Training with less medical data: A survey.

Raja Haseeb
raja@rit.kaist.ac.kr
Korean Advanced Institute of Science and Technology
Daejeon, South Korea

Abstract

Healthcare is one of the most important and essential components of human life. It is a very high-priority sector and people expect the best services and care possible. In recent years, due to their immense success in other real-world applications, machine learning, and Big Data have found their way into the medical sector as well. Machine learning helps improve the accuracy of the diagnosis and also speeds up the overall process. With the advent of Big Data and AI techniques, it became easier to deal with the huge amounts of patient data. Researchers have been working on applying machine learning techniques in combination with Big Data in the medical diagnosis of various diseases like a tumor, dementia, cancer, and so on. However, in many cases, these algorithms still require a huge amount of annotated data for training. In reality, we don't always have large data available, more so an annotated one, as it is costly and you need to hire medical experts for that purpose. Recently, researchers have been working on dealing with this problem of small training data. Based on some of the recent works, this paper presents a survey of various techniques for medical data, which can be used to mitigate the curse of small datasets.

Keywords: Healthcare, machine learning, deep learning, small dataset, deep learning for medical diagnosis

1 Introduction

This is the age of machine learning and artificial intelligence. These are the most exciting fields for computer science and engineering these days. It is the simulation of human perception and intelligence by computer systems. These days, everyone is utilizing modern machine learning tools somehow. The motivation behind this is pretty obvious; computers and machines will perform repetitive tasks that can be fatiguing and tiresome for human beings. Recently, machines have shown tremendous success in the implementation and learning of very complex tasks, and have shown promising results in the decision-making process. In all this, the most exciting part is the finding that in some cases, machines and computers can identify patterns that are even beyond the human perception [35]. The paradigm of machine learning mainly involves learning from experience. Given a set of training examples, the algorithm can learn the underlying features of data and can predict new future data. In recent decades, various algorithms have been designed for the implementation of machine learning. These algorithms are broadly classified into four main categories i.e. supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.

In supervised learning, examples are used to teach the machine. The scientist prepares a set of training examples called the training dataset and feeds it into the network. The correct answer to the problem is known to the operator, and the algorithm identifies patterns in data, learns from observations, and makes predictions. Based on the training dataset, the output for the new input value is predicted [60]. The operator then tries to tune the parameters or improve the data based on the model predictions, and this whole process continues until the algorithm achieves a high level of accuracy/performance. The output of the network can be a discrete value or continuous based on the task. Discrete value is for the classification tasks and continuous values are in the case of regression problems. In classification tasks, the model is usually provided with a set of examples along with the true labels for each category. After training the model is then applied to test data and it predicts which category the new example belongs to. Regression is more like estimation. The program tries to understand the relationship between different variables. Regression analysis usually involves a dependant variable and series of other changing variables and is particularly useful for prediction and forecasting problems. Examples of supervised learning algorithms include Decision Tree, Random Forest, KNN, Logistic Regression, and so on.

Unsupervised learning, as the name suggests, has no supervision i.e. data provided to the network is not labeled. So the data is organized by the network itself, and it identifies the underlying patterns and correlations by analysis of the given data. Unsupervised Learning algorithms predict the results based on the similarities between the input [109]. So the algorithm usually tries to sort out the data in some kind of useful structure. There are two main categories for unsupervised learning tasks i.e. clustering and dimensionality reduction. Clustering involves grouping the data into different groups based on their similarities and differences. Dimension reduction is used to reduce the number of variables in case of complex data, and to extract the most useful information which makes the task simpler.

Semi-supervised learning is somewhat between the supervised and unsupervised method. It utilized both labeled and unlabeled data. So the model can use the labeled data to develop labels for the unlabeled data [193]. It is mainly used for supervised learning tasks when there is a shortage of labeled data or the labeling process is costly. The oldest form of this algorithm is the self-training model [23]. Examples include generative models, graph-based models, or some two-step models with unsupervised learning followed by a supervised learning algorithm.

In reinforcement learning, the machine is training for decision-making purposes. the machine is presented with an environment, where it explores and exploits the surroundings and trains itself in a continuous manner by trial and error. The machine is provided with a set of actions, and a reward is associated with each action. However, the agent is also penalized in case of a wrong action. After experience and time, and a certain number of actions, the agent finally figures out the optimal set of actions that result in the desired output. the algorithm follows the Markov Decision Process (MDP).

Deep Learning [172][144][92][50][103] is the latest paradigm in machine learning domain where involves high-level abstraction of features. In the past few years, deep learning techniques completely transformed the way researchers and data scientists thought and achieved state-of-the-art results in tasks across multiple domains. It is a type of neural network, instead of linear mapping utilizes convolution filters and many hidden layers along with various activations. Numerous deep learning architectures have been developed in recent years including convolutional neural networks (CNNs) [93][88], recurrent neural networks (RNNs) [99], recursive neural networks [161], autoencoders [114] and Generative Adversarial Networks (GANs) [51]. The details of deep learning models and architectures are out of the scope of this paper, however, some important and papers are referenced in this section.

Deep learning can be used for various tasks. These tasks include classification [88][155][167][62][170], object detection [138][135], object tracking [191], depth estimation [101][90], semantic and instance segmentation [104][25][115][56], speech recognition [53][5], natural language processing (NLP)[29][162][49] and so on. Deep learning models have a wide range of applications accorss various domains like autonomous driving, robotics, fault detection, visual recognition, natura llanguage processing, fraud detection, anomaly detection, image captioning, pixel restoration, surveillance, healthcare and so on.

Such discoveries aroused the interest of data scientists and researchers in the field of machine learning and also adapting it to the medical domain. Deep learning caught the eye of many healthcare practitioners, and it did not take long for researchers to realize yet another opportunity for the application of deep learning. The Healthcare sector is not small anymore. Due to the enormous in imaging and medical devices, the amount of medical data is quite large. Hence, it makes medical diagnosis and analysis of the patient result more challenging. Such rapid growth of medical data requires tedious efforts by doctors and medical experts and are usually susceptible to human error, and can have various interpretations among different experts. There is a need for efficient and effective tools and methods to help speed up the diagnosis process by proper management and processing of this large data while maintaining high levels of accuracy and performance [131].

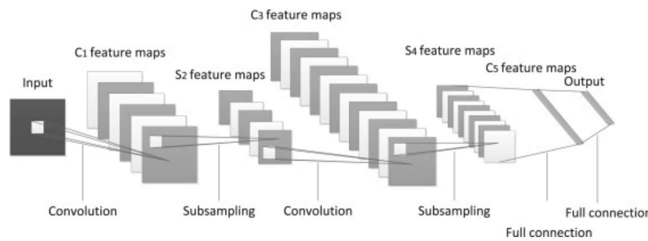

Figure 1. A schematic diagram of standard CNN [103].

Machine learning and Big Data help address the above issue. The term 'Big Data' evokes images of large datasets, both structured and unstructured, having a variety of formats and multiple sources. The term Big Data does not only depend on the volume of data. There are 4V's that help classify data as Big Data. This includes Volume, Variety, Velocity, and Veracity [121]. When we talk about data volume in the case of Big Data, it is in the range of terabytes to petabytes. Various industries like retail, banking, healthcare, insurance, media, and so on are dependent upon Big Data infrastructure and technology. However, in the healthcare sector, the progress and application of Big Data were delayed which seems a bit strange given that there have been some earlier predictions regarding the need and usefulness of Big Data in the healthcare domain [121].

One of the most important issues is the increasing gap between healthcare costs and the outcomes [94]. This poses an essential need for the improvement of healthcare quality and patient outcomes, and also with the increasing amount of medical data, Big Data is necessary for the healthcare sector to citerumsfeld2016big. There can be many possible sources for medical big data such as clinical registries, administrative claim record, digital health records, biometric data, patient data, the internet, medical imaging, large clinical trials, biomarker data, and so on [94][158]. Due to the high complexity, large volume, and velocity of medical data, it is really difficult for the experts to manage, analyze and accumulate the large volumes of data for medical assessment and diagnosis [157]. Plain visual inspection by experts is not a reliable and satisfactory procedure for the diagnosis. Thus the medical sector demands the development of, automated, time, memory, and computationally efficient frameworks and systems for efficient diagnosis and treatment process. Such solutions are provided by the use of Bid Data analytics in the healthcare sector, and it has shown promising results and have taken the burden off the shoulders of medical experts [18][139][7].

Deep learning has been used extensively along with Big Data analytics in health sector in recent years. There has been continuous development of automated image and signal analysis tools and models, which are the key factor in improving the analysis and diagnosis of various medical ailments. There are many applicatins of deep learning in medical domain now such as cancer and tumor detection and screening [176][31][36][67][76][78], detection of Diabetic Retinopathy (DR) [**?**][44], Histological and Microscopical Elements detection [100][133][151][150][156], Gastrointestinal Disease detection [181][192], Cardiac Imaging [106][34], Alzheimer's and Parkinsons diseases detection [141][143][66][126] and so on. Mostly the data in all the mentioned application is obtained in the form of radiological images via X-Ray, CT or MRI scans.

Despite the tremendous success of deep learning techniques across multiple domains, one important issue still needs to be addressed, and researchers and data scientists are making continuous efforts to address that issue i.e. the need for a large training dataset for deep learning models. Although deep learning models show promising results with a very high level of accuracy, this requires a very careful training process along with a large number of training examples. It is a common belief among all the researchers that the more data you feed into your network, the better the performance will be. So you need a sufficiently large volume of data to train a successful machine learning algorithm [92][32]. Data with high-class imbalance or insufficient data can lead to the poor performance of the algorithm [148]. Similar is the case for medical data. However, despite the rapid growth of medical and patient data, there are still many cases when the training data is not sufficient and it is expensive and time-consuming to obtain more training data especially the annotated one. In clinical settings, usually small datasets are obtained. Creating large datasets is usually time-consuming and might not be feasible specially incase of some rare diseases or new imaging modalities [111]. For example, in the case of the recent COVID-19 infection, the training data is still an issue for many researchers [122] and practitioners are in continuous efforts to obtain more data through clinical assessments, which is going to take quite some time because the infection is novel and trials are in early stages.

This problem is more significant, especially when employing supervised learning algorithms that require large labeled training examples. As mentioned earlier, the collection of large medical data is a complex, expensive, and time-consuming process which requires the collaboration of radiologists and researchers [54]. In recent years, many researchers have tried to overcome the issue of small datasets by using various techniques. It has been and still a very important research area for many researchers and data scientists across the world. There have been many attempts to resolve this issue in different domains by using various networks and techniques like data augmentation and synthetic data using GANs [51][171], Meta-learning [174], using surrogate data, n-shot learning (zero-shot, one-shot and few-shot) [43][186][160][183] and Siamese Networks [89][112]. For the case of limited labeled data (since the annotations are done by experts which are costly), there are other learning methods like unsupervised learning [109][13], semi-supervised learning [193] and more recently self-supervised learning [75], which do not require much labeled data for training purposes and can still achieve high accuracies.

Such methods also got the attention of researchers and experts working with medical data and the healthcare sector. Recently there have been several attempts to work with small medical data. Researchers have employed methods used in other domains in the healthcare sector. Such work can help mitigate the need for huge medical data required for training the algorithm and can help reduce the time and cost of the annotation process. This paper highlights some such efforts made in the medical and healthcare sector.

2 Method of study

This paper mainly surveys the literature in academic journals and conferences. The surveyed papers are based on some of the recent attempts in the medical field to deal with the curse of small datasets. These works adopt some of the methods mentioned in the previous section. Some methods deal with the issue when there is just not enough medical data to train with mainly because the disease is rare or new. Then we have other techniques which provide a solution when we have sufficient data, but not all of it is labeled i.e we lack labeled training data, and annotations can be time-consuming and expensive. For the sake of simplicity, this survey is not going to contain a lot of mathematical and technical details.

3 Machine learning techniques to deal with small training data in healthcare

Data is the driving force for any machine learning model. More training data means better training hence better performance. However, as mentioned in previous sections, large datasets are not always available, especially in medical diagnosis. This section mentions some of the recent works and methodologies to deal with the issue of having small training data in medical cases.

3.0.1 N-shot learning methods. One way to deal with small datasets is to use n-shot learning techniques [37][40]. N-shot learning usually refers to zero, one, or few-shot learning (FSL) methods. The aim of few-shot learning (FSL) is to learn good representations given a very small amount of training examples [159][175][137][61][142]. Hence, contrary to the classic method of supervised learning in which model is fed with a large number of training samples, here we only feed the model with a limited number of training samples per category [179]. One-shot learning (OSL) and zero-shot learning (ZSL) are special cases of n-shot learning in which the model is fed with only one/few training samples or with zero training samples respectively. So zero-shot learning deals with situations where at inference time, the model observes a data sample which was not present in the training dataset and still needs to predict the correct category to which the new example belongs to [178]. FSL can help reduce the efforts that are put into gathering data for data-intensive applications. Examples include object tracking [17], gesture recognition [128], image classification [175], image captioning, neural architecture search [20], language modeling [175] and so on [179].

