Beyond Supervised Learning for Pervasive Healthcare

Xiao Gu, Fani Deligianni, Jinpei Han, Xiangyu Liu,
Wei Chen, Guang-Zhong Yang*, and Benny Lo*

***Abstract*—The integration of machine/deep learning and sensing technologies is transforming healthcare and medical practice. However, inherent limitations in healthcare data, namely *scarcity*, *quality*, and *heterogeneity*, hinder the effectiveness of supervised learning techniques which are mainly based on pure statistical fitting between data and labels. In this paper, we first identify the challenges present in machine learning for pervasive healthcare and we then review the current trends beyond fully supervised learning that are developed to address these three issues. Rooted in the inherent drawbacks of empirical risk minimization that underpins pure fully supervised learning, this survey summarizes seven key lines of learning strategies, to promote the generalization performance for real-world deployment. In addition, we point out several directions that are emerging and promising in this area, to develop data-efficient, scalable, and trustworthy computational models, and to leverage multi-modality and multi-source sensing informatics, for pervasive healthcare.**

***Index Terms*—Pervasive Sensing, Healthcare, Deep Learning, Machine Learning, Real-World Applications.**

I. Introduction

Over the past decade, paradigm shifts have been made towards the way healthcare is delivered and managed. In particular, emerging pervasive sensing technologies, ranging from flexible, ultra-thin sensors, to low-power modules, coupled with advanced data analytics [1], [2], have enabled real-time personalized health monitoring, which transforms healthcare practice in terms of diagnostics [3], preventive healthcare [4], and rehabilitative and assistive technologies [5], as shown in Figure 1. Systems that are capable of acquiring physiological and behavioral signals are established, and they gradually turn into a common practice to draw information that indicates the states of human health and well being [6].

In particular, artificial intelligence (AI), especially deep learning, has enabled the analysis of large complex sensing information, with a series of advanced computational models being proposed and developed. The combination of these models and large datasets has resulted in systems that can potentially outperform clinical experts in a variety of healthcare applications [3], [7].

Albeit the superior performance demonstrated by artificial intelligence in sensing informatics, the success of existing approaches is mostly driven by statistical fitting between input data and given labels. This fully supervised learning paradigm largely relies on the diversity and size of curated data that assumes well-balanced datasets with identical distributions between training and testing data. However, for real-world healthcare applications utilizing sensing informatics, these prior assumptions seldom hold, making the robustness of AI methodologies poor, especially when the algorithm development and deployment stage are mostly separated [8]. Specifically, we have raised the following three questions regarding the challenges in the context of deploying deep learning in data obtained by emerging, innovative sensing technologies,

- **Scarcity** - Although pervasive sensing enables continuous long-term data collection upon sensor deployment, *can we get enough data-label pairs for training?*
- **Quality** - Although pervasive sensing enables data collection in free-living environments, beyond controlled laboratory settings, *can we guarantee the quality of both the sensing data and associated labels?*
- **Heterogeneity** - Although pervasive sensing enables collecting data from multiple sources (subjects, devices, hospitals, etc.), *can we make sure that the data is collected from the same data space, independently and identically distributed?*

In fact, the answers to the above three questions are mostly unfavorable in the real world, posing significant obstacles in leveraging sensing informatics to meet practical healthcare needs, by pure fully supervised learning paradigms. In this paper, we provide theoretical insights into these three issues underlying existing machine/deep learning methodologies for healthcare sensing applications, summarize the problem's nature and challenges, followed by in-depth discussions on typical examples, and present a comprehensive review of potential solutions.

It should be noted that existing surveys have provided insightful ideas in terms of different aspects of model generalization and robustness [9], [10], [11], [12], [13]. Meanwhile, a series of surveys have discussed the application of deep learning in typical sensing modalities [2], [14], [15], [16], [17].

Xiao Gu and Benny Lo are with the Hamlyn Centre, Imperial College London, London SW7 2AZ, U.K. (e-mail: xiao.gu17@imperial.ac.uk; benny.lo@imperial.ac.uk).

Fani Deligianni is with the School of Computing Science, University of Glasgow, Glasgow G12 8RZ, U.K. (e-mail: fani.deligianni@glasgow.ac.uk).

Jinpei Han is with the Brain & Behaviour Lab, Imperial College London, London SW7 2AZ, U.K. (e-mail: j.han20@imperial.ac.uk).

Xiangyu Liu is with the College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: liuxiangyu@usst.edu.cn).

Wei Chen is with the Center for Intelligent Medical Electronics, School of Information Science and Technology, Fudan University, Shanghai 200433, China (e-mail: w_chen@fudan.edu.cn).

Guang-Zhong Yang is with the Institute of Medical Robotics and School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: gzyang@sjtu.edu.cn).

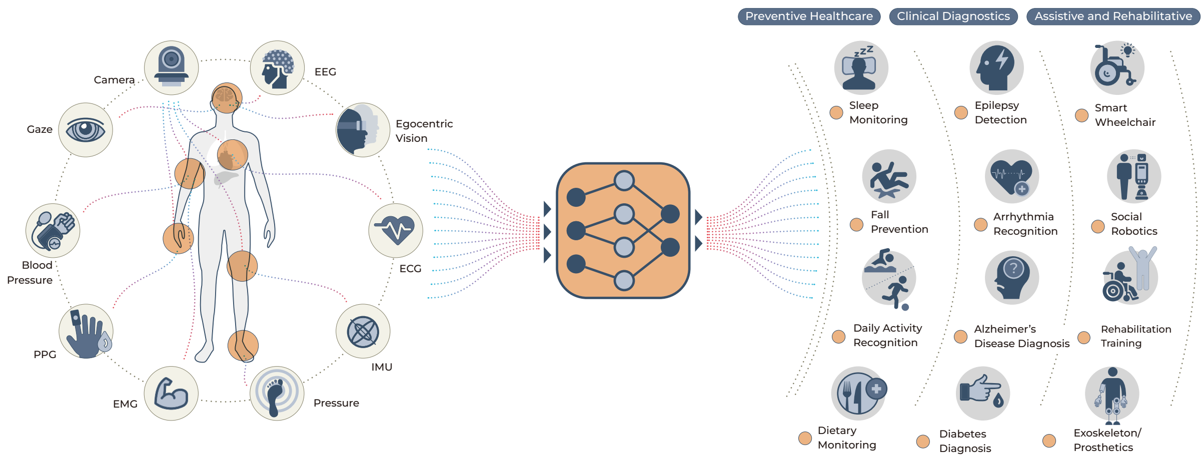

Fig. 1. The integration of advanced sensing informatics and machine/deep learning technologies benefits healthcare applications, ranging from preventive healthcare, clinical diagnostics, to rehabilitation and assistive technologies.

In contrast, this survey is not focused on discussing specific applications, and instead, provides insights into the machine learning challenges associated with conventional learning paradigms when handling pervasive sensing informatics in health care. Our contributions are listed below.

- **Problem Nature and Challenges.** We present a summary that provides theoretical insights into the inherent issues associated with applying deep/machine learning to sensory data in healthcare. We emphasize the common issues across varied healthcare applications, and provide context with concrete examples for intuitive understanding.
- **Comprehensive Review of Technical Approaches.** A comprehensive survey of the state-of-the-art machine/deep learning approaches to address the above issues in the existing literature. This provides a guide to key methodologies and inspiration for research on pervasive healthcare.
- **Promising and Emerging Research Directions.** An overview of promising and emerging research directions in pervasive healthcare, to foster research that well addresses real-world practical needs.

II. Challenges

Over the course of pervasive healthcare model development, without loss of generality, we denote the input data as X, and the output label as Y. The input data X represents the pervasive sensing informatics, with the data types ranging from videos/images (e.g., ambient/wearable cameras), time series (e.g., ECG, EEG, PPG, IMU), or multi-modality combinations. The output labels Y vary across different applications, thus leading to different machine learning tasks, such as classification, regression, segmentation, etc. The computational model $f : \mathcal{X} \rightarrow \mathcal{Y}$, infers the target based on the input. The purely fully supervised paradigm aims to minimize the average error across all the training samples $\{\boldsymbol{x}_i, y_i\}_{i=1}^{n}$, by commonly applied classification/regression loss functions $\mathcal{L}$ for measuring the discrepancy between the prediction and label. This is known as the *Empirical Risk Minimization* (ERM), formulated as below,

$$\mathcal{R} = \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(f(\boldsymbol{x}_i), y_i).^{1} \tag{1}$$

This is used to approximate the expected risk on the training data distributions P_{tr}, i.e., $\mathbb{E}_{(\boldsymbol{x},y)\sim p_{tr}}\mathcal{L}(f(\boldsymbol{x}), y)$.

Subsequently, the empirical error ϵ estimated on the testing data is, therefore, formulated as in Equation (2).

$$\begin{aligned}\epsilon &= \mathbb{E}_{(\boldsymbol{x},y)\sim p_{te}}\mathcal{L}(f(\boldsymbol{x}), y)\\ &= \mathbb{E}_{(\boldsymbol{x},y)\sim p_{tr}}\frac{p_{te}(\boldsymbol{x}, y)}{p_{tr}(\boldsymbol{x}, y)}\mathcal{L}(f(\boldsymbol{x}), y),\end{aligned} \tag{2}$$

where p_{tr} and p_{te} refer to the distribution of training and testing data, respectively.

As indicated in Equation (2), such a learning paradigm requires large amounts of diverse and plausible paired $\{\boldsymbol{x}, y\}$ to derive the model f, and its success on predicting labels of the testing data mostly lies in the assumption of the identical distribution between $P_{te}(X, Y)$ and $P_{tr}(X, Y)$. However, in sensing-based healthcare applications, there are three main factors deteriorating the performance of model deployment. We summarize these key issues as **Scarcity**, **Quality**, **Heterogeneity**. Below, we give detailed explanations with concrete healthcare applications as examples.

A. *Scarcity - Limited amount of paired* $\{\boldsymbol{x}, y\}$

1) Label Scarcity - Limited Availability of Y: Recent advances in sensor development have made the acquisition of continuous physiological and behavioral data in the wild possible. This could generate vast volumes of health informatics data upon deployment. However, handling and annotating the collected raw data is not trivial.

Manual annotation of streaming data is often not practical, which requires domain-related medical knowledge and/or considerable annotation time. This applies especially to timestamp-wise annotation tasks, such as temporal event segmentation [18], or cardiac characteristic waveform recognition [19]. Typically, in lab settings, medical experts or highly specialized devices are used to acquire gold standards labeling [20], [21], [22]. For example, in sleep-studies,

[1] In Section III, for some approaches, the difference between their equations and this ERM, is highlighted in blue.

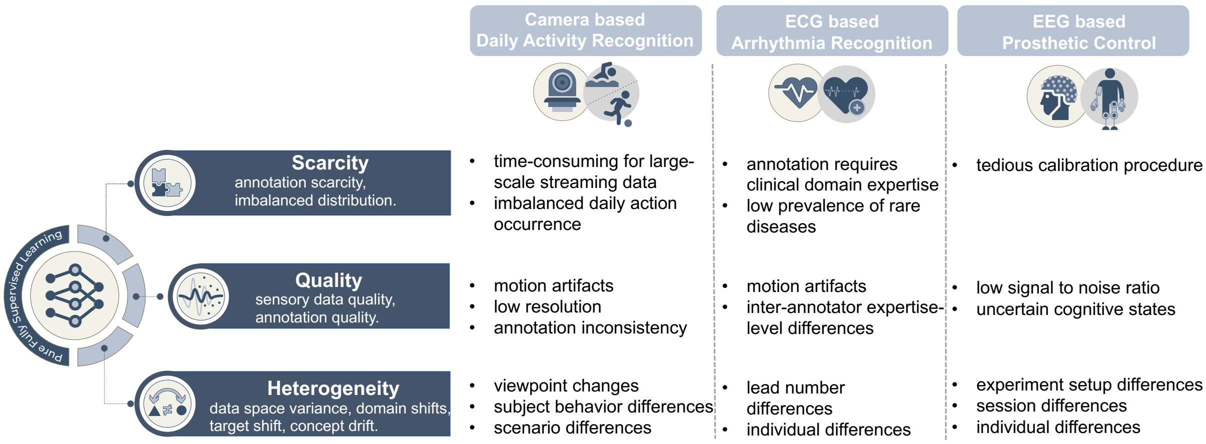

Fig. 2. Examples of the three main issues in sensing data acquisition and annotation for pervasive healthcare applications. We selected three representative applications that apply sensing techniques to healthcare practice, and list the concrete examples of the three issues inherent in these applications.

Polysomnography is used to acquire complex multi-modal data of brain/eye/muscle activities [23]. However, in free-living environments, it is not possible to get access to those gold-standard devices without tedious calibration procedures and complex experimental settings.

Furthermore, some tasks, such as affective recognition and fatigue/pain assessment, rely on retrospective self-assessment by the subjects [24]. Such pipeline is limited, in practice, to be applied in free-living environments. This results in sparse datasets and missing data, thus inducing biases over the course of modeling and inference.

2) Low Prevalence - Imbalanced $P(Y)$: The challenges associated with acquiring enough $\{X, Y\}$ training pairs result from not only the underlying difficulty in acquiring the labels Y, but also from insufficient X samples across certain Y categories [25]. For example, the prevalence of rare diseases and related adverse events is relatively low. This would also cause the total label space distribution $P(Y)$ imbalanced.

One typical example has been noted in established arrhythmia detection datasets [26], [27], with an extremely low prevalence of certain rare beat types, such as supraventricular premature [28], [29]. On the other hand, in daily routine based applications, it has also been argued that the label space follows severely imbalanced distributions, such as in human action recognition [18], and affective behavior analysis [30]. For instance, in action recognition datasets, there are only a few actions being frequently executed in daily routines, whereas a large number of actions only happen rarely. Such sample scarcity of minor classes contributes to an overall challenging long-tailed classification problem [31].

In fact, recent work has shown that the distribution of $P(Y)$ can be used as a posteriori to calibrate the model output probability so that the dominant classes are assigned with higher probabilities [32]. However, this process does not favor the classes with fewer samples. In fact, this would lead to a significant limitation for healthcare applications, where models are expected to perform consistently well across all classes, in order to be clinically useful. Furthermore, this posterior calibration would, rather, result in poorer performance, if there is any discrepancy of $P(Y)$ between training and testing data.

B. ***Quality*** - *Low-quality of $\boldsymbol{x}$ or y*

The capability of modeling the relationship between $\mathcal{X}$ and $\mathcal{Y}$ can be affected by the quality of either the raw data or labeling. The quality of sensor signals has been a long-lasting issue for real-world deployment. On the other hand, the inherent ambiguity/noises existing in acquiring healthcare-related ground-truth labels have also attracted significant research attention. Both issues have to be carefully handled during training, in order to avoid spurious correlations caused by the noises in X/Y.

