_{n=1}^N s(n)} \right] \right)^2 \right) \frac{s(n)}{N \sum_{n=1}^N S(n)} \right]^4 \quad (30)$	(30)

Those twenty-four indices can serve as prior feature group, *FG*, which can be illustrated by the following:

$$FG = \{ p_1^i, p_2^i, p_3^i, \dots, p_{24}^i \}_{i=1}^N \quad (31)$$

The twenty-four feature indicators can be utilized relying on the data standardization process, which can be achieved by the following mathematical expression:

$$p^i = \frac{p^i - \text{mean}(p^i)}{\text{STD}(p^i)} \quad (32)$$

#### 4. Statistical Figures on Critical SSL Rationale

To provide elaborating statistical facts pertaining to SSL importance in handling robust data detection and efficient data classification crucial for industrial disciplines, two comparative analyses were implemented; the first one is correlated with fault diagnostics in actual industrial applications. In the meantime, the second comparative study is concentrated on the essential prediction of health issues in the real medical context.

Table 3 summarizes the major findings and major limitations of SSL models involved in real industrial scenarios.

From Table 3, the material significance of SSL models can be noticed from their considerable practicalities in carrying out precise machine failure prediction, supporting the maintenance team in executing the necessary repair procedures without encountering the common problems of massive data annotation and time-consuming identification of wide failure mode databases, which are essential for DL and ML models.

Besides these noteworthy findings, it is crucial to point out that in spite of the constructive prediction success of those SSL paradigms, there are a couple of issues that could restrict their broad prediction potential, including instability, imbalance, noise, and random variations in the data, which may cause uncertainties and a reduction in their overall prediction performance. Correspondingly, it is hoped that these barriers can be handled efficiently in future work.

On the other hand, Table 3 provides some concluding remarks pertaining to the imperative outcomes and prevalent challenges of SSL paradigms utilized in real medical health diagnosis.

It is inferred from the outcomes explained in Table 4 that SSL also offered a collection of noteworthy implications in favor of better medical treatment that can support healthcare providers in classifying swiftly and durably the sort of clinical problem in patients. Therefore, the most appropriate therapeutic process can be successfully prescribed. Similar to what was discussed previously pertaining to the industrial domain, performing the prognosis of rotational machinery is not an easy task since failure modes and machinery faults are diversified and they are not necessarily identical to past failure situations. In the medical context, diagnosis may sometimes be complicated as different patients have various illness conditions and disorder features that do not necessarily simulate historical patient databases.

**Table 3.** Summary of the core findings and critical limitations respecting SSL engagement in different real industrial scenarios for machine health prognostics and fault prediction.

#	Author(s) (Year)	Industrial SSL Application Sort	Dataset Category	Encountered Research Limitations	Critical Contributions and Positive SSL Impacts
1	Yuan and Lin (2020) [107]	Generative SSL Recognition	SITS-BERT	Not available (N/A)	The classification accuracy of a transformer, 1D CNN, and bidirectional long short-term memory (LSTM) network is significantly improved by the proposed pre-training approach in experimental data.
2	Zhao et al. (2020) [108]	Scene SSL Classification	NWPU, AID, UC Merced, and WHU-RS19	The loss function forced the primary classifier to be invariant with respect to the transformations. Therefore, the utilization of additional labeling in the SSL did not guarantee performance improvement in fully supervised classification conditions.	Their results related to the NWPU, AID, UC Merced, and WHU-RS19 dataset classifications revealed state-of-the-art average accuracy levels, recording 94.21%, 96.89%, 99.11%, and 98.98%, respectively. Their suggested strategy enhanced the accuracy of remote sensing scene categorization, as evidenced by experimental findings and visual representations, by learning additional discriminative features while simultaneously encoding orientation information.
3	Tao et al. (2023) [109]	Remote Sensing Image Understanding	DLRSD and AID	N/A	Based on their numerical simulations, it was found that utilizing their TOV model to help facilitate the classification process of information related to RSIU using SSL principles contributed to enhanced levels of classification accuracy.
4	Stojnic and Risojevic (2018) [110]	SSL Classification of Visual Dataset Considering LAB and RGB Color Spaces	AID	N/A	Their simulation outcomes confirmed that near-state-of-the-art performance was attained, registering a classification accuracy of 89.27% on the AID dataset, needing a minimal amount of unlabeled training images.
5	Jung and Jeon (2021) [111]	SSL Classification of Visual Database	National Agriculture Imagery Program (NAIP) and CDL	The original Tile2Vec classification model faced a degradation issue when the epoch reached the maximum number of epochs, which was 50. On the other hand, the Tile2Vec classification model, which had one randomized layer, contributed to a slight degradation in the classification process.	The scholars found that obtaining more robust representations was facilitated by not updating the completely connected layers. Their proposed Tile2Vec algorithm provided more significant performance in terms of classification accuracy compared with random forest (RF), logistic regression (LR), and multi-layer classifiers (MLC).

Table 3. Cont.

#	Author(s) (Year)	Industrial SSL Application Sort	Dataset Category	Encountered Research Limitations	Critical Contributions and Positive SSL Impacts
6	Hahn and Mechefske (2021) [9]	SSL Forecasting in the Context of Machinery Health Monitoring	Milling and CNC Dataset	<p>Vibrations from the table and spindle would cause sources of error in carrying out the necessary detection process, influencing the final outcomes of the anomaly model. Also, the trained models, which correlated with the CNC data, did not generalize well across all the unique parts in the dataset.</p> <p>Furthermore, the investigated models did not generalize well across multiple metal-cutting variables.</p> <p>When more variations of faults occur, RSS models may not perform robust identification of segments free from noise.</p>	<p>The approach got the best PR-AUC score of 0.80 for shallow-depth cuts and a score of 0.45 for all cutting parameters on a milling dataset. The best PR-AUC score of this SSL method attained an ultimate PR-AUC score of roughly 0.41 based on a real-world industrial CNC dataset.</p>
7	Jiang et al. (2017) [112]	SSL Process Monitoring of Chemical Processes	Finite Discrete Dataset consisting of 100 Samples	<p>Comparatively, conventional models reduce reconstruction errors, contributing to lower sensitivity to fault variations. When Gaussian noise is considered, the sensitivity correlated with the RSS model could be increased in processing drifts.</p> <p>The complicated operating conditions and variant fault behaviors in industrial environments may result in multiple difficulties and further challenges to achieving maximum accuracy in fault diagnosis and identification.</p>	<p>Their theoretical analysis revealed that their SSL models offered more sensitive aspects of fault occurrence in the analytical process.</p> <p>The efficiencies of both robust autoencoders and robust principal component analysis (PCA) monitoring provided enhanced performance and optimum and active monitoring levels of chemical processes.</p>
8	Yu et al. (2023) [113]	SSL Estimation of the RUL	C-MAPSS		<p>Their approach could successfully enhance the model's feature extraction capacity. Hidden characteristics were preferable to raw data when the clustering process was applied.</p>

Table 3. Cont.

#	Author(s) (Year)	Industrial SSL Application Sort	Dataset Category	Encountered Research Limitations	Critical Contributions and Positive SSL Impacts
9	Hu et al. (2023) [114]	SSL Fault Diagnosis and Defect Prediction	Open-Source and Self-Designed Datasets	The single-task dominance problem did exist in the multitask algorithm that conducted necessary fault diagnosis and failure identification in the industrial context. (To solve this issue, an uncertainty-correlated dynamic weighting tactic was utilized to automatically distribute weight for every task referring to its uncertainty level, helping ensure better stability in multi-task optimization.)	Their proposed SSL model provided more superiority in performing crucial machine fault prognostics, which could help handle efficient fault maintenance more flexibly with upgraded levels of accuracy and performance compared with other semi-supervised and supervised models.
10	Huang et al. (2022) [115]	SSL Distilling Process for Recommender Systems from Ubiquitous, Sparse, And Noisy Data	N/A	N/A	SSL models in recommender systems could support engineers in minimizing the rates of noisy data and ineffective information that lower the performance and reliability of recommender systems.
11	Wang et al. (2020) [116]	SSL Fault Detection in Prognostics and Health Management (PHM)	The LAM 9600 Metal Etcher	Their suggested method, which formulated an SS algorithm and relied on the Kernel PCA (KPCA), was only trained utilizing normal samples. At the same time, fault detection was merely accomplished by KPCA rather than a combination of various ML models.	SSL offered relevant fault detection findings, outperforming existing fault detection methods with enhanced efficacy.
12	Nair et al. (2017) [117]	SSL Manipulation of Deformable Objects	(Self-Collected Datasets by Robotic System) Raw Images of a Rope	When the changes in rope orientation are not very sharp, the robot might perform better. Furthermore, because the researchers did not have a comparable number of additional randomly collected databases, they were not capable of identifying the levels of improvement. This issue is correlated with the higher quality and quantity of the collected databases.	Robots could successfully manipulate a rope into a broad range of goal shapes with only a sequence of photographs offered by a person by merging the high- and low-level plans.

Table 3. Cont.