Matching networks [175] propose query set labels for new classes by learning an attention-based mechanism over the provided set of classes. Another form of such networks, Prototypical networks [159] learn to classify novel examples based on the Euclidean distances after jointly learning centroid representations and embeddings. In [46], the main motivation is to perform few-shot learning without forgetting i.e. without catastrophic forgetting [84]. The authors of [134] propose a few-shot based method for representational learning by using generative networks. In [61], the authors create an imbalanced benchmark from ImageNet to study few-shot learning.

One such method has been adopted by the authors of [81]. In their research, the author mainly focuses on applying deep learning techniques to small-sized datasets of high-quality medical images. The way they overcome this limitation of a small dataset is by leveraging the few-shot learning. Inspiration for this came from the earlier work which combined deep learning with few-shot learning and demonstrated high potential [142]. There are several different problems in medical image analysis, but the authors focused their research mainly on the early diagnosis of glaucoma, which is an eye disease and is considered to be one of the leading causes of vision loss in humans [129].

Ophthalmologists use various kinds of eye imaging techniques for the diagnosis of glaucoma like fundus imaging (FUNDUS), Optical coherence tomography (OCT), Retinal Nerve Fibre Layer (RNFL), and so on. By combining these different imaging techniques, they can correctly diagnose glaucoma. However, it is very time-consuming for radiologists to obtain these various kinds of images and examine them. Mijung Kim *et al.* [81] attempted to develop an effective computational model by only using one type of images. Such a model can help in the early diagnosis of glaucoma and can save many patients from vision loss by leveraging state-of-the-art deep learning and few-shot learning techniques.

Two main challenges were addressed in their method. The first one is to deal with small datasets, for which they employed Matching Networks (MNs) introduced in [175] along with few-shot learning [142]. The second is to preserve the quality of high-resolution input images as much as possible. The authors use high-resolution CNNs similar to [45]. the intuition behind this is that the use of downscaling in medical images can cause the loss of some key features, which can affect the performance of deep learning algorithms. Attention mechanism similar to [175] is used that leverages cosine similarity and softmax. The embedding function used consists of two parts; a CNN component which is a stack of deep convolutional layers and utilizes high-resolution input images, and a bidirectional LSTM layer. At the end of the network, a flattened feature vector is obtained for both negative and positive images. These feature vectors are then used to predict the new unseen data using the attention mechanism similar to [175].

The dataset consists of 1080 high-resolution RGB images obtained using FUNDUS, provided by Samsung Medical Center in Korea. inference results indicate that Inception ResNet V2 obtains the highest accuracy in all experiments, but such architecture usually requires some heavy data augmentation [166]. Also, the authors observed that using more samples per class resulted in better accuracies. A similar trend can be found in the input image resolution i.e. the high the resolution of input images, the better the performance. The authors provide experimental results based on 20-shot learning only, as it outperformed 5- and 10-shot learning by a good margin. the experimental results indicate that the method used is indeed promoting and can help in early detection of glaucoma by just making use of a small-sized medical dataset. In the future, such a method can be used for the detection and diagnosis of other kinds of diseases as well and also for different types of images.

Model	Input size	Data Aug.	Acc.
VGG-16	$224 \times 224 \times 3$	Yes	65.2%
Inception ResNet V2	$239 \times 239 \times 3$	Yes	89.5%
Our model (low)[a]	$256 \times 256 \times 3$	No	79.0%
Our model (low)	$256 \times 256 \times 3$	Yes	77.2%
Our model (mid)[b]	$512 \times 512 \times 3$	No	81.2%
Our model (mid)	$512 \times 512 \times 3$	Yes	83.4%
Our model (high)[c]	$1024 \times 1024 \times 3$	No	88.1%
Our model (high)	$1024 \times 1024 \times 3$	Yes	87.9%

[a] low denotes a center-cropped image down-sized to 256×256.
[b] mid denotes a center-cropped image down-sized to 512×512.
[c] high denotes the use of the original center-cropped image.

Figure 2. Experimental results indicating the effectiveness of proposed architecture and methodology [81].

Another such approach makes use of a deep siamese neural network [86]. The intuition behind the approach is that the siamese network will learn the distance and feature representation from the textures present in the source domain, and then apply this learned knowledge to classify the tumoral and healthy tissues in a target by using few-shot learning. A three-headed siamese network is used in [111] as shown in Figure 3. A VGG16 [155] network is used as a base-network with slight modifications to prevent overfitting by reducing the number of parameters. This is done by using the method introduced in [146], in which low dimensional representation of the input images are obtained in the form of embeddings by using a 128 neuron layer. During training, embeddings for images that belong to similar classes are forced closer and those that belong to different classes have their embedding vectors pulled apart. This embedding vector is the set to train a classifier like Support Vector Machine (SVM) [85]. After that, we transfer this learned knowledge to the target domain. A subset from the training dataset of the target domain is prepared for few-shot learning. The selected images from each class are then used as input to the siamese network and embedding vectors are obtained which are then used to train a linear SVM classifier that classifies the images. Results indicate that this method outperforms the traditional fine-tuned neural networks and achieves very high accuracy with one few samples per class.

One important medical task in drug discovery, which unfortunately also lacks the amount of data available for predictions. hance one-shot learning is introduced to resolve this issue [4][37][40]. Some of the recent works demonstrated that standard deep learning networks and random forests are capable of learning meaningful chemical features with only a few hundred compounds, but for the drug discovery case, even a few hundred compounds are often too resource-intensive. The authors of [4] introduce a new network architecture called iterative refinement long short-term memory (LSTM) with some modifications to the matching-networks and residual networks architecture [175][63]. Furthermore, instead of two-dimensional convolutions on images, graph-convolutional networks have been used [83]. The intuition is that the molecules can be represented as undirected graphs, with bonds as edges and atoms as nodes. In this way, graph-convolutional networks are useful for the transformation of small molecules into vector representations and allow the learning of complex molecular structures.

Segmentation of the medical results is also an important application that can help radiologists in identifying various diseases. With enough labeled data, supervised deep learning-based segmentation methods have shown state-of-the-art results. However, manual segmentation labeling of medical images is resource and time-intensive. Also, the use of different image acquisition procedures can cause wide variations in tissue appearance, image noise and resolution [95]. One approach to mitigate the issue of less labeled data is to synthesize realistic and diverse labeled examples by leveraging unlabeled examples [189], which results in an automated data augmentation method for synthesizing labeled medical images. One labeled scan is used along with unlabeled data in a semi-supervised manner. Transformations are learned from the unlabeled data which are used along with the labeled example to synthesize more labeled examples. These transformations include spatial and intensity transformations. Convolutional neural networks are used to learn models of these transformations, which are later utilized to develop new training examples. Since only one labeled example is being used, this method comes under one-shot learning methods.

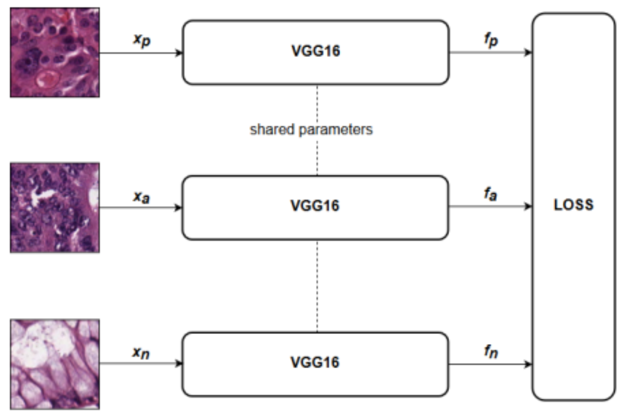

Figure 3. *Histology Siamese Network* training phase [111].

Separate appearance and spatial transform models are learned to capture the underlying differences between the atlas (labeled examples) and unlabeled examples. New labeled examples are then synthesized by applying these transformations to the atlas. After that, these new examples can be used along with atlas to train the segmentation network in a supervised manner. One thing pointed out by the authors is the choice of the atlas. It is usually selected to be close to the anatomical average of the distribution. So it should be the example which is most similar to the computed anatomical average [12]. The proposed segmentation model outperforms existing methods for one-shot medical image segmentation [189].

One way to deal with the issue of the small annotated dataset is to develop surrogate tasks that make use of unlabeled data to provide labels. A novel few-shot learning framework has been presented in [39]. Training is done in an episodic manner where unlabeled data is also presented during each episode. Recently, self-supervised learning has emerged as the new paradigm [75] where data itself provides supervision. various techniques have been proposed for it like solving jigsaw puzzles [119], rotation prediction [47] or restoring missing parts in an image [187]. Image denoising has been adopted as the surrogate task in [39]. This surrogate task is used to integrate an auxiliary loss during training, which compensated for the lack of annotated data. Qualitative and quantitative analysis of the results indicates the effectiveness of utilizing a close connection between

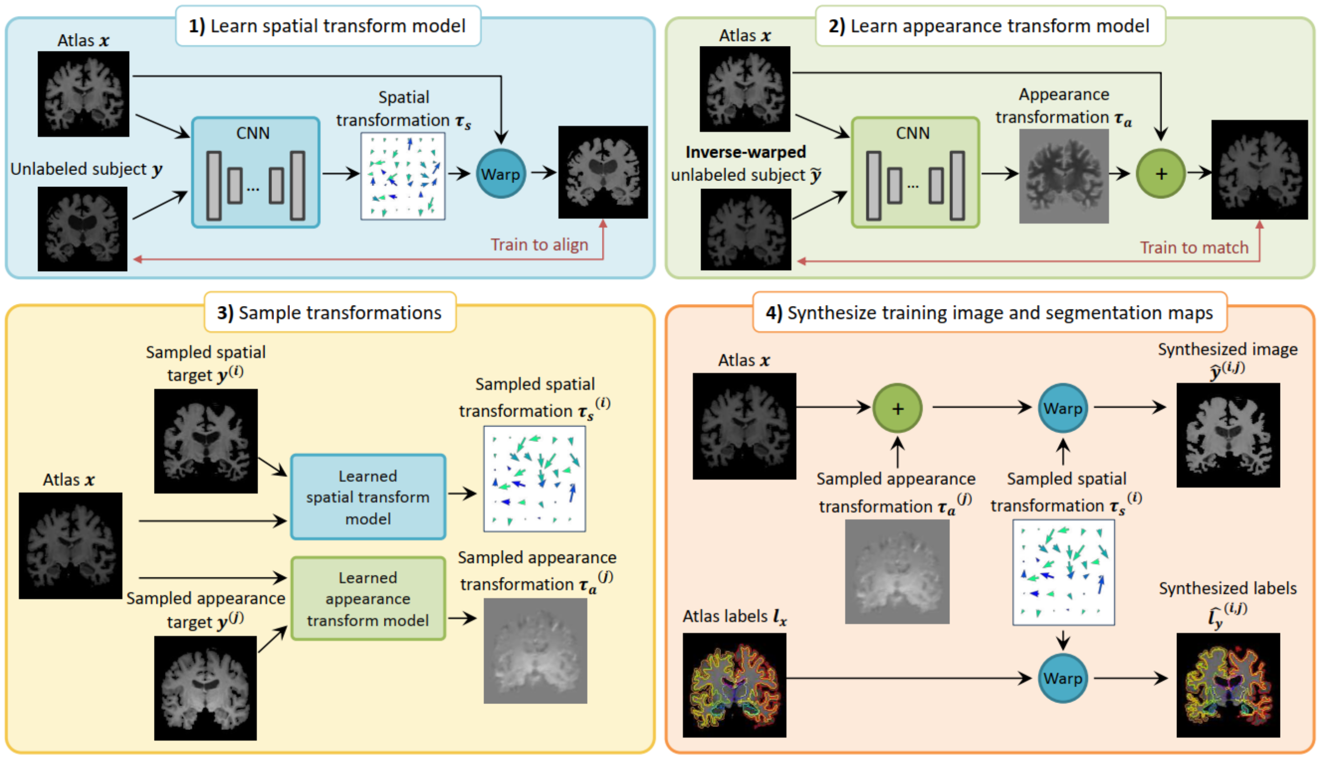

Figure 4. An overview of the proposed method [189].

self-supervised and few-shot learning and improves upon the previous few-shot learning segmentation methods for medical data.

