Review

Survey of Transfer Learning Approaches in the Machine Learning of Digital Health Sensing Data

Lina Chato *[ID] **and Emma Regentova**

Department of Electrical and Computer Engineering, University of Nevada, Las Vegas, NV 89154, USA;
emma.regentova@unlv.edu

* Correspondence: lina.chato@unlv.edu

Abstract: Machine learning and digital health sensing data have led to numerous research achievements aimed at improving digital health technology. However, using machine learning in digital health poses challenges related to data availability, such as incomplete, unstructured, and fragmented data, as well as issues related to data privacy, security, and data format standardization. Furthermore, there is a risk of bias and discrimination in machine learning models. Thus, developing an accurate prediction model from scratch can be an expensive and complicated task that often requires extensive experiments and complex computations. Transfer learning methods have emerged as a feasible solution to address these issues by transferring knowledge from a previously trained task to develop high-performance prediction models for a new task. This survey paper provides a comprehensive study of the effectiveness of transfer learning for digital health applications to enhance the accuracy and efficiency of diagnoses and prognoses, as well as to improve healthcare services. The first part of this survey paper presents and discusses the most common digital health sensing technologies as valuable data resources for machine learning applications, including transfer learning. The second part discusses the meaning of transfer learning, clarifying the categories and types of knowledge transfer. It also explains transfer learning methods and strategies, and their role in addressing the challenges in developing accurate machine learning models, specifically on digital health sensing data. These methods include feature extraction, fine-tuning, domain adaptation, multitask learning, federated learning, and few-/single-/zero-shot learning. This survey paper highlights the key features of each transfer learning method and strategy, and discusses the limitations and challenges of using transfer learning for digital health applications. Overall, this paper is a comprehensive survey of transfer learning methods on digital health sensing data which aims to inspire researchers to gain knowledge of transfer learning approaches and their applications in digital health, enhance the current transfer learning approaches in digital health, develop new transfer learning strategies to overcome the current limitations, and apply them to a variety of digital health technologies.

Keywords: digital health; domain adaptation; feature extraction; federated learning; fine-tune; inductive transfer learning; portable devices; transductive transfer learning; transfer learning; wearable devices

1. Introduction

Digital health (DH) refers to the use of information and communication technologies in healthcare and medicine to enhance healthcare services and outcomes [1,2]. DH technologies encompass both hardware and software services and applications, including telemedicine, wearable devices, and augmented/virtual reality [3]. Hardware components include: *(a) sensors* for measuring vital signs and detecting physiological events, *(b) communication and network tools* for transmitting and exchanging information between patients/users and healthcare providers, and *(c) mobile devices* (such as computers and smartphones) for storing, displaying, and processing collected and transmitted information [3].

In healthcare, a multitude of technologies have been developed to enhance diagnosis and prognosis outcomes, as well as to support decision-making and treatment plan selections. The primary objectives of DH are centered on improving the accuracy of diagnosis and predictions, expediting the diagnosis process, and reducing diagnosis and treatment costs [1–3]. Additionally, DH technologies aim to empower users, including patients, to track their health status and wellness, while simultaneously enhancing the overall healthcare experience for both providers and patients [2]. Furthermore, DH provides continuous, automatic, and mobile health monitoring, which have the potential to significantly improve patients' quality of life [3].

Embedded and integrated sensors found in portable and wearable devices are central to the DH system, particularly in the new telemedicine paradigm developed to enhance the quality of healthcare services [4,5]. These devices benefit from the integration of Artificial Intelligence technology (AI). The latter aims for providing sophisticated end-to-end solutions that are technologically intensive and cost-efficient [6]. By combining these technologies, healthcare providers can receive comprehensive and accurate information of patient health, and provide personalized treatment plans and diagnoses, thus leading to improving patients' outcomes and the more efficient use of healthcare resources [7].

The portable mobile monitoring technology can be classified into three main categories based on the number of embedded sensers: *homogenous-based technology*, *dual-based technology*, and *heterogenous-based technology* [8]. Homogenous-based technology consists of a single type of sensor, such as electrocardiography (ECG), electroencephalography (EEG), electromyography (EMG), the global positioning system (GPS), photoplethysmography (PPG), or an accelerometer. Dual-based technology employs two different types of sensors for various forms of health monitoring to increase the reliability and accuracy, such as the accelerometer and gyroscope, the accelerometer and PPG, ECG and PPG, blood pressure and temperature, and ECG and EEG. Heterogenous-based technology embeds multiple sensors in a single device to provide multifunction monitoring services, as well as to improve the quality and precision of disease diagnosis. Anikwe et al. presented and discussed various DH applications based on the above three technologies [8]. Most heterogenous-based technology applications utilize Internet of Things (IoT) technology to provide multidimensional features and real-time services in mobile health (mHealth) [3,9]. Involving IoT for medical applications and services is generally called the Internet of Medical Things (IoMT) and refers to a network of Internet-connected medical devices, sensors, and software apps that provide online, flexible analysis and monitoring services. There are various IoMT applications based on their purpose and location of use, such as in-home IoMT, on-body IoMT, community IoMT, and in-hospital IoMT. For example, a smart health monitoring system was developed using Internet of Things (IoT) technology as a contactless tracing and treatment method for patients with COVID-19 to monitor blood pressure, heart rate, oxygen level, and temperature [10].

Wearable devices can be classified, according to the worn/mounted location, into the following groups: (a) wrist-mounted devices, (b) head-mounted devices, (c) E-textiles, and (d) smart jewelry and accessories [11]. Figure 1 presents examples of wearable and attachable technologies in digital healthcare. Wrist-mounted devices, such as fitness bands, smart watches, and stretchable patches, are noninvasive monitoring devices that are developed for physiological monitoring [11]. For example, wrist bands, watches, and arm-mounted stretchable patches are used for monitoring cardiovascular signals (such as heart rate and blood pressure) and sweat biomarkers (such as glucose, sodium, uric acid, lactate, etc.) [8,11]. The most popular wearable devices for health monitoring and tracking are the digital electrocardiogram (ECG) devices that are featured in smart watches and other fitness trackers, as well as wearable patches and chest straps. Head-mounted devices, such as mouthguards, are used for salivary-content monitoring (lactate, uric acid, and glucose) [12–15], while eyeglasses are used for sweat-content monitoring (lactate and potassium), as well as for cardiovascular signal monitoring (heart rate) [16,17]. Smart glasses based on pulse-sensing were used to continuously monitor the heart rate by a photo-