#	Author(s) (Year)	Industrial SSL Application Sort	Dataset Category	Encountered Research Limitations	Critical Contributions and Positive SSL Impacts
13	Ren et al. (2022) [118]	SSL Monitoring and Prediction of Production Status	Real-World Froth Flotation Dataset	<p>Due to their recursive nature or deep network structure, SSFAN and LSTM have relatively higher computational costs. In addition, since GSTAE adopts ‘tanh’ and ‘sigmoid’ activation functions to control the information flow, it has the most considerable computational cost.</p> <p>Statistical tests have not been adopted for the comparison since there is no established guideline to select a proper test in this powertrain application. Further, constructing a proper statistical test requires a performance analysis on multiple datasets. Also, the effect of unbalanced datasets is out of scope in their study.</p>	<p>Considering the real-world mining dataset, their proposed LSTM-DeepFM technique achieved state-of-the-art performance contributions compared with other stacked autoencoder-based models, semisupervised parallel DeepFM, and variational autoencoder-based models.</p>
14	Senanayaka et al. (2020) [119]	SSL Defect Categorization and Fault Diagnosis	Balanced and Training Dataset	<p>Due to reliability problems with the RealSense camera, their model utilized only depth images from the Ensenso N-10. It was trained on wooden primitive shapes having side lengths of 4 cm. Additionally, the prediction process for the displacement of the cross shape during clamping was difficult to accomplish.</p> <p>More neurons or layers in DNNs might encounter lengthy training processes connected with significant convergence scales. Unfortunately, pre-training might not offer remarkable practicalities since the evaluation of the remaining useful life (RUL) could vary. If more unlabeled data is offered, classification findings would be enhanced when limited annotated data is available.</p>	<p>Their SSL principles, applied to the proposed CNN algorithm, allowed an improved online diagnosis scheme to learn features according to the latest data. The effectiveness of their paradigm was validated via comparative analysis, explaining the significant practicality of the trained CNN model in detecting defects.</p>
15	Berscheid et al. (2020) [120]	Better Decision-Making Process for Robots Regarding Flexible Pick-and-Place	Self-Defined Visual Dataset: (A) RGBD-Images Handling Screws on Around 3500 Pick-and-Place Actions) and (B) Depth-Images of 25,000 Pick-and-Place Actions.	<p>Based on the SSL enhancements on the CNN model, it was determined that their robot could infer several pick-and-place operations from a single objective state, learn to select the correct thing when presented with multiple object kinds, accurately insert objects within a peg game, and pick screws out of dense clutter.</p>	
16	Akrim et al. (2023) [6]	SSL Detection of Fatigue Damage Prognostic Problems	Synthetic Datasets of Strain Data	<p>Based on the SSL enhancements on the CNN model, it was determined that their robot could infer several pick-and-place operations from a single objective state, learn to select the correct thing when presented with multiple object kinds, accurately insert objects within a peg game, and pick screws out of dense clutter.</p>	<p>SSL pre-trained models were capable of significantly outperforming the non-pre-trained models in the downstream RUL prediction challenge with lower computational expenses.</p>

Table 3. Cont.

#	Author(s) (Year)	Industrial SSL Application Sort	Dataset Category	Encountered Research Limitations	Critical Contributions and Positive SSL Impacts
17	Zhang et al. (2022) [106]	Active and Intelligent SSL Diagnosis to Mine the Health Information of Machines	Self-Collected Faulty Monitoring Dataset	N/A	<p>Their proposed SSL framework has successfully extracted more detailed monitoring information. Two experiments that simulated mechanical faults confirmed the remarkable efficacy of their suggested model. Their new approach gave workable inspections necessary for industrial fault issues by cognitive diagnosis of machines.</p> <p>Their model proved its practicality, especially for imbalanced data, where imbalance and instability would exist between normal data and faulty data in realistic industrial scenarios. Their SSL model, which considered few-shot learning principles with the help of intrinsic relationships in unlabeled wafer maps, achieved significant enhancements.</p>
18	Geng et al. (2021) [121]	Wafer Failure Pattern Detection to Prevent Yield Loss Excursion Events Linked to Semiconductor Manufacturing	N/A	N/A	<p>Their suggested approach outperformed various state-of-the-art tactics for wafer defect classification. SSL could alleviate the imbalance problem of data distribution in real industrial failure patterns since it utilizes the comprehensive advantage of the massive unlabeled wafer failure data.</p>
19	Yoa et al. (2021) [122]	SSL Anomaly Detection of Visual Databases with the Help of Dynamic Local Augmentation SSL Domain	MVTec Anomaly Detection Dataset	Dynamic local augmentation was helpful, but conventional local augmentation interferes with the performance.	Competitive performance was achieved for pixel-wise anomaly segmentation. A variety of combinations of four losses affected the performance.
20	Li, et al. (2022) [123]	Adaptation-Based Fault Diagnosis	A Gearbox Dataset	N/A	SSL could help achieve significant rates of effectiveness, accuracy, and reliability in detecting faults related to industrial engineering activities.
21	Lu et al. (2022) [124]	Intelligent SSL Fault Diagnosis via Feature Clustering	Industrial Training and Testing Datasets	N/A	SSL offered elevated performance and efficacy to detect faults related to industrial applications.

Table 3. Cont.

#	Author(s) (Year)	Industrial SSL Application Sort	Dataset Category	Encountered Research Limitations	Critical Contributions and Positive SSL Impacts
22	Ding et al. (2022) [125]	Practical Fault Prediction of Incipient Failure in Bearings	FEMTO-ST Datasets	<p>Hyperparameter manual modifications of the SSPCL model are needed. Integrating SSL could improve its practicality for further machine-prognostic tasks. A minimal data annotation rate could alleviate the model's machine prognostic performance.</p> <p>MixMatch could offer lower prediction performance compared with FFT + SVM.</p>	<p>Superior effectiveness was realized in SSL contrast learning (SSLCL). SSL pre-training was helpful for achieving better identification. Momentum contrast learning (MCL) was addressed to distinguish beneficial data from unlabeled datasets, overcoming time-consuming and expensive labeling processes.</p>
23	Yan, and Liu (2022) [126]	Practical Fault Diagnosis of Bearings Related to Aerspace Industry under Limited Databases	Two Independent Bearing Datasets from Paderborn University and the Polytechnic University of Turin for Experimental Verification	<p>More noise levels do exist because of the gap between the unannotated dataset and its corresponding data distribution. This gap might broaden in a gradual manner if the noise escalates, contributing to the worse capability of MixMatch to benefit from the unlabeled data necessary for diagnostic precision enhancement.</p>	<p>SMoCo performed feature extraction of vibration signals, considering frequency and time scales. SMoCo could learn potent and efficient feature extraction by pre-training using artificially injected fault-bearing data, improving data diagnosis accuracy regardless of the types of equipment, failure modes, noise magnitude, or working circumstances, and achieving prediction in a considerably shorter interval.</p>

**Table 4.** Crucial outcomes and major obstacles related to SSL involvement in medical diagnosis.

#	Author(s) (Year)	Medical SSL Application	Dataset Category	Encountered Research Limitations	Critical Contributions and Preferable SSL Impacts
1	Chen et al. (2019) [127]	SSL Medical Image Categorization via Image Context Restoration	Fetal 2D Ultrasound Images Linked to Abdominal Organs in CT Images and Brain Tumors in Multi-Modal MR Images	Not available (N/A)	The created SSL context restoration model strategy learned semantic image features, which are beneficial for various categories of corresponding visual dataset analysis. Their SSL algorithm was flexible to implement. Its applicability in three situations of data recognition—(1) segmentation, (2) localization, and (3) classification was competitive.
2	Nguyen et al. (2020) [128]	SSL Medical Image Recognition Utilizing Spatial Awareness	Struct-Seg-2019 Dataset	The overall SSL models achieved escalated performance outcomes because of abundant information for the pretext task. Nonetheless, the margin of enhancement is not considerably satisfying. (They showed that if more unannotated data is utilized, then the performance would enhance.) The SSL model needed a few more labeled training samples to attain an equivalent efficiency to that linked to the network trained from scratch.	Organ segmentation and cerebral bleeding detection via SSL models tested in their work demonstrated remarkable efficacy compared with other ML methods.
3	Jamaludin et al. (2017) [129]	Longitudinal Spinal MRI	A Dataset of 1016 subjects, 423 Possessing Follow-Up Scans.		Relying on the SSL involvement in longitudinal spinal MRI categorization, it was found that the effectiveness of the pre-trained SSL CNN model outperformed the performance of other models that were trained from scratch.
4	Zhu et al. (2020) [130]	3D Medical Image Classification; Task-Related Contrastive Prediction Coding (TCPC) Extensive Evaluation on	Brain Hemorrhage Dataset: 1,486 Brain CT Volumes	N/A	Their experimental analysis confirmed remarkable effectiveness correlated with lesion-related embedding before knowledge into NNs for 3D medical image classification.
5	Xie et al. (2020) [131]	Four Public Computerized Tomography (CT) Datasets of 11 Kinds of Major Human Organs and Two Tumors.	Pretext Task Dataset, including 1808 CT Scans from 5 Public Datasets	N/A	The results indicated that utilizing a pre-trained SSL PGL model could help initialize the downstream network, contributing to a preferable effectiveness compared with random initialization and the initialization by global consistency-based models.

Table 4. Cont.

#	Author(s) (Year)	Medical SSL Application	Dataset Category	Encountered Research Limitations	Critical Contributions and Preferable SSL Impacts
6	Li et al. (2020) [132]	Diagnosis of Retinal Diseases from Fundus Images	Two Public Benchmark Datasets for Retinal Disease Diagnosis	N/A	The experimental results revealed that the SSL model had clearly outperformed other SSL feature learning mechanisms and was comparable to the supervised level. The SSL models operated by MoCo-CXR-pre-training outperformed other non-MoCo-CXR-pre-training models. The MoCo-CXR pre-training provided the most benefit with a few labeled training datasets. Simultaneously, similar high-performance outputs were attained on the target Tuberculosis dataset, confirming that MoCo-CXR-pre-training endowed other superior models for chest X-ray classification.
7	Sowrirajan et al. (2021) [133]	Utilizing a proposed SSL Model (MoCo-CXR) to Classify Problems in Patients' Chest	Visual Chest X-ray Datasets	There were fewer unlabeled chest X-ray images than natural images. This aspect could limit the applicability of contrastive SSL to the necessary classification of chest X-rays.	Their approach was not practical for datasets that lack patient meta-data altogether. In addition, their strategies for negative pair selection did not enhance pre-trained representations. Their SSL models leveraged data on image laterality. However, future work is needed to determine whether negative pair selection strategies utilize other meta-data, notably image view (anteroposterior or posteroanterior), patient age, or patient gender.
8	Vu et al. (2021) [134]	Selecting positive pairs coming from views of possibly different images by the patient metadata	Visual Chest X-ray Datasets	Their approach was not practical for datasets that lack patient meta-data altogether. In addition, their strategies for negative pair selection did not enhance pre-trained representations. Their SSL models leveraged data on image laterality. However, future work is needed to determine whether negative pair selection strategies utilize other meta-data, notably image view (anteroposterior or posteroanterior), patient age, or patient gender.	Their contrastive SSL model achieved a performance upgrade of 14.4% in mean AUC from the ImageNet pre-trained baseline. Also, their controlled experiments showed that the ways to improve downstream performance on patient disease classification include (a) utilizing patient meta-data to properly create positive pairs from variant images with the same underlying pathologies and (b) maximizing the number of various images utilized in query pairing.
9	Sriram et al. (2021) [135]	Classification of Clinical Diseases Correlated with Large Mortality Rate Due to COVID-19 and Chest X-ray	Dataset of Chest X-ray Linked to COVID-19 and non-COVID-19 Risks	It was found that the inversion of the trend for oxygen requirement prediction (ORP) could not be illustrated by label scarcity. It could be that there were image features that became readily apparent for the ORP task closer to the actual moment of raised oxygen needs.	Their SSL model utilization showed that an improved AUC of 0.786 could be attained for predicting an adverse event at 96 h and an AUC of 0.848 for predicting mortalities at 96 h.