Zero-shot learning also found its way into the medical sector recently. By making use of information from two different modalities i.e. CT and x-ray, generalized zero-shot learning can prove valuable for the diagnosis of chest radiographs [125]. Generalized zero-shot learning (GZSL) corresponds to the type of zero-shot learning (ZSL) in which the test dataset may can both unseen and seen examples. First of all features are extracted from x-ray images using DenseNet-121 feature extractor [132]. CT radiology reports are used to generate signatures using Intelligent Word Embedding [14]. Semantic embedding is performed using autoencoders on the output feature vector of CNN based extractor to extract saliency from the noisy feature vectors [87]. These hidden space representations are then trained to be close to the provided semantic signatures of their class labels. Figure 5 represents the overall pipeline of the proposed model. Experiments show the robustness of the algorithm however performance highly depends on the extraction of salient features for which clear x-ray images are required.

More recently, there has been an outbreak of novel coronavirus (COVID-19) which is affecting the whole world. More than 5.4 million cases of COVID-19 have been confirmed as of May w4th 2020, with a fatality rate of roughly 6.3 around the world, according to World Health Organization [1]. People are also suffering from lockdown and social distancing is being practiced everywhere. To deal with this pandemic, people across various domains are working hard and providing different solutions to help reduce the effects of the virus and keep people safe. Research on vaccine discovery is also underway in various institutes throughout the world. However, so far no effective treatment has been found. Radiologists are trying their best to diagnose the virus using CT scans and other methods, but the procedure and diagnosis are hard due to the novelty of this disease. Hence, data scientists and researchers have been trying to help speed up this process via machine learning tools.

Recently, there have been many works for the detection of COVID-19 using deep learning tools. However, most of these methods require huge datasets. Due to the rare nature of this virus, large datasets are still unavailable. Hence due to the shortage of annotated datasets, traditional classification methods may not be able to work properly for COVID-19 detection. To deal with this situation, a few-shot based approach has been presented to make an accurate prediction by using a small number of chest CT images [26]. Self-supervised learning [123] is employed along with momentum contrastive learning for better performance. Momentum contrast is an algorithm based on self-supervised learning which utilizes a contrastive loss. Such methods are analogous to building dynamic dictionaries where the keys of the dictionaries are sampled from the data itself and their representation is then obtained using an encoder network. Representational learning and data augmentation are done at the pre-training stage. An instance discrimination task is used as the pretext task. Pretext task in self-supervised learning is solved to learn visual representations, which are then used for the training of downstream task which is the main supervised learning task. After that, the query image and support images are encoded by the pre-trained encoder for the classification purpose. The authors employ the prototypical network [159] to apply few-shot classification. Experimental results indicate that the proposed methods achieve better performance than ResNet-50 in limited data case.

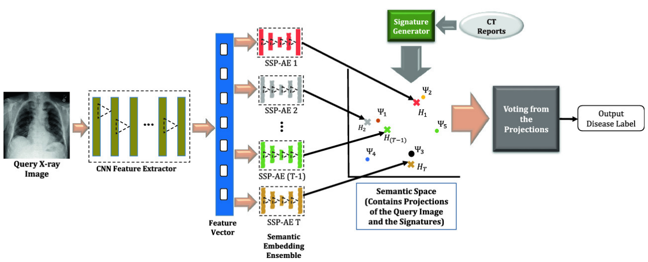

Figure 5. The block diagram of proposed model. [125].

3.0.2 GAN-based methods. Generative Adversarial Networks or GANs have been used extensively for data synthesis and augmentation in recent years [51][171]. They became quite popular due to their immense success across various tasks. Their most significant breakthrough has been in the domain of computer vision where they helped in dealing with challenges among various tasks like image generation, facial attribute manipulation, style transfer, and so on [180]. GANs have shown remarkable data augmentation performance and can synthesize realistic and novel samples e.g. SimGAN has shown over 21% improvement in eye-gaze estimation [153]. Such a network consists of two models, a generator and a discriminator. The generate generates samples of images. These samples are then passed to the discriminator network along with the original images, and then it classifies the generated examples as being real or fake. In this way training is done; the generator keeps better at generating more realistic images and tries to fool the discriminator whereas the discriminator gets better at distinguishing real images from the synthesized ones. In this way, we can get more realistic synthesized images in the end. Compared to the standard data augmentation techniques like cropping, translations, rotations, and so on, GANs also introduce more variability in the training dataset by generating more diversified samples.

Due to their immense popularity, GANs have also been used in the medical domain nowadays. Research has been conducted in the health sector to make use of these networks for data augmentation and synthesis to help deal with the issue of small and less annotated data. In this section, we are going to take a look at some of the recent works done in the medical sector using GANs to deal with the need for large datasets.

Recently, GAN frameworks have been applied to several medical imaging applications. For example, a method has been proposed in [30] to generate new eye fundus images from the existing data using adversarial learning i.e. GANs. In another work, the authors of [118] attempted to generate CT (computed tomography) images given the MR (magnetic resonance) images by using GANs. AnoGAN, a deep convolutional generative adversarial network has been proposed for the identification of anomalies in unseen data. Most of these techniques have employed segmentation-to-image translation, label-to-segmentation translation, or medical cross-modality translations. As mentioned earlier, GANs can also be used to synthesize high quality and diverse medical images for data augmentation like in the liver lesion classification task [42]. Along with bone and lungs, the liver is a common site for metastatic cancer. In 2012 alone, 8.2 million deaths were caused by cancer worldwide, out of which 745,000 are due to liver cancer [38]. Work has been done by researchers automate the classification process using CT images and help radiologists in the diagnosis procedure [48][22].

As mentioned earlier simple augmentation techniques like translations, rotation, rescaling, flip, and so on are not good enough for medical datasets due to a lack of diversity in the generated examples. Motivated by this, the authors of [42] adopted GAN for generating synthetic liver lesion in combination with standard image perturbation techniques for the improvement of liver lesion classification. Applying GAN helps enrich the training data by realistic image synthesis. However, to train the GAN, we need a lot of examples as well. So in the first step, standard data augmentation techniques are used to increase the dataset which is then used for GAN training. The examples generated by GAN are then used as an additional source for data augmentation. Then finally original and synthetic data is used to train a lesion classifier. Deep Convolution GAN (DCGAN) proposed in [130], has been employed for the synthesis of labeled lesions. By this combination of classical data augmentation and data synthesis using GAN, the authors were able to increase classification accuracy from 78.6% to 85.7% with a significant 7% improvement. Assessment of synthesized data by two different experts showed similar results for both real and generated data. This suggests that the generated examples are indeed meaningful.

An innovative GAN based method has been proposed in [57] for image augmentation. This is done by adding the bounding box conditions in the original PGGANs [77]. In PGGANs, generator and discriminator are progressively growing i.e. they start from generating low-resolution images and capture more details as training progresses. This network is modified by Han textitet al. [57] to further generate realistic MR images of 256×256 size, include randomly shaped tumors within bounding boxes. Figure 6 displays examples of generated images, including tumor bounding boxes. The authors use YOLOv3 [136] architecture for real-time detection of the tumor. Results indicate that by utilizing additional 4000 synthetic images, sensitivity is improved by 0.10 with an IoU threshold of 0.5. These generated images provide increased robustness during training and also further fill the distribution of real images, resulting in a high sensitivity of 0.91 with moderate IoU threshold 0.25, even with small MRI training data and coarse bounding box annotations.

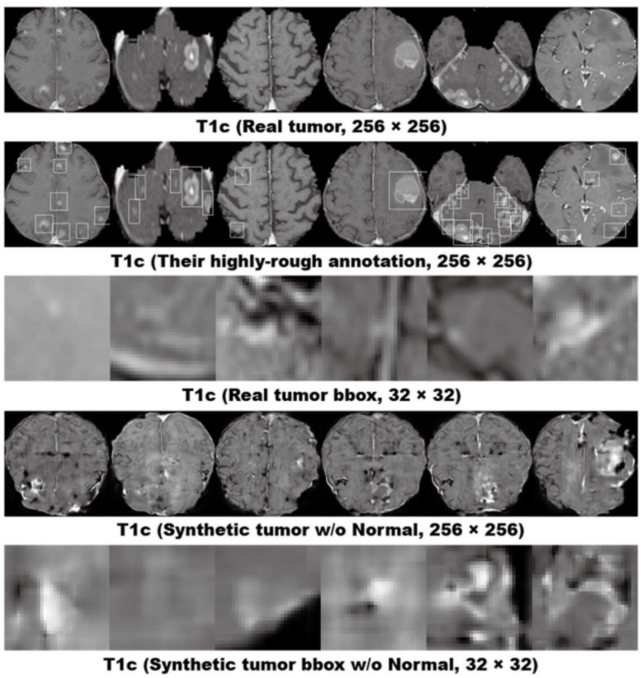

Figure 6. Examples of 256×256 real/synthetic MR images and resized 32×32 tumor bounding boxes [57].

Besides providing a diverse and additional form of data augmentation, GANs can also serve as an effective model for the anonymization of medical data [149], in this case specifically for tumor segmentation tasks using MRI images. Two publicly available datasets of brain MRI have been used by Shin *et al.* [149] including Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) dataset. To translate MRI to labels and vice versa, image-to-image translation conditional GAN (`pix2pix`) model introduced in [71] has been utilized. First of all this `pix2pix` model is trained on T1- weighted images of ADNI dataset to segment normal brain anatomy. The tumor labels provided in the BRATS dataset are then merged with this brain atlas which results in complete segmentation of the brain with the tumor. Synthetic multi-parametric MRI with a brain tumor can then be obtained by using these segmented images as an input to label-to-MRI GAN. The authors also claim that since tumor and brain anatomy labels are separate, they can alter either of those to obtain synthetic images with desired characteristics [149]. Another claim is that the GAN produces images by using aggregation to draw individuals and hence does not refer to underlying patients as individuals. This helps in the generation of anonymized synthetic images with variations and allows the sharing of training data.

In recent years, researchers have been utilizing noise-to-image GANs [58] or image-to image GANs [182] to improve the classification results with augmented images. An innovative method has been presented by Han *et al.* [59]. A two-step GAN-based data augmentation technique is presented. First, PGGANs [77] generates 256×256 diverse and realistic images. In the second step, a Multimodal Unsupervised image-to-image Translation (MUNIT) [69] model which is a combination of GANs/Variational AutoEncoders (VAEs) [82] or SimGAN [153], is used to refine the shape and texture of the images generated in the first step which furthers fits them into the distribution of real images. Visual Turing test and t-SNE (t-distributed stochastic neighbor embedding) results indicate that the presented combination of noise-to-image and image-to-noise GANs for augmentation of medical images improves the accuracy of CNN-based classification for tumor detection, especially when combined with classic DA techniques.

3.0.3 Meta-learning methods.

In recent years, meta-learning has gained a lot of popularity and more meta-learning techniques are being developed. The term *meta-learning* or *learning to learn* refers to techniques that leverage prior experience gained on other tasks [174]. We can leverage more types of meta-data depending on the similarity index between the new task and the prior task. It helps in learning the new tasks quickly than otherwise possible with a fewer number of examples. However, if the new task is different than the experience or random noise than leveraging prior knowledge might not be effective. Luckily that is not the case in most real-world applications. There are different types of meta-learning like metric-based, model-based, and optimization-based. The details of these types are out of scope for this paper.

Learning GAN-based synthesis, meta-learning also found its way into the medical regime. Maicas *et al.* [107] learned good initialization on a series of tasks by utilizing a meta-learning based method. In general, different randomly-sampled tasks are treated equally in terms of gradient directions during meta-learning. So often what happens is that the meta-learning process stops at a point where the easy tasks are learned well, but the misclassification remains in the case of difficult tasks. A novel difficulty aware meta-learning (DAML) method has been proposed to deal with this issue especially for rare disease classification [98]. First, a meta-classifier is trained on a series of related tasks like the classification of common diseases. Then these learned internal representations are transferred to the new task of rare disease classification. For the better optimization of meta-classifier, the authors formulated a dynamic modulating function over the learning tasks. The well-learned takes are automatically down-weighted by the function which puts more focus on the hard tasks. Evaluation is done on ISIC 2018 Skin Lesion Dataset [28]. Promising result shave been obtained by training the classifier on only four skin lesion classes and performing classification on the other three unseen classes with an AUC of 83.3%. The proposed methods also outperform standard fine-tuning methods and widely used few-shot learning approaches like Relation Net [165], MAML [41] and Task sample [107] for the task of rare disease classification. The result comparison of the proposed DAML with other methods is shown in Figure 7.

	backbone	Sample #	AUC	Sample #	AUC	Sample #	AUC
ConvFeature + KNN	DenseNet	1	50.00%	3	56.07%	5	62.69%
ConvFeature + SVM	DenseNet	1	61.46%	3	61.68%	5	67.44%
Finetune + Aug	DenseNet	1	58.57%	3	68.05%	5	73.65%
Finetune + Aug	4 Conv Blocks	1	58.49%	3	68.13%	5	75.90%
Relation Net [14]	4 Conv Blocks	1	59.97%	3	62.87%	5	72.40%
MAML [2]	4 Conv Blocks	1	63.77%	3	77.98%	5	81.20%
Task sampling [10]	4 Conv Blocks	1	64.21%	3	78.40%	5	82.05%
DAML (ours)	4 Conv Blocks	1	**67.33%**	3	**79.60%**	5	**83.30%**

Figure 7. The AUC performance of different methods on skin lesion dataset [98].