1) Sensing Data Quality - Raw data X: Advances in sensing technologies have enabled the development of miniaturized, wireless sensor systems that are able to amplify the signal while suppressing unwanted noises. Nevertheless, deploying sensors in an unconstrained environment would attenuate the signal quality acquired compared to those in controlled experiments in various ways. For example, motion artifacts, loose contact, illumination changes, and undesirable sensor positioning would result in unpredictable deterioration of data quality [33], [34]. In addition, there might be crosstalk effects for multi-modality/channel signals that are difficult to detect and resolve, e.g., multi-channel EMG [35], EEG [36], [37]. Furthermore, insufficient power, transmission interruptions, weak signal strength, and limited memory capacity [38], would unfortunately result in signal loss and thus degrade the effective resolution and quality of data.

2) Label Quality - Label Y: Clean and consistent labeling Y is also challenging in most healthcare applications. Normally, labeling is defined in two ways: Firstly, experimental sessions are assigned beforehand, and data are labeled subsequently. This is mostly adopted in applications related to the cognitive/affective states, which are difficult to quantify objectively, such as motor imagery classification [39], workload estimation [40], and emotion recognition [41], [42]. For instance, in basic motor-imagery brain-computer interface (BCI) paradigms, the subjects are given an instruction cue to imagine the corresponding body movements; in emotion recognition studies, videos or images are used as stimuli. Those instructional cues or stimuli types are adopted as the ground truth of recorded data. However, one inherent drawback is that the subjects may not follow the instructions exactly, which will result in data mislabeling.

Secondly, the data are annotated during post-processing by experts retrospectively. Recruiting multiple annotators for crowd-sourcing annotation [43], [44] has been a common practice for labeling massive amounts of data, especially streaming data. In this case, biases and inconsistencies arise due to the subjective nature of the annotation or the differences in expertise level. This is a well-known issue in clinical practice, and it has been already raised in the curation of several sensing datasets [3], [45], [44].

C. ***Heterogeneity*** *- Mismatch of $P_{te}(X,Y)$ and $P_{tr}(X,Y)$*

Another reason for failing to successfully deploy a model pretrained on $\mathcal{X} \times \mathcal{Y}$ following $P_{tr}(X,Y)$ to $P_{te}(X,Y)$ is due to their mismatch [46], [47]. The past decade has witnessed a surge of research activities focused on mitigating the shifts between training and testing in general computer vision / machine learning field [48]. A series of benchmark datasets such as DomainNet [49] have been developed, which focus mainly on the domain shifts in the conditional probability $P(Y|X)$. However, in pervasive healthcare, the factors contributing to the heterogeneity are more complicated. Rethinking the composition of $P(X,Y)$, we summarize the representative heterogeneity factors in healthcare sensing as below.

1) Data Space $Dim(X)$ Shifts: Sensor signals are mostly multi-channel, such as EEG/EMG/ECG. Experiments conducted in the wild involve different channel numbers, sensor configuration changes, sampling frequency differences, and data-collection protocol variations, which results in dimensionality mismatch of the original data space [50], [51], [52], [53]. For example, Reyna *et al.* [51] has explored the model generalization capability of varied-lead ECGs for cardiovascular diseases. The authors pointed out the fact that although the standard 12-lead ECG is widely used in clinical diagnosis, portable devices for daily monitoring tend to have limited accessibility to all the leads [51]. Another example is the difference in the electrode channel settings utilized in BCI experiments both in terms of relative position and number. Xu *et al.* [52] noted differences in motor-imagery datasets, which hinders direct data aggregation, and it requires ad-hoc channel selection [52]. However, this eliminates potentially useful information.

2) Domain Shifts $P(X|Y)$: Similar to the general computer vision tasks, the conditional distribution $P(X|Y)$ shift is also an outstanding urgent issue for healthcare related sensing informatics. There are a variety of factors that can lead to the heterogeneity of $P(X|Y)$, ranging from inter-subject variability to differences in sensor positions, devices, and contexts.

In fact, the poor generalization performance in cross-subject settings has been noted in many healthcare studies [54], [21], [55], [56], [57]. In [21], the authors have suggested that the biometric information encoded in certain human-centric signals, such as gait, would bias the pathology related representation learning.

Furthermore, for most sensing applications, the relative sensor positions are critical to sensor characteristics, e.g., the body-worn positions of wearable sensors [58], and the perspectives of ambulatory sensors [59], [60]. For the wearable movement and physiological sensors, such as IMU, EMG, and EEG, it is impractical to constrain the same sensor position across sessions [58] and across subjects [61], [62]. Although some classic methods have introduced anatomical calibration strategies to mitigate this issue, it is not practical to translate them into free-living environments for pervasive sensing. Furthermore, in ambulatory sensing [59], [4], the relative perspective of the subject captured by the sensor is variable and hard to constrain.

In addition, the sensing context also plays an important role in data representation. For example, for vision signals collected in daily environments, the backgrounds and illuminations of the collected images might change across scenes [63]. Gait cycle characteristics would also vary between different conditions such as fixed-speed treadmill and self-paced [64].

3) Target Shifts $P(Y)$: Another issue results from the shift of the target distribution of $P(Y)$. These differences in the prevalence of the labels can occur when data are sampled from different subjects, contexts, population groups, etc. The prevalence of different disease categories would change along with personal characteristics and behaviors; For instance, the prevalence of some neurodegenerative diseases has been reported to exhibit gender differences [65]. This poses challenges to performing personalized prediction, diagnosis and treatment.

On the other hand, the probability of different classes also depends on the context. This results from the variances in the prevalence of a certain class across different contexts or population of interest. For example, for daily activity recognition, some actions like “turn on/off tap” can rarely occur in the living room but are quite common in the bathroom.

4) Concept Drifts $P(X,Y)_t$: In pervasive sensing, concept drifts [66], [67], i.e. $P(X,Y)_t \neq P(X,Y)_{t+1}$ manifest as the result of inherent sensor drifts over time. This non-stationary setting is a significant challenge in continuous online health monitoring tasks. Representative examples are the low-frequency drifts and measurement bias in physiological sensors [68] along with measurement error accumulation in inertial sensors [69]. Whereas this issue is not present in commonly applied offline training-testing paradigms of model deployment, in online scenarios they will compromise the classification/regression performance [70].

In summary, we provide a few examples regarding the aforementioned three issues in Figure 2 to illustrate these issues in real-world pervasive healthcare applications. A more detailed discussion based on specific applications is available in the Supplementary Material.

III. Technical Approaches

To handle the aforementioned three issues, we summarize the advanced learning approaches, covering effective data augmentations, sampling strategies, loss function engineering, supplementary tasks, adaptive learning strategies, mixture of multiple models, and domain knowledge guidance. A visual summary is shown in Figure 3.

A. Learn with Effective Augmentations

The learning theory underpinning ERM training has proved that one trivial solution to minimizing ERM is to memorize all the samples, especially when the sample size is limited, and the network is over-parameterized. This results in poor generalization when the testing samples are, even slightly, different from the training data [71]. Data augmentation can be seen as a regularization strategy, also known as the Vicinal Risk Minimization (VRM) principle [72]. It aims to create virtual training samples that are in close proximity of the original samples but not identical. This results in an increased amount and diversity of the original training samples. In this way, it aims to achieve generalization of the machine learning model and prediction invariance across datasets sampled from the same distribution. Hence, it is of vital importance to investigate effective augmentations that can lead to ideal vicinity space.

The creation and training of virtual data lying in vicinity space is formulated as below,

$$
\begin{aligned}
P_{tr}^{\mathcal{V}}(\hat{\boldsymbol{x}}, \hat{y}) &= \frac{1}{n}\sum_{i}^{n} \mathcal{V}(\hat{\boldsymbol{x}}, \hat{y} | \boldsymbol{x}_i, y_i), \\
\arg\min \frac{1}{m} &\sum_{i=1}^{m} \mathcal{L}(f(\hat{\boldsymbol{x}}_i), \hat{y}_i),
\end{aligned}
\tag{3}
$$

where $P_{tr}^{\mathcal{V}}$ refers to data distribution in the vicinity space.

1) Mixup: Existing research has demonstrated that by just linearly interpolating training samples in terms of both $\boldsymbol{x}$ and y, the model complexity can be implicitly controlled. This results in a robust augmentation strategy, known as Mixup, which introduces linear behavior as useful inductive bias to the model [71]. Mixup has been adopted widely and it has shown promising results in ECG [73], EEG [74], images/videos [75]. Following the success of mixup, several studies have further investigated the principle in class imbalance settings [76] and domain shifts [77].

2) Domain-Knowledge Guided Augmentation: Instead of data-agnostic linear interpolation, the vicinity of the training data could be based on domain-knowledge of the specific dataset. This results in commonly applied data augmentation strategies in the spatial domain (e.g., flipping, scaling, and rotation), temporal domain (e.g., temporal shuffling), and spectral domain (e.g., frequency filters and wavelet transformations) in varied types of sensing modalities [78], [79]. Nevertheless, the way these methods are applied can have a profound impact on the performance of the models. For example, imposing temporal consistency of spatio-temporal augmentations has been suggested to facilitate a significant improvement of model performance compared to naive augmentations in spatial or temporal dimensions alone [80].

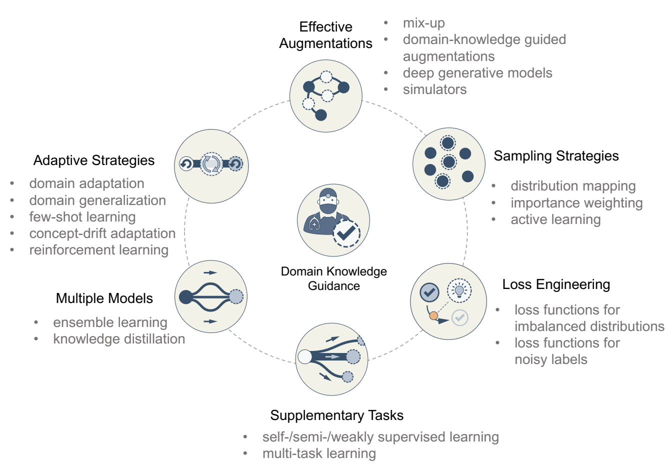

Fig. 3. Illustration of lines of learning approaches in existing literature that go beyond pure fully supervised learning, ranging from effective data augmentations, sampling strategies, loss engineering, supplementary tasks, adaptive strategies, mixture of multiple models, to domain knowledge guidance.

Different from the above commonly applied augmentation strategies, several works investigate augmentation techniques that are tailored to specific sensing modalities [81], [82], [83]. They were mostly inspired by application-specific underlying factors leading to the data variations in real-world sensing data. For instance, Gopal *et al.* [81] proposed a physiologically-inspired augmentation method over 12-lead ECG signals, which firstly maps multi-lead ECGs to vectorcardiogram (VCG) space, followed by 3D spatial augmentations and back-projection to the original ECG space. This is motivated by the observation that each ECG lead represents a different view of the same cardiac cycle and the VCGs derived from ECGs represent the heart's electrical activity along three orthogonal spatial axes. Another similar example is the rotational distortions applied on EEG or High-Density sEMG signals, which shift the electrode positions to generate artificial data [83], [84]. This can improve the model robustness against the variations of EEG-cap sensor positions.

3) Deep Generative Models: Instead of direct transformations of the input, a popular trend is to apply deep generative models to synthesize novel data by learning the data distribution in a generative manner. They approximate the data distribution in the latent space, based on the original training data, and subsequently realize the generation of virtual samples by drawing novel points from the latent space. Recent advances in deep generative models, such as variational autoencoder (VAE), generative adversarial network (GAN), flow, and diffusion models, have resulted in powerful methodologies for realistic synthetic data generation.

This line of work has been explored in a wide range of modalities, such as videos [85], ECG [86], EEG [87], EMG [88], etc. Performing data synthesis to induce more diversity has attracted considerable research efforts, such as [86], [87]. Recent research has also investigated the possibility of synthesizing data from different views/modalities, which is particularly useful when the data from one view/modality is sufficient and exhibits better quality, whereas that from another is limited. For example, Liu *et al.* [85] proposed a cross-view synthesis framework to generate first-person videos from those recorded from third-person viewpoints. Rey *et al.* [89] attempted to generate wearable IMU data from monocular RGB videos, simulating artificial yet semantically meaningful data.

On the other hand, the generative methodologies present an unsupervised solution to performing signal denoising/cleaning. Existing deep learning based signal denoising methods are mainly based on supervised training, and they mainly require paired clean-noisy sensing data; however it is practically difficult to acquire paired clean data for training in the real world. Several works in the existing literature have applied generative methods to generate clean samples from noisy samples [90], [91] or to learn artifacts [92], in an unsupervised manner.

4) Simulators: Another alternative is to use simulators to generate realistic synthetic data [93]. For vision signals, off-the-shelf simulators for novel viewpoint image generation exist and they incorporate several modalities (RGB, Depth, Infra-Red) [94]. Among existing literature, Srivastava *et al.* [95] developed a benchmark for embodied AI that can simulate a range of real-world activities. Tome *et al.* [96] developed a large-scale egocentric pose estimation dataset with the help of graphic render software Maya, leading to large variations across characters, environments, actions, and lighting conditions.

On the other hand, apart from off-the-shelf graphic simulators, a series of physiological and biomechanical parametric models have been proposed to explain or model the underlying function of real-world biological systems. For example, a system of ordinary differential equations (ODEs) was proposed by McSharry *et al.* [19] to describe the dynamics of characteristic ECG waveforms, which has been adopted in [97], [98] for ECG synthesis. Additionally, the parametric human model SMPL [99] and its follow-ups [100], [101], [102] have enabled a series of works on human-centric data synthesis. Recent works, such as synthesizing image data from different viewpoints [21], movement data (IMU) from different positions [103], [104], have demonstrated enhanced generalization capability enabled by diversity augmentation, thus mitigating the distribution gap between training-testing.

However, as mentioned, real-world sensing informatics are often contaminated with large noises, which cannot be easily modeled in a simulator based on a purely model-driven, bottom-to-top approach. Hence, significant domain shifts between those simulator-generated data and real-world data arise, which will impact real-world deployment. Further work on more advanced strategies should be applied to mitigate such real-synthetic domain gap issues [105], [21].

B. Learn with Sampling Strategies

Another commonly adopted strategy for machine learning model training is to deliberately change the original data distribution $P_{tr}(X, Y)$, so as to achieve identical distributions between training and testing or to amplify the importance of representative/hard samples, while minimizing the importance of noisy samples.