Table 4. Cont.

#	Author(s) (Year)	Medical SSL Application	Dataset Category	Encountered Research Limitations	Critical Contributions and Preferable SSL Impacts
10	Chen et al. (2021) [136]	Classification of Chest CT Images Linked to the COVID-19 To leverage structural similarity across volumetric medical images (domain-specific cue) and to learn distinctive representations of local regions that are practical for per-pixel segmentation (problem-specific cue).	Dataset of Chest CT images Correlated with COVID-19 Problems	The resizing operation done in their classification process slightly affected the overall identification performance.	Their SSL model classification results affirmed superior performance of accuracy in classifying COVID-19 problems related to chest CT images.
11	Chaitanya et al. (2020) [137]		Three Magnetic Resonance Imaging (MRI) datasets	N/A	The created SSL model yielded substantial improvement compared with other SSL and semi-supervised learning (SiSL) techniques. When combined with simple data augmentation, the created model reached within 8% of benchmark performance utilizing solely two annotated MRI data for training.

### 5. Discussion

Supportive information and elaborative details on modern technologies and the latest innovations are integrated into SSL classification models to improve their potential and efficacy in monitoring various data recognition, forecasting, or distinguishing with perfect levels of precision and reliability. The discussion includes a critical explanation and evaluation of the following SSL-supportive technologies:

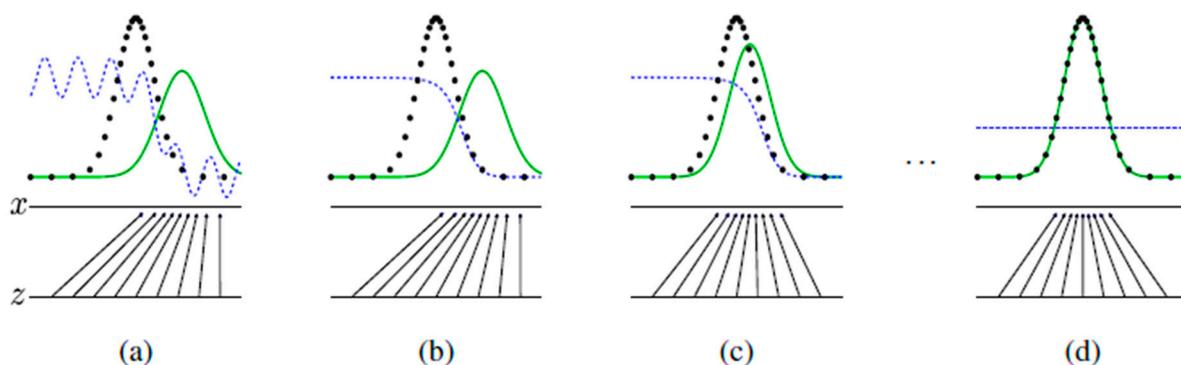
1. Generative Adversarial Networks (GAN);
2. Deep InfoMax (DIM);
3. Pre-trained Language Models (PTM);
4. Contrastive Predictive Coding (CPC);
5. Autoencoder and its associated extensions.

#### 5.1. Generative Adversarial Networks (GAN)

One category of DL architecture is the GAN. A GAN is commonly adopted to create new data based on the training process carried out by two neural networks, which compete with each other to generate the necessary authentic data. Images, movies, and text are examples of databases that can be handled and analyzed flexibly using the output of a GAN.

The concept of GANs was first addressed and investigated in an article published by [138]. An alternative paradigm for USL was created in their study, in which two neural networks were trained to compete with one another. Since then, GANs have emerged as powerful tools for generative modeling. Recently, GANs have proved essential in generative modeling, showcasing impressive skills.

GANs have a significant influence on various activities, including improving data augmentation strategies, enhancing reinforcement learning algorithms, and strengthening SSL methodologies. GANs are a fundamental concept in modern ML, enabling progress in different fields due to their adaptability. Simultaneous training is conducted for GANs considering the update of the distinctive distribution, which can be expressed as a dashed blue line,  $D$ , in Figure 16. Thus, this blue dashed line can distinguish between data samples related to the generative distribution,  $G$ ,  $p_g$ , which is characterized by a solid green line. The horizontal line down the photo expresses the domain and a source of  $z$  that can be uniformly sampled. At the same time, the horizontal line located in the upper area of the image indicates a part of the  $x$  domain. The mapping of  $x$  that equals  $G(z)$ , can impose the non-uniform distribution,  $G$ , on transformed samples. In areas with higher density,  $G$  could contract and enlarge in zones with lower density levels that are correlated with  $p_g$  [138].



**Figure 16.** Configurations of (a)  $D$  expresses a partial precise classifier and  $p_{data}$  is identical to  $p_g$ , (b)  $D$  was converged to  $D^*(x)$ , (c) when  $G$  was updated, the  $D$  gradient helped  $G(z)$  to transfer to areas, which are approximately considered as data, and (d) when various training processes have been conducted, if both  $D$  and  $G$  have sufficient potential, they would attain a position in which they could not enhance since  $p_{data}$  is identical to  $p_g$  [138].

From Figure 16,  $D^*(x)$  can be expressed by the following formula:

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \quad (33)$$

### 5.2. Deep InfoMax (DIM)

This new concept was first introduced by [139], who conducted a numerical analysis to examine novel means of unsupervised learning of representations. Researchers have optimized encoding by decoding mutual information. They confirmed the importance of structure by showing how including information about the input locality in an aim can significantly enhance the fitness of a representation for subsequent tasks. Adversarial matching to a prior distribution allows researchers to control representational features.

DIM outperforms numerous well-known unsupervised learning approaches and is competitive with fully supervised learning in typical architectures across a variety of classification problems. Furthermore, according to the numerical analysis of these researchers, DIM paved the way for more creative formulations of representation learning objectives to address specific end goals, and it also provided new opportunities for the unsupervised learning of representations, particularly in addition to other vital DL models involving SSL and semi-supervised learning procedures [140,141]. The researcher has implemented a higher-level DIM concept to enhance information representation [142].

### 5.3. Pre-Trained Language Models (PTM)

Regarding the beneficial merits of PTM for SSL models, Han et al. (2021) [143] explained that large-scale pre-trained language models (PTMs), such as BERT and generative pre-trained transformers (GPT), have become a benchmark in developing AI. Knowledge from large amounts of labeled and unlabeled data can be efficiently captured by large-scale PTMs owing to their advanced pretraining objectives and large model parameters. The rich knowledge implicitly contained in numerous parameters can help in a range of downstream activities, as has been thoroughly established through experimental verification and empirical analysis. This is achieved by storing knowledge in large parameters and fine-tuning the individual tasks. The AI community agrees that PTMs, rather than developing models from scratch, should serve as the foundation for subsequent tasks. In this study, they extensively examined the background of pre-training, focusing on its unique relationship with transfer learning and self-supervised learning, to show how pivotal PTMs are in the evolution of AI. In addition, the authors examined PTMs' most recent developments in PTMs in depth. Advances in these four key areas—effective architecture design, context use, computing efficiency, interpretation, and theoretical analysis—have been made possible by the explosion in processing power and the growing availability of data. Figure 17 illustrates the time profile of the emergence of various language-understanding benchmarks linked to the PTM [143].

### 5.4. Contrastive Predictive Coding (CPC)

The CPC can be described as an approach implemented for SSL models to support them in understanding and learning representations in latent embedding spaces using autoregressive models. The CPC seeks to learn from a global, abstract representation of the signal rather than a high-dimensional, low-level representation [144].

Through further investigations on CPC, some scholars, such as [144], explored modified versions of CPC, which was CPCv2 to replace the auto-regressive aspects in the RNN model of CPC, taking into consideration CNN, helping promote the quality of the learned representations for image classification tasks [43,45,145,146].

Ye and Zhao employed CPC for the SSL-based intrusion detection system [147], as illustrated in Figure 18.

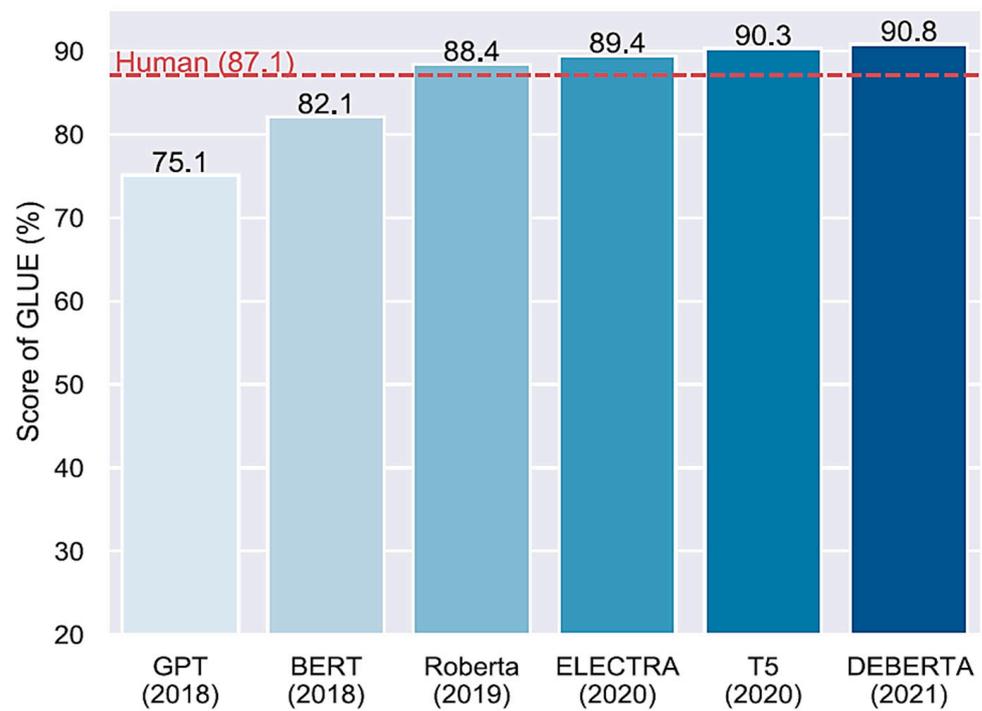


Figure 17. Emergence of various language understanding benchmarks linked to PTM [143].