Domain generalization is one of the most crucial tasks in deep learning i.e learning a model from multi-domain data and generalizing it to a new and unseen target domain. Various attempts have been made to address the domain generalization problem [11] [33] [96]. Such tasks become even more complex in the healthcare sector due to the complexity of medical images. To tackle this issue, recently a novel shape-aware meta-learning (SAML) scheme is presented by Liu *et al.* [102] for the improved model generalization in prostate MRI segmentation. The presented model is based on gradient-based meta-learning [79]. The source domain is first randomly split into meta-train and meta-test sets, which are then used to simulate the domain shift. This helps in robust optimization during the model training. The authors also addressed the specific deficiencies like ambiguous prediction boundaries and incomplete shapes, that are faced when applying the learned models to an unseen domain for segmentation. This is done by utilizing two complementary shape-aware loss functions for the regularization of the optimization process in meta-learning. The first regularization is applied to the predictions of meta-test data for *shape compactness*. This makes the sure complete shape of the segmentation masks is well preserved by the model in unseen domains. Second, a novel objective is presented to promote domain-invariant contour embeddings in the latent space which enhances the *shape smoothness* at the boundary under domain shift. The proposed method has been validated through extensive experimentation and shows effective results. This can be extended to other segmentation tasks as well which suffer from the issue of domain shift. For the sake of simplicity, mathematical formulation and theory are not mentioned here. It can be found in the original paper [102].

3.0.4 Surrogate methods.

Some other methods utilize some surrogate data or surrogate supervision to handle small datasets in the medical sector. However, this is not a properly defined category, still, some work is done by researchers to apply these methods to mitigate the need for large datasets. Shaikhina *et al.* [147] employ surrogate data analysis and method of multiple runs for the application of artificial neural network to regression problems involving limited medical datasets. A simple feed-forward neural network (NN) is used with one hidden layer. They introduced the method of multiple runs to handle the issue of a small dataset, in which multiple NNs of similar design are trained simultaneously. The networks are similar in terms of design and topology, however, differences lie in the weight and bias initialization, training and validation split, the order in which data is fed into the NN. However, in the case of small datasets, the learning of NN is not efficient and even a poorly designed NN can achieve good results on test examples at random. To deal with these random occurrences and to evaluate the performance of NN in presence of such events, a *surrogate data test* is proposed [147]. Such a dataset resembles the original dataset in terms of the statistical properties. However, the surrogates do not keep the complex interrelationships among various components of the original dataset. Various methods [145][173][97] can be used for the generation of surrogate data and can be validated via NN techniques. To generate robust surrogate data, this technique utilizes multiple runs of 2000 neural networks and provides adequate data volumes for the predictions of modern machine learning algorithms. The highest performing surrogate data NN defines a new lower threshold for the original data, and to pass the surrogate data test, the NN must outperform this threshold on the real dataset. Therefore, the proposed surrogate dataset also provides a strategy for the evaluation of a NN when there is no additional test data is available. Using the proposed framework, small-dataset NN can be developed which are comparable in performance to the large-dataset NN. This also results in an overall reduction of 18 times in the amount of data required for the training of NN. The proposed framework is shown in Figure 8.

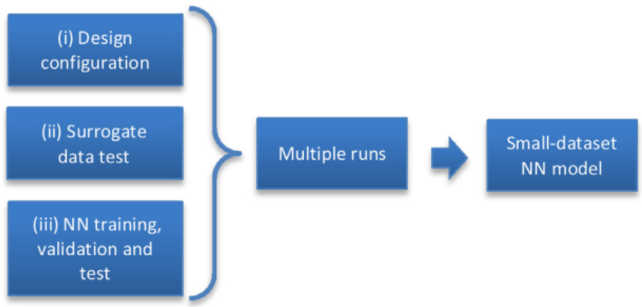

Figure 8. Proposed framework for the application of NN to regression task in medical analysis [147].

Motivated by the work done by Shaikhina *et al.* [147], Sabay *et al.* also utilize surrogate data generation method to overcome the problem of the limited dataset in the prediction of heart disease. However, the authors claim that the number of neurons and computational resources required in the prior work is large. Hence, an alternate solution for the generation of surrogate data is presented. There various tools available for that purpose. One such library is Synthpop, which is an R language library and provides functionality for data comparison and synthesis [120]. Methodology and experimentation are divided into three stages. During the first stage, the authors mainly performed the validation of their machine learning models i.e. logistic regression, random forest, and decision tree models, and compared results with the prior works [152][108][164]. The Cleveland 14 dataset is used for that purpose. Stage 2 is based on the improvement of the stability of the designed models. Due to the small size of the Cleveland 14 dataset (297 records only), the performance of ML models is not stable. Therefore, the authors generated a surrogate dataset of 50,000 observations based on the characteristics present in the Cleveland 14 dataset. This is done by using the Synthpop library in R studio. Statistical comparison between the original Cleveland 14 data and the generated surrogate dataset show similar properties with huge overlap, which means that the generated observations indeed resemble the real set and can be used for the stable training of machine learning models.

Experimental analysis shows that the results of accuracy, precision, and recall are similar for the original dataset and surrogate dataset. However, the main difference lies in the fact that the variance of the synthetic data is very small due to a large number of observations, which guarantees the stable performance of machine learning algorithms like the logistic regression model. Finally, the objective of the last stage was to improve the accuracy, precision, and recall results obtained in stage 2 by utilizing a Perceptron Neural Network model. The model consisted of 3 hidden layer units and 3 dropout layers. Since NN models require a large amount of data for training, 60,000 records were generated as the surrogate dataset from the original Cleveland dataset. Results indicate the effectiveness of proposed methods and superior results of the Perceptron model compared to the traditional logistic regression models.

3.0.5 Some miscellaneous methods. In this section, I am briefly going to highlight some of the very recent works which are unique based on their method and lack much prior and follow-up work. However, in the future, it is more research is expected based on these methods. Without diving into too much detail, the section is mainly just going to provide a basic overview and concept used in the research.

One such method is presented in [27]. To learn the representations of medical data with less supervision and training data, the authors propose a Deep Siamese CNN architecture. Siamese networks were originally developed by Bromley *et al.* [21] for the task of signature verification. They have also been used for image classification using the one-shot learning method [86]. Later, they have also been used along with few-shot learning in the medical domain [111] as mentioned earlier. Deep SCNN architecture is a variant of neural networks that can find correlations and similarities between the members or input data. Two identical CNN sharing the same weights have been used for the construction of deep SCNN. ResNet-50 architecture with pre-training on ImageNet dataset have been employed to build the CNN models [62]. To avoid overfitting on training data, dropout layers and batch normalization have been used [163][70]. Euclidean distance is used as the metric to calculate the similarity between the images and a contrastive loss function is utilized [55]. Compared to the multi-class problem with a single CNN architecture, deep SCNN transforms the multi-class problem into a binary classification task. Experimental analysis on the CBMIR task demonstrates that the performance of the prosed SCNN architecture is comparable to the single deep CNN architecture, which requires a huge amount of labeled data for training.

Layer	MAP	MRR
CNN (third-last)	0.6209	0.7608
CNN (second-last)	0.6369	0.7691
CNN (softmax)	0.6673	0.7745
SCNN (last layer)	0.6492	0.7737

Figure 9. Performance comparison of single CNN model and proposed SCNN architecture [27].

Due to the small size of medical datasets, generalizing on new samples is a hard task. Deep learning models can easily overfit the training data because of the limited size. Improving model generalization is one of the most crucial tasks in deep learning and various techniques have been presented over the years like data augmentation, dropout, batch normalization, regularization, and so on [163][70][184].

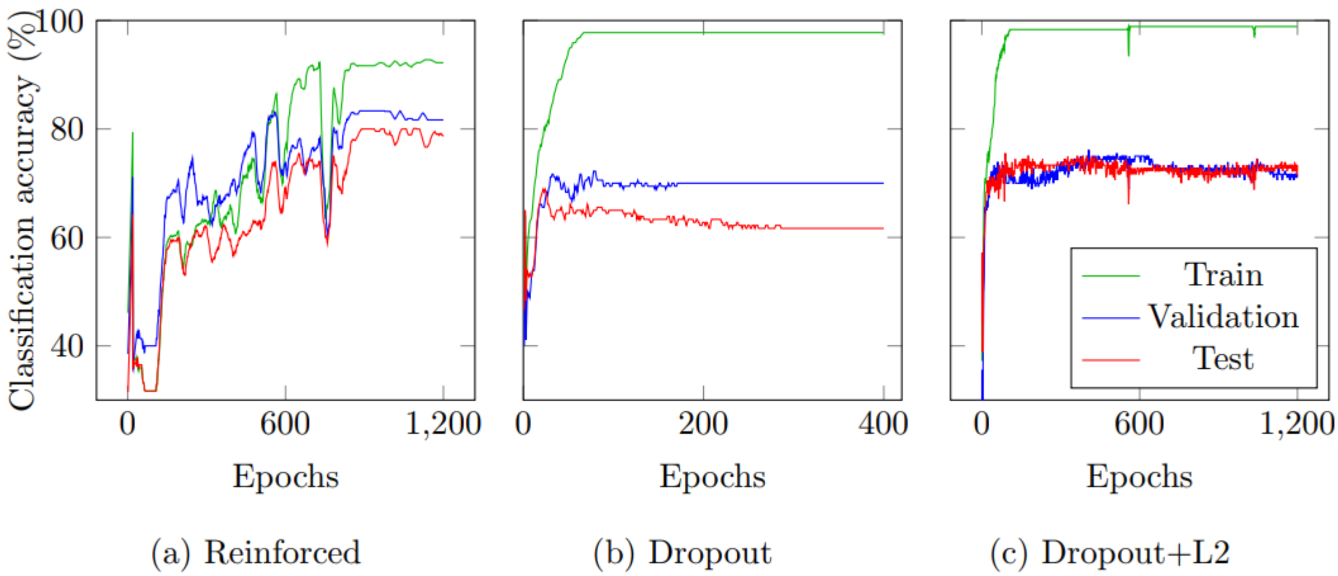

Figure 10. Learning curve comparison of proposed reinforced classifier with other approaches [2].

However, all these techniques are based on supervised learning methods, which require a huge amount of training data. The network is trained to fit on the training set and there is no additional control for generalization on the test data, except tuning the network architecture and parameters.

To prevent the overfitting of the classifier on small medical data and to improve generalization, a novel reinforced classifier is proposed in [2]. Instead of trying to hard-fit its output to the training example, the proposed model tries to directly improve the generalization performance on the unseen data based on the policy gradient methods in reinforcement learning (RL) [9]. The general goal of RL is to increase the cumulative reward of an agent by optimizing its policy using a sequential MDP (Markov Decision Process) [116]. Supervised learning performs a direct matching of outputs to the known labels. Whereas in RL, we try to optimize the model output through exploration and exploitation and it may not have an immediate effect on the actual task at hand. Therefore, the training process in RL can be slow and difficult. The authors propose to embed the long-term consequence i.e. generalization, of supervised learning methods into the objective function and provide a way to control the performance on the validation dataset. With the new proposed objective, it is not possible to perform a comparison between the model predictions and the target labels since the class distribution is different for both cases now. In RL, some future feedback or reward is used as a measure of how good an action is. Therefore, we can formulate the generalization problem as a reinforcement learning task of an agent by incorporating feedback. Given an input image, the agent tries to update its policy to improve the performance of the validation dataset. Learning curves and experimental results on different datasets indicate the improved generalization power of the proposed framework compared to the previous deep learning regularization like L2 and dropout. The two datasets used consisted of 100 CT and 60 MR images respectively. Comparison with other overfitting prevention approaches s shown in Figure 10.

As mentioned earlier, recently COVID-19 has become a global pandemic last year. It has been eleven months since the start of this disease, and researchers and scientists have been working on finding ways to deal with this situation along with the government and other agencies. Just like any other domain these days, machine learning is also being applied for the detection of corona virus disease 2019 (COVID 2019). Deep learning /9DL) approaches by utilizing chest X-ray have been actively explored for the classification of COVID-19 [177][117][6][65]. Inspired by these early works, the authors of [122] further explore the feasibility of deep learning methods for corona virus detection. However, due to the small amount of available data, it is difficult to train large deep learning models. The intuition behind the presented approach is that the patch-wise intensity distribution has statistically significant differences in COVID-19 CXR (Chest X-ray). These findings led to the formation of a novel patch-based deep neural network architecture with patch cropping in a random manner. The presented approach reduces the network complexity and in each image, multiple patches can be used to augment the training dataset. Hence, the neural network can be trained efficiently even with such small data without overfitting.