1) Distribution Mapping: In sensing applications, we often have to tackle situations where the training and test distribution are not the same, or the training distribution is imbalanced whereas the evaluation metrics treat each class equally. Resample-based approaches such as Random Over Sampler (ROS), Random Under Sampler (RUS), Cluster Centroids (CC), Near Miss (NM), Edited Nearest Neighbours (ENN), Repeated Edited Nearest Neighbors (RENN), Neighborhood Cleaning Rule (NCR) and One-Sided Selection (OSS) have been commonly applied to cope with the data imbalance issue [78], [106], [107], [108]. This strategy can also be integrated with the data augmentation algorithms in Section III-A to enforce changes in the distribution [109].

2) Importance Weighting:

$$\arg\min \frac{1}{n} \sum_{i=1}^{n} w(\boldsymbol{x}_i) \mathcal{L}(f(\boldsymbol{x}_i), y_i) \tag{4}$$

Recent research has found that not all samples are equally informative. For example, some data points contain noisy annotations, or they sample a small homogeneous neighborhood of the manifold space and thus they are trivial. Adjusting weights to different samples based on their importance (informativeness or noise degree) can facilitate obtaining better decision boundaries.

To mine hard informative samples that facilitate better model training, Lin *et al.* [110] considered those inputs with large MSE losses as hard samples when predicting hand kinematics from surface EMG, and only performed updates on these hard samples. Li *et al.* [111] jointly explored the sample importance in the task of EEG-based emotion recognition. It treated hard samples as those with higher losses and utilized a sample importance vector to identify them while minimizing the loss aggregated from all samples. Subsequently, a self-paced function was introduced to gradually train the model from easy to hard samples. The selection criteria of these works were mostly based on large training losses to find hard samples.

On the other hand, one line of research for learning with noisy labels is centered around identifying noisy samples and removing them during training [112], [113]. Most existing research built upon the assumption that clean samples would be associated with small training losses. Representative methods have been developed to select noisy samples. For example, co-Teaching [113] and JoCoR [114] utilized two networks with different learning abilities where only clean samples would exhibit consistent learning performance. They have been applied in ECG arrhythmia detection [108] and BCI [115]. Another work SCN [116] derived the importance of each sample by a self-attention mechanism, and treated those low-weighted samples as noisy ones in affective analysis.

It is noteworthy that for hard samples, their importance is normally targeted to be augmented, whilst for noisy samples, cases are opposite. However, it is often challenging to identify whether the sample is hard or noisy, since both are based on large-loss assumptions during training. Alternative solutions that do not explicitly pick up noisy samples by large training losses, are necessary to avoid such confusion [117]. For instance, in [117], the authors assumed that noisy labels in facial expression recognition would lead to a model only focusing on part of the relevant features, and subsequently applied attention-based erasing strategies to avoid memorizing the noisy samples.

3) Active Learning: On top of the above, another line of approach, active learning, tackles the annotation cost issue. It aims to minimize the involvement of human annotators by only selecting the most useful unlabeled samples for annotating, and thus the selection criteria for unlabeled samples are quite important for this approach. In most cases, the existing methodologies are either based on uncertainty or diversity/representativeness. The uncertainty-based approach aims to select the samples with the highest probability of being wrongly classified, i.e., those samples with high-entropy predictions. For instance, Wang *et al.* [118] proposed that the ECG samples that are near the classification hyperplane are the most informative samples because they tend to have high entropy. Similar ideas have also been applied in [119] to select the most informative samples for BCI inference model adaptation on new subjects, thus reducing calibration efforts. On the other hand, diversity/representativeness-based approach aims to select samples that could cover a diverse input and output spaces. On straightforward solution is to perform unsupervised clustering, and subsequently selects samples evenly distributed across clusters. Bi *et al.* [120] took into account both uncertainty and diversity/representativeness, when performing active learning based activity recognition with the presence of novel activities.

C. Learn with Loss Engineering

In classification tasks, commonly utilized functions of loss $\mathcal{L}$ are cross-entropy, whereas in regression tasks are $\mathcal{L}_1$ or $\mathcal{L}_2$ losses. These types of losses are not powerful enough when faced with the challenges described in Section II. There are several works dedicated to engineer loss functions to handle those issues.

1) Loss Functions for Imbalanced Distributions: In the context of imbalanced distribution, one of the most fundamental works is to design loss functions to balance the contribution of each class. A series of classification loss functions such as focal loss, label-distribution-aware margin loss, equalization loss, balanced softmax cross-entropy loss and so on have been developed to circumvent the imbalanced distribution [121]. These losses proposed varied strategies to balance the distribution of each class during training by logit adjustment. They have already been applied in various healthcare applications suffering from data imbalance issues, such as facial expression understanding [106], intention prediction [122], and anomalous ECG recognition [123], [124].

Compared to classification, regression tasks have received less attention in terms of their inherent imbalance issue [125], [126]. The study of Yang *et al.* [125] is the pioneering work that investigated imbalanced regression problems based on deep learning, followed by other works like [126] that developed a balanced regression loss function. They were mostly inspired by works from imbalanced classification; Yang *et al.* [125] focused on reweighting loss for samples lying in rare ranges, whilst Ren *et al.* [126] aimed to approximate a surrogate training loss on balanced distribution, statistically converted from the imbalanced training data.

2) Loss Functions for Noisy Labels: Loss function customization has been an important direction in handling label noises [12]. The goal is to design noise-robust or noise-aware loss functions, or estimates the label confusion matrix as caused by label noises to be injected into the loss function [127].

The core idea underpinning noise-robust loss function design is to achieve an identical global minimum of empirical loss (cfg. Eq.2) between training with noisy labels and training with clean labels. Existing research [12] indicated that simply applying cross-entropy loss cannot realize this appropriately. Instead, a series of loss functions such as generalized cross-entropy loss [128], symmetric cross entropy loss [129], and active passive loss [130], have been proposed to mitigate the limitation of cross-entropy loss.

Another way to deal with noisy labels is by approximating inter-class relationships to model the probability of mislabeling. Zhang *et al.* [131] developed an emotion-aware distribution loss to smooth the targeted label for EEG-based affect recognition. It smooths the probability of similar emotions yet keeps the probability of opposite emotions to zero. Zheng *et al.* [41] considered the problem of multi-dataset annotation inconsistency in facial expression datasets, and utilized multiple noise transition matrices, to maximize the log-likelihood of inconsistent annotations.

D. Learn with Supplementary Tasks

On top of direct end-to-end training for the single specific label Y of the targeted applications, a popular direction is to seek help from supplementary tasks.

1) Self-/Semi-/Weakly Supervised Learning: Since training a fully supervised learning model requires the availability of large amounts of data X with high-quality labels Y, which are not available in most healthcare applications, a series of self-/semi-/weakly supervised approaches have been proposed.

a) Self-Supervised Learning: Self-supervised learning aims to learn semantically meaningful representations by self-generated proxy tasks [10], and subsequently to benefit the downstream tasks (actual tasks) by minimizing the need for accessing a large number of data-label pairs for supervised training. It is formulated as below,

$$\arg\min \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(f(\boldsymbol{x}_i), p(\boldsymbol{x}_i)) \tag{5}$$

where $p(\boldsymbol{x}_i)$ refers to the corresponding label of $\boldsymbol{x}_i$ in the self-generated proxy tasks.

In this way, self-supervised methods exploit the inherent ability of deep neural networks (DNNs) to automatically learn hierarchical representations of the input data by self-curated labels. This has been an increasingly popular research topic in general image recognition and language understanding fields [10]. These works have inspired a surge of research activities, as well as posed novel challenges on self-supervised learning for pervasive sensing informatics.

Existing self-supervised learning approaches are generative, contrastive, or adversarial [132]. The generative approaches aim to perform reconstruction or context-related generation tasks to learn semantic representations [133], [134], whereas the contrastive ones maximize the similarity between the latent representations of positive-paired inputs, indexed from pseudo labels [135], [136]. On the other hand, adversarial approaches can be seen as an amalgamation of generative and contrastive learning, since they combine the generative process with a distributional divergence loss. Existing self-supervised pretexts on natural-image-recognition [135] or natural-language-processing [134] applications may not be directly applicable to sensing informatics, since their low-level representations are fundamentally different, and far more complicated. Instead, they emphasized the necessity of leveraging domain knowledge to build novel self-supervised approaches [137], [138], [139], [140], [141], [142].

For instance, Cheng *et al.* [137] suggested that biosignal representations include both disease/task information and biometrics. To learn the former in a self-supervised manner, they developed a subject-aware contrastive learning framework, which incorporated an adversarial subject identifier to learn subject-invariant representations from physiological signals such as EEG and ECG. Banville *et al.* [138] focused on exploring EEG representations with self-supervised learning tasks. Subsequently, classifiers developed based on the self-supervised features outperformed fully supervised DNNs, whereas the latent structure revealed clinically meaningful physiological parameters such as age effects. Wagh *et al.* [140] leveraged domain-guided knowledge to learn effective EEG representations, where three tasks including hemispheric symmetry, behavioral state estimation, and age contrastive loss, were proposed. These proposed tasks have demonstrated benefits to specific downstream EEG classification tasks (such as eye state detection, and gender classification). Additionally, Kiyasseh *et al.* [142] proposed a patient-specific contrastive learning method, exploiting both temporal and spatial invariance for ECG arrhythmia detection.

On the other hand, in connected healthcare settings, sensor signals are often collected from several modalities, and they reflect different aspects/views of the same underlying neuro-physiological process. Therefore, self-supervised learning can be facilitated by tasks that take into account the correlation across these co-occurring modalities. For example, Ehsani *et al.* [143] leveraged the human interaction and attention cues encoded in egocentric videos, including body part movement (IMU) and gaze, to help learn visual scene representations from egocentric videos. Shukla *et al.* [144] explored the interaction between audio and visual modalities for self-supervised emotion representation learning.

b) Semi-Supervised Learning: In semi-supervised learning, training is performed on both a small amount of labeled data and a large volume of unlabeled data [145], [146]. The general form of semi-supervised learning is as below,

$$\arg\min \frac{1}{n_L}\sum_{i=1}^{n_L}\mathcal{L}(f(\boldsymbol{x}_i), y_i) + \frac{1}{n_U}\sum_{i=1}^{n_U}\mathcal{L}'(f(\boldsymbol{x}_i), R(\boldsymbol{x}_i, \mathcal{X}^L)), \quad (6)$$

where n_L and n_U refer to the amount of labeled and unlabeled samples, and R refers to the task built upon the relationship between unlabeled and labeled datasets. A wide variety of methods have been developed to perform semi-supervised learning, and we refer the readers to [145], [147] for comprehensive theoretical surveys.

The key idea of semi-supervised learning is to find a surrogate loss that can be applied to unlabeled data. One representative line of work is the supervised-training based on pseudo labeling. Wei *et al.* [148] targeted at a motor imagery classification problem and proposed a pseudo-labeling strategy based on the feature distance to the class-wise prototypes generated from labeled samples. Li *et al.* [149] resorted to multi-branch prediction consistency to derive pseudo labels from ECG recordings. In action recognition field, researchers have exploited complementary modalities [150] or models [151] for more reliable pseudo labels.

Another line of work pays more attention to the principle of consistency regularization, assuming that the prediction should be similar under different augmentations and adversarial perturbations. Several works have integrated self-supervised learning solutions into semi-supervised learning frameworks by considering the principle of consistency regularization [152]. For example, FixMatch [153] was a successful combination of self-training and consistency regularization. It derived pseudo labels based on samples with weak augmentations and subsequently applies the pseudo labels as ground truth for the same samples superimposed with strong augmentations. Due to its simple yet effective strategy, it has been applied to a series of applications, such as activity recognition [151] and emotion recognition [154]. In some cases, since the pseudo labels may not be correct, the importance weighting strategies introduced in Section III-B2 can be applied to select those unlabeled samples with more convincing pseudo labels [155].

c) Weakly Supervised Learning: To tackle the lack of large, high-quality labeled data, weakly-supervised learning relies on "weakly"-annotated data that is much easier to acquire. Weakly supervised learning is categorized into inexact, incomplete and inaccurate supervision that corresponds, respectively, to scenarios where high-quality labels exist only for a small subset of training data, only indistinct labels exist for the training data and labels are inaccurate [156]. In this section, we would like to focus more on inexact supervision, where coarse-gained annotations or descriptions are given, whereas the solutions for the other two types have been covered in other sections including self-/semi-supervised learning (Section III-D1), importance weighting (Section III-B2)/loss functions (Section III-C2) for noisy labels.

The formula representing coarse-gained based weakly supervised learning is as below,

$$\arg\min \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(f(\boldsymbol{x}_i), c_i), \quad (7)$$

where c_i refers to the coarse-grained annotation of $\boldsymbol{x}_i$.

Instead of interpreting and annotating data at each frame, the coarse-grained annotating procedure is much less tedious. For example, in human activity recognition tasks, it is much easier for an annotator to infer whether the targeted activity occurred in a long sequence, rather than strictly annotating each timestamp [157]. Wang *et al.* [157] exploited a recurrent attention framework to perform training on sequential weakly-labeled multi-activity and location tasks. Liu *et al.* [158] proposed a weakly supervised framework for arrhythmia detection, which permits training a fine-grained (beat-by-beat) arrhythmia detector with the use of large amounts of coarsely annotated ECG data (labels are given to each recording) to improve the generalization ability. Saab *et al.* [159] explored weak supervision for EEG seizure onset detection. Instead of relying on experienced clinicians to provide labels, the authors resorted to the annotations already being produced within existing clinical workflows, given by a mixture of technicians, fellows, students, and board-certified epileptologists during initial data screening [160]. These annotations, although less accurate, can significantly improve the detection of the seizure onset.

2) Multi-Task Learning:

$$\arg\min \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{1}^{n_t}\mathcal{L}(f_t(\boldsymbol{x}_i), y_i^t), \tag{8}$$

where t stands for a single task, T the total task number, and $\{\boldsymbol{x}_i, y_i\}_{i=1}^{n_t}$ is the data with annotations under the task t.

Multi-task learning stands for incorporating additional tasks into the overall network optimization to regulate the representations. It differs from self-supervised learning, in the fact that the additional tasks are normally trained simultaneously for the purposes of regularization, and in a fully supervised manner.

Multi-tasking imposes inter-task feature similarity constraints that can effectively improve the generalization performance of tasks that are semantically related [161], [162], [163], [164], [165]. In pervasive sensing for healthcare applications, the collected data can represent multiple different yet related tasks. For example, in affective analysis, human emotion comes with varying representation forms, including category expression, action units and valence/arousal. They are inherently related, and based on this, several multi-tasking frameworks have been exploited [166], [167]. Besides, in [168], the authors simultaneously predicted the systolic and diastolic blood pressure from PPG, achieving better performance than directly estimating each individual parameter. The heterogeneity between different tasks can also be transductive (e.g., the same target space), considering, for example, the inference on each subject as an individual task [169], [169], [163], [162].