plethysmography (PPG) sensor placed on the nose pad [18]. E-textiles include textiles with electrodes which are used for sweat-content monitoring (glucose and lactate) [19], those worn on the calf for cardiovascular signal monitoring (heart rate and temperature) [11], and footwear for physical activity monitoring (foot motion) [20]. Smart jewelry and accessories, such as rings, necklaces, and clips, are used for physical activity monitoring (sleep, daily activity) [11], and belts worn on waist and chest are used for physical activity monitoring (step count and sitting time), as well as for physiological signal monitoring (ECG and direct current) [11,21]. Most ECG devices in smartphones and fitness trackers are based on a single-lead ECG and are connected with apps to store the ECG tracing. The sensors that adhere to the skin, such as patches, are mostly wireless and water-resistant; they can monitor and collect large amounts of continuous data for cardiologists for up to 30 days. Smart continuous glucose monitoring (CGM) systems are common portable devices that allow patients to measure the glucose levels in their blood [22]. This smart device is small in size and can be connected to a smartphone to monitor blood sugar levels over time, and have the ability to share the information with healthcare providers [22]. There are three CGM modalities based on the method of placing the sensors: non-invasive (based on optical sensors, such as contact lenses that do not require skin puncture), minimally invasive (based on a microneedle-type sensor), and invasive (based on an implantable sensor that is inserted under the skin) [22,23].
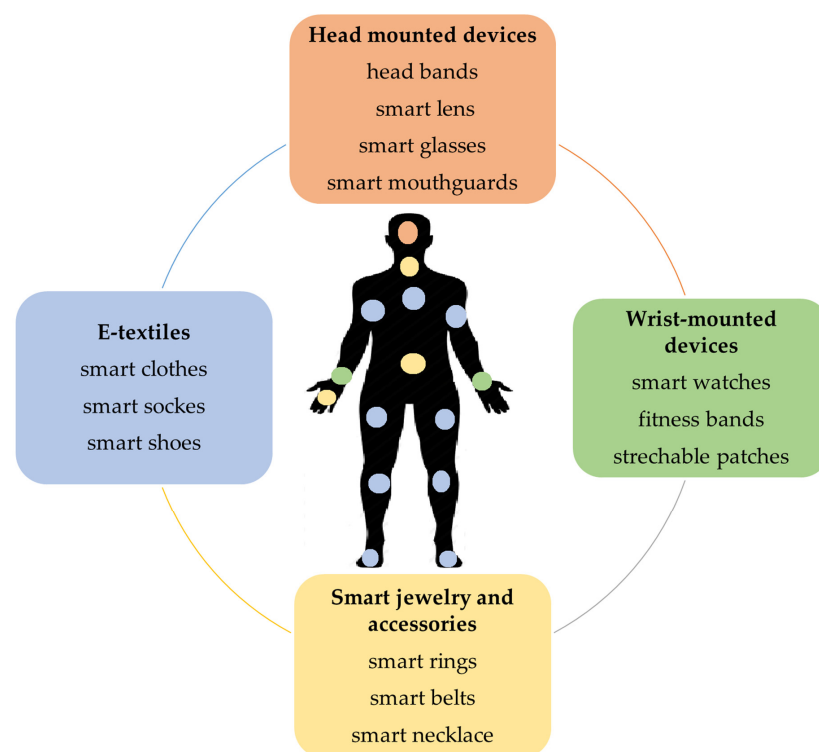

Figure 1. Wearable and attachable technologies for health monitoring based on the worn/mounted location.

Machine Learning (ML), including Deep Learning (DL), methods have been widely employed across diverse domains, particularly in healthcare, to enhance the well-being and safety of individuals. The influence of these methods within DH has been substantial, revolutionizing the analysis and utilization of patient data. As a result, there has been a notable improvement in the precision and efficiency of diagnoses, anomaly detection, and the prediction of potential health concerns [24–26]. ML methods include both traditional ML algorithms (such as the decision tree (DT), support vector machine (SVM), K-nearest neighbor (KNN), and artificial neural network (ANN)) and DL models, which are just ANNs with a hierarchical deep structure of multiple layers. These algorithms are trained

using sufficient data to develop reliable automatic prediction models. By training on vast amounts of electronic health records, medical images, sensor data, and genomic data, ML/DL can develop high-performance predictive models for accurate diagnoses and prognoses, and personalized patient care, thereby lowering healthcare costs [26–28]. Powered by ML/DL, DH will further revolutionize healthcare services [27,29].

The application of robust and effective ML/DL algorithms demands substantial amounts of high-quality data collected and processed by experts [24,30], and the sufficiency of the data depends on the adequate size of the data in terms of the number of samples (i.e., the number of control and patient subjects), meaningful descriptors in each sample, and/or accurate annotations. Processing the data is commonly performed at the preprocessing stage to provide high-quality data that contain meaningful attributes, and it is therefore essential to develop reliable ML/DL models. The most common preprocessing data methods are: data transmission, data storing, data cleaning/denoising, data transformation, as well as data fusion. Data-transformation methods include sampling/resampling, rescaling, augmentation, feature selection, and feature extraction. These methods are important to develop robust attributes from the raw data, improve the prediction accuracy, as well as to speed up the learning performance. Information fusion is also widely used to develop accurate ML models, which includes: (1) data fusion or multimodal data from various sensors and resources, for example, using medical imaging data as well as wearable-based sensor data to; (2) feature fusion, which includes combining various types of features that can be extracted from the data, such as texture, shape, histogram, and DL features; (3) decision fusion, performed as a postprocessing step to increase the performance and reduce the prediction error rate. The effectiveness of the complex computations in ML/DL methods depends on the number of samples, the sample size, the type of data, and the size and type of hardware (i.e., physical and cloud memory to store data and perform complex computations) [24,30,31]. Although the accuracy of DL methods outperforms traditional ML methods, most DL methods require big data and a huge amount of physical or cloud memory to deal with the complex and deep architectures that are required of the expensive computations, as shown in Figure 2. Using traditional ML methods is the ideal choice when the memory size is small and/or the data size is small in terms of the number of samples. However, traditional ML methods require robust features to overcome the small data size, a common situation in healthcare. Extracting meaningful features is a time-consuming and complex process that might require conducting extensive experiments to produce the best model configuration. The state-of-the-art ML models are frequently used across various domains; however, they do not necessarily produce high-quality outcomes due to the differences in the tasks and/or the domains that make the same model less efficient when adopted [32–34].

Transfer learning (TL) is a ML approach that has been developed for leveraging previously acquired knowledge in one domain to enhance the performance in a different but related domain [32–34]. TL relies on generalization theory, which posits that individuals can apply their knowledge and skills to new contexts by recognizing the underlying relatedness [34]. In general, TL methods are employed as an ML optimization tool to improve the performance, provide better generalization, reduce overfitting, and mitigate bias [32]. Suppose that a DL model was developed and trained on a large dataset of colored images containing thousands of instances from four groups: car, bike, ship, and airplane. The purpose was to automatically classify images into one of these four groups. Now, consider a research group with 200 MRI images containing tumors (benign or malignant). They aim to develop an automatic classification DL model, but the limited number of medical images (200 samples) is insufficient for creating an accurate model. Instead, the researchers can utilize transfer knowledge from a well-trained model, such as the one developed for colored images, to create a new model for a healthcare task; for instance, they can classify MRI images into benign or malignant tumors. This can be done by using the pretrained model, updating only the output layer to classify the images into two classes,