The CXR images are first pre-processed for data normalization. Dataset is obtained from various resources and may introduce heterogeneity during training. After this, the preprocessed data is fed into the segmentation network. In the proposed architecture, FC-DenseNet103 is used as the backbone of the segmentation network [74] and extracts lung areas from the entire image. These extracted lung areas are then used as input to the classification network, which classifies the diseases based on the patch-by-patch training and inference mechanism. The authors [122] used a simple ResNet-18 architecture for classification task to avoid overfitting, and it also allows to use of the pre-trained ImageNet weights for the small training set and makes the overall training process stable. Classification network consists of two different levels i.e. global and local approaches. In the global approach, the entire image is fed into the network after resizing to 224×224 and focuses on global representations of the CXR data. The second part is the proposed local patch-based approach in which the masked images were cropped randomly and the resulting patches are used as input to the classification network. Each patch is fed into the network and output is generated. The number of output is equal to the number of input patches used, and the final classification decision is made by counting maximum votes i.e. the class that appears most is declared as the final overall class. The proposed framework is depicted in Figure 11. Experimental analysis shows that the network performs well even on small training data and provides comparative results to the SOTA methods.

3.0.6 Self-supervised learning methods. Self-supervised learning enables us to utilize a variety of labels that are already present in the data. As we already know that generating a set of annotated examples is both time and resource consuming. However, a huge amount of unlabeled is available most of the time. Therefore, an efficient solution would be to use this unlabelled data somehow to generate labels for the training process. Self-supervised learning enables us to do so by setting learning objectives in a way such that the data itself provides supervision.

Generally, self-supervised learning consists of two tasks i.e. a pretext task and a real (downstream task). The downstream task is the main task which we want to accomplish like classification or detection, and in this case with insufficient labeled data. The pretext task is the self-supervised task which is usually invented based on the type of data we have or the designer's choice. The best idea is to choose a task that if solved, would provide some understanding of the actual data that would be used for the real task. Such tasks may include color prediction [91], rotation prediction [47], predicting noise [19] guessing position of input patches in an image, inpainting, distortion, solving a jigsaw puzzle, and so on. The main goal of the pretext task is to learn the visual representations, to use these learned representations and weights for the actual downstream task. Here we usually don't care about the performance or accuracy of the invented task because our goal is not to achieve high performance but to learn representations that can carry good semantic and structural meaning and are helpful in the downstream tasks.

Self-supervised learning has emerged as the new paradigm in recent years due to its effectiveness [72][75], especially for the case of small annotated datasets. It is clear from our previous discussions, that one domain the suffers most from the curse of small datasets and annotations is the medical sector. Therefore, in recent years, researchers have made attempts to utilize self-supervised learning methods to mitigate the need for large annotated datasets. One such approach has been presented recently in [168]. The underlying idea presented in the paper is to utilize the large available unlabeled medical data and assign labels to it. This is done by a process knows as surrogate supervision. This is mainly employed at the pre-training stage, in which a deep learning model is pre-trained using surrogate supervision including tasks like rotation, reconstruction, and colorization. It is a form of self-supervised learning [75]. The findings of the research indicated that pre-training with surrogate supervision is effective and can outperform the models trained from scratch or transfer learning from an unrelated domain, especially in the case of medical data [168].

Some of the limitations of using transfer learning like pre-training on ImageNet data is that it can restrict the designer to architectures that are needlessly complex for the medical dataset, and can cause performance degradation during training and inference. Also, transfer learning does not work most of the time in the case of 3D medical analysis applications, since the 3D and 2D kernel shapes are not compatible. Hence, transfer learning has its limitations in the case of medical data. This brings us to the problem of assigning artificial labels to labeled medical data and utilizing this large amount of unlabeled data for medical image analysis. This can be achieved by developing surrogate tasks. The key idea is to assign artificial labels to the unlabeled data through a surrogate task and employ this surrogate supervision for the training of the main network for the real task. This scheme is very useful in the case of medical data, where the annotation is time-consuming and expensive.

Some of the prior research prosed various surrogate task to apply self-supervised learning with surrogate supervision in the medical domain. However, the literature is still very limited, since it is still a new trend and requires more attention. Jamaluddin *et al.* [73] proposed a surrogate task based on the longitudinal relationships between medical images. Alex *et al.* [3] used small image patches for noise removal as the surrogate task. Ross *et al.* [140] employed image colorization as the surrogate task. In [168], rotation prediction [47] is used as the surrogate task in the cases when the geometry is consistent and data has adequate landmarks. GAN-based surrogate supervision is where rotation might not be the

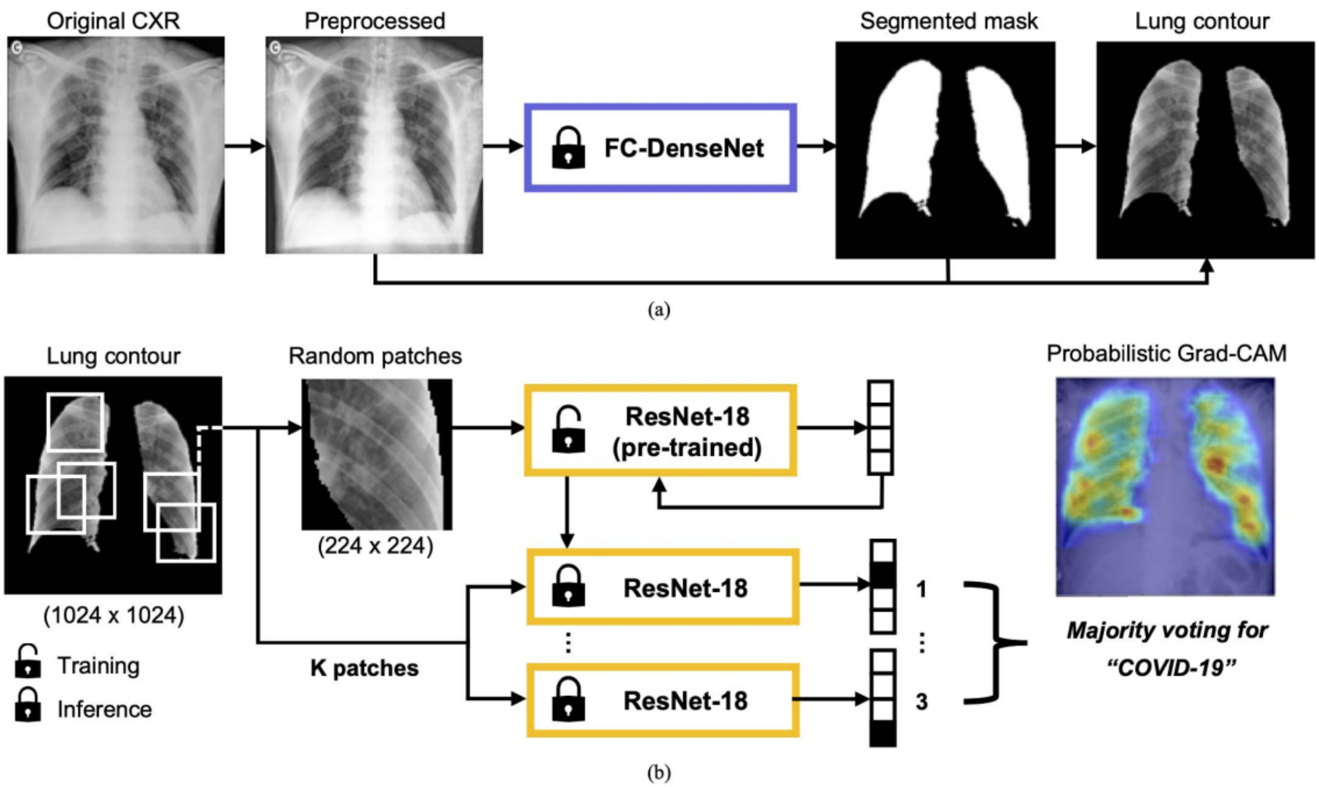

Figure 11. Proposed neural network approach: (a) Segmentation network (b) Classification network. [122].

appropriate choice. Such schemes include image colorization and patch reconstruction tasks by using conditional and Wasserstein GAN respectively [91][8]. To avoid expensive computations in the case of 3D data, the authors resort to flipping around the axes instead of rotations. Such surrogate tasks enable the model to learn the relative location of various structures in medical data, distinguish between them, and learn the overall geometry of the data. Results indicate that the proposed methodology outperforms training models from scratch, and also demonstrates that pre-training directly on the medical domain by using surrogate supervision is more effective than utilizing transfer learning from an unrelated domain like ImageNet.

An early example of the adaptation of self-supervised learning to medical diagnosis was provided by Ammara *et al.* [110]. There has been a rapid increase in the observed cases of melanoma, which is the deadliest form of skin cancer, across various regions [154]. Computer-aided systems enforced with machine learning models can help in the diagnosis of skin cancer. Such systems usually required labeled data to perform correct classification and diagnosis. Skin cancer data can be costly and time-consuming to label as it requires experts opinion. To deal with this issue [110] proposed a classification model in which a deep belief net is used as the architecture for classification and another SA-SVM (self-advised support vector machine) alongside the first classifier. Deep belief network is trained in a greedy layer-wise manner to citebengio2006greedy. Training data consists of both labeled and unlabeled examples. SVM is a well-known algorithm and is utilized for various machine learning tasks. However, if some data points are not linearly separated by the kernel during the training phase, then SVM ignores them. During inference, if data similar to the ignored misclassified examples appears, it is going to be classified wrongly. The authors adopt a self-advising support vector machine [105] that does not ignore the misclassified examples. Using the misclassified training data, it generates the advice weights. These weights are used along with decision values from the SVM during the test phase. Another benefit of using these weights is the elimination of outlier data. In the end, the classifier is aggregated using least square error based weighting. Classification accuracies are used to determine the weights of the classifiers in the overall architecture [80]. The proposed method shows effective results and is an important application of machine learning to skin cancer diagnosis especially in the case of less labeled data.

A novel pretext task for self-supervised learning has been presented by Chen *et al.* [24] which they termed as *context restoration*. Specifically, they presented this approach for medical image analysis. Input and output images are paired together for training, and then the mapping is learned between them. Two steps are involved in the context restoration approach. Two small patches are randomly selected and swapped in the given input image. Repeating this process multiple times results in the formation of a new altered image. We can model this as a corruption function that changes the context of the original image and results in a new version of the image. Afterward, CNN is used to approximate an inverse function that restores the altered image to the original version. CNNs used to learn content restoration tasks learn useful semantic features which can later be utilized for main downstream tasks like classification, localization, and segmentation. The proposed network [24] consists of two main parts i.e. analysis part and reconstruction part. The purpose of the analysis part is to encode the disordered input images into feature maps which are used later in the reconstruction part to produce the correct output images. The convolution blocks used in the analysis part can be of any choice i.e. single convolution layers, residual blocks [62], inception modules [167], dense [68] and so on. The filter weights learned in this part are then used for the initialization of the following tasks. The reconstruction part is composed of upsampling and convolutional layers and is used to restore the original context in the produced output images. Again here we have the freedom of choosing any desired architecture. For the training of content restoration task, L2 loss is used. The algorithm is tested for the classification, localization, and segmentation tasks on fetal ultrasound, abdominal CT, and brain MR images respectively. Results indicate that the proposed self-supervised learning method based on the content restoration tasks outperforms previous pretraining methods. The proposed network is shown in Figure 12.

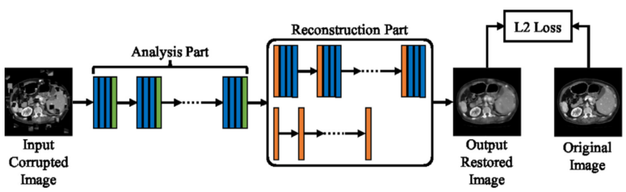

Figure 12. General CNN architecture self-supervised learning based on context restoration pretext task [24].

According to [10], an efficient and effective pretext task for medical data is to predict the anatomical positions defined by some bounding boxes. The intuition behind this approach is that in this way, the network not only learns features related to particular parts but also starts to recognize their neighboring regions. Then these learned features can be passed to the actual task like performing segmentation of different anatomical structures. Solving Jigsaw Puzzles is also another valuable proxy task for self-supervised learning [119]. This makes sense as the model not only has to learn the objects inside the images, but also the parts of those objects. This approach has been extended to multimodal medical data [169] with the intuition that by mixing data from various modalities, we can learn more informative features and representations. However, this multimodal puzzle-solving approach also leads to increased model complexity. To deal with this computational burden, the authors employ two techniques namely Sinkhorn operator [113] and a feed-forward network. Since we don't have a large amount of multimodal data always. However single modal data is still available in a large amount. The authors [169] also make use of Pix2Pix GAN-based image-to-image translation network [71] to generate more data for the modality with less amount of data, by utilizing data from another larger modality as shown in Figure 13. Experiments show that using data in this way produces better results compared to using data from a single modality for creating puzzles. The comparison indicates that the proposed model outperforms other "single-modal" and "from scratch" approaches.