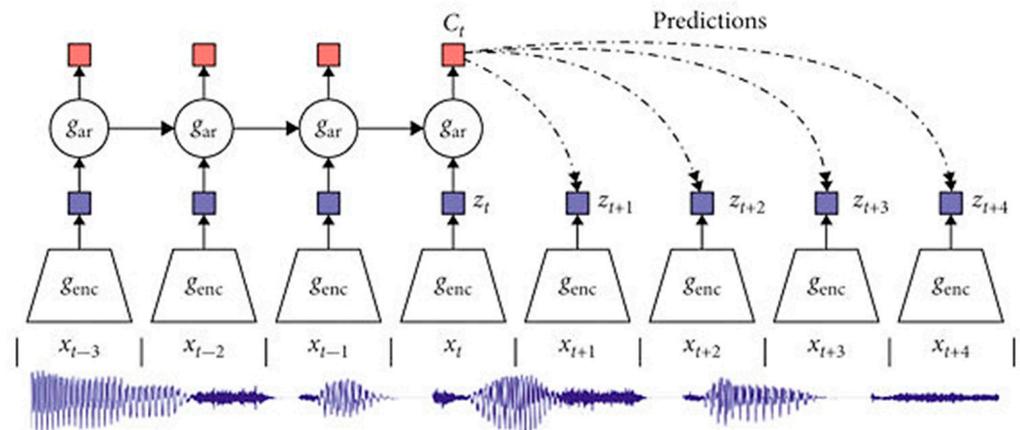


Figure 18. An example of a CPC process adopted for SSL classification task. From Ref. [147], used under Creative Commons CC-BY license.

On the other hand, Henaff (2020) [145] elucidated some prominent merits of CPC in recognizing certain visual data more efficiently compared with SSL models that are trained by raw pixels, which can be explained in Figure 19. In this figure, when a low volume of labeled data is offered, trained SSL models based on raw pixels may fail to generalize, which is indicated by the red line. By training SSL models with the unsupervised representations that are learned by CPC, those models could retain considerable levels of precision within this lower data domain. Those trained SSL models can be expressed as a blue line in the same figure. The precision of SSL models could be attained with a remarkably lower number of labels, which are expressed with horizontal arrows.

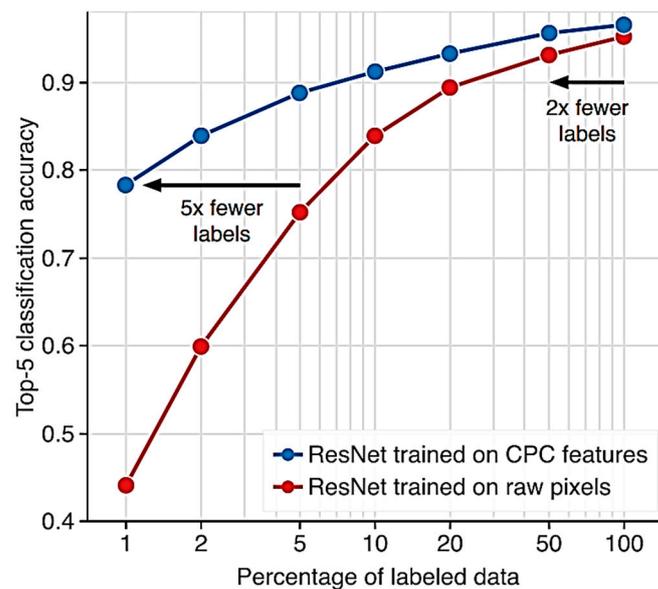


Figure 19. Involving CPC to recognize visual data efficiently [145].

5.5. Autoencoder and Its Associated Extensions

Autoencoders (AEs) and their corresponding extensions are other examples of modern techniques that enable the active implementation of SSL models. Some researchers, including Wang et al. (2020) [148], examined the beneficial impacts of autoencoder integration into the SSL classification task. They reported that by utilizing SSL models, single-channel speech could be enhanced by feeding the network with a noisy mixture and training it to output data closer to the ideal target.

According to Jiang et al. (2017) [112], the AE seeks to learn the function, expressed by the following formula:

$$r(x) = x \tag{34}$$

where  $x$  is the input vector.

The AE learning action is correlated with two major phases: (a) encoding and (b) decoding. In the first phase, the encoder can map the vector, which expresses the data input, into a code vector. The latter can express the input. After this action, the decoder will try to utilize this code vector of the information input to restructure the input vector, providing a lower level of error. In their working principles, the decoder and encoder rely on ANN to complete their tasks. As a result, the output target pertaining to the AE would express the AE itself. The major configurations of the encoder and decoder in the AE could be expressed, respectively, as follows:

$$t_i = f(W_i x_i + b_i) \tag{35}$$

$$r(x_i) = g(W_2 t_i + b_2) \tag{36}$$

where  $i = 1, 2, 3, \dots, L$ .  $L$  expresses the sample number of the raw data.  $x_i \in R^{J \times 1}$ , where  $i$  is  $i^{th}$  sample vector.  $z_i \in R^{K \times 1}$ , and it expresses the pattern or code taken from  $x_i$ .  $W_1 \in R^{K \times J}$ .  $b_1 \in R^{K \times 1}$ .  $b_1$  expresses the weight matrix and bias level between the hidden layer (layer No. 2) and the input layer (layer No. 1).  $b_2 \in R^{J \times 1}$  and  $W_2 \in R^{J \times K}$ .  $b_2$  indicates the bias existing between layers two and three.  $W_2$  is the weight matrix between those two layers as well.

From Figure 20,  $L(x,r)$  is the squared error,  $\theta(t)$  is the reconstruction function.  $\varphi(x + \varepsilon)$  is the projection function that can map the input to the feature space.  $\varepsilon$  expresses a vector through which each index is independent and behaves similarly to the Gaussian distribution that has a variance,  $\sigma_\varepsilon^2$ .

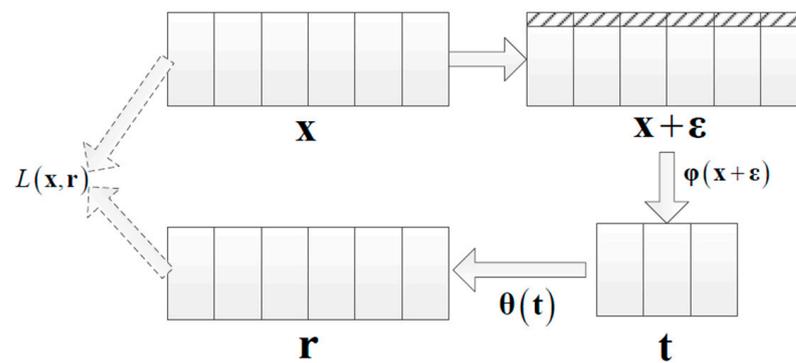


Figure 20. An extensive training process conducted for  $x$  [112].

## 6. Conclusions

This study was carried out in response to the poor classification robustness and weak categorization efficiency of conventional DL and ML models and even modern DL algorithms that have been involved recently in medicine and industry to conduct practical prediction processes. However, because of the huge cost, effort, and time corresponding to data annotation in those two domains, the ML and DL prediction procedures would be considerably challenging. Remarkable R&D revealed a noteworthy SSL that was evolved to enable flexible and efficient classification without referring to arduous data annotation. In addition, SSL was created to overcome another problem reflected in the varying trends and behavior of new data that do not necessarily simulate past documented data. Therefore, when data annotation is fully applied, ML and DL models may not provide important prediction outcomes or classification capabilities.

To shed light on the constructive benefits and substantial contributions of SSL models in facilitating prediction tasks, this paper adopted a comprehensive overview through which variant efficacious applications of two necessary scientific fields were explored, including (a) industry and manufacturing and (b) medicine. Within those two domains, industrial engineers and healthcare providers encounter repetitive obstacles in predicting certain types of faults in machines and ailment situations in patients, respectively. As illustrated here, even if historical databases of machine fault behavior and patient disorders are fully annotated, most ML and DL models fail to perform precise data identification. Relying on the thorough overview implemented in this article, the imperative research findings can be summarized in the following aspects:

- I. Involving SSL algorithms in industrial engineering and clinical contexts could support manufacturing engineers and therapists in carrying out efficient classification procedures and predictions of the current machine fault and patient problems with remarkable levels of performance, accuracy, and feasibility.
- II. Profitable savings in the computational budget, time, storage, and effort needed in the annotation and training of unlabeled data can be eliminated when SSL is utilized, maintaining approximately optimum prediction efficacy.
- III. Functional human thinking, learning approaches, and cognition are utilized in SSL models, contributing to upgraded machine classification and computer prediction outcomes correlated with different fields.

## 7. Future Work

Based on the statistical numerical outcomes and noteworthy ideas obtained from the extensive overview in this paper, the current work proposes some crucial future work perspectives and essential ideas that can help promote SSL prediction potential. The remarkable suggestions that can be taken into consideration are as follows:

- To review the importance of SSL in carrying out accurate predictions pertaining to other scientific domains.

- To overcome some problems not addressed carefully in the literature encountering most SSL models, reflected in SSL trials, analyze and take into consideration solely semantic characteristics linked to the investigated dataset. They do not benefit from critical features existing in visual medical databases.
- To classify other crucial applications of SSL, including either recognition or categorization, not correlated with the relevance of the predictions addressed in this paper.
- To identify other remarkable profitabilities and workable practicalities of SSL other than their contributions to cutting much computational time, budget, and effort for necessary data annotation in the same prediction context.
- To expand this overview with a few case studies in which contributory SSL predictions are carefully explained.