and updating the pretrained model parameters by training it on the MRI data, as shown in Figure 3.
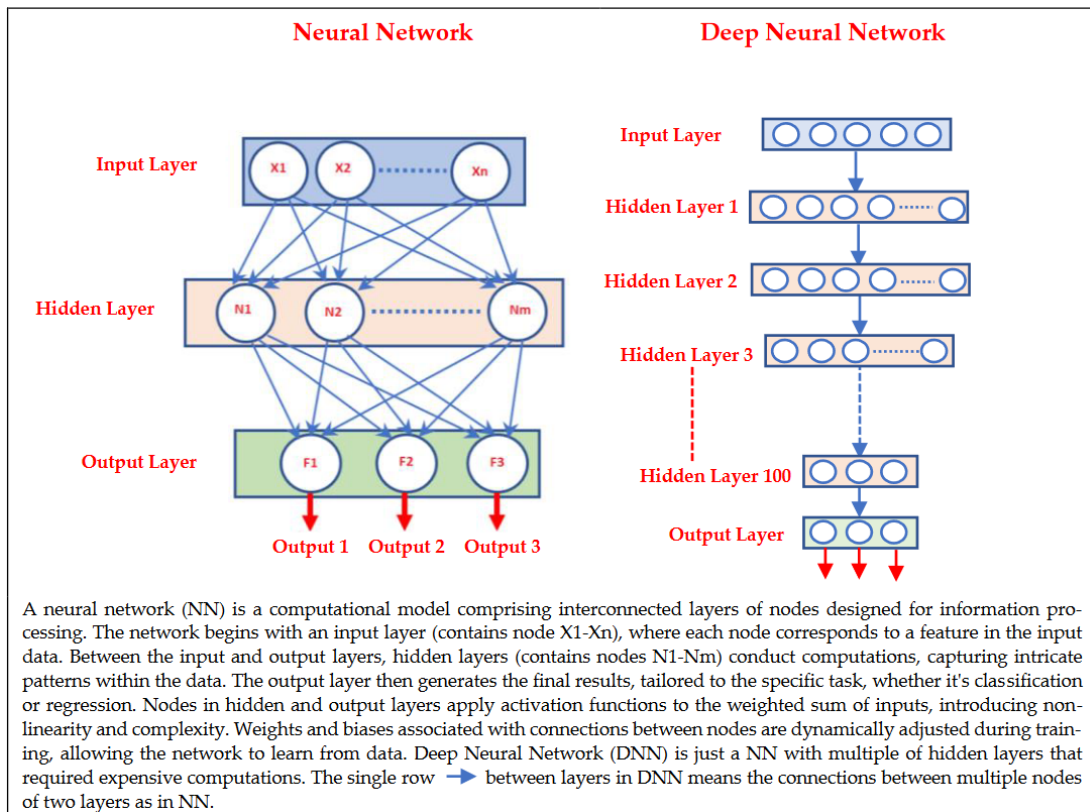

A neural network (NN) is a computational model comprising interconnected layers of nodes designed for information processing. The network begins with an input layer (contains node X1-Xn), where each node corresponds to a feature in the input data. Between the input and output layers, hidden layers (contains nodes N1-Nm) conduct computations, capturing intricate patterns within the data. The output layer then generates the final results, tailored to the specific task, whether it's classification or regression. Nodes in hidden and output layers apply activation functions to the weighted sum of inputs, introducing nonlinearity and complexity. Weights and biases associated with connections between nodes are dynamically adjusted during training, allowing the network to learn from data. Deep Neural Network (DNN) is just a NN with multiple of hidden layers that required expensive computations. The single row ➔ between layers in DNN means the connections between multiple nodes of two layers as in NN.

Figure 2. Examples of neural network and deep neural network architectures.
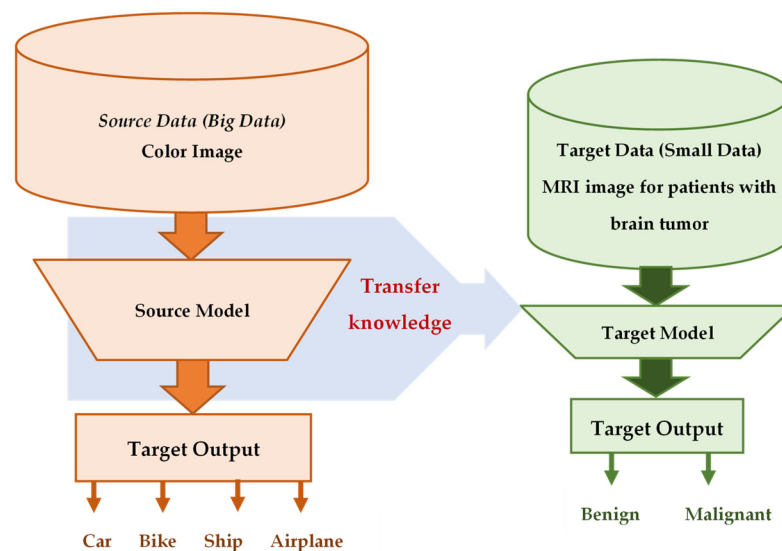

Figure 3. Example of transferred knowledge from a color nonmedical image prediction model (source) that was developed based on big data to the medical image prediction model (target) based on small data.

There are several criteria have been used to categorize TL methods and strategies. In [33], the TL methods were grouped into three main categories with respect to the data annotation status, as shown in Figure 4: (a) inductive transfer learning, (b) transductive transfer learning, and (c) unsupervised transfer learning. The domains of the source and

target models in inductive TL are the same, and the tasks can be different but related. In transductive transfer learning, the tasks are the same, and the domains can be different but related. Domains and tasks are different but related in unsupervised learning. Zhuang et al. discussed two different approaches to categorize the TL methods: data-based interpretation and model-based interpretation [34]. The data-based interpretation approach is centered around transferring knowledge through the adjustment and transformation of data. In this approach, TL methods can be categorized into two distinct categories based on their objectives and strategies. On the other hand, model-based interpretation focuses on the development of accurate prediction models in the target task by transferring knowledge based on the model control strategy, parameter control strategy, and model ensemble strategy [34].
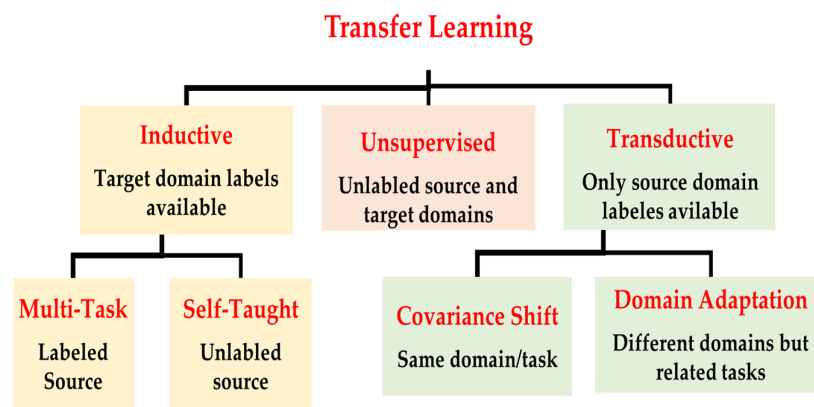

Figure 4. Transfer learning approaches in the perspective of the data labeling status in the source and target domains.