3.0.7 Unsupervised learning methods.

Unsupervised machine learning models mainly deal with the methods of utilizing unlabeled data for training. The main goal is to identify and learn hidden structures, similarities, and relationships in the provided data samples [185]. Different clusters are formed based on the distance or similarity measures, hence samples that are alike are clustered together. Besides clustering algorithms, unsupervised learning algorithms can also use some kind of association rule which is used to find similarities and relationships between samples. Some common unsupervised algorithms include K-means clustering, K-nearest neighbors (KNN), hierarchical clustering, anomaly detections, Principal Component Analysis (PCA), singular value decomposition, and neural networks.

Unsupervised learning has been adopted in various domains in recent years and also in the past. However, the amount of work done in medical diagnosis via unsupervised algorithms is still very limited. Goswami *et al.* [52] utilized a neural network based on unsupervised learning for the detection of brain tumor using MRI images. Input images are first pre-processed to improve the quality of data by enhancing the desired features in the data and suppress the undesired noise and distortions. After that features are extracted by utilizing Independent component analysis, which is a technique for dimensionality reduction. This helps in the extraction of the most useful features and reduces redundancy which further results in improved performance and fewer computations. Finally, an unsupervised learning-based classification technique "Self-organizing map" (SOM) is used to classify the MRI brain images along with the K-means clustering algorithm. SOM is a form of competitive unsupervised learning and comprises an artificial neural network. Results indicate that the presented approach achieves higher accuracy than the previous methods.

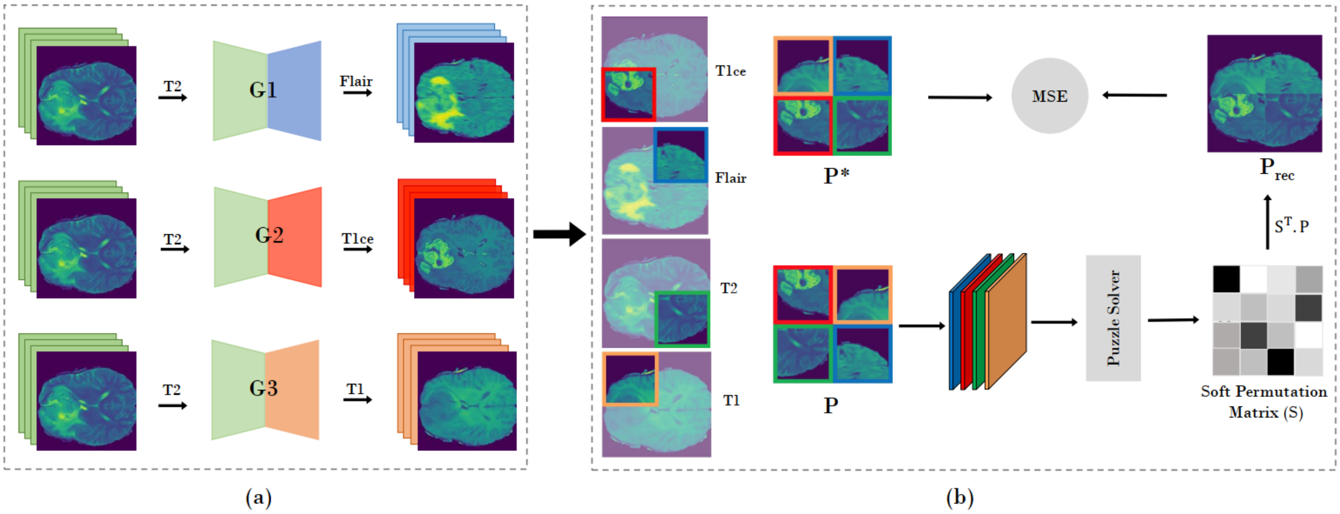

Figure 13. Proposed method for data generation: (a) Generating more data for smaller modalities by using data from larger modality (b) Synthetic and real images used to create multimodal jigsaw puzzles [169].

Unsupervised domain adaptation [64][124] transfer knowledge learned on a source domain with a large number of labeled examples to a target domain with unlabeled data. This is very useful in medical diagnosis when there is limited annotated data. However, in the case of medical images, there are two main issues. First, given the input images are from the same domain, there can still be significant discrepancies among those images. Second, a significant amount of medical annotations turns out to be noisy which can lead to poor performance. To deal with these issues, a novel Collaborative Unsupervised Domain Adaptation method (CoUDA) is proposed in [188]. Two separate peer networks are utilized to overcome the annotation noise and domain discrepancy collaboratively with more focus on hard-to-transfer images. Extensive experiments on real-world data show the generalization and effectiveness of the proposed network.

3.0.8 Semi-supervised learning methods. Semi-supervised learning [193] is another approach for machine learning which uses a combination of labeled and unlabeled i.e. it utilizes a small number of labeled examples with a large number of unlabeled data. Hence, it is in between supervised and unsupervised learning. Semi-supervised machine learning also makes use of large unlabeled data like unsupervised learning and helps in reducing the cost of annotations. Due to their success in various domains, semi-supervised algorithms have been used in the medical domain as well in recent years.

A semi-supervised learning-based algorithm for dimensionality reduction can be used to extract useful information from unlabeled data which helps in improving the classification accuracy on medical images [15]. To reduce the dimensionality, regularized matrix factorization formalism is used [16]. Two or more matric is obtained by the decomposition of a larger matrix such that the decomposition is as accurate as possible. This method also retains the semantics of the input images and hence produces useful results. A semi-supervised regularization is also utilized which controls the complexity of the model and also encourages the classifier to generate labels for an image based on the labels of its nearest neighbors. The proposed method outperforms supervised learning-based dimensionality reduction methods for the classification tasks.

Zhao *et al.* [190] proposed a compact graph-based semi-supervised learning approach for the medical diagnosis specifically for Alzheimer's disease. Authors claim that the newly proposed model can represent the data manifold in a more compact form and also predict whether a patient is suffering from dementia or not. To their knowledge, this is the first work that applied a semi-supervised graph-based method for medical diagnosis and shows better accuracy and performance compared to other graph-based semi-supervised learning methods. An immune-inspired semi-supervised algorithm has been presented for breast cancer diagnosis [127]. Breast cancer is one of the most serious and frequent forms of cancer among women which leads to many deaths among women. An automated algorithm for breast cancer diagnosis is presented (referred to as Aisl), which organically integrates semi-supervised learning with artificial immune. The authors try to model the breast cancer diagnosis problem as the mechanism of the biological immune system and regard the individuals with breast cancer as antigens. Semi-supervised learning is employed to make use of large unlabeled data. The unlabeled antigens are divided into different parts and for each part, classification is done based on the majority voting by k nearest neighbor cells to the presented test antigen. Evaluation is done on two known breast cancer datasets and shows satisfactory results.

Semi-supervised learning based GAN model is proposed for automated diagnosis of cardiac disease [106]. The loss function is comprised of three different loss functions namely supervised, unsupervised, and generated loss. Training the model involves one pass through the generator and three passes through the discriminator during each iteration. The labeled image is passed through the discriminator network to calculate the cross-entropy loss for the discriminator. After that, a fake image and an unlabeled image are passed through the discriminator and binary cross-entropy loss is computed for both. The summation is performed on all three losses which are then used in the backpropagation step for the discriminator. Different split levels between labeled and unlabeled data were utilized for training to determine the relationship of model accuracy with the amount of labeled data. Results indicate that model can learn with a very limited amount of labeled data i.e. only 4%, while remaining data is kept unlabeled and achieves an accuracy of up to 80% for classification task as shown in Figure 14.

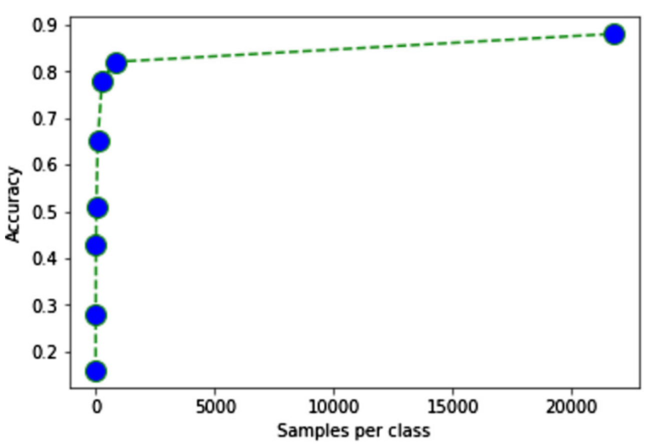

Figure 14. View classification performance of the semi-supervised generative adversarial network for varying amounts of labels as input [106].

4 Conclusion

Dealing with the need of huge training data for training machine learning models has been an active research area for past decades. Researchers have been trying to explore ways and methods to train machine learning models with less data and less supervision. It is a known fact that a machine learning model achieves better performance if it is provided with large amount of labeled data for training. However, obtaining such huge data becomes a hard task especially in domains like medical and healthcare where data is limited and annotations can be costly. This paper presented a survey of some of the recent attempts to deal with this issue of small data in medical diagnosis. Various techniques like GANs, few-shot learning, self-supervised learnign and so on have been adopted by researchers and have shown to improve the performance of machine learning models on tasks like classification, segmentation and detection by utilizing small training data and unlabeled data. These works also opened a gateway for further research on this topic and provided interesting topics for future research.

References

[1] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. 2020. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* (2020), 200642.

[2] Walid Abdullah Al and Il Dong Yun. 2019. Reinforcing Medical Image Classifier to Improve Generalization on Small Datasets. *arXiv preprint arXiv:1909.05630* (2019).

[3] Varghese Alex, Kiran Vaidhya, Subramaniam Thirunavukkarasu, Chandrasekharan Kesavadas, and Ganapathy Krishnamurthi. 2017. Semisupervised learning using denoising autoencoders for brain lesion detection and segmentation. *Journal of Medical Imaging* 4, 4 (2017), 041311.

[4] Han Altae-Tran, Bharath Ramsundar, Aneesh S Pappu, and Vijay Pande. 2017. Low data drug discovery with one-shot learning. *ACS central science* 3, 4 (2017), 283–293.

[5] Dario Amodei, Sundaram Ananthanarayanan, Rishita Anubhai, Jingliang Bai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Qiang Cheng, Guoliang Chen, et al. 2016. Deep speech 2: End-to-end speech recognition in english and mandarin. In *International conference on machine learning*. 173–182.

[6] Ioannis D Apostolopoulos and Tzani A Mpesiana. 2020. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* (2020), 1.

[7] Hidetaka Arimura, Taiki Magome, Yasuo Yamashita, and Daisuke Yamamoto. 2009. Computer-aided diagnosis systems for brain diseases in magnetic resonance images. *Algorithms* 2, 3 (2009), 925–952.

[8] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein gan. *arXiv preprint arXiv:1701.07875* (2017).

[9] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. 2017. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine* 34, 6 (2017), 26–38.

[10] Wenjia Bai, Chen Chen, Giacomo Tarroni, Jinming Duan, Florian Guitton, Steffen E Petersen, Yike Guo, Paul M Matthews, and Daniel Rueckert. 2019. Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 541–549.

[11] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. 2018. Metareg: Towards domain generalization using meta-regularization. In *Advances in Neural Information Processing Systems*. 998–1008.

[12] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. 2018. An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 9252–9260.

[13] Pierre Baldi. 2012. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning*. 37–49.

[14] Imon Banerjee, Sriraman Madhavan, Roger Eric Goldman, and Daniel L Rubin. 2017. Intelligent word embeddings of free-text radiology reports. In *AMIA Annual Symposium Proceedings*, Vol. 2017. American Medical Informatics Association, 411.

[15] Kayhan N Batmanghelich, H Ye Dong, Kilian M Pohl, Ben Taskar, Christos Davatzikos, et al. 2011. Disease classification and prediction via semi-supervised dimensionality reduction. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 1086–1090.

[16] Nematollah Batmanghelich, Ben Taskar, and Christos Davatzikos. 2009. A general and unifying framework for feature construction, in image-based pattern classification. In *International Conference on Information Processing in Medical Imaging*. Springer, 423–434.

[17] Luca Bertinetto, João F Henriques, Jack Valmadre, Philip Torr, and Andrea Vedaldi. 2016. Learning feed-forward one-shot learners. In *Advances in neural information processing systems*. 523–531.