## 8. Research Limitations

In spite of the successful achievement of the meta-analysis and thorough review of various robust SSL applications in industrial and medical contexts, the study encountered a few research constraints that restricted the broad implementation of the extensive review. Those limitations are translated into the following aspects:

- a. Some newly published academic papers (more than 2022) have no direct access to download the overall document. Additionally, some web journals do not have full access to researchers, even for oldly published papers. For this reason, the only extracted data from those articles were the abstract.
- b. There is a lack of abundant databases correlated with the direct applications involved in SSL in machinery prognostics and medical diagnosis.
- c. There were no direct explanations or abundant classifications of major SSL limitations that needed to be addressed and handled.

**Author Contributions:** Conceptualization, M.M.A., N.T.A.R., N.L.F., M.S. (Muhammad Syafrudin), and S.W.L.; methodology, M.M.A., N.T.A.R., N.L.F., M.S. (Muhammad Syafrudin) and S.W.L.; validation, M.M.A., N.T.A.R., A.A.H. and M.S. (Mohammad Salman); formal analysis, N.L.F., D.K.Y. and S.W.L.; investigation, D.K.Y., N.L.F., M.S. (Muhammad Syafrudin) and S.W.L.; data curation, M.M.A., N.T.A.R., A.A.H. and M.S. (Mohammad Salman); writing—original draft preparation, M.M.A., N.T.A.R., A.A.H. and M.S. (Mohammad Salman); writing—review and editing, D.K.Y., N.L.F., M.S. (Muhammad Syafrudin) and S.W.L.; funding acquisition, M.S. (Muhammad Syafrudin) and S.W.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the National Research Foundation of Korea (grant number: NRF2021R1I1A2059735).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Nomenclature

<b>AI</b>	Artificial Intelligence
<b>AE</b>	Autoencoder
<b>AP</b>	Average Precision
<b>APC</b>	Autoregressive Predictive Coding
<b>AUCs</b>	Area under the Curve
<b>AUROC</b>	Area Under the Receiver Operating Characteristic
<b>BERT</b>	Bidirectional Encoder Representations from Transformers
<b>BoW</b>	Bag-of-Visual-Words
<b>BYOL</b>	Bootstrap Your Own Latent
<b>CaiD</b>	Context-Aware instance Discrimination
<b>CERT</b>	Contrastive self-supervised Encoder Representations through Transformers
<b>CNNs</b>	Convolutional Neural Networks
<b>CPC</b>	Contrastive Predictive Coding
<b>CT</b>	Computed Tomography
<b>DCL</b>	Dense Contrastive Learning
<b>DIM</b>	Deep InfoMax

<b>DL</b>	Deep Learning
<b>DNN</b>	Deep Neural Network
<b>DSC</b>	Dice Similarity Coefficient
<b>EHRs</b>	Electronic Health Records
<b>EMA</b>	Exponentially Moving Average
<b>EVs</b>	Electric Vehicles
<b>GAN</b>	Generative Adversarial Network
<b>GPT</b>	Generative Pre-trained Transformer
<b>HVAC</b>	Heating, Ventilation, And Air-Conditioning
<b>IoU</b>	Intersection over Union
<b>Li-ion</b>	Lithium-ion
<b>LMLM</b>	Label-Masked Language Model
<b>LMs</b>	Language Models
<b>LR</b>	Logistic Regression
<b>LSTM</b>	Long Short-Term Memory
<b>MAE</b>	Mean-Absolute-Error
<b>ML</b>	Machine Learning
<b>MLC</b>	Multi-Layer Classifiers
<b>MLM</b>	Masked Language Model
<b>MoCo</b>	Momentum Contrast
<b>MPC</b>	Model Predictive Control
<b>MPQA</b>	Multi-Perspective Question Answering
<b>MRs</b>	Movie Reviews
<b>MSAs</b>	Multiple Sequence Alignments
<b>NAIP</b>	National Agriculture Imagery Program
<b>NLP</b>	Natural Language Processing
<b>NSA</b>	Natural Synthetic Anomalies
<b>PdL</b>	Predictive Learning
<b>pLMs</b>	protein LMs
<b>PPG</b>	Phoneme Posteriorogram
<b>PTM</b>	Pre-trained Language Models
<b>PxL</b>	Pretext Learning
<b>RF</b>	Random Forest
<b>RMSE</b>	Root-Mean-Square-Error
<b>RNN</b>	Recurrent Neural Network
<b>ROC</b>	Receiver Operating Characteristic
<b>RUL</b>	Remaining Useful Life
<b>SL</b>	Supervised Learning
<b>SOC</b>	State of Charge
<b>SSEDD</b>	Self-Supervised Efficient Defect Detector
<b>SSL</b>	Self-Supervised Learning
<b>SST2</b>	Stanford Sentiment Treebank-2
<b>SwAV</b>	Swapping Assignments across Views
<b>TREC</b>	Text Retrieval Conference
<b>USL</b>	Unsupervised Learning
<b>VAE</b>	Variational Auto-Encoders
<b>VC</b>	Voice Conversion
<b>VICReg</b>	Variance, Invariance, and Covariance Regularization

## References

1. Lai, Y. A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition. *J. Phys. Conf. Ser.* **2019**, *1314*, 012148. [[CrossRef](#)]
2. Rezaeianjouybari, B.; Shang, Y. Deep learning for prognostics and health management: State of the art, challenges, and opportunities. *Measurement* **2020**, *163*, 107929. [[CrossRef](#)]
3. Thoppil, N.M.; Vasu, V.; Rao, C.S.P. Deep Learning Algorithms for Machinery Health Prognostics Using Time-Series Data: A Review. *J. Vib. Eng. Technol.* **2021**, *9*, 1123–1145. [[CrossRef](#)]
4. Zhang, L.; Lin, J.; Liu, B.; Zhang, Z.; Yan, X.; Wei, M. A Review on Deep Learning Applications in Prognostics and Health Management. *IEEE Access* **2019**, *7*, 162415–162438. [[CrossRef](#)]

5. Deng, W.; Nguyen, K.T.P.; Medjaher, K.; Gogu, C.; Morio, J. Bearings RUL prediction based on contrastive self-supervised learning. *IFAC-PapersOnLine* **2023**, *56*, 11906–11911. [[CrossRef](#)]
6. Akrim, A.; Gogu, C.; Vingerhoeds, R.; Salatin, M. Self-Supervised Learning for data scarcity in a fatigue damage prognostic problem. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105837. [[CrossRef](#)]
7. Zhuang, J.; Jia, M.; Ding, Y.; Zhao, X. Health Assessment of Rotating Equipment with Unseen Conditions Using Adversarial Domain Generalization Toward Self-Supervised Regularization Learning. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 4675–4685. [[CrossRef](#)]
8. Melendez, I.; Doelling, R.; Bringmann, O. Self-supervised Multi-stage Estimation of Remaining Useful Life for Electric Drive Units. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; IEEE: Amsterdam, The Netherlands, 2019; pp. 4402–4411. [[CrossRef](#)]
9. Von Hahn, T.; Mechefske, C.K. Self-supervised learning for tool wear monitoring with a disentangled-variational-autoencoder. *Int. J. Hydromechatronics* **2021**, *4*, 69. [[CrossRef](#)]
10. Wang, R.; Chen, H.; Guan, C. A self-supervised contrastive learning framework with the nearest neighbors matching for the fault diagnosis of marine machinery. *Ocean. Eng.* **2023**, *270*, 113437. [[CrossRef](#)]
11. Jing, L.; Tian, Y. Self-Supervised Visual Feature Learning with Deep Neural Networks: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 4037–4058. [[CrossRef](#)]
12. Kong, D.; Zhao, L.; Huang, X.; Huang, W.; Ding, J.; Yao, Y.; Xu, L.; Yang, P.; Yang, G. Self-supervised knowledge mining from unlabeled data for bearing fault diagnosis under limited annotations. *Measurement* **2023**, *220*, 113387. [[CrossRef](#)]
13. Chowdhury, A.; Rosenthal, J.; Waring, J.; Umeton, R. Applying Self-Supervised Learning to Medicine: Review of the State of the Art and Medical Implementations. *Informatics* **2021**, *8*, 59. [[CrossRef](#)]
14. Nadif, M.; Role, F. Unsupervised and self-supervised deep learning approaches for biomedical text mining. *Brief. Bioinform.* **2021**, *22*, 1592–1603. [[CrossRef](#)] [[PubMed](#)]
15. Fang, H.; Wang, S.; Zhou, M.; Ding, J.; Xie, P. Cert: Contrastive self-supervised learning for language understanding. *arXiv* **2020**, arXiv:2005.12766.
16. Jaiswal, A.; Babu, A.R.; Zadeh, M.Z.; Banerjee, D.; Makedon, F. A Survey on Contrastive Self-Supervised Learning. *Technologies* **2020**, *9*, 2. [[CrossRef](#)]
17. Shurrab, S.; Duwairi, R. Self-supervised learning methods and applications in medical imaging analysis: A survey. *PeerJ Comput. Sci.* **2022**, *8*, e1045. [[CrossRef](#)] [[PubMed](#)]
18. Ohri, K.; Kumar, M. Review on self-supervised image recognition using deep neural networks. *Knowl. Based Syst.* **2021**, *224*, 107090. [[CrossRef](#)]
19. He, Y.; Carass, A.; Zuo, L.; Dewey, B.E.; Prince, J.L. Autoencoder based self-supervised test-time adaptation for medical image analysis. *Med. Image Anal.* **2021**, *72*, 102136. [[CrossRef](#)] [[PubMed](#)]
20. Huang, S.-C.; Pareek, A.; Jensen, M.; Lungren, M.P.; Yeung, S.; Chaudhari, A.S. Self-supervised learning for medical image classification: A systematic review and implementation guidelines. *NPJ Digit. Med.* **2023**, *6*, 74. [[CrossRef](#)]
21. Baek, S.; Yoon, G.; Song, J.; Yoon, S.M. Self-supervised deep geometric subspace clustering network. *Inf. Sci.* **2022**, *610*, r235–r245. [[CrossRef](#)]
22. Zhang, X.; Mu, J.; Zhang, X.; Liu, H.; Zong, L.; Li, Y. Deep anomaly detection with self-supervised learning and adversarial training. *Pattern Recognit.* **2022**, *121*, 108234. [[CrossRef](#)]
23. Ciga, O.; Xu, T.; Martel, A.L. Self supervised contrastive learning for digital histopathology. *Mach. Learn. Appl.* **2022**, *7*, 100198. [[CrossRef](#)]
24. Liu, Y.; Zhou, S.; Wu, H.; Han, W.; Li, C.; Chen, H. Joint optimization of autoencoder and Self-Supervised Classifier: Anomaly detection of strawberries using hyperspectral imaging. *Comput. Electron. Agric.* **2022**, *198*, 107007. [[CrossRef](#)]
25. Hou, Z.; Liu, X.; Cen, Y.; Dong, Y.; Yang, H.; Wang, C.; Tang, J. GraphMAE: Self-Supervised Masked Graph Autoencoders. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 14–18 August 2022; ACM: New York, NY, USA, 2022; pp. 594–604. [[CrossRef](#)]
26. Li, Y.; Lao, Q.; Kang, Q.; Jiang, Z.; Du, S.; Zhang, S.; Li, K. Self-supervised anomaly detection, staging and segmentation for retinal images. *Med. Image Anal.* **2023**, *87*, 102805. [[CrossRef](#)]
27. Wang, T.; Wu, J.; Zhang, Z.; Zhou, W.; Chen, G.; Liu, S. Multi-scale graph attention subspace clustering network. *Neurocomputing* **2021**, *459*, 302–314. [[CrossRef](#)]
28. Li, J.; Ren, W.; Han, M. Variational auto-encoders based on the shift correction for imputation of specific missing in multivariate time series. *Measurement* **2021**, *186*, 110055. [[CrossRef](#)]
29. Sun, C. HAT-GAE: Self-Supervised Graph Auto-encoders with Hierarchical Adaptive Masking and Trainable Corruption. *arXiv* **2023**, arXiv:2301.12063. [[CrossRef](#)]
30. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R. Masked Autoencoders Are Scalable Vision Learners. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*; Ernest N. Morial Convention Center: New Orleans, LA, USA; IEEE: Amsterdam, The Netherlands, 2022; pp. 16000–16009.
31. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.

32. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]
33. Fekri, M.N.; Ghosh, A.M.; Grolinger, K. Generating Energy Data for Machine Learning with Recurrent Generative Adversarial Networks. *Energies* **2019**, *13*, 130. [[CrossRef](#)]
34. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
35. Berg, P.; Pham, M.-T.; Courty, N. Self-Supervised Learning for Scene Classification in Remote Sensing: Current State of the Art and Perspectives. *Remote Sens.* **2022**, *14*, 3995. [[CrossRef](#)]
36. Doersch, C.; Gupta, A.; Efros, A.A. Unsupervised Visual Representation Learning by Context Prediction. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; IEEE: Amsterdam, The Netherlands, 2015; pp. 1422–1430.
37. Zhang, R.; Isola, P.; Efros, A.A. Colorful Image Colorization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 649–666. [[CrossRef](#)]
38. Gidaris, S.; Singh, P.; Komodakis, N. Unsupervised Representation Learning by Predicting Image Rotations. In Proceedings of the Sixth International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018; ICLR 2018. Cornell University: Ithaca, NY, USA, 2018.
39. Noroozi, M.; Favaro, P. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2016; pp. 69–84. [[CrossRef](#)]
40. Dosovitskiy, A.; Springenberg, J.T.; Riedmiller, M.; Brox, T. Discriminative unsupervised feature learning with convolutional neural networks. *Adv. Neural Inf. Process Syst.* **2014**, *27*, 1–9. [[CrossRef](#)] [[PubMed](#)]
41. Lee, C.P.; Lim, K.M.; Song, Y.X.; Alqahtani, A. Plant-CNN-ViT: Plant Classification with Ensemble of Convolutional Neural Networks and Vision Transformer. *Plants* **2023**, *12*, 2642. [[CrossRef](#)]
42. Dong, X.; Shen, J. Triplet Loss in Siamese Network for Object Tracking. In *European Conference on Computer Vision (ECCV)*; Springer: Munich, Germany, 2018; pp. 459–474.
43. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In *37th International Conference on Machine Learning, PMLR 119*; PLMR: Vienna, Austria, 2020; pp. 1597–1607.
44. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226. [[CrossRef](#)]
45. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Seattle, WA, USA, 2020; pp. 9729–9738.
46. Li, X.; Zhou, Y.; Zhang, Y.; Zhang, A.; Wang, W.; Jiang, N.; Wu, H.; Wang, W. Dense Semantic Contrast for Self-Supervised Visual Representation Learning. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual, 20–24 October 2021; ACM: New York, NY, USA, 2021; pp. 1368–1376. [[CrossRef](#)]
47. Fini, E.; Astolfi, P.; Alahari, K.; Alameda-Pineda, X.; Mairal, J.; Nabi, M.; Ricci, E. Semi-Supervised Learning Made Simple with Self-Supervised Clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; IEEE: New York, NY, USA, 2023; pp. 3187–3197.
48. Khan, A.; AlBarri, S.; Manzoor, M.A. Contrastive Self-Supervised Learning: A Survey on Different Architectures. In Proceedings of the 2022 2nd International Conference on Artificial Intelligence (ICAI), Islamabad, Pakistan, 30–31 March 2022; IEEE: New York, NY, USA, 2022; pp. 1–6. [[CrossRef](#)]
49. Liu, Y.; Zhu, L.; Yamada, M.; Yang, Y. Semantic Correspondence as an Optimal Transport Problem. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2020; pp. 4462–4471. [[CrossRef](#)]
50. Shvetsova, N.; Petersen, F.; Kukleva, A.; Schiele, B.; Kuehne, H. Learning by Sorting: Self-supervised Learning with Group Ordering Constraints. *arXiv* **2023**, arXiv:2301.02009.
51. Li, H.; Liu, J.; Cui, L.; Huang, H.; Tai, X.-C. Volume preserving image segmentation with entropy regularized optimal transport and its applications in deep learning. *J. Vis. Commun. Image Represent.* **2020**, *71*, 102845. [[CrossRef](#)]
52. Li, R.; Lin, G.; Xie, L. Self-Point-Flow: Self-Supervised Scene Flow Estimation from Point Clouds with Optimal Transport and Random Walk. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 19–25 June 2021; IEEE: New York, NY, USA, 2021; pp. 15577–15586.
53. Scetbon, M.; Cuturi, M. Low-rank optimal transport: Approximation, statistics and debiasing. *Adv. Neural Inf. Process Syst.* **2022**, *35*, 6802–6814.
54. Zhang, C.; Zhang, C.; Zhang, K.; Zhang, C.; Niu, A.; Feng, J.; Yoo, C.D.; Kweon, I.S. Decoupled Adversarial Contrastive Learning for Self-supervised Adversarial Robustness. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2022; pp. 725–742. [[CrossRef](#)]
55. Liu, W.; Li, Z.; Zhang, H.; Chang, S.; Wang, H.; He, J.; Huang, Q. Dense lead contrast for self-supervised representation learning of multilead electrocardiograms. *Inf. Sci.* **2023**, *634*, 189–205. [[CrossRef](#)]
56. Wang, X.; Zhang, R.; Shen, C.; Kong, T. DenseCL: A simple framework for self-supervised dense visual pre-training. *Vis. Inform.* **2023**, *7*, 30–40. [[CrossRef](#)]

57. Liu, X.; Sinha, A.; Unberath, M.; Ishii, M.; Hager, G.D.; Taylor, R.H.; Reiter, A. Self-Supervised Learning for Dense Depth Estimation in Monocular Endoscopy. In *OR 2.0 Context-Aware. Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin. Image Analysis: First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5*; Springer International Publishing: Cham, Switzerland, 2018; pp. 128–138. [[CrossRef](#)]
58. Kar, S.; Nagasubramanian, K.; Elango, D.; Nair, A.; Mueller, D.S.; O’Neal, M.E.; Singh, A.K.; Sarkar, S.; Ganapathysubramanian, B.; Singh, A. Self-Supervised Learning Improves Agricultural Pest Classification. In *Proceedings of the AI for Agriculture and Food Systems*, Vancouver, BC, Canada, 28 February 2021.
59. Niizumi, D.; Takeuchi, D.; Ohishi, Y.; Harada, N.; Kashino, K. BYOL for Audio: Self-Supervised Learning for General-Purpose Audio Representation. In *Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, 18–22 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–8. [[CrossRef](#)]
60. Grill, J.B.; Strub, F.; Althé, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Pires, B.A.; Guo, Z.D.; Azar, M.G.; et al. Bootstrap your own latent—a new approach to self-supervised learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21271–21284.
61. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531.
62. Wang, J.; Zhu, T.; Gan, J.; Chen, L.L.; Ning, H.; Wan, Y. Sensor Data Augmentation by Resampling in Contrastive Learning for Human Activity Recognition. *IEEE Sens. J.* **2022**, *22*, 22994–23008. [[CrossRef](#)]
63. Wu, J.; Gong, X.; Zhang, Z. Self-Supervised Implicit Attention: Guided Attention by The Model Itself. *arXiv* **2022**, arXiv:2206.07434.
64. Haresamudram, H.; Essa, I.; Plötz, T. Investigating Enhancements to Contrastive Predictive Coding for Human Activity Recognition. In *Proceedings of the 2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Atlanta, GA, USA, 13–17 March 2023; IEEE: New York, NY, USA, 2023; pp. 232–241. [[CrossRef](#)]
65. Caron, M.; Touvron, H.; Misra, I.; Jegou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. Emerging Properties in Self-Supervised Vision Transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2021; pp. 9630–9640. [[CrossRef](#)]
66. Balestriero, R.; Ibrahim, M.; Sobal, V.; Morcos, A.; Shekhar, S.; Goldstein, T.; Bordes, F.; Bardes, A.; Mialon, G.; Tian, Y.; et al. A cookbook of self-supervised learning. *arXiv* **2023**, arXiv:2304.12210.
67. Chen, Y.; Liu, Y.; Jiang, D.; Zhang, X.; Dai, W.; Xiong, H.; Tian, Q. SdAE: Self-distilled Masked Autoencoder. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2022; pp. 108–124. [[CrossRef](#)]
68. Alfaro-Contreras, M.; Ríos-Vila, A.; Valero-Mas, J.J.; Calvo-Zaragoza, J. Few-shot symbol classification via self-supervised learning and nearest neighbor. *Pattern Recognit. Lett.* **2023**, *167*, 1–8. [[CrossRef](#)]
69. Lee, D.; Aune, E. VbCReg: Variance-invariance-better-covariance regularization for self-supervised learning on time series. *arXiv* **2021**, arXiv:2109.00783.
70. Mialon, G.; Balestriero, R.; LeCun, Y. Variance covariance regularization enforces pairwise independence in self-supervised representations. *arXiv* **2022**, arXiv:2209.14905.
71. Bardes, A.; Ponce, J.; LeCun, Y. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. *arXiv* **2021**, arXiv:2105.04906.
72. Chen, S.; Guo, W. Auto-Encoders in Deep Learning—A Review with New Perspectives. *Mathematics* **2023**, *11*, 1777. [[CrossRef](#)]
73. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
74. Liu, Z.; Miao, Z.; Zhan, X.; Wang, J.; Gong, B.; Yu, S.X. Open Long-Tailed Recognition in A Dynamic World. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *46*, 1836–1851. [[CrossRef](#)] [[PubMed](#)]
75. Liu, Y.; Jin, M.; Pan, S.; Zhou, C.; Zheng, Y.; Xia, F.; Yu, P. Graph Self-Supervised Learning: A Survey. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 5879–5900. [[CrossRef](#)]
76. Krishnan, R.; Rajpurkar, P.; Topol, E.J. Self-supervised learning in medicine and healthcare. *Nat. Biomed. Eng.* **2022**, *6*, 1346–1352. [[CrossRef](#)] [[PubMed](#)]
77. Azizi, S.; Mustafa, B.; Ryan, F.; Beaver, Z.; Freyberg, J.; Deaton, J.; Loh, A.; Karthikesalingam, A.; Kornblith, S.; Chen, T.; et al. Big Self-Supervised Models Advance Medical Image Classification. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2021; pp. 3458–3468. [[CrossRef](#)]
78. Zhang, Y.; Jiang, H.; Miura, Y.; Manning, C.D.; Langlotz, C.P. Contrastive learning of medical visual representations from paired images and text. In *Machine Learning for Healthcare Conference*; PMLR: Vienna, Austria, 2022; pp. 2–25.
79. Bozorgtabar, B.; Mahapatra, D.; Vray, G.; Thiran, J.-P. SALAD: Self-supervised Aggregation Learning for Anomaly Detection on X-rays. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, 4–8 October 2020, Proceedings, Part I 23*; Springer International Publishing: Cham, Switzerland, 2020; pp. 468–478. [[CrossRef](#)]
80. Tian, Y.; Pang, G.; Liu, F.; Chen, Y.; Shin, S.H.; Verjans, J.W.; Singh, R.; Carneiro, G. Constrained Contrastive Distribution Learning for Unsupervised Anomaly Detection and Localisation in Medical Images. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021, Proceedings, Part V 24*; Springer International Publishing: Cham, Switzerland, 2021; pp. 128–140. [[CrossRef](#)]