E. Learn with Adaptive Strategies

Previous discussions have highlighted the heterogeneity issue in human-centric sensing as a common yet challenging problem. Below we present several common lines of research based on adaptive strategies that are designed to cope with shifts in the underlying distribution.

1) Domain Adaptation: Domain adaptation is one particular area of transfer learning, which aims to transfer the knowledge learned from source domains to the target domain, assuming access to a reasonable amount of labeled/unlabeled data in the target domain during training [13]. Considering differences in the number and type of features between the source and target domain, existing deep learning based domain adaptation approaches can be divided into either homogeneous or heterogeneous.

Research in homogeneous domain adaptation mostly focuses on the distribution shifts of $P(X)$ caused by domain shift $P(X|Y)$. Existing solutions are either discrepancy based, adversarial or reconstruction based approaches. The discrepancy-based approach aims to eliminate the representation shifts across different domains. The discrepancy measures the difference between two distributions, and they often are used as loss functions for network optimization. For example, Zhang *et al.* [62] utilized the non-parametric measure of maximum mean discrepancy (MMD) to align feature representations of myoelectric patterns across subjects. Jin *et al.* [170] incorporated multi-layer, multi-kernel MMD constraints to minimize the distribution discrepancy across different domains, in ECG based atrial fibrillation.

The adversarial based algorithms aim to minimize the domain shifts directly with adversarial training. For instance, Zhang *et al.* [54] built upon one popular adversarial based domain adaptation method, maximum classifier discrepancy [171], to realize cross-subject human locomotion intent prediction. Li *et al.* [172] looked into EEG-based emotion recognition and applied gradient reversal layer for adversarial domain adaptation on the representation extracted from the early layers.

The reconstruction-based approaches are built on a reconstruction based structure, utilizing either the reconstructed samples [173] or hidden representations [174], [175] for domain adaptation [174], [175], [173]. The former performs low-level domain adaptation by transforming feature representations from source domains to target domains in order to synthesize labeled training samples in the target domain. Xu *et al.* [173] applied CycleGAN based network to generate motor imagery EEG signals from normal subjects to stroke patients, thus enabling domain adaptation across these two groups. The latter hidden-representation-based adaptation applies stack autoencoders to align the representation. Specialized training strategies and architectures are normally required to enforce the similarity across domains. For example, recent work on EEG-based emotion recognition [174] applied a subject-shared encoder and subject-private encoders for all subjects to address inter-subject variability and thus enabled domain adaptation.

On the other hand, heterogeneous domain adaptation is required to address changes in the dimensionality of input features X. This applies to the scenarios where data are collected from different sensing modalities and spatial-temporal resolutions [176], [177]. For instance, Gao *et al.* [176] investigated EEG signals collected from multiple devices with different numbers of channels. The authors firstly projected the sensor recordings to a common manifold space and subsequently performed domain adaptation based on the feature extracted from the manifold embeddings. Nevertheless, compared to homogeneous domain adaptation, this area is much less explored.

Recent works on domain adaptation have demonstrated a series of further promising trends. Multi-source or multi-target domain adaptation has been designed to address the problem where the training data or testing data are collected from multiple distributions [178], [179]. For instance, Wei *et al.* [178] treated data from different subjects as unique sources and argued that multi-source domain adaptation could introduce more task-related knowledge than just considering multiple sources as a whole combined source. Another approach is the target distribution shift. Focusing on shifts not only in $P(Y|X)$ but also in P(Y), [180] explored domain adaptation methods for brain-computer interface tasks with different label sets. Label alignment was achieved by aligning the EEG covariance matrices of the source space to the target space. The authors tackled the covariate shift along with the differences in label sets by re-centering the source per class at the corresponding class of the target space.

2) Domain Generalization: In contrast to domain adaptation which conducts training with access to the target data, another line of research, domain generalization, aims to generalize to domains that are totally unseen during training [48]. Domain generalization has received increasing attention in computer vision. Existing solutions for domain generalization cover augmenting data diversity [181], aligning intra-class inter-domain representations [182], simulating domain gaps during training [183], [31], as well as performing disentangled representation learning [184].

In sensor informatics, the adoption of domain generalization solutions has attracted a significant amount of research efforts [185], [186], [187], [188], [189], [190], [191], [192], [193], [194], [195]. Adversarial/discrepancy based training to realize domain-invariant feature representation have been utilized in several works [192], [194], [187]. Another popular line of strategies is based on disentangled representation learning. Gu *et al.* [185] proposed a disentangled representation learning framework to learn subject-invariant and clean sensory representations from noisy sensing modalities, for generalizable pathological gait analysis. Jeon *et al.* [188] attempted to maximize the mutual information intra-class inter-subject EEG samples, to realize feature disentanglement. In sensor-based human activity recognition, Qin *et al.* [187] also divided the latent representation into both domain-specific and domain-invariant. Instead of considering the domain-specific representations as redundant, the authors leveraged such features to enhance the model discrimination capability for each domain.

Existing research also studied the possibility of utilizing multi-modal sensor informatics to pursue domain generalization, since different sensing modalities may have varied degrees of sensitivity to domain shifts. For instance, for healthcare-related action recognition/monitoring, both audio and image can be used. More importantly, audio signals carry complementary information that is less affected by different environmental conditions. Inspired by this, Planamente *et al.* [190] leveraged the representation alignment between audio and visual modalities to eliminate intrinsic adverse environmental conditions.

3) Few-Shot Learning: Considering the scarce resources of the healthcare data, few-shot adaptation is an emerging technique that enables model adaptation to new realms (e.g., subjects and scenarios), or new tasks, with samples of limited sizes via a pre-trained model. The concept of few-shot learning was originally proposed and mainly developed for inductive transfer, to generalize the model to novel unseen label space with only a few samples [9]. It was dominated by meta-learning based approaches [196]. Rahimian *et al.* [197] applied a meta-learning framework to enable few-shot EMG based hand gesture recognition. Tang *et al.* [198] employed prototype to select representative samples, realizing few-shot wearable time-series classification.

It is noteworthy that a relatively large amount of existing research integrates few-shot learning to solve domain adaptation tasks, realizing fast cross-domain transductive transfer learning. These strategies enable seamless development of personalized healthcare computational model [9] and they have been applied to a variety of scenarios, such as motor imagery classification [199], sleep staging [200], activity recognition [201], and others [202], [203], [204].

4) Concept Drift Adaptation: Online adaptation to temporal changes of the input data and target variables is a unique and challenging problem for streaming data [205]. Therefore, concept drift adaptation is important for real-time healthcare applications to address data drifts. Although concept drift adaptation is a relatively unexplored research area, particularly in healthcare, it is expected to grow rapidly [206].

It is difficult, in practice, to realize online adaptation of a single model during training, since only a limited number of samples under the same distribution can be collected for training, under data drifts [70]. Du *et al.* [207] developed a novel strategy to characterize and recognize the temporal distribution shifts, and clustered temporal data into several groups to perform distribution matching during training. To make dynamic adaptations to continually drifting data streams, ensemble learning is dominating conventional algorithms, which adapts the learning process of the sub-models gradually. For instance, researchers attempted to ensemble multiple classifiers to deal with the temporal covariate shifts of EEG time series, by adaptively combining the most informative predictions [208], [209].

5) Reinforcement Learning: Beyond conventional supervised learning that typically performs statistical fitting over training pairs, reinforcement learning (RL) forms as decision-making process learning of an agent through the interaction (action and reward) with the environment. RL offers a unique approach for realizing adaptiveness in pervasive healthcare by iteratively learning from the environment and improving decision-making over time. The dynamic nature of healthcare situations, ranging from patient health status, treatment interventions, and progression of disease, fits well within the RL framework and has been proposed in a large number of healthcare applications including chronic diseases [210], mental disorders [211], and critical clinical care [212]. A detailed review of the theoretical formulation of RL and its application in healthcare can be found in recent reviews/guidelines [213], [214].

Nevertheless, scarcity and compromised data quality result in incomplete and noisy state space of the agent and pose significant challenges in RL efficiency. RL also requires assigning data with a "reward" label that might be impractical to obtain in real-world applications. Therefore, the ability to learn optimal policies with large unlabeled data has also attracted considerable interest in the RL community [215]. Furthermore, pervasive sensing brings unique challenges in terms of signals with large time horizons without direct outcome measures, as well as the demands for online estimation of optimal treatment strategies [216].

F. Learn with Multiple Models

Existing literature has found that the performance of only one single model may be limited. Leveraging the coherent performance boosted from multiple models has been a popular direction to circumvent this issue.

1) Ensemble Learning: Ensemble learning has been a commonly used strategy that averages (sometimes weighted) the output of multiple models trained with different data proportions [217], data distributions [218], model architectures [219] and so on. In healthcare applications, missing data/input variables and even the whole sensor modalities are common, which are difficult to handle in developing appropriate machine learning models. Prince *et al.* [217] studied this problem in Parkinson's disease (PD) monitoring. They developed a multi-source ensemble learning method to handle the scenario where different PD patients might undergo different PD tests, namely tapping, walking, voice, and memory. Rad *et al.* [220] combined multiple models trained under stochastic optimization algorithms to achieve a better stereotypical motor movement detection for Autism Spectrum Disorders with wearable sensors. In [221], the authors proposed a novel parallel ensemble framework to capture complementary EEG connectivity features for neuropsychiatric patients.

2) Knowledge Distillation: Knowledge distillation aims to transfer knowledge from well-trained teacher networks to student networks, so that the student network is able to master the higher capacity or unique knowledge from the teachers.

In existing works, Wang *et al.* [222] exploited the frame-to-frame correlation property in EEG recordings and proposed a frame-level distillation neural network to remove redundant and meaningless information in single frames. Wu *et al.* [223] studied the gap between the patient-specific model and patient-independent model for seizure prediction, and indicates that the models trained on all patients' data can capture more informative features yet may not be as sensitive as patient-specific models, whereas each patient includes only a few samples for training. In this work, the authors leveraged knowledge distillation to transfer knowledge from the network trained on all patients to patient-specific models, thus bridging the gap between patient-specific and patient-independent predictors. Sepahvand and Abdali-Mohammad [224] focused on single-lead ECG arrhythmia detection, and proposed a teacher-student architecture to bridge the gap between multi-lead and single-lead ECG signals. Jeon *et al.* [225] compared time-series data augmentation strategies that are effective for knowledge distillation, and summarized a general set of recommended strategies.

On the other hand, knowledge distillation can be applied to the model learned from different sensing modalities, to leverage the complementary/correlated information presented in different modalities. Liu *et al.* [226] suggested that in action recognition, vision-based approaches are prone to occlusion and appearance variances compared to wearable sensor-based approaches. This observation leads to the attempt of using networks trained on wearable sensors to guide training based on visual signals. Other works also investigated the possibility of cross-modality knowledge distillation for emotion recognition, which transferred knowledge derived from facial images to other sensory modalities, e.g., EEG [227], voice [228].

G. Learn with Domain Knowledge

Leveraging domain knowledge has been emphasized in several aspects of the approaches introduced above, such as data augmentations, loss engineering, and self-/weakly-supervised loss design. On top of these, we are going to discuss another two directions that have been explored in pervasive health applications to fully exploit domain knowledge, i.e., domain-knowledge guided features and explainability/interpretability.

As "black boxes", deep learning models are mostly trained end-to-end with direct statistical fitting. However, considering the severely small number of annotations and significant noise contamination, there are many potential local optimums or shortcuts that could be learned as the decision boundary.

1) Domain-Knowledge Guided Features: To mitigate this, existing research has considered incorporating feature engineering into model training. They exploited those features that have been commonly used in conventional methods, summarized with domain knowledge, as input, instead of extracting features directly with end-to-end training [229], [230]. For example, Zhang *et al.* [230] employed both wavelet and common-spatial-pattern transformation to extract discriminative features for training. In this way, they guided the development of more EEG-based efficient seizure prediction. For better ECG-based cardiac arrhythmia detection, Hong *et al.* [229] integrated both deep-learning extracted features and hand-crafted features, including statistical, signal, morphological, and unsupervised features.

2) Explainability/Interpretability: To incorporate domain knowledge into model learning with the aim of lifting its generalization capability, another line of research pays attention to the explainability/interpretability. It can facilitate the understanding of the whole workflow from raw representations to the final prediction/decisions, by uncovering the inner properties of "black-box" models.

Research into explainable machine learning models is complex since the nature of appropriate explanations depends both on the type of the application as well as on the system's user [231]. Explanations can also be categorized as model agnostic/black box explanations that normally depend on a mechanism that perturbs the input features, and models specific explanations that are tailored to specific architectures. Other types of explanations and taxonomies also exist and include but not limited to surrogate models, counterfactual explanations, post-hoc and ad-hoc explanations.

In pervasive healthcare, one typical explainability/interpretability related research uncovers the feature that a trained model pays attention to, thus helping speculate whether the attended feature is meaningful and subject to domain knowledge, or else merely spurious correlated [232], [233], [234]. For instance, regarding human skeleton data, researchers exploited the spatio-temporal attention to discover the unique characteristics of action recognition, pathological gait [235], and individual walking patterns [233]. To discover the interpretability of motor-related EEG data, Borra *et al.* [234] utilized gradient-based techniques to derive salient

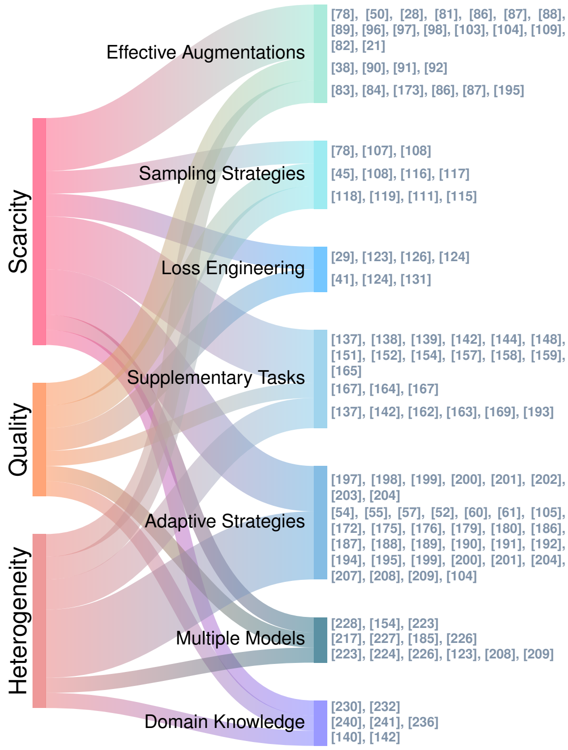

Fig. 4. Associations between the issues listed in Section II and the approaches in Section III. Representative works of pervasive healthcare are provided alongside each association links. In addition, a table summarizing the potential ways of how to apply these approaches to address each issue, is available in the Supplementary Material.

spectral-spatial features, and found that the learned feature is coherent with the motor-related EEG activity.