[18] Harald Binder and Maria Blettner. 2015. Big data in medical science—a biostatistical view: Part 21 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 112, 9 (2015), 137.

[19] Piotr Bojanowski and Armand Joulin. 2017. Unsupervised learning by predicting noise. *arXiv preprint arXiv:1704.05310* (2017).

[20] Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. 2017. Smash: one-shot model architecture search through hypernetworks. *arXiv preprint arXiv:1708.05344* (2017).

[21] Jane Bromley, James W Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. 1993. Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence* 7, 04 (1993), 669–688.

[22] Chin-Chen Chang, Hong-Hao Chen, Yeun-Chung Chang, Ming-Yang Yang, Chung-Ming Lo, Wei-Chun Ko, Yee-Fan Lee, Kao-Lang Liu, and Ruey-Feng Chang. 2017. Computer-aided diagnosis of liver tumors on computed tomography images. *Computer methods and programs in biomedicine* 145 (2017), 45–51.

[23] O Chapelle. 2006. B. Sch olkopf, and A. Zien, Semi-supervised Learning.

[24] Liang Chen, Paul Bentley, Kensaku Mori, Kazunari Misawa, Michitaka Fujiwara, and Daniel Rueckert. 2019. Self-supervised learning for medical image analysis using image context restoration. *Medical image analysis* 58 (2019), 101539.

[25] Liang-Chieh Chen, Alexander Hermans, George Papandreou, Florian Schroff, Peng Wang, and Hartwig Adam. 2018. Masklab: Instance segmentation by refining object detection with semantic and direction features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4013–4022.

[26] Xiaocong Chen, Lina Yao, Tao Zhou, Jinming Dong, and Yu Zhang. 2020. Momentum contrastive learning for few-shot covid-19 diagnosis from chest ct images. *arXiv preprint arXiv:2006.13276* (2020).

[27] Yu-An Chung and Wei-Hung Weng. 2017. Learning deep representations of medical images using siamese CNNs with application to content-based image retrieval. *arXiv preprint arXiv:1711.08490* (2017).

[28] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. 2018. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 168–172.

[29] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*. 160–167.

[30] Pedro Costa, Adrian Galdran, Maria Inês Meyer, Michael David Abràmoff, Meindert Niemeijer, Ana Maria Mendonça, and Aurélio Campilho. 2017. Towards adversarial retinal image synthesis. *arXiv preprint arXiv:1701.08974* (2017).

[31] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L Moreira, Narges Razavian, and Aristotelis Tsirigos. 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature medicine* 24, 10 (2018), 1559–1567.

[32] Pedro Domingos. 2012. A few useful things to know about machine learning. *Commun. ACM* 55, 10 (2012), 78–87.

[33] Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. 2019. Domain generalization via model-agnostic learning of semantic features. *Advances in Neural Information Processing Systems* 32 (2019), 6450–6461.

[34] Omar Emad, Inas A Yassine, and Ahmed S Fahmy. 2015. Automatic localization of the left ventricle in cardiac MRI images using deep

learning. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 683–686.

[35] Bradley J Erickson, Panagiotis Korfiatis, Zeynettin Akkus, and Timothy L Kline. 2017. Machine learning for medical imaging. *Radiographics* 37, 2 (2017), 505–515.

[36] Rasool Fakoor, Faisal Ladhak, Azade Nazi, and Manfred Huber. 2013. Using deep learning to enhance cancer diagnosis and classification. In *Proceedings of the international conference on machine learning*, Vol. 28. ACM New York, USA.

[37] Li Fei-Fei, Rob Fergus, and Pietro Perona. 2006. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence* 28, 4 (2006), 594–611.

[38] Jacques Ferlay, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer* 136, 5 (2015), E359–E386.

[39] Abdur R Feyjie, Reza Azad, Marco Pedersoli, Claude Kauffman, Ismail Ben Ayed, and Jose Dolz. 2020. Semi-supervised few-shot learning for medical image segmentation. *arXiv preprint arXiv:2003.08462* (2020).

[40] Michael Fink. 2005. Object classification from a single example utilizing class relevance metrics. In *Advances in neural information processing systems*. 449–456.

[41] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. *arXiv preprint arXiv:1703.03400* (2017).

[42] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. 2018. Synthetic data augmentation using GAN for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 289–293.

[43] Victor Garcia and Joan Bruna. 2018. Few-Shot Learning with Graph Neural Networks. arXiv:1711.04043 [stat.ML]

[44] Rishab Gargeya and Theodore Leng. 2017. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 124, 7 (2017), 962–969.

[45] Krzysztof J Geras, Stacey Wolfson, Yiqiu Shen, Nan Wu, S Kim, Eric Kim, Laura Heacock, Ujas Parikh, Linda Moy, and Kyunghyun Cho. 2017. High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv preprint arXiv:1703.07047* (2017).

[46] Spyros Gidaris and Nikos Komodakis. 2018. Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4367–4375.

[47] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. 2018. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728* (2018).

[48] Miltos Gletsos, Stavroula G Mougiakakou, George K. Matsopoulos, Konstantina S. Nikita, Alexandra S Nikita, and Dimitrios Kelekis. 2003. A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier. *IEEE transactions on information technology in biomedicine* 7, 3 (2003), 153–162.

[49] Yoav Goldberg. 2016. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research* 57 (2016), 345–420.

[50] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*. Vol. 1. MIT press Cambridge.

[51] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[52] Suchita Goswami and Lalit Kumar P Bhaiya. 2013. Brain tumour detection using unsupervised learning based neural network. In *2013 International Conference on Communication Systems and Network Technologies*. IEEE, 573–577.

[53] Alex Graves and Navdeep Jaitly. 2014. Towards end-to-end speech recognition with recurrent neural networks. In *International conference on machine learning*. 1764–1772.

[54] Hayit Greenspan, Bram Van Ginneken, and Ronald M Summers. 2016. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging* 35, 5 (2016), 1153–1159.

[55] Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2. IEEE, 1735–1742.

[56] Abdul Mueed Hafiz and Ghulam Mohiuddin Bhat. 2020. A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval* (2020), 1–19.

[57] Changhee Han, Kohei Murao, Tomoyuki Noguchi, Yusuke Kawata, Fumiya Uchiyama, Leonardo Rundo, Hideki Nakayama, and Shin'ichi Satoh. 2019. Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 119–127.

[58] Changhee Han, Leonardo Rundo, Ryosuke Araki, Yujiro Furukawa, Giancarlo Mauri, Hideki Nakayama, and Hideaki Hayashi. 2020. Infinite brain MR images: PGGAN-based data augmentation for tumor detection. In *Neural approaches to dynamics of signal exchanges*. Springer, 291–303.

[59] Changhee Han, Leonardo Rundo, Ryosuke Araki, Yudai Nagano, Yujiro Furukawa, Giancarlo Mauri, Hideki Nakayama, and Hideaki Hayashi. 2019. Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection. *IEEE Access* 7 (2019), 156966–156977.

[60] Jiawei Han, Jian Pei, and Micheline Kamber. 2011. *Data mining: concepts and techniques*. Elsevier.

[61] Bharath Hariharan and Ross Girshick. 2017. Low-shot visual recognition by shrinking and hallucinating features. In *Proceedings of the IEEE International Conference on Computer Vision*. 3018–3027.

[62] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[63] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Identity mappings in deep residual networks. In *European conference on computer vision*. Springer, 630–645.

[64] Tobias Heimann, Peter Mountney, Matthias John, and Razvan Ionasec. 2013. Learning without labeling: Domain adaptation for ultrasound transducer localization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 49–56.

[65] Ezz El-Din Hemdan, Marwa A Shouman, and Mohamed Esmail Karar. 2020. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055* (2020).

[66] Ehsan Hosseini-Asl, Robert Keynton, and Ayman El-Baz. 2016. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 126–130.

[67] Zilong Hu, Jinshan Tang, Ziming Wang, Kai Zhang, Ling Zhang, and Qingling Sun. 2018. Deep learning for image-based cancer detection and diagnosis- a survey. *Pattern Recognition* 83 (2018), 134–149.

[68] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

[69] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. 2018. Multimodal unsupervised image-to-image translation. In *Proceedings of*

the European Conference on Computer Vision (ECCV)*. 172–189.
[70] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).
[71] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.
[72] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. 2020. A Survey on Contrastive Self-supervised Learning. arXiv:2011.00362 [cs.CV]
[73] Amir Jamaludin, Timor Kadir, and Andrew Zisserman. 2017. Self-supervised learning for spinal MRIs. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 294–302.
[74] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. 2017. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 11–19.
[75] Longlong Jing and Yingli Tian. 2020. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).
[76] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis* 36 (2017), 61–78.
[77] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).
[78] SanaUllah Khan, Naveed Islam, Zahoor Jan, Ikram Ud Din, and Joel JP C Rodrigues. 2019. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters* 125 (2019), 1–6.
[79] Mikhail Khodak, Maria-Florina F Balcan, and Ameet S Talwalkar. 2019. Adaptive gradient-based meta-learning methods. In *Advances in Neural Information Processing Systems*. 5917–5928.
[80] Hyun-Chul Kim, Shaoning Pang, Hong-Mo Je, Daijin Kim, and Sung-Yang Bang. 2002. Support vector machine ensemble with bagging. In *International Workshop on Support Vector Machines*. Springer, 397–408.
[81] Mijung Kim, Jasper Zuallaert, and Wesley De Neve. 2017. Few-shot learning using a small-sized dataset of high-resolution FUNDUS images for glaucoma diagnosis. In *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*. 89–92.
[82] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
[83] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
[84] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences* 114, 13 (2017), 3521–3526.
[85] Stefan Knerr, Léon Personnaz, and Gérard Dreyfus. 1990. Single-layer learning revisited: a stepwise procedure for building and training a neural network. In *Neurocomputing*. Springer, 41–50.
[86] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. 2015. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, Vol. 2. Lille.
[87] Alex Krizhevsky and Geoffrey E Hinton. 2011. Using very deep autoencoders for content-based image retrieval.. In *ESANN*, Vol. 1. Citeseer, 2.
[88] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
[89] Vijay Kumar B G, Gustavo Carneiro, and Ian Reid. 2016. Learning Local Image Descriptors With Deep Siamese and Triplet Convolutional Networks by Minimising Global Loss Functions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
[90] Hamid Laga, Laurent Valentin Jospin, Farid Boussaid, and Mohammed Bennamoun. 2020. A Survey on Deep Learning Techniques for Stereo-based Depth Estimation. *arXiv preprint arXiv:2006.02535* (2020).
[91] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. 2017. Colorization as a proxy task for visual understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6874–6883.
[92] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.
[93] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.
[94] Choong Ho Lee and Hyung-Jin Yoon. 2017. Medical big data: promise and challenges. *Kidney research and clinical practice* 36, 1 (2017), 3.
[95] Kelvin K Leung, Matthew J Clarkson, Jonathan W Bartlett, Shona Clegg, Clifford R Jack Jr, Michael W Weiner, Nick C Fox, Sébastien Ourselin, Alzheimer's Disease Neuroimaging Initiative, et al. 2010. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *Neuroimage* 50, 2 (2010), 516–523.
[96] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. 2017. Learning to generalize: Meta-learning for domain generalization. *arXiv preprint arXiv:1710.03463* (2017).
[97] Der-Chiang Li, Chih-Sen Wu, Tung-I Tsai, and Yao-San Lina. 2007. Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge. *Computers & Operations Research* 34, 4 (2007), 966–982.
[98] Xiaomeng Li, Lequan Yu, Yueming Jin, Chi-Wing Fu, Lei Xing, and Pheng-Ann Heng. 2020. Difficulty-aware meta-learning for rare disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 357–366.
[99] Zachary C Lipton, John Berkowitz, and Charles Elkan. 2015. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019* (2015).
[100] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports* 6 (2016), 26286.
[101] Fayao Liu, Chunhua Shen, and Guosheng Lin. 2015. Deep convolutional neural fields for depth estimation from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5162–5170.
[102] Quande Liu, Qi Dou, and Pheng-Ann Heng. 2020. Shape-aware Meta-learning for Generalizing Prostate MRI Segmentation to Unseen Domains. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 475–485.
[103] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E Alsaadi. 2017. A survey of deep neural network architectures and their applications. *Neurocomputing* 234 (2017), 11–26.
[104] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.
[105] Yashar Maali and Adel Al-Jumaily. 2013. Self-advising support vector machine. *Knowledge-Based Systems* 52 (2013), 214–222.

[106] Ali Madani, Jia Rui Ong, Anshul Tibrewal, and Mohammad RK Mofrad. 2018. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ digital medicine* 1, 1 (2018), 1–11.

[107] Gabriel Maicas, Andrew P Bradley, Jacinto C Nascimento, Ian Reid, and Gustavo Carneiro. 2018. Training medical image analysis systems like radiologists. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 546–554.