81. Ouyang, J.; Zhao, Q.; Adeli, E.; Sullivan, E.V.; Pfefferbaum, A.; Zaharchuk, G.; Pohl, K.M. Self-supervised Longitudinal Neighbourhood Embedding. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021, Proceedings, Part II 24*; Springer International Publishing: Cham, Switzerland, 2021; pp. 80–89. [[CrossRef](#)]
82. Liu, F.; Tian, Y.; Cordeiro, F.R.; Belagiannis, V.; Reid, I.; Carneiro, G. Self-supervised Mean Teacher for Semi-supervised Chest X-ray Classification. In *International Workshop on Machine Learning in Medical Imaging*; Springer International Publishing: Cham, Switzerland, 2021; pp. 426–436. [[CrossRef](#)]
83. Li, H.; Xue, F.F.; Chaitanya, K.; Luo, S.; Ezhov, I.; Wiestler, B.; Zhang, J.; Menze, B. Imbalance-Aware Self-supervised Learning for 3D Radiomic Representations. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021, Proceedings, Part II 24*; Springer International Publishing: Cham, Switzerland, 2021; pp. 36–46. [[CrossRef](#)]
84. Manna, S.; Bhattacharya, S.; Pal, U. Interpretive Self-Supervised pre-Training. In *Twelfth Indian Conference on Computer Vision, Graphics and Image Processing*; ACM: New York, NY, USA, 2021; pp. 1–9. [[CrossRef](#)]
85. Zhao, Z.; Yang, G. Unsupervised Contrastive Learning of Radiomics and Deep Features for Label-Efficient Tumor Classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021, Proceedings, Part II 24*; Springer International Publishing: Cham, Switzerland, 2021; pp. 252–261. [[CrossRef](#)]
86. Esrafilian-Najafabadi, M.; Haghghat, F. Towards self-learning control of HVAC systems with the consideration of dynamic occupancy patterns: Application of model-free deep reinforcement learning. *Build. Environ.* **2022**, *226*, 109747. [[CrossRef](#)]
87. Long, J.; Chen, Y.; Yang, Z.; Huang, Y.; Li, C. A novel self-training semi-supervised deep learning approach for machinery fault diagnosis. *Int. J. Prod. Res.* **2023**, *61*, 8238–8251. [[CrossRef](#)]
88. Yang, Z.; Huang, Y.; Nazeer, F.; Zi, Y.; Valentino, G.; Li, C.; Long, J.; Huang, H. A novel fault detection method for rotating machinery based on self-supervised contrastive representations. *Comput. Ind.* **2023**, *147*, 103878. [[CrossRef](#)]
89. Wei, M.; Liu, Y.; Zhang, T.; Wang, Z.; Zhu, J. Fault Diagnosis of Rotating Machinery Based on Improved Self-Supervised Learning Method and Very Few Labeled Samples. *Sensors* **2021**, *22*, 192. [[CrossRef](#)]
90. Lei, Y.; Karimi, H.R.; Chen, X. A novel self-supervised deep LSTM network for industrial temperature prediction in aluminum processes application. *Neurocomputing* **2022**, *502*, 177–185. [[CrossRef](#)]
91. Xu, R.; Hao, R.; Huang, B. Efficient surface defect detection using self-supervised learning strategy and segmentation network. *Adv. Eng. Inform.* **2022**, *52*, 101566. [[CrossRef](#)]
92. Bharti, V.; Kumar, A.; Purohit, V.; Singh, R.; Singh, A.K.; Singh, S.K. A Label Efficient Semi Self-Supervised Learning Framework for IoT Devices in Industrial Process. *IEEE Trans. Ind. Inform.* **2023**, *20*, 2253–2262. [[CrossRef](#)]
93. Hannan, M.A.; How, D.N.T.; Lipu, M.S.H.; Mansor, M.; Ker, P.J.; Dong, Z.Y.; Sahari, K.S.M.; Tiong, S.K.; Muttaqi, K.M.; Mahlia, T.M.I.; et al. Deep learning approach towards accurate state of charge estimation for lithium-ion batteries using self-supervised transformer model. *Sci. Rep.* **2021**, *11*, 19541. [[CrossRef](#)] [[PubMed](#)]
94. Li, C.L.; Sohn, K.; Yoon, J.; Pfister, T. CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, North America*; IEEE: Washington, DC, USA, 2021; pp. 9664–9674.
95. Yang, C.; Wu, Z.; Zhou, B.; Lin, S. Instance Localization for Self-Supervised Detection Pretraining. In *CVF Conference on Computer Vision and Pattern Recognition, North America*; IEEE: Washington, DC, USA, 2021; pp. 3987–3996.
96. Schlüter, H.M.; Tan, J.; Hou, B.; Kainz, B. Natural Synthetic Anomalies for Self-supervised Anomaly Detection and Localization. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2022; pp. 474–489. [[CrossRef](#)]
97. Taher, M.R.H.; Haghghi, F.; Gotway, M.B.; Liang, J. CAiD: Context-Aware Instance Discrimination for Self-Supervised Learning in Medical Imaging. In *International Conference on Medical Imaging with Deep Learning*; MIDL Foundation: Zürich, Switzerland, 2022; pp. 535–551.
98. Gidaris, S.; Bursuc, A.; Puy, G.; Komodakis, N.; Cord, M.; Pérez, P. Obow: Online Bag-of-Visual-Words Generation for Self-Supervised Learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, North America*; IEEE: Washington, DC, USA, 2021; pp. 6830–6840.
99. Baevski, A.; Babu, A.; Hsu, W.N.; Auli, M. Efficient Self-Supervised Learning with Contextualized Target Representations for Vision, Speech and Language. In *Proceedings of the 40th International Conference on Machine Learning, PMLR 2023, Honolulu, HI, USA, 23–29 July 2023*; PMLR: Vienna, Austria, 2023; pp. 1416–1429.
100. Park, D.; Ahn, C.W. Self-Supervised Contextual Data Augmentation for Natural Language Processing. *Symmetry* **2019**, *11*, 1393. [[CrossRef](#)]
101. Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; et al. ProfTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 7112–7127. [[CrossRef](#)]
102. Lin, J.H.; Lin, Y.Y.; Chien, C.M.; Lee, H.Y. S2VC: A Framework for Any-to-Any Voice Conversion with Self-Supervised Pretrained Representations. *arXiv* **2021**, arXiv:2104.02901.
103. Chung, Y.-A.; Hsu, W.-N.; Tang, H.; Glass, J. An Unsupervised Autoregressive Model for Speech Representation Learning. In *Interspeech 2019*; ISCA: Singapore, 2019; pp. 146–150. [[CrossRef](#)]