On the other hand, the explainability/interpretability can involve the computational architectural design, resembling clinical workflows that depend on interpretations of unique physiological/biological characteristics underlying the data modalities along with medical knowledge. For example, in ECG reading, cardiologists will base their diagnosis on specific characteristics of the ECG waveform, such as the P wave, the QRS complex and the RR intervals [236]. One solution is to apply domain knowledge as supervisory signals for training and incorporating (clinical) decision logical flows into the architecture design [236], [140], [237]. For instance, Hong *et al.* [236] designed a hierarchical attention network by performing attention on multiple levels (including beat, rhythm, and frequency), to improve the prediction of ECG arrhythmia. Zhou *et al.* [238] attempted to discover the anomalous morphological changes of heartbeats based on generative adversarial training. For the anticipation of human actions, Zhang *et al.* [237] referred to the "two-system" assumption on human cognition [239], and developed a two-branch model (intuition and analysis based) for action anticipation.

3) Human-in-the-Loop Learning: Involving human-level control presents another way of incorporating domain knowledge during model training. The ability of deep RL to make decisions from unstructured data with minimal feature engineering and achieve human-level control has made it a very promising approach for human-in-the-loop systems [242]. There is a significant increase of applications that use deep RL in modeling human joint kinematics and joint moments for use in assistive robots and prosthetics [243]. Several human-in-the-loop systems have been developed with RL for facilitating decision making in BCIs [244], [245]. Neurophysiological signals, such as EEG can facilitate learning by providing feedback via brain signals elicited when an error has been detected (error-related potentials) [245]. This information can be incorporated into the reward function and facilitate faster convergence without the need of "labeling" the signal based on explicit human feedback. RL is also expected to play an important role in enabling human-AI interfaces in sensor networks that involve ambient and wearable sensing for monitoring patients in natural settings [246].

To conclude, in Figure 4, we provide a Sankey diagram to illustrate the associations between different approaches and issues utilized in existing works on pervasive healthcare applications, whereas a more detailed table summarizing the potential mapping between approaches and issues is given in the Supplementary Material. It should be noted that these above approaches are not standalone, independently, to solve certain single issues. Instead, in most cases, several approaches are combined together to function in a collaborative manner.

IV. Emerging and Promising Directions

Thus far, a detailed overview has been provided, on existing technical advances that go beyond pure fully supervised learning for pervasive healthcare. These provides fundamental solutions/operations to addressing the real-world generalization issues in pervasive healthcare. In fact, the rapid evolution of deep/machine learning as well as the growing interests of deploying them to real-world applications, has fostered a series of emerging/promising directions (cfg. Figure 5) that are aimed for robust and trustworthy models that can handle pressing real-world challenges and meets practical needs.

A. From Single-Modality to Multi-Modality

In collected health settings, sensing informatics can be collected from varied modalities [1], [247], presenting both modality-unique representations and task-common information. For instance, in gait analysis, inertial sensors reflect movement kinematics, whilst pressure sensors capture the ground reaction forces, which reflect kinetics [248]. For brain-computer interfaces, EEG represents brain electrical activities, whereas NIRS reflects brain oxygenation level [249]. Dedicated research is required to guide the feature extraction by taking into consideration their interaction across modalities. In fact, taking advantage of modality-unique representations can enhance representation power, whilst maximizing task-common information avoids overfitting or short-cuts on a single modality.

As discussed in Section III, there have been several lines of solutions on handling data from multiple modalities. These include, but are not limited to cross-modality data generation (Section III-A), cross-modality self-supervised learning (Section III-D1, multi-moality domain adaptation/generalization (Section III-E), multi-modality ensemble/knowledge distillation (Section III-F1)). They go beyond simplistic feature concatenation or ensemble modeling, and are promising directions

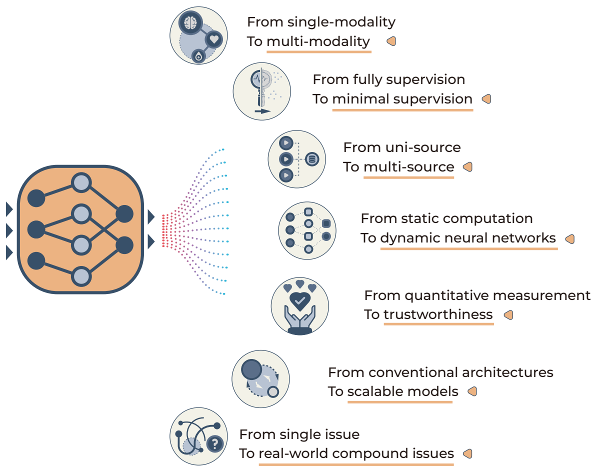

Fig. 5. Emerging and Promising Directions in Pervasive Healthcare.

that require increasing research attention. Among these, we would like to emphasize two directions that could underpin the development of multi-modality fusion methodologies.

1) Cross-Modality Self-Supervised Learning: Self-supervised vision-language pretraining, such as CLIP [136], has received rapidly increasing attention in computer/machine vision tasks. They have demonstrated superior capabilities in deriving semantically meaningful representations from both images and languages. It can implicitly mitigate the limitations in single-modality self-supervised learning, saving the need for designing effective data augmentations to transform data into another view. This is promising to be adopted in multi-modality sensing informatics, since acquisition of multi-modality data is trivial with off-the-shelf sensing technologies [1].

On the other hand, in some cases, different modalities of human-centric informatics reflect different stages throughout the loop of human perception, cognition, and action. Leveraging their inherent relationships [250], [143], can help facilitate self-supervised representation learning, and, in turn, obtain better models in general computer vision [143].

2) Multi-Modality for Domain Shifts: Data acquired from each individual modality reflect different aspects of the same healthcare task and thus inherently suffer from domain shifts. Nonetheless, the domain shifts underlying each modality are mostly unique, and not correlated. For instance, the wearable inertial sensors are sensitive to their body-worn positions, whilst wearable cameras suffer from the changes of camera viewpoints as caused by camera orientation variations. Such modality-unique bias can be eliminated by aligning representations across modalities; this would inspire works with an emphasis on domain generalization [251], [190], where domain-shift-prone sensing modality can be helped by those modalities that are less affected by domain shifts. In addition, it should be noted that the representation capability across modalities should be taken into consideration to avoid overlooking the less-representative modalities [190].

B. From Fully Supervision to Minimal Supervision

The trends in healthcare sensor informatics lead to greater amounts of raw sensor data with less clean annotations. In line with the scarcity and quality issue, this highlights the need of developing data efficient learning methodologies by minimizing the reliance on annotations. Beyond the key technical breakthroughs underpinning learning with minimal supervision (e.g., Section III-B3 active learning, Section III-D1 self-/semi-/weakly supervised learning, Section III-E3 few-shot learning), here we present two promising directions to facilitate retrieving meaningful representations from raw sensor informatics.

1) Incorporating Domain Knowledge: Domain knowledge is essential in self-supervised pretext design, which not only includes the physiological/biological mechanisms underlying signals, but also considers the fundamental challenges in associated tasks.

Designing self-supervised loss is data/modality-dependent. As discussed in Section III-D1, self-supervised contrastive learning is a promising direction. Its success highly depends on effective data augmentations resembling the way real-world sensor informatics could be perturbed, in time/spatial/frequency domains. This fails the strategy of naively transferring augmentation operations in general computer vision fields, and emphasizes the need of exploring effective augmentations for each specific modality [81], [252]. On the other hand, self-supervised loss should be relevant to the primary task. Task-unique domain knowledge can facilitate self-supervised loss design. For example, in human pose estimation [253] and facial expression recognition [254], the property of "chirality" has been investigated, where the mirroring of faces or bodies would impose equivariant transformations of the output or intermediate feature space.

2) Weak Annotation Retrieval: Beyond the conventional data annotation procedure, exploring weak annotations that are directly available from data collection workflows is a promising direction. For example, annotations provided by junior clinicians or trainees in original clinical workflows can carry related information [159], and they reduce the annotation workload [157]. Beyond these, leveraging human attention over the course of clinical experts screening the data based on eye tracking has been explored in medical imaging [255]. This approach can be potentially extended to general sensing informatics in clinical diagnostic settings to provide weak annotations.

C. From Uni-Source to Multi-Source

The characteristics of healthcare data differ across regions, hospitals, and individuals. This promotes data diversity during training, whilst it also increases the difficulty in handling the data heterogeneity to ensure model generalization/personalization.

1) Beyond Simple Domain Shifts: The domain shifts across multiple sources highlight the necessity of techniques such as domain generalization/adaptation (Section III-E1,III-E2). However, different from current multi-source benchmarks in natural image recognition, the heterogeneity associated with real-world healthcare data distributions is much more complicated. First of all, those natural-image benchmarks are constrained to a limited number of domains with distinct domain differences. However, in healthcare, the number of domains is much larger (e.g., patient number), and in most cases, existing solutions on explicitly eliminating data heterogeneity, cannot handle the data collected from tens or hundreds of sources, efficiently. Moreover, the heterogeneity factors in the real world are highly complex, and the discrepancy degrees across sources are varied [256]. These problems have been raised, with several promising directions being proposed and evaluated in computer vision, such as compound domain adaptation [256], source-free domain adaptation [257], test-time adaptation [258], open-set domain adaptation [259]. However, their efficacy in real-world pervasive healthcare applications is yet to be validated.

2) Learn in Federated Settings: Another emerging direction of training on multiple sources is federated learning [260]. It trains models on samples held at distributed edge devices or servers. One representative pipeline is to perform local training on each local edge, and then send the local model updates to the main server. Such decentralized training keeps data localized in remote edges, and increases data complexity by collecting data from massive number of sources. Nevertheless, it also poses novel challenges in optimizing the model, as well as performing privacy-preserved data analysis. On one hand, the heterogeneity across edges violates the independently and identically distributed assumption significantly. Existing approaches on handling data heterogeneity in non-federated settings could be extended to federated learning. However, it is challenging to develop scalable models when distributed among a large number of edge devices and meanwhile guarantee the training convergence [260]. On the other hand, although data are kept in the edges, the model updates can also result in leaking sensitive information. Hence, it remains a challenge to guarantee/enhance privacy in federated training [261].

D. From Static Computation to Dynamic Neural Networks

Considering the varied representation forms of the sensing informatics over the course of real-world deployment, such as feature permutation and dimensionality changes (which results in the issue of quality and heterogeneity), automatically making adaptions of the network structure or parameters is one promising direction [262]. Some ideas underlying the approaches (e.g., Section III-E4,III-F1) introduced above have already touched the concepts of "dynamics". We refer the readers to [262] for a detailed recipe of existing dynamic neural network techniques. In particular, the following three aspects are meaningful for the generalization issue in pervasive healthcare applications.

1) Scalable to Input Variations: Real-world sensing data is prone to the changes of sensor spatio-temporal configurations. Adapting the network architectures or parameters can enable a unified model that is able to handle real-world sensing data under different representation forms, such as feature permutation, dimensionality changes [53], concept drifts [208], etc. Recent research has already attempted to implement dynamic convolution kernels on corrupted images [263].

2) Representation Capability: On the other hand, dynamic neural networks are able to derive more meaningful representations by a series of strategies, such as attention mechanisms and mixture of experts [262]. These can be applied in pervasive healthcare applications with sophisticated designs, to boost the inference performance.

3) Inference Efficiency: There is a pressing need for deploying deep learning in computational-resource-limited settings, such as edge devices. Dynamic neural networks have proved to be able to promote efficient inference via varied strategies, such as early-exiting [264], partial model activation [265], and so on.

E. From Quantitative Measurement to Trustworthiness

The rapid and widespread adoption of deep/machine learning techniques on pervasive sensing applications has the potential to provide objective measures of well-being and disease progression, along with rehabilitation strategies. This approach would empower patients with chronic diseases/disorders as well as elderly people to live an independent life at home while managing their conditions. To support this goal there is a strong initiative to promote AI advances to provide valuable insight in the field of healthcare informatics. This requires building human-centered technologies that are trustworthy by enabling intuitive explanations, uncertainty estimation and privacy preservation in both centralized and distributed frameworks [266], [267].

1) Explainable Machine Learning Models: Instead of rigidly adopting existing deep learning architectures to conduct end-to-end training, more efforts are necessary to uncover the "blackbox". This can enhance the model capability by more effectively modeling the relationship between data and labels and minimizing spurious correlations. Meanwhile, the improved interpretability can help the experts trust the model inference performance, and select trustworthy and reasonable deployment models. Researchers envision human-in-the-loop systems (as discussed in Section III-G3) that not only learn from data but also incorporate direct users' feedback [268].

2) Privacy and Security: Privacy and security concerns also challenge the translation of recent advances in sensing technology and artificial intelligence in health care [269]. Several ethical concerns arise in relation to the data and privacy of the users. Deep learning is a powerful method that inherently memorizes data and therefore it can be exploited in several ways that can result in leakage of sensitive information at inference time. A recent review has highlighted adversarial attacks along with model and data poisoning as major vulnerabilities during model training that can compromise both privacy and safety of the patient [269]. To mitigate privacy concerns, deep learning capabilities can be also implemented based on hybrid in-situ&in-silico algorithms [270]. This recent breakthrough provides a unique capability to develop smart sensors for privacy-preserved and secure machine learning solutions that are able to be trained efficiently.

3) Fairness: Fairness in machine learning refers to identifying and correcting algorithmic biases that can result in gender and racial neglect and/or discrimination. Algorithmic biases can emerge unintentionally and thus are difficult to identify.

Among the most common reasons is the lack of representative samples from sensitive groups and biases already existing in historic data [271], [272]. Although the need to oversee machine learning pipelines for fairness has been understood, the research on how to guarantee this is still in its infancy [273]. For example, when the prevalence of a disease is inherently higher in one sensitive group over another, it is yet unclear how to guarantee fairness without compromising accuracy.

F. From Conventional Architectures to Scalable Models

Although traditional architectures, are typically developed, trained, and deployed using a modest number of graphic cards, currently the increasing volume and complexity of healthcare data underpin the need for evolving scalable deep learning models. Notably, both ends of the scale spectrum, from small models suitable for edge device deployment to large foundational models at the server level, are receiving increasing attention in the research landscape.

1) Small Models on Edge Devices: The advent of pervasive sensing technologies has led to the development of edge devices capable of generating a wealth of sensing informatics in real-time. Meanwhile, the continual progression of hardware miniaturization and optimization makes the incorporation of machine/deep learning models into these devices a promising prospect. This incorporation can enable real-time analysis and decision-making at the point of data collection, in addition to addressing privacy concerns by retaining sensitive health data under the control of the user.