[108] Hamid Reza Marateb and Sobhan Goudarzi. 2015. A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences* 20, 3 (2015), 214.

[109] Stephen Marshland. 2014. *Machine Learning: An Algorithmic Perspective*. CRC press.

[110] Ammara Masood, Adel Al-Jumaily, and Khairul Anam. 2015. Self-supervised learning model for skin cancer diagnosis. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 1012–1015.

[111] Alfonso Medela, Artzai Picon, Cristina L Saratxaga, Oihana Belar, Virginia Cabezón, Riccardo Cicchi, Roberto Bilbao, and Ben Glover. 2019. Few shot learning in histopathological images: reducing the need of labeled data on biological datasets. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 1860–1864.

[112] I. Melekhov, J. Kannala, and E. Rahtu. 2016. Siamese network features for image matching. In *2016 23rd International Conference on Pattern Recognition (ICPR)*. 378–383. https://doi.org/10.1109/ICPR.2016.7899663

[113] Gonzalo Mena, David Belanger, Scott Linderman, and Jasper Snoek. 2018. Learning latent permutations with gumbel-sinkhorn networks. *arXiv preprint arXiv:1802.08665* (2018).

[114] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernocký, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010* 2, 1045–1048.

[115] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. 2020. Image segmentation using deep learning: A survey. *arXiv preprint arXiv:2001.05566* (2020).

[116] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *nature* 518, 7540 (2015), 529–533.

[117] Ali Narin, Ceren Kaya, and Ziynet Pamuk. 2020. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849* (2020).

[118] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. 2017. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 417–425.

[119] Mehdi Noroozi and Paolo Favaro. 2016. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European Conference on Computer Vision*. Springer, 69–84.

[120] Beata Nowok, Gillian M Raab, Chris Dibben, et al. 2016. synthpop: Bespoke creation of synthetic data in R. *J Stat Softw* 74, 11 (2016), 1–26.

[121] A Oguntimilehin and EO Ademola. 2014. A review of big data management, benefits and challenges. *A Review of Big Data Management, Benefits and Challenges* 5, 6 (2014), 1–7.

[122] Yujin Oh, Sangjoon Park, and Jong Chul Ye. 2020. Deep learning covid-19 features on cxr using limited training data sets. *IEEE Transactions on Medical Imaging* (2020).

[123] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).

[124] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.

[125] Angshuman Paul, Thomas C Shen, Niranjan Balachandar, Yuxing Tang, Yifan Peng, Zhiyong Lu, and Ronald M Summers. 2020. COMe-SEE: Cross-modality Semantic Embedding Ensemble for Generalized Zero-Shot Diagnosis of Chest Radiographs. In *Interpretable and Annotation-Efficient Learning for Medical Image Computing*. Springer, 103–111.

[126] Adrien Payan and Giovanni Montana. 2015. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. *arXiv preprint arXiv:1502.02506* (2015).

[127] Lingxi Peng, Wenbin Chen, Wubai Zhou, Fufang Li, Jin Yang, and Jiandong Zhang. 2016. An immune-inspired semi-supervised algorithm for breast cancer diagnosis. *Computer methods and programs in biomedicine* 134 (2016), 259–265.

[128] Tomas Pfister, James Charles, and Andrew Zisserman. 2014. Domain-adaptive discriminative one-shot learning of gestures. In *European Conference on Computer Vision*. Springer, 814–829.

[129] Harry A Quigley and Aimee T Broman. 2006. The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology* 90, 3 (2006), 262–267.

[130] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).

[131] Wullianallur Raghupathi and Viju Raghupathi. 2014. Big data analytics in healthcare: promise and potential. *Health information science and systems* 2, 1 (2014), 3.

[132] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. 2017. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017).

[133] Muhammad Imran Razzak and Bandar Alhaqbani. 2015. Automatic detection of malarial parasite using microscopic blood images. *Journal of Medical Imaging and Health Informatics* 5, 3 (2015), 591–598.

[134] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. 2017. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2001–2010.

[135] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[136] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[137] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B Tenenbaum, Hugo Larochelle, and Richard S Zemel. 2018. Meta-learning for semi-supervised few-shot classification. *arXiv preprint arXiv:1803.00676* (2018).

[138] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.

[139] Joachim Roski, George W Bo-Linn, and Timothy A Andrews. 2014. Creating value in health care through big data: opportunities and policy implications. *Health affairs* 33, 7 (2014), 1115–1122.

[140] Tobias Ross, David Zimmerer, Anant Vemuri, Fabian Isensee, Manuel Wiesenfarth, Sebastian Bodenstedt, Fabian Both, Philip Kessler, Martin Wagner, Beat Müller, et al. 2018. Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. *International journal of computer assisted radiology and surgery* 13, 6

(2018), 925–933.
[141] Masaharu Sakamoto and Hiroki Nakano. 2016. Cascaded neural networks with selective classifiers and its evaluation using lung x-ray ct images. *arXiv preprint arXiv:1611.07136* (2016).
[142] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. 2016. One-shot learning with memory-augmented neural networks. *arXiv preprint arXiv:1605.06065* (2016).
[143] Saman Sarraf, Ghassem Tofighi, Alzheimer's Disease Neuroimaging Initiative, et al. 2016. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *BioRxiv* (2016), 070441.
[144] Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural networks* 61 (2015), 85–117.
[145] Thomas Schreiber and Andreas Schmitz. 1996. Improved surrogate data for nonlinearity tests. *Physical review letters* 77, 4 (1996), 635.
[146] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 815–823.
[147] Torgyn Shaikhina and Natalia A Khovanova. 2017. Handling limited datasets with neural networks in medical applications: A small-data approach. *Artificial intelligence in medicine* 75 (2017), 51–63.
[148] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging* 35, 5 (2016), 1285–1298.
[149] Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine P Andriole, and Mark Michalski. 2018. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *International workshop on simulation and synthesis in medical imaging*. Springer, 1–11.
[150] Syed Hamad Shirazi, Arif Iqbal Umar, Nuhman Ul Haq, Saeeda Naz, and Muhammad Imran Razzak. 2015. Accurate microscopic red blood cell image enhancement and segmentation. In *International Conference on Bioinformatics and Biomedical Engineering*. Springer, 183–192.
[151] Syed H Shirazi, Arif Iqbal Umar, Saeeda Naz, and Muhammad I Razzak. 2016. Efficient leukocyte segmentation and recognition in peripheral blood image. *Technology and Health Care* 24, 3 (2016), 335–347.
[152] Mai Shouman, Tim Turner, and Rob Stocker. 2012. Integrating decision tree and k-means clustering with different initial centroid selection methods in the diagnosis of heart disease patients. In *Proceedings of the International Conference on Data Science (ICDATA)*. The Steering Committee of The World Congress in Computer Science, Computer . . . , 1.
[153] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. 2017. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2107–2116.
[154] Rebecca Siegel, Deepa Naishadham, and Ahmedin Jemal. 2012. Cancer statistics, 2012. *CA: a cancer journal for clinicians* 62, 1 (2012), 10–29.
[155] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
[156] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, David RJ Snead, Ian A Cree, and Nasir M Rajpoot. 2016. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging* 35, 5 (2016), 1196–1206.
[157] Siuly Siuly and Yanchun Zhang. 2016. Medical big data: neurological diseases diagnosis through medical data analysis. *Data Science and Engineering* 1, 2 (2016), 54–64.
[158] Gerard P Slobogean, Peter V Giannoudis, Frede Frihagen, Mary L Forte, Saam Morshed, and Mohit Bhandari. 2015. Bigger data, bigger problems. *Journal of orthopaedic trauma* 29 (2015), S43–S46.
[159] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in neural information processing systems*. 4077–4087.
[160] Richard Socher, Milind Ganjoo, Christopher D Manning, and Andrew Ng. 2013. Zero-Shot Learning Through Cross-Modal Transfer. In *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 935–943. http://papers.nips.cc/paper/5027-zero-shot-learning-through-cross-modal-transfer.pdf
[161] Richard Socher, Cliff Chiung-Yu Lin, Andrew Y Ng, and Christopher D Manning. 2011. Parsing natural scenes and natural language with recursive neural networks. In *ICML*.
[162] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*. 1631–1642.
[163] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.
[164] Kirubakaran Sumathi and E Kirubakaran. 2012. Enhanced Weighted K-Means Clustering Based Risk Level Prediction for Coronary Heart Disease. *European Journal of Scientific Research* 71, 4 (2012), 490–500.
[165] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. 2018. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1199–1208.
[166] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. 2016. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261* (2016).
[167] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.
[168] Nima Tajbakhsh, Yufei Hu, Junli Cao, Xingjian Yan, Yi Xiao, Yong Lu, Jianming Liang, Demetri Terzopoulos, and Xiaowei Ding. 2019. Surrogate supervision for medical image analysis: Effective deep learning from limited quantities of labeled data. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 1251–1255.
[169] Aiham Taleb, Christoph Lippert, Tassilo Klein, and Moin Nabi. 2019. Multimodal self-supervised learning for medical image analysis. *arXiv preprint arXiv:1912.05396* (2019).
[170] Mingxing Tan and Quoc V Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946* (2019).
[171] Fabio Henrique Kiyoiti dos Santos Tanaka and Claus Aranha. 2019. Data augmentation using GANs. *arXiv preprint arXiv:1904.09135* (2019).
[172] Yichuan Tang. 2013. Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239* (2013).
[173] J Timmer. 1998. Power of surrogate data testing with respect to nonstationarity. *Physical Review E* 58, 4 (1998), 5153.
[174] Joaquin Vanschoren. 2018. Meta-learning: A survey. *arXiv preprint arXiv:1810.03548* (2018).
[175] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. In *Advances in neural information processing systems*. 3630–3638.

[176] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. 2016. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718* (2016).

[177] Linda Wang, Zhong Qiu Lin, and Alexander Wong. 2020. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports* 10, 1 (2020), 1–12.

[178] Wei Wang, Vincent W Zheng, Han Yu, and Chunyan Miao. 2019. A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, 2 (2019), 1–37.

[179] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. 2020. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)* 53, 3 (2020), 1–34.

[180] Zhengwei Wang, Qi She, and Tomas E. Ward. 2020. Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. arXiv:1906.01529 [cs.LG]

[181] Georg Wimmer, Sebastian Hegenbart, Andreas Vécsei, and Andreas Uhl. 2016. Convolutional neural network architectures for the automated diagnosis of celiac disease. In *International Workshop on Computer-Assisted and Robotic Endoscopy*. Springer, 104–113.

[182] Eric Wu, Kevin Wu, David Cox, and William Lotter. 2018. Conditional infilling GANs for data augmentation in mammogram classification. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Springer, 98–106.

[183] Yongqin Xian, Bernt Schiele, and Zeynep Akata. 2017. Zero-Shot Learning - the Good, the Bad and the Ugly. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[184] Lingxi Xie, Jingdong Wang, Zhen Wei, Meng Wang, and Qi Tian. 2016. Disturblabel: Regularizing cnn on the loss layer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4753–4762.

[185] Rui Xu and Donald Wunsch. 2005. Survey of clustering algorithms. *IEEE Transactions on neural networks* 16, 3 (2005), 645–678.

[186] Ruixiang ZHANG, Tong Che, Zoubin Ghahramani, Yoshua Bengio, and Yangqiu Song. 2018. MetaGAN: An Adversarial Approach to Few-Shot Learning. In *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.). Curran Associates, Inc., 2365–2374. http://papers.nips.cc/paper/7504-metagan-an-adversarial-approach-to-few-shot-learning.pdf

[187] Richard Zhang, Phillip Isola, and Alexei A Efros. 2017. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1058–1067.

[188] Yifan Zhang, Ying Wei, Qingyao Wu, Peilin Zhao, Shuaicheng Niu, Junzhou Huang, and Mingkui Tan. 2020. Collaborative unsupervised domain adaptation for medical image diagnosis. *IEEE Transactions on Image Processing* 29 (2020), 7834–7844.

[189] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V Guttag, and Adrian V Dalca. 2019. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8543–8553.

[190] Mingbo Zhao, Rosa HM Chan, Tommy WS Chow, and Peng Tang. 2014. Compact graph based semi-supervised learning for medical diagnosis in alzheimer's disease. *IEEE signal processing letters* 21, 10 (2014), 1192–1196.

[191] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. 2019. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems* 30, 11 (2019), 3212–3232.

[192] Rongsheng Zhu, Rong Zhang, and Dixiu Xue. 2015. Lesion detection of endoscopy images based on convolutional neural network features. In *2015 8th International Congress on Image and Signal Processing (CISP)*. IEEE, 372–376.

[193] Xiaojin Zhu and Andrew B Goldberg. 2009. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning* 3, 1 (2009), 1–130.