There are various TL approaches that have been developed and examined to improve healthcare services and patients' health, such as fine-tuning [35–40], feature extraction [41–44], multitask learning [45,46], domain adaptation [40,47,48], federate learning [49–51], as well as meta learning methods (such as zero-shot [52], one-shot [53], and few-shot learning [53,54]). In this paper, we discuss twenty-seven studies in detail, distributed as presented in Figure 5, to highlight the applications of TL to enhance healthcare services and outcomes based on DH sensing data, as shown in Figure 6.
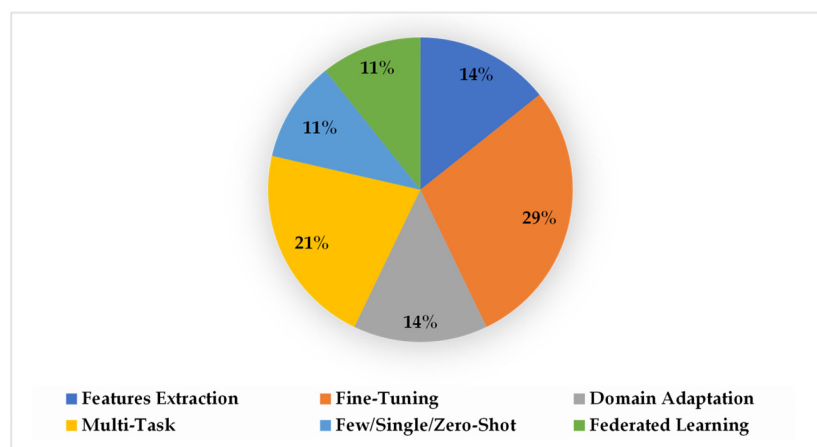

Figure 5. The distribution of the twenty-seven studies that are discussed and clarified in this paper based on the following TL methods: feature extraction, fine-tuning, domain adaptation, multitask learning, few-/single-/zero-shot learning, and federated learning.
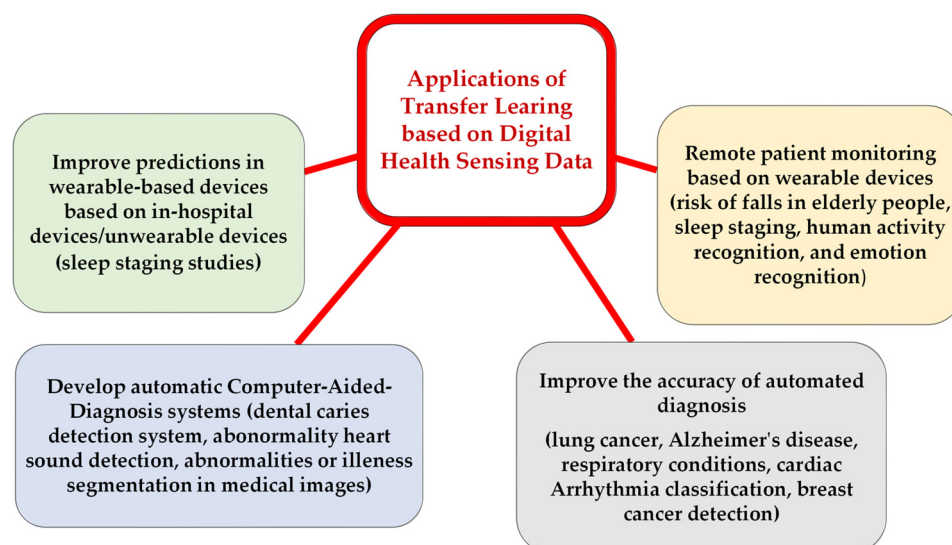
Figure 6. Applications of TL on DH sensing data to enhance healthcare services and outcomes.

This survey paper aims to present a comprehensive study of the applications of TL techniques to enhance DH services and advance healthcare outcomes. The primary motivation for this study stems from the necessity to address challenges, such as limited data availability, data-sharing restrictions, and the high computational demands, within the healthcare sector, all of which can hinder the development of effective ML prediction models. Furthermore, this survey paper explores a variety of DH sensing technologies that can serve as abundant data sources for the development of automated and continuous health monitoring and diagnostic methods, primarily based on ML and DL techniques. This study is positioned to become a valuable resource for both practitioners and researchers in the field of digital healthcare, offering insights and resources for researchers and practitioners in the field of digital healthcare to the application of TL techniques for empowering digital healthcare technologies. Our paper is primarily focused on demonstrating how to overcome ML challenges using a TL approach. The literature included in this survey paper has been sourced from diverse research databases, including, but not limited to, IEEE Xplore, MDPI, PubMed, Google Scholar, ACM Digital Library, and others; in our search, we used the included keywords in this survey paper, and more that are related to the applications of TL on DH sensing data and DH sensing technologies, such as TL for brain tumor detection, applications of multitask learning on medical imaging data, TL for sleep staging, TL for mental health, applications of TL on wearable sensor data, attachable and wearable devices for health monitoring, etc.

The remainder of this survey paper is organized as follows: Section 2 presents the various sensing technologies that are currently available or that can be utilized in the future for DH applications based on ML. These technologies provide efficient methods for health monitoring and disease diagnosis, and they also generate sufficient data that can be utilized for ML and DL applications; Section 3 explains the meaning of TL and illustrates the categories; Section 4 presents the methods, strategies, and applications of TL on DH sensing data to improve healthcare services and outcomes; Section 5 summarizes the use of TL methods and strategies to address the challenges related to the developments of accurate ML prediction models within the field of DH. In addition, it highlights potential challenges that could result in negative transfer.

2. Digital Health Sensing Technologies

Sensing technologies play a critical role in DH, enabling the collection of various physiological, behavioral, and environmental data to monitor and manage health conditions, along with enhancing the diagnosis and prognosis outcomes [1–3,16]. These technologies

include wearable sensors, biosensors, environmental sensors, and imaging sensors [8]. Wearable sensors are widely used in activity monitoring [8,36,47,48], sleep tracking [37–40], fitness tracking [8,16], and health monitoring [8,11,22,23]. Biosensors, on the other hand, can measure various biomarkers, such as glucose, cholesterol, blood pressure, and other vital signs [12,15–17]. Environmental sensors can detect air quality, temperature, and humidity levels, which can impact health conditions [16]. Imaging sensors can provide a noninvasive way to visualize internal organs and diagnose various diseases. ML/DL methods have been used widely to analyze patients' data that have been collected from DH sensing technologies to enhance diagnoses and prognoses by providing early disease detection, automatic and fast diagnosis, personalized medicine, decision support, patient monitoring, and user self-monitoring. In this paper, we classify the sensing technologies in DH into two main categories: (a) portable technologies and (b) nonportable technologies.

2.1. Portable Sensing Technologies

Portable DH devices and technologies refer to the use of portable (at-home and in-hospital) electronic devices and technologies that allow individuals to monitor and manage their health and well-being, as well as the healthcare provider to diagnose and manage individuals' illnesses. The at-home devices can range from simple fitness trackers that monitor the steps taken and calories burned to more complex medical devices that can measure vital signs, such as blood pressure, heart rate, and blood glucose levels; these technologies are designed to be easy to use and accessible to individuals in a variety of settings, including at home, in the workplace, and on-the-go. The in-hospital portable devices can be small or big size, and some of them require experts and trained people to setup and use them. Below, we categorize the portable DH technologies into five groups.

2.1.1. Wearable and Attachable Sensing Technologies

With the advent of digital healthcare, wearable and attachable devices have gained popularity as tools for health monitoring [8,16]. These devices offer a convenient and easy method for individuals to monitor their health in real-time [8]. Most of these devices provide the continuous and noninvasive monitoring of key biological parameters and vital signs, such as blood pressure, heart rate, cholesterol, glucose level, and oxygen saturation levels [8,13,17]. This real-time health data can help individuals identify potential health problems before they become serious and take proactive steps to manage their health. Wearable devices are electronic devices that are designed to be worn on the body, such as smart watches, fitness trackers, smart rings, smart shoes, and smart glasses [8]. One advantage of wearable devices is their convenience and ease of use. Some of these devices are designed to be worn continuously, and often come equipped with sensors that collect and transmit data about the wearer's health and physical activity [8,11]. In contrast, attachable devices are electronic devices that can be attached to the body [8], such as heart rate monitors and blood glucose meters. These devices are typically used for short periods and are designed to collect specific health information. They are designed to be worn throughout the day, providing continuous health monitoring without requiring any extra effort on the part of the user [11]. However, their continuous use can also be a disadvantage, as they may require frequent charging and can sometimes be uncomfortable to wear. Attachable devices, on the other hand, are generally more precise and accurate in measuring specific health indicators. They are often used by healthcare professionals to monitor patients with specific conditions, such as heart disease or diabetes [11]. However, they may be less convenient for everyday use, as they require more effort to attach and remove, and may not be suitable for long-term monitoring [55]. The patch is a major attachable step towards the continuous, real-time, and noninvasive health monitoring of chronic conditions, as well as early-warning signs of disease development [16]. Devices capable of the noninvasive sensing of health status offer significant improvements in the management of chronic conditions, such as diabetes and hypertension. Ultimately, the

choice between wearable and attachable devices depends on the specific health needs and preferences of the user [16].