104. Pan, T.; Chen, J.; Zhou, Z.; Wang, C.; He, S. A Novel Deep Learning Network via Multiscale Inner Product with Locally Connected Feature Extraction for Intelligent Fault Detection. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5119–5128. [[CrossRef](#)]
105. Chen, J.; Wang, C.; Wang, B.; Zhou, Z. A visualized classification method via t-distributed stochastic neighbor embedding and various diagnostic parameters for planetary gearbox fault identification from raw mechanical data. *Sens. Actuators A Phys.* **2018**, *284*, 52–65. [[CrossRef](#)]
106. Zhang, T.; Chen, J.; He, S.; Zhou, Z. Prior Knowledge-Augmented Self-Supervised Feature Learning for Few-Shot Intelligent Fault Diagnosis of Machines. *IEEE Trans. Ind. Electron.* **2022**, *69*, 10573–10584. [[CrossRef](#)]
107. Yuan, Y.; Lin, L. Self-Supervised Pretraining of Transformers for Satellite Image Time Series Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 474–487. [[CrossRef](#)]
108. Zhao, Z.; Luo, Z.; Li, J.; Chen, C.; Piao, Y. When Self-Supervised Learning Meets Scene Classification: Remote Sensing Scene Classification Based on a Multitask Learning Framework. *Remote Sens.* **2020**, *12*, 3276. [[CrossRef](#)]
109. Tao, C.; Qi, J.; Zhang, G.; Zhu, Q.; Lu, W.; Li, H. TOV: The Original Vision Model for Optical Remote Sensing Image Understanding via Self-Supervised Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4916–4930. [[CrossRef](#)]
110. Stojnic, V.; Risojevic, V. Evaluation of Split-Brain Autoencoders for High-Resolution Remote Sensing Scene Classification. In Proceedings of the 2018 International Symposium ELMAR, Zadar, Croatia, 16–19 September 2018; IEEE: New York, NY, USA, 2018; pp. 67–70. [[CrossRef](#)]
111. Jung, H.; Jeon, T. Self-supervised learning with randomised layers for remote sensing. *Electron. Lett.* **2021**, *57*, 249–251. [[CrossRef](#)]
112. Jiang, L.; Song, Z.; Ge, Z.; Chen, J. Robust Self-Supervised Model and Its Application for Fault Detection. *Ind. Eng. Chem. Res.* **2017**, *56*, 7503–7515. [[CrossRef](#)]
113. Yu, Z.; Lei, N.; Mo, Y.; Xu, X.; Li, X.; Huang, B. Feature extraction based on self-supervised learning for RUL prediction. *J. Comput. Inf. Sci. Eng.* **2023**, *24*, 021004. [[CrossRef](#)]
114. Hu, C.; Wu, J.; Sun, C.; Yan, R.; Chen, X. Interinstance and Intratemporal Self-Supervised Learning With Few Labeled Data for Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2023**, *19*, 6502–6512. [[CrossRef](#)]
115. Huang, C.; Wang, X.; He, X.; Yin, D. Self-Supervised Learning for Recommender System. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; ACM: New York, NY, USA, 2022; pp. 3440–3443. [[CrossRef](#)]
116. Wang, T.; Qiao, M.; Zhang, M.; Yang, Y.; Snoussi, H. Data-driven prognostic method based on self-supervised learning approaches for fault detection. *J. Intell. Manuf.* **2020**, *31*, 1611–1619. [[CrossRef](#)]
117. Nair, A.; Chen, D.; Agrawal, P.; Isola, P.; Abbeel, P.; Malik, J.; Levine, S. Combining Self-Supervised Learning and Imitation for Vision-Based Rope Manipulation. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; IEEE: New York, NY, USA, 2017; pp. 2146–2153. [[CrossRef](#)]
118. Ren, L.; Wang, T.; Laili, Y.; Zhang, L. A Data-Driven Self-Supervised LSTM-DeepFM Model for Industrial Soft Sensor. *IEEE Trans. Ind. Inform.* **2022**, *18*, 5859–5869. [[CrossRef](#)]
119. Senanayaka, J.S.L.; Van Khang, H.; Robbersmyr, K.G. Toward Self-Supervised Feature Learning for Online Diagnosis of Multiple Faults in Electric Powertrains. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3772–3781. [[CrossRef](#)]
120. Berscheid, L.; Meisner, P.; Kroger, T. Self-Supervised Learning for Precise Pick-and-Place Without Object Model. *IEEE Robot. Autom. Lett.* **2020**, *5*, 4828–4835. [[CrossRef](#)]
121. Geng, H.; Yang, F.; Zeng, X.; Yu, B. When Wafer Failure Pattern Classification Meets Few-shot Learning and Self-Supervised Learning. In Proceedings of the 2021 IEEE/ACM International Conference on Computer Aided Design (ICCAD), Munich, Germany, 1–4 November 2021; IEEE: New York, NY, USA, 2021; pp. 1–8. [[CrossRef](#)]
122. Yoa, S.; Lee, S.; Kim, C.; Kim, H.J. Self-Supervised Learning for Anomaly Detection with Dynamic Local Augmentation. *IEEE Access* **2021**, *9*, 147201–147211. [[CrossRef](#)]
123. Li, J.; Huang, R.; Chen, J.; Xia, J.; Chen, Z.; Li, W. Deep Self-Supervised Domain Adaptation Network for Fault Diagnosis of Rotating Machine with Unlabeled Data. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3510509. [[CrossRef](#)]
124. Lu, N.; Xiao, H.; Ma, Z.; Yan, T.; Han, M. Domain Adaptation with Self-Supervised Learning and Feature Clustering for Intelligent Fault Diagnosis. *IEEE Trans. Neural Netw. Learn Syst.* **2022**, 1–14. [[CrossRef](#)]
125. Ding, Y.; Zhuang, J.; Ding, P.; Jia, M. Self-supervised pretraining via contrast learning for intelligent incipient fault detection of bearings. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108126. [[CrossRef](#)]
126. Yan, Z.; Liu, H. SMOCo: A Powerful and Efficient Method Based on Self-Supervised Learning for Fault Diagnosis of Aero-Engine Bearing under Limited Data. *Mathematics* **2022**, *10*, 2796. [[CrossRef](#)]
127. Chen, L.; Bentley, P.; Mori, K.; Misawa, K.; Fujiwara, M.; Rueckert, D. Self-supervised learning for medical image analysis using image context restoration. *Med. Image Anal.* **2019**, *58*, 101539. [[CrossRef](#)]
128. Nguyen, X.-B.; Lee, G.S.; Kim, S.H.; Yang, H.J. Self-Supervised Learning Based on Spatial Awareness for Medical Image Analysis. *IEEE Access* **2020**, *8*, 162973–162981. [[CrossRef](#)]
129. Jamaludin, A.; Kadir, T.; Zisserman, A. Self-supervised Learning for Spinal MRIs. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*; Springer International Publishing: Cham, Switzerland, 2017; pp. 294–302. [[CrossRef](#)]

130. Zhu, J.; Li, Y.; Hu, Y.; Zhou, S.K. Embedding task knowledge into 3D neural networks via self-supervised learning. *arXiv* **2020**, arXiv:2006.05798.
131. Xie, Y.; Zhang, J.; Liao, Z.; Xia, Y.; Shen, C. PGL: Prior-guided local self-supervised learning for 3D medical image segmentation. *arXiv* **2020**, arXiv:2011.12640.
132. Li, X.; Jia, M.; Islam, M.T.; Yu, L.; Xing, L. Self-Supervised Feature Learning via Exploiting Multi-Modal Data for Retinal Disease Diagnosis. *IEEE Trans. Med. Imaging* **2020**, *39*, 4023–4033. [[CrossRef](#)] [[PubMed](#)]
133. Sowrirajan, H.; Yang, J.; Ng, A.Y.; Rajpurkar, P. MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models. In *Machine Learning Research 143*; Stanford University: Stanford, CA, USA, 2021; pp. 728–744.
134. Vu, Y.N.T.; Wang, R.; Balachandar, N.; Liu, C.; Ng, A.Y.; Rajpurkar, P. Medaug: Contrastive Learning Leveraging Patient Metadata Improves Representations for Chest X-ray Interpretation. In Proceedings of the 6th Machine Learning for Healthcare Conference, Virtual, 6–7 August 2021; Doshi-Velez, F., Ed.; PMLR: Vienna, Austria, 2021; pp. 755–769.
135. Sriram, A.; Muckley, M.; Sinha, K.; Shamout, F.; Pineau, J.; Geras, K.J.; Azour, L.; Aphinyanaphongs, Y.; Yakubova, N.; Moore, W. COVID-19 prognosis via self-supervised representation learning and multi-image prediction. *arXiv* **2021**, arXiv:2101.04909.
136. Chen, X.; Yao, L.; Zhou, T.; Dong, J.; Zhang, Y. Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. *Pattern Recognit.* **2021**, *113*, 107826. [[CrossRef](#)]
137. Chaitanya, K.; Erdil, E.; Karani, N.; Konukoglu, E. Contrastive learning of global and local features for medical image segmentation with limited annotations. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12546–12558.
138. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–9.
139. Hjelm, R.D.; Fedorov, A.; Lavoie-Marchildon, S.; Bachman, P.; Trischler, A.; Bengio, Y. Learning Deep Representations by Mutual Information Estimation and Maximization. In Proceedings of the Seventh International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019. ICLR Committee 2019.
140. Oord, A.V.D.; Li, Y.; Vinyals, O. Representation Learning with Contrastive Predictive Coding. *arXiv* **2019**, arXiv:1807.03748.
141. Bachman, P.; Hjelm, R.D.; Buchwalter, W. Learning representations by maximizing mutual information across views. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
142. Veličković, P.; Fedus, W.; Hamilton, W.L.; Liò, P.; Bengio, Y.; Hjelm, R.D. Deep Graph Infomax. In Proceedings of the Seventh International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019. ICLR Committee 2019.
143. Han, X.; Zhang, Z.; Ding, N.; Gu, Y.; Liu, X.; Huo, Y.; Qiu, J.; Yao, Y.; Zhang, A.; Zhang, L.; et al. Pre-trained models: Past, present and future. *AI Open* **2021**, *2*, 225–250. [[CrossRef](#)]
144. Deldari, S.; Smith, D.V.; Xue, H.; Salim, F.D. Time Series Change Point Detection with Self-Supervised Contrastive Predictive Coding. In *Web Conference 2021*; ACM: New York, NY, USA, 2021; pp. 3124–3135. [[CrossRef](#)]
145. Henaff, O. Data-Efficient Image Recognition with Contrastive Predictive Coding. In Proceedings of the 37th International Conference on Machine Learning, PMLR, Vienna, Austria, 13–18 July 2020; PMLR: Vienna, Austria, 2020; pp. 4182–4192.
146. Tian, Y.; Krishnan, D.; Isola, P. Contrastive Multiview Coding. *arXiv* **2019**, arXiv:1906.05849.
147. Ye, F.; Zhao, W. A Semi-Self-Supervised Intrusion Detection System for Multilevel Industrial Cyber Protection. *Comput. Intell. Neurosci.* **2022**, *2022*, 4043309. [[CrossRef](#)]
148. Wang, Y.C.; Venkataramani, S.; Smaragdis, P. Self-supervised learning for speech enhancement. *arXiv* **2020**, arXiv:2006.10388.

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