However, it is often impossible to directly deploy a model trained in laboratory computers/servers to edge devices, or perform model training on edge devices themselves, given their limited data processing power and storage capability. As such, efficient learning strategies, from efficient model design to on-device training/processing, are paramount for developing models that can operate in these resource-constrained environments while maintaining high predictive accuracy [274].

2) Large Foundation Models: On the other hand, large foundation models typically trained on massive amounts of data at extensive server scales [275], promise to be an effective tool for discerning complex patterns and extracting valuable insights from sensing informatics. The recent success of these models hinges on the adept deployment of self-supervised learning paradigms on substantial volumes of unlabeled data. Emerging works on large language models [276] and vision models [277] have underscored their considerable potential in abstracting high-level knowledge from natural languages/images.

The pretrained large foundation model has been featured by its rapid adaptability to downstream tasks with minimal labeled samples [275], as well as the emerging training paradigms based on multiple modalities [278]. Considering the multi-modality and the high-dimensionality of the sensing data, along with the diverse nature of healthcare tasks, it presents substantial opportunities for developing and leveraging foundation models for various healthcare scenarios. Despite their potential, issues typically associated with model training and deployment persist, in addition to other challenges like interpretability and the ethical usage of data.

G. From Single Issue to Real-World Compound Issues

Existing research mostly focuses solely on one single issue at a time, which is rarely the case. In fact, in the open real world, situations concerning sensing applications are complicated and multiple issues would co-occur. The solution to addressing one single issue may be built upon the assumption that another co-occurring issue is not encountered [31]. One representative example is the coexistence of heterogeneity and scarcity. Although domain adaptation/generalization has been developed to solve the heterogeneity across multi-source data, the difficulty of collecting enough data in each domain cannot be overlooked as well [181]. For instance, for arrhythmia detection, both data scarcity issue (arrhythmia class imbalance) and heterogeneity issue (cross-subject morphological variations) exist. These compound issues in the real world significantly limit the model performance upon real-world deployment. In the future, novel algorithms and frameworks that are robust against compound realistic issues, are expected.

V. Conclusions

Machine/deep learning has demonstrated unprecedented performance in data analytics, and they have been successfully applied in handling sensor informatics for healthcare applications. However, the fundamental empirical risk minimization theory underpinning the purely fully supervised learning paradigm, is associated with severe issues in real-world healthcare sensing. The scarcity of labeled data pairs, the quality of both data and label, as well as the heterogeneity across different sources, result in poor generalization of the computational models statistically fitted in training data. This has raised much attention in the research community and several lines of learning approaches have been proposed to tackle the above issues, ranging from data augmentations, sampling strategies, loss engineering, supplementary tasks, adaptive strategies, mixture of multiple models, to incorporating domain knowledge.

We overview and critically appraise key trends in tackling the aforementioned limitations. Existing state-of-the-art approaches focus on one issue and they are not designed and appropriately evaluated in real-word scenarios. More emphasis should be given in human-centered approaches that promote trust. This requires taking into consideration several aspects that include the privacy of the user, the ability to provide intuitive explanations and reliably estimate the uncertainty of the models. Furthermore, it has become evident that minimally supervised models should be developed to handle continuous streams of heterogeneous, multi-modal data both in centralized and federated settings. Dynamic neural networks that adapt their structure or parameters during inference is a promising direction to enable scalable and efficient inference on computational-resource-limited devices while they enhance representation capabilities. In addition, considering the increasing volume and complexity of healthcare sensing informatics, developing deep learning towards both small-scale and large foundational models is necessary.

ACKNOWLEDGMENTS

The authors would like to acknowledge Engineering and Physical Sciences Research Council (EP/W01212X/1), the Royal Society Research Grant (RGS/R2/212199), Shanghai Municipal Science and Technology International R&D Collaboration Project (Grant No.20510710500), and Shanghai Sailing Program (22YF 1430800).

REFERENCES

[1] G.-Z. Yang, Ed., *Body Sensor Networks*. London: Springer London, 2014. [Online]. Available: http://link.springer.com/10.1007/978-1-4471-6374-9

[2] D. Ravì *et al.*, "Deep learning for health informatics," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 4–21, 2016.

[3] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, pp. 65–69, 2019.

[4] A. Haque, A. Milstein, and L. Fei-Fei, "Illuminating the dark spaces of healthcare with ambient intelligence," *Nature*, vol. 585, no. 7824, pp. 193–202, 2020.

[5] Y. Guo, X. Gu, and G.-Z. Yang, "Human–robot interaction for rehabilitation robotics," in *Digitalization in Healthcare*. Springer, 2021, pp. 269–295.

[6] K. T. Johnson and R. W. Picard, "Advancing neuroscience through wearable devices," *Neuron*, vol. 108, no. 1, pp. 8–12, 2020.

[7] Ł. Kidziński *et al.*, "Deep neural networks enable quantitative movement analysis using single-camera videos," *Nature communications*, vol. 11, no. 1, pp. 1–10, 2020.

[8] A. Zhang *et al.*, "Shifting machine learning for healthcare from development to deployment and from models to data," *Nature Biomedical Engineering*, pp. 1–16, 2022.

[9] Y. Wang *et al.*, "Generalizing from a few examples: A survey on few-shot learning," *ACM computing surveys (csur)*, vol. 53, no. 3, pp. 1–34, 2020.

[10] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 11, pp. 4037–4058, 2020.

[11] G. Pang *et al.*, "Deep learning for anomaly detection: A review," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021.

[12] H. Song *et al.*, "Learning from noisy labels with deep neural networks: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[13] W. M. Kouw and M. Loog, "A review of domain adaptation without target labels," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 3, pp. 766–785, 2019.

[14] X. Chen *et al.*, "Toward open-world electroencephalogram decoding via deep learning: a comprehensive survey," *IEEE Signal Processing Magazine*, vol. 39, no. 2, pp. 117–134, 2022.

[15] Y. Wu, "Robust learning-enabled intelligence for the internet of things: A survey from the perspectives of noisy data and adversarial examples," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9568–9579, 2020.

[16] S. Hong *et al.*, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Computers in Biology and Medicine*, vol. 122, p. 103801, 2020.

[17] W. Li *et al.*, "A comprehensive survey on machine learning-based big data analytics for iot-enabled smart healthcare system," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 234–252, 2021.

[18] D. Damen *et al.*, "Rescaling egocentric vision: Collection, pipeline and challenges for epic-kitchens-100," *International Journal of Computer Vision*, vol. 130, no. 1, pp. 33–55, 2022.

[19] P. E. McSharry *et al.*, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE transactions on biomedical engineering*, vol. 50, no. 3, pp. 289–294, 2003.

[20] B. M. Maweu *et al.*, "Generating healthcare time series data for improving diagnostic accuracy of deep neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2021.

[21] X. Gu *et al.*, "Cross-domain self-supervised complete geometric representation learning for real-scanned point cloud based pathological gait analysis," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[22] G. Averta *et al.*, "Optimal reconstruction of human motion from scarce multimodal data," *IEEE Transactions on Human-Machine Systems*, 2022.

[23] K. S. Park and S. H. Choi, "Smart technologies toward sleep monitoring at home," *Biomedical engineering letters*, vol. 9, no. 1, pp. 73–85, 2019.

[24] N. R. Adão Martins *et al.*, "Fatigue monitoring through wearables: A state-of-the-art review," *Frontiers in physiology*, p. 2285, 2021.

[25] D. Josephs *et al.*, "semg gesture recognition with a simple model of attention," in *Machine Learning for Health*. PMLR, 2020, pp. 126–138.

[26] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[27] G. D. Clifford *et al.*, "Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[28] J. Jiang *et al.*, "A novel multi-module neural network system for imbalanced heartbeats classification," *Expert Systems with Applications: X*, vol. 1, p. 100003, 2019.

[29] J. Gao *et al.*, "An effective lstm recurrent network to detect arrhythmia on imbalanced ecg dataset," *Journal of healthcare engineering*, vol. 2019, 2019.

[30] D. Kollias, "Abaw: Valence-arousal estimation, expression recognition, action unit detection & multi-task learning challenges," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2022, pp. 2328–2336.

[31] X. Gu *et al.*, "Tackling long-tailed category distribution under domain shifts," in *ECCV*, 2022.

[32] C.-Y. Wu *et al.*, "Long-term feature banks for detailed video understanding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 284–293.

[33] S. Chatterjee *et al.*, "Review of noise removal techniques in ecg signals," *IET Signal Processing*, vol. 14, no. 9, pp. 569–590, 2020.

[34] H. Zhang *et al.*, "Eegdenoisenet: A benchmark dataset for deep learning solutions of eeg denoising," *Journal of Neural Engineering*, vol. 18, no. 5, p. 056057, 2021.

[35] D. Farina *et al.*, "Surface emg crosstalk evaluated from experimental recordings and simulated signals," *Methods of information in medicine*, vol. 43, no. 01, pp. 30–35, 2004.

[36] M. Congedo *et al.*, "On the "dependence" of "independent" group eeg sources; an eeg study on two large databases," *Brain topography*, vol. 23, no. 2, pp. 134–138, 2010.

[37] M. Miao *et al.*, "Automated cca-mwf algorithm for unsupervised identification and removal of eog artifacts from eeg," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[38] T. Hossain *et al.*, "A method for sensor-based activity recognition in missing data scenario," *Sensors*, vol. 20, no. 14, p. 3811, 2020.

[39] A. K. Porbadnigk *et al.*, "Extracting latent brain states—towards true labels in cognitive neuroscience experiments," *NeuroImage*, vol. 120, pp. 225–253, 2015.

[40] M. Schultze-Kraft *et al.*, "Unsupervised classification of operator workload from brain signals," *Journal of neural engineering*, vol. 13, no. 3, p. 036008, 2016.

[41] J. Zeng, S. Shan, and X. Chen, "Facial expression recognition with inconsistently annotated datasets," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 222–237.

[42] F. Deligianni, Y. Guo, and G.-Z. Yang, "From emotions to mood disorders: A survey on gait analysis methodology," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2302–2316, 2019.

[43] D. Damen *et al.*, "The epic-kitchens dataset: Collection, challenges and baselines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4125–4141, 2020.

[44] S. Zhang *et al.*, "Facial emotion recognition with noisy multi-task annotations," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 21–31.

[45] J. Zhang *et al.*, "Automatic annotation correction for wearable eeg based epileptic seizure detection," *Journal of Neural Engineering*, vol. 19, no. 1, p. 016038, 2022.

[46] J. Quiñonero-Candela *et al.*, *Dataset shift in machine learning*. Mit Press, 2008.

[47] D. C. Castro, I. Walker, and B. Glocker, "Causality matters in medical imaging," *Nature Communications*, vol. 11, no. 1, pp. 1–10, 2020.

[48] J. Wang *et al.*, "Generalizing to unseen domains: A survey on domain generalization," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[49] X. Peng *et al.*, "Moment matching for multi-source domain adaptation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1406–1415.

[50] P. L. Rodrigues, M. Congedo, and C. Jutten, "Dimensionality transcending: a method for merging bci datasets with different dimensionalities," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 2, pp. 673–684, 2020.

[51] M. A. Reyna *et al.*, "Will two do? varying dimensions in electrocardiography: the physionet/computing in cardiology challenge 2021," in *2021 Computing in Cardiology (CinC)*, vol. 48. IEEE, 2021, pp. 1–4.

[52] L. Xu *et al.*, "Cross-dataset variability problem in eeg decoding with deep learning," *Frontiers in human neuroscience*, vol. 14, p. 103, 2020.

[53] X. Gu *et al.*, "Generalizable movement intention recognition with multiple heterogenous eeg datasets," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, May 2023.

[54] K. Zhang *et al.*, "Unsupervised cross-subject adaptation for predicting human locomotion intent," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 3, pp. 646–657, 2020.

[55] P. Zanini *et al.*, "Transfer learning: A riemannian geometry framework with applications to brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp. 1107–1116, 2017.

[56] X. Zhai, Z. Zhou, and C. Tin, "Semi-supervised learning for ecg classification without patient-specific labeled data," *Expert Systems with Applications*, vol. 158, p. 113411, 2020.

[57] U. Côté-Allard *et al.*, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 4, pp. 760–771, 2019.

[58] C. Chen *et al.*, "Motiontransformer: Transferring neural inertial tracking between domains," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 8009–8016.

[59] A. Haque *et al.*, "Towards viewpoint invariant 3d human pose estimation," in *European conference on computer vision*. Springer, 2016, pp. 160–177.

[60] Y. Liu *et al.*, "Hierarchically learned view-invariant representations for cross-view action recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 8, pp. 2416–2430, 2018.

[61] A. Ameri *et al.*, "A deep transfer learning approach to reducing the effect of electrode shift in emg pattern recognition-based control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 2, pp. 370–379, 2019.

[62] X. Zhang *et al.*, "Domain adaptation with self-guided adaptive sampling strategy: Feature alignment for cross-user myoelectric pattern recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1374–1383, 2022.

[63] Z. Sun *et al.*, "Human action recognition from various data modalities: A review," *IEEE transactions on pattern analysis and machine intelligence*, 2022.

[64] K. Theunissen *et al.*, "Self-paced and fixed speed treadmill walking yield similar energetics and biomechanics across different speeds," *Gait & Posture*, vol. 92, pp. 2–7, 2022.

[65] T. C. van der Hoek *et al.*, "Prevalence of depression in parkinson's disease: effects of disease stage, motor subtype and gender," *Journal of the neurological sciences*, vol. 310, no. 1-2, pp. 220–224, 2011.

[66] D. D. Patil and V. M. Wadhai, "Adaptive real time data mining methodology for wireless body area network based healthcare applications," *Advanced Computing*, vol. 3, no. 4, p. 59, 2012.

[67] J. Gama *et al.*, "A survey on concept drift adaptation," *ACM computing surveys (CSUR)*, vol. 46, no. 4, pp. 1–37, 2014.

[68] A. Kilicarslan, R. G. Grossman, and J. L. Contreras-Vidal, "A robust adaptive denoising framework for real-time artifact removal in scalp eeg measurements," *Journal of neural engineering*, vol. 13, no. 2, p. 026013, 2016.

[69] F. A. Garcia *et al.*, "Adaptive algorithm for gait segmentation using a single imu in the thigh pocket," *IEEE Sensors Journal*, 2022.

[70] H. Guo, S. Zhang, and W. Wang, "Selective ensemble-based online adaptive deep neural networks for streaming data with concept drift," *Neural Networks*, vol. 142, pp. 437–456, 2021.

[71] H. Zhang *et al.*, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*, 2018.

[72] O. Chapelle *et al.*, "Vicinal risk minimization," *Advances in neural information processing systems*, vol. 13, 2000.