Table A1 presents the most common wearable and attachable sensing technologies in DH, and highlights their applications and features. Wearable/attachable sensing technology is divided into two main categories based on the number of embedded sensors: homogenous-based sensors (containing only a single sensor) and heterogenous-based sensors (containing multiple sensors) [8]. Wearable-based sensors include the EEG, EMG, PPG, and GPSA. The EEG sensor is used to detect abnormalities in heart functions, irregular heart rhythm, and cardiovascular disease [8]. The EEG sensor is used to continuously measure and monitor the electrical activities of the human brain through scalp-wearable devices, and they are commonly used to identify brain health issues, such as epileptic seizures, brain injuries, antidepressant treatments, and sleep-stage analysis [8]. The EMG sensor is widely used for measuring the electric signal from muscular activities to diagnose neurological disorders. The global positioning system (GPS) sensor is used for activity classifications by detecting the location and velocity of a subject, and it was utilized to manage contact tracing to minimize the spread of COVID-19 [56,57]. The PPG sensor is an optical sensor which is utilized to measure the volumetric variation in blood circulation to study sleep disorders [58]. The accelerometer sensor measures the human acceleration of force and dynamically assesses muscle movement, and it is used in human activity identification studies, such as fall detections in the elderly [59,60].

A wearable heterogeneous-based sensor is a type of sensor that is integrated into a wearable device with the purpose of measuring multiple physiological parameters simultaneously [8]. It can be utilized to monitor changes in vital signs, identify the early symptoms of illness, and provide feedback on lifestyle choices, such as exercise and diet.

In the following, we present various attachable/wearable DH sensing technologies:

- **Blood-Pressure-Monitoring (BPM) Technology**

BPM devices are used to monitor blood pressure. Wireless BPMs are highly portable and utilize smart technology to record and monitor patients' blood pressure and send the obtained information to the healthcare provider. There are two main types of out-of-office BPMs: the arm-cuff and wrist-cuff. Other blood pressure devices used through a finger (such as blood pressure monitors in smartphones) are less reliable. These wearable devices are developed to be used as out-of-office blood pressure measures in order to optimize the management of hypertensive individuals [61]. However, these devices mostly accept a limited number of measurements to be recorded, and have discomfort limitations, specifically at night [61]. The microelectromechanical system (MEMS) blood pressure technology is a chip-based low-cost system with low-nonlinearity error and high-precision inertial sensors [62]. The smartphone-based technology is an extension of the oscillometric principle for cuffless blood pressure monitoring [63]; the smartphone is embedded with a PPG and force transducers that are used as a blood pressure sensor to measure the blood volume oscillations and applied pressure [63]. To activate the sensor, the user presses her/his finger against the sensor location in the smartphone to increase the external pressure of the underlying artery.

- **Cardiac Monitor Technology**

The most popular wearable devices for health monitoring and tracking are digital electrocardiogram (ECG) devices that are featured in smart watches and other fitness trackers, as well as wearable patches and chest straps. The ECG records the electrical signal from the heart to detect abnormalities and different heart conditions. The ECG devices that are used in hospitals and healthcare centers contain a standard 12-lead electrocardiogram, which is a representation of the heart's electrical activity recorded from the electrodes on the body's surface. There are many ECG attachable/attachable devices that are produced to be worn or used by people as a flexible portable monitoring method. Most ECG devices in smartphones and fitness trackers are based on a single-lead ECG and are connected with apps to store the ECG tracing. Wearable ECG monitoring devices are used as low-cost

devices to store and plot ECG data in real-time [64–66]. Some of these devices are adapted with IoT remote monitoring technology to transmit the measured data online to healthcare providers [67]. For cardiac monitoring, compact ECG patches are commonly used. The Zio Patch, measuring $123 \times 53 \times 10.7$ mm and weighing just 34 g, is water-resistant and captures up to 14 days of continuous single-lead ECG data [68]. Technicians apply it to the patient's left pectoral region using a skin adhesive. Users can initiate monitoring with a button press, then send the device and diary to a data processing center for analysis, generating a report for the ordering physician [68]. In [69], a patch-based mobile cardiac telemetry system was developed for COVID-19 patients outside intensive care and telemetry units, proving useful for in-patient management and arrhythmia detection.

- **Wearable Mental-Health-Monitoring Technology**

Various wearable devices have been developed to be used in various crucial applications in mental health and panic disorder research studies due to the simplicity of collecting continuous online data and the availability of multisensory data that are related to understanding people's mental health conditions and statuses [70]. For example, wearable sensors are used to track physiological parameters, such as heart rate and breathing patterns, and the changes in heart rate are found to be associated with stress or anxiety [71,72]. Some other wearable devices can track behavioral signals and parameters, such as sleep patterns, physical activity, and social interactions, that are connected to depression and anxiety [73–75]. In [74], the authors used Lief, a smart patch, as a wearable device and placed it beneath the left breast to collect physiological signals to manage stress remotely so to reduce the symptoms of anxiety. In addition, smart wearable devices are a good source of real-time monitoring and can provide real-time feedback to people related to their mental health [74,76]. Some wearable devices can collect data as well as deliver personalized interventions and recommendations based on the collected and processed measures and parameters to improve people's sleeping habits [73] and activity [77]. Moreover, wearable devices can be used to deliver online therapy recommendations and treatment decisions [78]. The most common types of mobile wearable and portable devices that are used in mood and anxiety disorder applications are: blood pressure cuffs, patches, headsets, headbands, wrist bands, smartphones, electronic textiles, and smart watches [79], and the most common types of sensors that are embedded in these devices are: accelerometers, actigraphs, ECGs, EEGs, EGGs, EMGs, GPSs, PPGs, glucometer magnetometers, microphones, pedometers, as well as temperature and infrared proximity [79].