[73] H. Han *et al.*, "Towards high generalization performance on electrocardiogram classification," in *2021 Computing in Cardiology (CinC)*, vol. 48. IEEE, 2021, pp. 1–4.

[74] D. Kostas and F. Rudzicz, "Thinker invariance: enabling deep neural networks for bci across more people," *Journal of Neural Engineering*, vol. 17, no. 5, p. 056008, 2020.

[75] Z. Huang *et al.*, "Tada! temporally-adaptive convolutions for video understanding," in *ICLR*, 2022.

[76] H.-P. Chou *et al.*, "Remix: rebalanced mixup," in *European Conference on Computer Vision*. Springer, 2020, pp. 95–110.

[77] M. Xu *et al.*, "Adversarial domain adaptation with domain mixup," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 6502–6509.

[78] J. Fan *et al.*, "Eeg data augmentation: towards class imbalance problem in sleep staging tasks," *Journal of Neural Engineering*, vol. 17, no. 5, p. 056017, 2020.

[79] X. Liu *et al.*, "Deep learning in ecg diagnosis: A review," *Knowledge-Based Systems*, vol. 227, p. 107187, 2021.

[80] R. Qian *et al.*, "Spatiotemporal contrastive video representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6964–6974.

[81] B. Gopal *et al.*, "3kg: Contrastive learning of 12-lead electrocardiograms using physiologically-inspired augmentations," in *Machine Learning for Health*. PMLR, 2021, pp. 156–167.

[82] P. K. Gyawali *et al.*, "Learning to disentangle inter-subject anatomical variations in electrocardiographic data," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 2, pp. 860–870, 2021.

[83] M. M. Krell and S. K. Kim, "Rotational data augmentation for electroencephalographic data," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 471–474.

[84] L. Wu *et al.*, "Improved high-density myoelectric pattern recognition control against electrode shift using data augmentation and dilated convolutional neural network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 12, pp. 2637–2646, 2020.

[85] G. Liu *et al.*, "Cross-view exocentric to egocentric video synthesis," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 974–982.

[86] T. Golany and K. Radinsky, "Pgans: Personalized generative adversarial networks for ecg synthesis to improve patient-specific deep ecg classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 557–564.

[87] F. Fahimi *et al.*, "Generative adversarial networks-based data augmentation for brain–computer interface," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 9, pp. 4039–4051, 2020.

[88] J. J. Bird *et al.*, "Synthetic biological signals machine-generated by gpt-2 improve the classification of eeg and emg through data augmentation," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3498–3504, 2021.

[89] V. F. Rey *et al.*, "Let there be imu data: Generating training data for wearable, motion sensor based activity recognition from monocular rgb videos," in *Adjunct proceedings of the 2019 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2019 ACM international symposium on wearable computers*, 2019, pp. 699–708.

[90] X. Wang *et al.*, "An ecg signal denoising method using conditional generative adversarial net," *IEEE Journal of Biomedical and Health Informatics*, 2022.

[91] P. Sawangjai *et al.*, "Eeganet: removal of ocular artifact from the eeg signal using generative adversarial networks," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[92] X. Gu *et al.*, "Coupled real-synthetic domain adaptation for real-world deep depth enhancement," *IEEE Transactions on Image Processing*, 2020.

[93] R. J. Chen *et al.*, "Synthetic data in machine learning for medicine and healthcare," *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 493–497, 2021.

[94] J. Duan *et al.*, "A survey of embodied ai: From simulators to research tasks," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2022.

[95] S. Srivastava *et al.*, "Behavior: Benchmark for everyday household activities in virtual, interactive, and ecological environments," in *Conference on Robot Learning*. PMLR, 2022, pp. 477–490.

[96] D. Tome *et al.*, "Selfpose: 3d egocentric pose estimation from a headset mounted camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[97] T. Golany, K. Radinsky, and D. Freedman, "Simgans: Simulator-based generative adversarial networks for ecg synthesis to improve deep ecg classification," in *International Conference on Machine Learning*. PMLR, 2020, pp. 3597–3606.

[98] T. Golany, D. Freedman, and K. Radinsky, "Ecg ode-gan: Learning ordinary differential equations of ecg dynamics via generative adversarial learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, 2021, pp. 134–141.

[99] M. Loper *et al.*, "Smpl: A skinned multi-person linear model," *ACM transactions on graphics (TOG)*, vol. 34, no. 6, pp. 1–16, 2015.

[100] A. A. A. Osman, T. Bolkart, and M. J. Black, "STAR: A sparse trained articulated human body regressor," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 598–613. [Online]. Available: https://star.is.tue.mpg.de

[101] G. Pavlakos *et al.*, "Expressive body capture: 3d hands, face, and body from a single image," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 10 975–10 985.

[102] M. Keller *et al.*, "Osso: Obtaining skeletal shape from outside," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 492–20 501.

[103] Y. Huang *et al.*, "Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–15, 2018.

[104] L. Pei *et al.*, "Mars: Mixed virtual and real wearable sensors for human activity recognition with multidomain deep learning model," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9383–9396, 2021.

[105] C. Chen *et al.*, "Unsupervised sim-to-real adaptation for environmental recognition in assistive walking," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022.

[106] N. Churamani, O. Kara, and H. Gunes, "Domain-incremental continual learning for mitigating bias in facial expression and action unit recognition," *IEEE Transactions on Affective Computing*, 2022.

[107] W. Lu, H. Hou, and J. Chu, "Feature fusion for imbalanced ecg data analysis," *Biomedical Signal Processing and Control*, vol. 41, pp. 152–160, 2018.

[108] T. Pokaprakarn *et al.*, "Sequence to sequence ecg cardiac rhythm classification using convolutional recurrent neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 572–580, 2021.

[109] D. Kiyasseh *et al.*, "Plethaugment: Gan-based ppg augmentation for medical diagnosis in low-resource settings," *IEEE journal of biomedical and health informatics*, vol. 24, no. 11, pp. 3226–3235, 2020.

[110] C. Lin *et al.*, "A bert based method for continuous estimation of cross-subject hand kinematics from surface electromyographic signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022.

[111] X. Li *et al.*, "Efficient sample and feature importance mining in semi-supervised eeg emotion recognition," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2022.

[112] E. Malach and S. Shalev-Shwartz, "Decoupling" when to update" from" how to update"," *Advances in neural information processing systems*, vol. 30, 2017.

[113] B. Han *et al.*, "Co-teaching: Robust training of deep neural networks with extremely noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.

[114] H. Wei *et al.*, "Combating noisy labels by agreement: A joint training method with co-regularization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 726–13 735.

[115] D.-K. Han *et al.*, "Confidence-aware subject-to-subject transfer learning for brain-computer interface," in *2022 10th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2022, pp. 1–4.

[116] K. Wang *et al.*, "Suppressing uncertainties for large-scale facial expression recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6897–6906.

[117] Y. Zhang *et al.*, "Learn from all: Erasing attention consistency for noisy label facial expression recognition," in *Proceedings of the European conference on computer vision (ECCV)*, 2022.

[118] G. Wang *et al.*, "A global and updatable ecg beat classification system based on recurrent neural networks and active learning," *Information Sciences*, vol. 501, pp. 523–542, 2019.

[119] D. Wu *et al.*, "Switching eeg headsets made easy: Reducing offline calibration effort using active weighted adaptation regularization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 11, pp. 1125–1137, 2016.

[120] H. Bi *et al.*, "Human activity recognition based on dynamic active learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 922–934, 2020.

[121] Y. Zhang *et al.*, "Deep long-tailed learning: A survey," *arXiv preprint arXiv:2110.04596*, 2021.

[122] X. Gu *et al.*, "Transaction: Icl-sjtu submission to epic-kitchens action anticipation challenge 2021," *arXiv preprint arXiv:2107.13259*, 2021.

[123] Z. Zhu *et al.*, "Identification of 27 abnormalities from multi-lead ecg signals: An ensembled se_resnet framework with sign loss function," *Physiological Measurement*, vol. 42, no. 6, p. 065008, 2021.

[124] C. G. Vázquez *et al.*, "Two will do: Cnn with asymmetric loss, self-learning label correction, and hand-crafted features for imbalanced multi-label ecg data classification," in *2021 Computing in Cardiology (CinC)*, vol. 48. IEEE, 2021, pp. 1–4.

[125] Y. Yang *et al.*, "Delving into deep imbalanced regression," in *International Conference on Machine Learning*. PMLR, 2021, pp. 11 842–11 851.

[126] J. Ren *et al.*, "Balanced mse for imbalanced visual regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 7926–7935.

[127] G. Patrini *et al.*, "Making deep neural networks robust to label noise: A loss correction approach," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1944–1952.

[128] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.

[129] Y. Wang *et al.*, "Symmetric cross entropy for robust learning with noisy labels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 322–330.

[130] X. Ma *et al.*, "Normalized loss functions for deep learning with noisy labels," in *International conference on machine learning*. PMLR, 2020, pp. 6543–6553.

[131] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, 2020.

[132] X. Liu *et al.*, "Self-supervised learning: Generative or contrastive," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[133] K. He *et al.*, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 000–16 009.

[134] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

[135] T. Chen *et al.*, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[136] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8748–8763.

[137] J. Y. Cheng *et al.*, "Subject-aware contrastive learning for biosignals," *arXiv preprint arXiv:2007.04871*, 2020.

[138] H. Banville *et al.*, "Uncovering the structure of clinical eeg signals with self-supervised learning," *Journal of Neural Engineering*, vol. 18, no. 4, p. 046020, 2021.

[139] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz, "Bendr: using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data," *Frontiers in Human Neuroscience*, vol. 15, 2021.

[140] N. Wagh *et al.*, "Domain-guided self-supervision of eeg data improves downstream classification performance and generalizability," in *Machine Learning for Health*. PMLR, 2021, pp. 130–142.

[141] J. Ye *et al.*, "Cosleep: A multi-view representation learning framework for self-supervised learning of sleep stage classification," *IEEE Signal Processing Letters*, vol. 29, pp. 189–193, 2021.

[142] D. Kiyasseh, T. Zhu, and D. A. Clifton, "Clocs: contrastive learning of cardiac signals across space, time, and patients," in *International Conference on Machine Learning*. PMLR, 2021, pp. 5606–5615.

[143] K. Ehsani *et al.*, "What can you learn from your muscles? learning visual representation from human interactions," in *International Conference on Learning Representations*, 2020.

[144] A. Shukla, S. Petridis, and M. Pantic, "Does visual self-supervision improve learning of speech representations for emotion recognition," *IEEE Transactions on Affective Computing*, 2021.

[145] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 542–542, 2009.

[146] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *Proceedings of the 20th International conference on Machine learning (ICML-03)*, 2003, pp. 912–919.

[147] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[148] Y. Wei *et al.*, "A semi-supervised progressive learning algorithm for brain computer interface," *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*.

[149] J. Li *et al.*, “Mixup asymmetric tri-training for heartbeat classification under domain shift,” *IEEE Signal Processing Letters*, vol. 28, pp. 718–722, 2021.

[150] B. Xiong *et al.*, “Multiview pseudo-labeling for semi-supervised learning from video,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7209–7219.

[151] Y. Xu *et al.*, “Cross-model pseudo-labeling for semi-supervised action recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2959–2968.

[152] H. Fan *et al.*, “Semi-supervised time series classification by temporal relation prediction,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3545–3549.

[153] K. Sohn *et al.*, “Fixmatch: Simplifying semi-supervised learning with consistency and confidence,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 596–608, 2020.

[154] S. Zhang *et al.*, “Combining cross-modal knowledge transfer and semi-supervised learning for speech emotion recognition,” *Knowledge-Based Systems*, vol. 229, p. 107340, 2021.

[155] M. Chen *et al.*, “Semi-supervised learning with multi-head co-training,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 6, 2022, pp. 6278–6286.

[156] Z.-H. Zhou, “A brief introduction to weakly supervised learning,” *National science review*, vol. 5, no. 1, pp. 44–53, 2018.

[157] K. Wang, J. He, and L. Zhang, “Sequential weakly labeled multiactivity localization and recognition on wearable sensors using recurrent attention networks,” *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 355–364, 2021.

[158] Y. Liu *et al.*, “Generalizable beat-by-beat arrhythmia detection by using weakly supervised deep learning,” *Frontiers in Physiology*, vol. 13, 2022.

[159] K. Saab *et al.*, “Weak supervision as an efficient approach for automated seizure detection in electroencephalography,” *NPJ digital medicine*, vol. 3, no. 1, pp. 1–12, 2020.

[160] N. S. Abend, A. A. Topjian, and S. Williams, “How much does it cost to identify a critically ill child experiencing electrographic seizures?” *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, vol. 32, no. 3, p. 257, 2015.

[161] Y. Zhang and Q. Yang, “A survey on multi-task learning,” *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[162] S. Taylor *et al.*, “Personalized multitask learning for predicting tomorrow’s mood, stress, and health,” *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 200–213, 2017.

[163] H. Kaji, H. Iizuka, and M. Sugiyama, “Ecg-based concentration recognition with multi-task regression,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 1, pp. 101–110, 2018.

[164] Y. Zhang *et al.*, “Improving eeg decoding via clustering-based multitask feature learning,” *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[165] Y. Song *et al.*, “Eeg-based motor imagery classification with deep multi-task learning,” in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[166] W. Zhang *et al.*, “Prior aided streaming network for multi-task affective analysis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3539–3549.

[167] S. Zhang *et al.*, “Facial emotion recognition with noisy multi-task annotations,” in *Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2021.

[168] Q. Hu, D. Wang, and C. Yang, “Ppg-based blood pressure estimation can benefit from scalable multi-scale fusion neural networks and multi-task learning,” *Biomedical Signal Processing and Control*, vol. 78, p. 103891, 2022.

[169] D. Lopez-Martinez and R. Picard, “Multi-task neural networks for personalized pain recognition from physiological signals,” in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2017, pp. 181–184.

[170] Y. Jin *et al.*, “A novel domain adaptive residual network for automatic atrial fibrillation detection,” *Knowledge-Based Systems*, vol. 203, p. 106122, 2020.

[171] K. Saito *et al.*, “Maximum classifier discrepancy for unsupervised domain adaptation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3723–3732.

[172] J. Li *et al.*, “Domain adaptation for eeg emotion recognition based on latent representation similarity,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 2, pp. 344–353, 2019.

[173] F. Xu *et al.*, “Classification of left-versus right-hand motor imagery in stroke patients using supplementary data generated by cyclegan,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2417–2424, 2021.

[174] L.-M. Zhao, X. Yan, and B.-L. Lu, “Plug-and-play domain adaptation for cross-subject eeg-based emotion recognition,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, pp. 863–870, May 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16169

[175] M. Han *et al.*, “Disentangled adversarial autoencoder for subject-invariant physiological feature extraction,” *IEEE signal processing letters*, vol. 27, pp. 1565–1569, 2020.