- **Wearable Sleep Technology**

Sleep medicine experts utilize polysomnography (PSG) systems to record and analyze sleep studies performed in a sleep laboratory. These PSG systems use sensors to measure things such as eye movements, oxygen levels in an individual's blood, heart and breathing rates, snoring, and body movements. PSG systems are used to diagnose sleep disorders, such as sleep apnea, narcolepsy, insomnia, REM sleep behavior disorder, and sleepwalking. These systems are high-cost and require complicated setup with trained professional healthcare staff. In addition, they are inconvenient for sleep monitoring. Current developments in wearable devices help to overcome the PSG system limitations and complexities. Several sensors are utilized in PSG wearable-based systems, such as electrodes to measure the EEG for brain waves (location: forehead, scalp, and ear), and the ECG and impedance cardiography (ICG) to measure the heart activities (location: chest) [80]. In [81], the authors developed a wearable monitoring device based on multisensors for sleep studies as a comfortable and reliable technology. They used an accelerometer, light sensor, sound sensor, temperature sensors, as well as an optical PPG sensor. Microsleep is a short sleep episode that lasts for few seconds and is caused by sleep deprivation, sleep apnea, and narcolepsy [82]. These episodes have very dangerous effects on communities and people's lives, such as a reduction in work performance, traffic accidents, and work injuries. Pham et al. developed WAKE, a behind-ear wearable technology, to detect microsleep by monitoring biomarkers from eye movements (using an electro-oculogram), brain waves

(using an EEG), facial muscle contractions (using electromyography), and sweat gland activities on the skin (using the electrodermal activity score) [82]; this microsleep device was developed as a flexible, accurate, comfortable, cost-consuming, and continuous monitoring trend that can be involved in a wide range of cognitive monitoring and improvement applications [82]. It contains ten embedded electrodes, adheres comfortably behind the ear, and requires only 20 min for setup, which is much quicker than the traditional PSG, which takes around 45 min. Moreover, textile-based sensors have been developed as comfortable, wearable, smart physiological monitoring devices to be used for noninvasively diagnosing various diseases, such as obstructive sleep apnea–hypopnea syndrome and cardiovascular diseases [83,84]. In [83], the authors developed a small-sized flower-shaped textile (which contained two layers of a silver-coated fabric as the base and electrode flower shapes as the superstructure) to be stitched/sewn on clothes, such as bands, to measure pulse waves at the forehead, wrist, arm, and chest [83]. In [84], the authors developed a small-sized smart waterproof textile based on a triboelectric sensor that was adhered to the waterproof Band-Aid bandage for ambulatory cardiovascular monitoring; they combined their smart textile with an ANN to continuously and precisely measure the systolic and diastolic pressure [84]. In [85], the authors developed and fabricated a wearable graphene-textile-based strain sensor with negative resistance variation through a simple thermally reduced graphene oxide (GO) to be knitted directly on clothing or to be adhered in various body locations to detect various physiology signals and monitor various subtle motions; for example, attached on the side of the mouth for various motion detections and facial expressions; on the finger, head, and wrist for pulse monitoring and handwriting recognition; on the neck for pulse monitoring as well as vocal vibration detections; near the abdomen for abdominal breathing detection and analysis; on various body joints to detect the bending of joints [85]. In [86], the authors reported the manufacturing process of a silicone–textile composite resistive strain sensor for monitoring human motion and physiological parameters; the wearable sensor can be worn on the chest and elbow to monitor respiratory activity and joint motion, respectively; it has a high sensitivity, low hysteresis, and ease for shaping custom designs, while also being flexible, skin-safe, and moisture-resistant.

- **Wearable Noninvasive Continuous-Glucose-Monitoring Technology**

Continuous-glucose-monitoring (CGM) systems are a commonly portable device that allows patients to measure their glucose levels in real-time [87]. The most common glucose monitoring methods are invasive, based on finger-prick testing [13]. However, invasive methods can cause a physical and mental burden and an infection risk for diabetes patients, and circadian fluctuations are also reported [13,14]. Thus, noninvasive continuous-glucose-monitoring methods have been developed to reduce the risks and burdens in measuring and monitoring glucose levels. These noninvasive devices are small in size and can be easily connected to a smartphone to monitor blood sugar levels over a period of time [88]. Additionally, the collected data can be shared with healthcare providers online, allowing for better management and adjustment of treatment plans. Because of plasma leakage from blood into tears via the blood–tear barrier, glucose levels in tears are related with blood glucose [89]. Smart contact lenses based on optical sensors are developed as a noninvasive glucose monitoring system to measure glucose levels in the tear fluid [89]. Contact leans are included with various features to be used as an ideal medical device for biosensing applications [90]. Another type of noninvasive glucose monitoring system is the sweat glucose monitoring patch [91]; this system uses sweat sensors that can be worn on the skin, typically on the arm or wrist, to measure glucose levels in the sweat and provide a real-time reading. The detection of the glucose levels based on the contact lenses is based on electrochemical sensors that consist of hydrogels with immobilized glucose oxidases (GOx) [92]. Nanomaterials, such as gold-doped graphene and gold porous structures [93,94], and carbon nanotubes [91], are utilized to enhance the glucose sensor sensitivity. However, the most common challenges in these methods are the inaccurate detection of glucose levels and the low sensitivity due to the low-glucose concentration in the small volume of tears [95]. To overcome these limitations, Kim et al. proposed and

developed smart contact lenses based on HA-Au@Pt BiNCs immobilized in the nanoporous hydrogels of the glucose sensor for long-term and robust continuous glucose monitoring to capture rapid changes in glucose levels [95].

- **Wearable Activity-Recognition Technology**

  Activity recognition is a valuable tool that can provide insights into an individual's physical activity levels and patterns, which can have significant impacts on overall health and well-being. Accurately tracking physical activity can help individuals make informed decisions about their lifestyle and exercise habits. In addition, healthcare professionals can use this information to develop personalized treatment plans. There are several applications of activity recognition, including fitness tracking, healthcare monitoring, elderly care, movement disorder, and sport performance analysis [59,60,96]. Fitness tracking monitors physical activity levels and provides feedback on progress towards fitness goals. Healthcare monitoring can track patients with chronic conditions, such as heart disease or diabetes, and provide healthcare professionals with real-time data on physical activity levels and health metrics [97]. Elderly care involves monitoring elderly individuals and alerting caregivers or emergency services in the event of a fall or a sudden change in physical activity levels [59,60]. Sport performance analysis can provide athletes with valuable insights into areas of improvement and injury prevention. The type of sensors commonly used in activity recognition include accelerometers, gyroscopes, and magnetometers [35,37,49,96]. These sensors can detect various types of movement and changes in body orientation, allowing for the identification and tracking of physical activities, such as walking, running, or cycling. Accelerometers measure the changes in the linear acceleration, gyroscopes measure the changes in the rotational velocity, and magnetometers detect the changes in the Earth's magnetic field. By combining the data from these sensors, wearable devices can accurately recognize and classify different types of physical activities. Smart watches, fitness trackers, smart clothing, and smart shoes are the most common activity-tracking wearable devices [97–100]. These devices can be mounted on different body locations, such as the arms, legs, wrists, chest, and more, to collect data from various sensors [97,100].

- **Wearable Mouth-Based Systems Technology**

  Smart mouthguard monitoring systems have been embedded with biosensors for health monitoring and diagnosis [12,15]. In [12], the authors developed a smart wearable mouthguard as a highly sensitive, selective, continuous, and stable noninvasive monitoring biosensor to detect the salivary uric acid levels in a real-time and wireless fashion [12]. It was embedded with an enzyme (uricase)-modified screen-printed electrode system, a microcontroller, and a Bluetooth low-energy transceiver to transmit the measured information to mobile computers (such as smartphones and laptops) so to be displayed and stored for diagnosis and monitoring purposes. In [15], the authors developed a smart noninvasive wearable oral-monitoring glucose biosensor to measure saliva glucose [15]; this biosensor was integrated in a mouthguard to be installed in the oral cavity. The sensor circuit has a small size and contains a glucose biosensor, a battery, and a telemetry system to sample the saliva, continuously measure glaucous levels, and transmit the readings wirelessly to mobile devices (smartphones/tablets) [15]. The observation of experienced dentists or X-ray are the best ways to diagnose dental caries; however, dental caries is hard to detect in its early stages, and it is mostly detected when the cavity or decayed surface appears [101]. Li et al. developed a wearable fluorescent mouthguard, which consisted of a zinc oxide–polynanocomposite, to precisely identify the locations of lesion sites in humans [102]. This mouthguard displayed a highly sensitive and selective response to volatile sulfur compounds in oral cavities, and showed high fluorescent stability, perfect biocompatibility, and low biological effects. A wireless electronic solution for orthodontic bond failure diagnosis was presented in [103], and it was based on developing a low-power-capacitive-humidity implanted microchip that contained a humidity sensor to detect the bond failure between the tooth and orthodontic braces. Tongue impairments in the elderly impact swallowing, speech, and nutrition. A low-power smart wireless intra-oral wearable

device with six capacitive coplanar sensors was developed to monitor tongue movements and strength, making it suitable for long-lasting rehabilitation without the need for X-rays or internal mouth cables [104].

- **Smart Shoes Technology**

Various types of sensors have been integrated with smart shoes [105], such as (a) a pressure sensor to measure foot pressure, commonly used for diabetic patients, (b) an ultrasonic sensor to measure the distance to an object, mostly used by blind people, (c) an accelerometer sensor to track movements, widely used for gait analysis, (d) a temperature sensor to measure the body temperature and also the atmospheric temperature, (e) an altitude sensor to provide an early warning to climbers or trekkers while at high altitudes, (f) a piezoelectric pedometer to count the number of steps and speed for a specific time, and (g) a gyroscopic sensor to track the angular movement for gait walking pattern identifications [105]. Smart wearable shoes serve two key purposes: enhancing sports and well-being, and enabling medical monitoring and diagnosis. Users utilize them to track daily activities, such as step count and speed [106], gait analysis, and joint stress detection, for improved lifestyle choices.

- **Tear Biomarker Monitoring Using Eyeglasses-Nose-Bridge Pad Technology**

The authors developed a noninvasive real-time tear alcohol and glucose biosensor monitoring device that is placed outside the eyes [17]; wireless electronic circuitry was embedded on the eyeglasses frame to provide a fully portable and convenient-to-use sensing device. These eyeglasses monitoring devices based on the nose-bridge pad was developed to overcome the downsides of a direct contact of the embedded sensor of the contact lens with the eye, such as potential vision impairment [89,92], unsteady mechanical stability, and non-biocompatibility due to immune response and toxic reactions [87].

- **Attachable Patch/Bands for Sweat-Biomarker-Monitoring Technology**

Sweat glands, primarily found in the hands, feet, lower back, and underarms, have led to the development of various portable technologies for measuring sweat biomarkers to diagnose diseases and monitor health. These include head/wrist bands [91], head/arm patches [107], touchpad–fingertip sensors [108], and smart clothing (underwear, socks, gloves, and finger cots). These technologies detect biomarkers, like lactate for fatigue, glucose for diabetes, cortisol for mental stress, creatinine and urea for kidney disorders, and caffeine and lactate for dosage tracking and metabolic monitoring. Bae et al. introduced a stretchable patch with an omnidirectionally stretchable nanoporous gold (NPG) electrochemical biosensor and a stretchable passive microfluidic device for accurate glucose monitoring from sweat [107]. Emaminejad et al. developed a smart wearable head/wrist band platform for multiplexed in situ perspiration analysis, measuring sweat metabolites, electrolytes, and skin temperature for personalized diagnostics and physiological monitoring [91]. Bo Wang et al. devised a thin hydrogel micro patch on the fingertip to sample sweat and monitor biomarkers, like caffeine and lactate, using an electrochemical sensor [108].

2.1.2. Implantable Sensing Technology

Implantable sensing technology involves the use of small devices implanted within the body to measure and monitor various physiological parameters, such as the blood glucose levels, heart rate, blood pressure, and oxygen saturation [109,110]. They can also be used to detect and monitor the presence of specific substances in the body, such as drugs, hormones, and neurotransmitters. These devices can be used to diagnose and manage a range of medical conditions, from chronic diseases like diabetes and heart disease to neurological disorders like epilepsy. One of the key benefits of implantable sensing technology is that it allows for the continuous monitoring of physiological parameters, providing more accurate and reliable data than intermittent testing [110,111]. This can be particularly important for people with chronic conditions that require ongoing management. Another advantage

of implantable sensors is that they can be used to deliver targeted therapies directly to the affected area of the body. For example, implantable pumps can be used to deliver medications to treat pain, spasticity, and other symptoms associated with neurological disorders. Implantable sensors can also be used to monitor the effectiveness of treatments and adjust dosages as needed. This can help to optimize treatment outcomes and reduce the risk of complications. Implantable sensors can be categorized into three distinct types based on their functionality [112]: biopotential sensors that are designed to measure electrical activity, mechanical sensors that respond to changes in mechanical parameters, and chemical biosensors that are specifically engineered to transduce the concentration of a targeted molecule of interest. Overall, implantable sensing technology has a wide range of applications in the field of medicine. Constant efforts are being made to develop new devices to improve patient outcomes and enhance their quality of life. Presented below are some examples and applications of implantable sensing technology:

1. Glucose Monitoring: Implantable glucose sensors can be used to monitor blood sugar levels in people with diabetes [110]. These devices can continuously measure glucose levels and send data to a handheld device or smartphone, allowing patients to adjust their insulin dosages as needed.
2. Cardiac Monitoring: Implantable cardiac monitors can be used to track heart rhythm and detect abnormalities, such as arrhythmias [113]. These devices can also monitor the heart rate, blood pressure, and other vital signs to help doctors diagnose and manage heart disease [110,113].
3. Neurological Monitoring: Implantable sensors can be used to monitor the brain activity in people with epilepsy, helping doctors to diagnose and treat the condition [112]. They can also be used to monitor intracranial pressure in people with traumatic brain injuries.
4. Drug Delivery: Implantable sensors can be used to monitor drug levels in the body, allowing doctors to adjust dosages as needed [110,112,114]. They can also be used to deliver medications directly to the affected area of the body, reducing the risk of side effects [112].
5. Cancer Treatment: Implantable sensors can be used to monitor tumor growth and response to treatment, helping doctors to adjust treatment plans as needed [114,115]. They can also be used to deliver targeted therapies directly to the tumor site, minimizing the damage to healthy tissue.

2.1.3. Ingestible Sensing Technology

Ingestible sensing technology refers to the use of miniature electronic devices that are swallowed or ingested in the form of pills or capsules to monitor various physiological parameters within the gastrointestinal tract [116,117]. These devices contain sensors that can detect and transmit information about the pH levels, temperature, pressure, and other relevant indicators, and can provide valuable insights into digestive processes, medication effectiveness, and disease progression [116]. Ingestible sensing technology relies on a variety of sensors to measure physiological parameters within the body. Dagdeviren et al. developed an ingestible sensor that can be placed on the lining of the stomach to monitor vital signs and mechanical changes in the gastric cavity [118] for diagnosing and treating motility disorders, and monitoring food intake in individuals with obesity. In another study [119], researchers developed an ingestible device that combined probiotic sensor bacteria with microelectronics, which can communicate with external devices like smartphones. They engineered heme-sensitive probiotic biosensors, and showed the precise detection of gastrointestinal (GI) bleeds in pigs, with a remarkable sensitivity of 100% after 120 min. Below, the most common types of sensors used in this technology are presented [117]:

1. **pH sensors** are used to measure the acidity or alkalinity of the digestive system. These sensors can be used to diagnose conditions like acid reflux, gastroesophageal reflux disease (GERD), and Helicobacter pylori infection.