[176] D. Gao *et al.*, “Hhhfl: Hierarchical heterogeneous horizontal federated learning for electroencephalography,” *arXiv preprint arXiv:1909.05784*, 2019.

[177] C. Ju *et al.*, “Federated transfer learning for eeg signal classification,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 3040–3045.

[178] W. Wei *et al.*, “Reducing calibration efforts in rsvp tasks with multi-source adversarial domain adaptation,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 11, pp. 2344–2355, 2020.

[179] F. Mu *et al.*, “Unsupervised domain adaptation for position-independent imu based gait analysis,” in *2020 IEEE SENSORS*. IEEE, 2020, pp. 1–4.

[180] H. He and D. Wu, “Different set domain adaptation for brain-computer interfaces: A label alignment approach,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 5, pp. 1091–1108, 2020.

[181] F. Qiao, L. Zhao, and X. Peng, “Learning to learn single domain generalization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 556–12 565.

[182] Y. Li *et al.*, “Deep domain generalization via conditional invariant adversarial networks,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 624–639.

[183] Q. Dou *et al.*, “Domain generalization via model-agnostic learning of semantic features,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[184] X. Peng *et al.*, “Domain agnostic learning with disentangled representations,” in *International Conference on Machine Learning*. PMLR, 2019, pp. 5102–5112.

[185] X. Gu *et al.*, “Cross-subject and cross-modal transfer for generalized abnormal gait pattern recognition,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 546–560, 2020.

[186] H. Qian, S. J. Pan, and C. Miao, “Latent independent excitation for generalizable sensor-based cross-person activity recognition,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 13, 2021, pp. 11 921–11 929.

[187] X. Qin *et al.*, “Domain generalization for activity recognition via adaptive feature fusion,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2022.

[188] E. Jeon *et al.*, “Mutual information-driven subject-invariant and class-relevant deep representation learning in bci,” *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[189] T. Dissanayake *et al.*, “Domain generalization in biosignal classification,” *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 6, pp. 1978–1989, 2020.

[190] M. Planamente *et al.*, “Domain generalization through audio-visual relative norm alignment in first person action recognition,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1807–1818.

[191] B.-Q. Ma *et al.*, “Depersonalized cross-subject vigilance estimation with adversarial domain generalization,” in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[192] ——, “Reducing the subject variability of eeg signals with adversarial domain generalization,” in *International Conference on Neural Information Processing*. Springer, 2019, pp. 30–42.

[193] P. Autthasan *et al.*, “Min2net: End-to-end multi-task learning for subject-independent motor imagery eeg classification,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 6, pp. 2105–2118, 2021.

[194] Z. Jia *et al.*, “Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1977–1986, 2021.

[195] W. Lu *et al.*, “Semantic-discriminative mixup for generalizable sensor-based cross-domain activity recognition,” *Proceedings of the*

ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 2, pp. 1–19, 2022.

[196] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.

[197] E. Rahimian *et al.*, "Fs-hgr: Few-shot learning for hand gesture recognition via electromyography," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 29, pp. 1004–1015, 2021.

[198] W. Tang, L. Liu, and G. Long, "Interpretable time-series classification on few-shot samples," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[199] S. An *et al.*, "Few-shot relation learning with attention for eeg-based motor imagery classification," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 933–10 938.

[200] N. Banluesombatkul *et al.*, "Metasleeplearner: A pilot study on fast adaptation of bio-signals-based sleep stage classifier to new individual subject using meta-learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 1949–1963, 2020.

[201] T. Gong *et al.*, "Metasense: few-shot adaptation to untrained conditions in deep mobile sensing," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019, pp. 110–123.

[202] X. Liu *et al.*, "Metaphys: few-shot adaptation for non-contact physiological measurement," in *Proceedings of the conference on health, inference, and learning*, 2021, pp. 154–163.

[203] A. Wijekoon *et al.*, "A knowledge-light approach to personalised and open-ended human activity recognition," *Knowledge-based systems*, vol. 192, p. 105651, 2020.

[204] A. Akbari, J. Martinez, and R. Jafari, "A meta-learning approach for fast personalization of modality translation models in wearable physiological sensing," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[205] J. Lu *et al.*, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, 2018.

[206] H. Yu and G. I. Webb, "Adaptive online extreme learning machine by regulating forgetting factor by concept drift map," *Neurocomputing*, vol. 343, pp. 141–153, 2019.

[207] Y. Du *et al.*, "Adarnn: Adaptive learning and forecasting of time series," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 402–411.

[208] H. Raza *et al.*, "Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related eeg-based brain-computer interface," *Neurocomputing*, vol. 343, pp. 154–166, 2019.

[209] H. Raza and S. Samothrakis, "Bagging adversarial neural networks for domain adaptation in non-stationary eeg," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–7.

[210] J. Watts *et al.*, "Optimizing individualized treatment planning for parkinson's disease using deep reinforcement learning," in *2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*. IEEE, 2020, pp. 5406–5409.

[211] J. Pineau *et al.*, "Treating epilepsy via adaptive neurostimulation: a reinforcement learning approach," *International journal of neural systems*, vol. 19, no. 04, pp. 227–240, 2009.

[212] M. Komorowski *et al.*, "The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care," *Nature medicine*, vol. 24, no. 11, pp. 1716–1720, 2018.

[213] C. Yu *et al.*, "Reinforcement learning in healthcare: A survey," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–36, 2021.

[214] O. Gottesman *et al.*, "Guidelines for reinforcement learning in healthcare," *Nature medicine*, vol. 25, no. 1, pp. 16–18, 2019.

[215] T. Yu *et al.*, "How to leverage unlabeled data in offline reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2022, pp. 25 611–25 635.

[216] F. Zhu *et al.*, "Group-driven reinforcement learning for personalized mhealth intervention," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer, 2018, pp. 590–598.

[217] J. Prince, F. Andreotti, and M. De Vos, "Multi-source ensemble learning for the remote prediction of parkinson's disease in the presence of source-wise missing data," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1402–1411, 2018.

[218] Y. Zhang *et al.*, "Self-supervised aggregation of diverse experts for test-agnostic long-tailed recognition," in *Advances in Neural Information Processing Systems*, 2022.

[219] X. Fan *et al.*, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ecg recordings," *IEEE journal of biomedical and health informatics*, vol. 22, no. 6, pp. 1744–1753, 2018.

[220] N. M. Rad *et al.*, "Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders," *Signal Processing*, vol. 144, pp. 180–191, 2018.

[221] C.-R. Phang *et al.*, "A multi-domain connectome convolutional neural network for identifying schizophrenia from eeg connectivity patterns," *IEEE journal of biomedical and health informatics*, vol. 24, no. 5, pp. 1333–1343, 2019.

[222] Z. Wang *et al.*, "Fldnet: Frame-level distilling neural network for eeg emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2533–2544, 2021.

[223] D. Wu, J. Yang, and M. Sawan, "Bridging the gap between patient-specific and patient-independent seizure prediction via knowledge distillation," *Journal of Neural Engineering*, 2022.

[224] M. Sepahvand and F. Abdali-Mohammadi, "A novel method for reducing arrhythmia classification from 12-lead ecg signals to single-lead ecg with minimal loss of accuracy through teacher-student knowledge distillation," *Information Sciences*, vol. 593, pp. 64–77, 2022.

[225] E. S. Jeon *et al.*, "Role of data augmentation strategies in knowledge distillation for wearable sensor data," *IEEE Internet of Things Journal*, 2021.

[226] Y. Liu *et al.*, "Semantics-aware adaptive knowledge distillation for sensor-to-vision action recognition," *IEEE Transactions on Image Processing*, vol. 30, pp. 5573–5588, 2021.

[227] S. Zhang, C. Tang, and C. Guan, "Visual-to-eeg cross-modal knowledge distillation for continuous emotion recognition," *Pattern Recognition*, p. 108833, 2022.

[228] S. Albanie *et al.*, "Emotion recognition in speech using cross-modal transfer in the wild," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 292–301.

[229] S. Hong *et al.*, "Combining deep neural networks and engineered features for cardiac arrhythmia detection from ecg recordings," *Physiological measurement*, vol. 40, no. 5, p. 054009, 2019.

[230] Y. Zhang *et al.*, "Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 465–474, 2019.

[231] A. B. Arrieta *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82–115, 2020.

[232] J. Yoo, T. J. Jun, and Y.-H. Kim, "xecgnet: Fine-tuning attention map within convolutional neural network to improve detection and explainability of concurrent cardiac arrhythmias," *Computer Methods and Programs in Biomedicine*, vol. 208, p. 106281, 2021.

[233] F. Horst *et al.*, "Explaining the unique nature of individual gait patterns with deep learning," *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.

[234] D. Borra, S. Fantozzi, and E. Magosso, "Interpretable and lightweight convolutional neural network for eeg decoding: Application to movement execution and imagination," *Neural Networks*, vol. 129, pp. 55–74, 2020.

[235] C. Dindorf *et al.*, "Interpretability of input representations for gait classification in patients after total hip arthroplasty," *Sensors*, vol. 20, no. 16, p. 4385, 2020.

[236] S. Hong *et al.*, "Mina: Multilevel knowledge-guided attention for modeling electrocardiography signals," in *IJCAI*, 2019.

[237] T. Zhang *et al.*, "An egocentric action anticipation framework via fusing intuition and analysis," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 402–410.

[238] B. Zhou *et al.*, "Beatgan: Anomalous rhythm detection using adversarially generated time series." in *IJCAI*, 2019, pp. 4433–4439.

[239] D. Kahneman, *Thinking, fast and slow*. Macmillan, 2011.

[240] X. Liu *et al.*, "Multiscale space-time-frequency feature-guided multitask learning cnn for motor imagery eeg classification," *Journal of Neural Engineering*, vol. 18, no. 2, p. 026003, 2021.

[241] A. Vijayvargiya *et al.*, "Hardware implementation for lower limb surface emg measurement and analysis using explainable ai for activity recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022.

[242] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[243] W. Wu, K. R. Saul, and H. Huang, "Using reinforcement learning to estimate human joint moments from electromyography or joint kinematics: An alternative solution to musculoskeletal-based biomechanics," *Journal of Biomechanical Engineering*, vol. 143, no. 4, p. 044502, 2021.

[244] Y. Ming *et al.*, "Eeg-based drowsiness estimation for driving safety using deep q-learning," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 4, pp. 583–594, 2020.

[245] D. Xu *et al.*, "Accelerating reinforcement learning using eeg-based implicit human feedback," *Neurocomputing*, vol. 460, pp. 139–153, 2021.

[246] M. Taherisadr, S. A. Stavroulakis, and S. Elmalaki, "adaparl: Adaptive privacy-aware reinforcement learning for sequential decision making human-in-the-loop systems," in *Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation*, 2023, pp. 262–274.

[247] J. N. Acosta *et al.*, "Multimodal biomedical ai," *Nature Medicine*, vol. 28, no. 9, pp. 1773–1784, 2022.

[248] F. Deligianni *et al.*, "A fusion framework to estimate plantar ground force distributions and ankle dynamics," *Information Fusion*, vol. 41, pp. 255–263, 2018.

[249] J. Shin *et al.*, "Simultaneous acquisition of eeg and nirs during cognitive tasks for an open access dataset," *Scientific data*, vol. 5, no. 1, pp. 1–16, 2018.

[250] E. Orhan, V. Gupta, and B. M. Lake, "Self-supervised learning through the eyes of a child," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9960–9971, 2020.

[251] Y. Zhang *et al.*, "Audio-adaptive activity recognition across video domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 13 791–13 800.

[252] Y. Shu *et al.*, "Revisiting self-supervised contrastive learning for facial expression recognition," in *BMVC*, 2022.

[253] R. Yeh, Y.-T. Hu, and A. Schwing, "Chirality nets for human pose regression," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[254] L. Lo *et al.*, "Facial chirality: From visual self-reflection to robust facial feature learning," *IEEE Transactions on Multimedia*, 2022.

[255] N. Khosravan *et al.*, "A collaborative computer aided diagnosis (c-cad) system with eye-tracking, sparse attentional model, and deep learning," *Medical image analysis*, vol. 51, pp. 101–115, 2019.

[256] Z. Liu *et al.*, "Open compound domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 406–12 415.

[257] J. N. Kundu *et al.*, "Universal source-free domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4544–4553.

[258] D. Wang *et al.*, "Tent: Fully test-time adaptation by entropy minimization," in *International Conference on Learning Representations*, 2020.

[259] P. Panareda Busto and J. Gall, "Open set domain adaptation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 754–763.

[260] T. Li *et al.*, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[261] G. A. Kaissis *et al.*, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.

[262] Y. Han *et al.*, "Dynamic neural networks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[263] G. Liu *et al.*, "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 85–100.

[264] L. Qendro *et al.*, "Early exit ensembles for uncertainty quantification," in *Machine Learning for Health*. PMLR, 2021, pp. 181–195.

[265] N. Rashid, B. U. Demirel, and M. A. Al Faruque, "Ahar: Adaptive cnn for energy-efficient human activity recognition in low-power edge devices," *IEEE Internet of Things Journal*, 2022.

[266] A. A. de Hond *et al.*, "Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review," *npj Digital Medicine*, vol. 5, no. 1, pp. 1–13, 2022.

[267] D. Kaur *et al.*, "Trustworthy artificial intelligence: a review," *ACM Computing Surveys (CSUR)*, vol. 55, no. 2, pp. 1–38, 2022.

[268] S. Bruckert, B. Finzel, and U. Schmid, "The next generation of medical decision support: A roadmap toward transparent expert companions," *Frontiers in artificial intelligence*, vol. 3, p. 507973, 2020.

[269] A. Qayyum *et al.*, "Secure and robust machine learning for healthcare: A survey," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 156–180, 2020.

[270] P. Korus and N. Memon, "Computational sensor fingerprints," *IEEE Transactions on Information Forensics and Security*, 2022.

[271] N. Mehrabi *et al.*, "A survey on bias and fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021.

[272] J. Kleinberg, S. Mullainathan, and M. Raghavan, "Inherent trade-offs in the fair determination of risk scores," *arXiv preprint arXiv:1609.05807*, 2016.

[273] Z. Obermeyer *et al.*, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019.

[274] H. Cai *et al.*, "Enable deep learning on mobile devices: Methods, systems, and applications," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 27, no. 3, pp. 1–50, 2022.

[275] R. Bommasani *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[276] T. B. Brown *et al.*, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.

[277] A. Kirillov *et al.*, "Segment anything," *arXiv:2304.02643*, 2023.

[278] R. Girdhar *et al.*, "Omnivore: A single model for many visual modalities," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 102–16 112.