2. **Temperature sensors** are used to measure the temperature of the digestive system. These sensors can be used to monitor body temperature and detect fever, as well as to diagnose conditions like Barrett's esophagus and inflammatory bowel disease.

3. **Pressure sensors** are used to measure the pressure within the digestive system. These sensors can be used to diagnose conditions like gastroparesis, achalasia, and other motility disorders.

4. **Electrolyte sensors** are used to measure the levels of various electrolytes within the body, including sodium, potassium, and chloride. These sensors can be used to monitor electrolyte imbalances and diagnose conditions like dehydration and electrolyte disorders.

5. **Glucose sensors** are used to measure blood sugar levels within the body. These sensors are commonly used to monitor glucose levels in people with diabetes.

6. **Drug sensors** are used to monitor the absorption and distribution of medications within the body. These sensors can be used to optimize drug formulations and dosages for better treatment outcomes.

7. **Magnetic sensors** are used to detect the presence of magnetic particles within the digestive system. These sensors can be used to diagnose conditions like gastrointestinal bleeding.

2.1.4. Smartphones

The use of smartphones in DH has revolutionized the way we approach healthcare, enabling individuals to monitor their health and wellness anytime and anywhere. Smartphones are increasingly being utilized as portable devices for a wide range of health-related applications, including fitness tracking, medication reminders, and telemedicine [120–122]. Through the use of various sensors and applications, smartphones can track important health metrics, such as heart rate, blood pressure, and sleep quality, providing users with real-time insights into their physical and mental well-being [123,124]. In addition, smartphones can be used to store and share medical records [120], access educational resources, and connect with healthcare professionals via telemedicine services [7,8,120]. The widespread availability and affordability of smartphones make them a powerful tool for improving health outcomes, particularly in underserved and remote areas, where access to traditional healthcare services may be limited. However, the use of smartphones in DH also raises concerns regarding privacy, data security, and the accuracy and reliability of health-related information. As such, it is important to ensure that appropriate measures are in place to safeguard user privacy and data security, and to verify the accuracy and reliability of health-related data obtained through smartphone-based applications. Smartphones are equipped with various sensors that can be used for health monitoring and DH applications [121,123]. Table 1 displays the common sensors found in smartphones with their features and applications.

2.1.5. Others

There are several other portable sensing technologies that are not considered in the previous categories, such as portable smart inhalers [125,126], ultrasound devices [127], and in-hospital ECG devices [128], EEGs [129], PPGs [130], spirometers [131], blood analyzers [132], oximeters [133], gas sensors, and smart pill bottles [134]. Smart inhalers are a type of medical device that incorporates electronic sensors and wireless connectivity to provide additional features beyond traditional inhalers [125,126]. They are used to treat respiratory conditions, such as asthma and chronic obstructive pulmonary disease.

Table 1. Smartphones' sensors and their applications in digital healthcare [121,123].

Sensor	Function	Application
Accelerometer	Measures the phone's movement and orientation	• tracking physical activity • monitoring sleep quality
GPS	Provides location information	• movement tracking • physical activity monitoring • detecting location-based health information
Gyroscope	Measures the phone's rotation to detect changes in position	• tracking physical activity
Photoplethysmography (PPG)	Measures the heart rate	• fitness tracking to monitor heart health • stress management
Photodiode sensor (ambient light sensor)	Measures the amount of light in the user's environment	• adjusting the phone's display brightness to reduce eye strain • improving sleep quality
Infrared (IR) sensor (proximity sensor)	Detects the presence of nearby objects or surfaces, as well as the contactless monitoring of vital signs	• remote patient monitoring • sleep apnea detection • stress management • fall detection

2.2. Nonportable Sensing Technologies

Nonportable DH technologies refer to those devices that are not easily transportable and usually require a fixed installation. These technologies can be used in various settings, such as hospitals, clinics, and smart homes, to provide continuous monitoring and improve patient outcomes. Below are the most prevalent forms of nonportable sensing technology:

1. **Stationary medical imaging technologies**: Imaging technologies are noninvasive methods to visualize internal organs and diagnose various diseases [135]. Examples include X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). Owing to the extensive literature available on medical imaging methods and their applications in detecting and diagnosing various diseases and abnormalities, we have not provided detailed features of each method. Instead, we have referenced key review articles, such as Hosny et al., which presented a comprehensive overview of imaging technologies that have been enhanced with artificial intelligence techniques to diagnose various diseases [136]. Guluma et al. also reviewed DL methods in the detection of cancers using medical imaging data [137]. Additionally, Rana et al. discussed the use of ML and DL as medical imaging analysis tools for disease detection and diagnosis [138]. These articles provide valuable insights into the types of medical imaging data and applications of advanced computational techniques in medical imaging, and demonstrate their potential in improving disease diagnosis and patient outcomes.

2. **Environmental sensing technologies**: They are used to detect and monitor environmental factors that can impact health conditions. Examples include air quality sensors, temperature sensors, and humidity sensors [139]. These sensors are used in smart homes. By combining these sensors with other DH technologies, they can play significant roles in improving the quality of care, reducing healthcare costs, and enhancing the independence and well-being of individuals [140].

3. **Monitoring and diagnostic technologies**: Monitoring and diagnostic technologies based on biosensors are used to monitor and diagnose health conditions [141]. These devices are used to measure various biomarkers, such as glucose, cholesterol, and

other vital signs, such as ECG, EEG, electro-oculography (EOG), and electroretinography (ERG).

4. **Robotic surgery systems**: They are advanced medical devices that utilize robotic arms and computer-controlled instruments to assist surgeons in performing minimally invasive surgeries [141–143]. Examples of common robotic surgery systems include: (1) the da Vinci Surgical System [141], which is comprised of a console for the surgeon, and several robotic arms that hold surgical instruments and a camera; (2) MAKOplasty [142], utilized for orthopedic surgeries, such as knee and hip replacements; (3) the CyberKnife [143], employed for radiation therapy to treat cancer; (4) the ROSA Surgical System, utilized for neurosurgery procedures.

The most prevalent sensors utilized in digital healthcare aimed at developing robust ML/DL models for health monitoring and diagnosis are presented in Table A2. In this table, we have outlined the data types and the ML/DL applications associated with each DH sensor technology [144–203].

3. Transfer Learning: Strategies and Categories

3.1. Why the Transfer Learning Technique

In this paper, the term classical learning (CL) denotes a learning approach based on traditional ML or DL methods which emphasizes the design and development of prediction models from scratch using labeled or unlabeled collected data to perform predictions on future data.

Any prediction problem based on ML can be categorized into three categories according to annotation status of the train and test datasets: (a) supervised, (b) semi-supervised, and (c) unsupervised [162]. In the supervised learning approach, both the train and test datasets are labeled and suitable to generalize an accurate prediction model; in supervised learning, the prediction model performs mapping between inputs (features) and outputs (labeled targets) [204]. Various prediction tasks can be performed in the supervised learning approach: classification, detection, segmentation, and regression. In semi-supervised learning, the available data contain small labeled data and large unlabeled data, and both labeled and unlabeled data samples are used to generate a prediction model [205]. The unsupervised learning approach utilizes unlabeled data only, and it is used widely in dimensionality reductions, feature selections, and clustering applications. In addition, there is reinforcement learning (RL), which aims to achieve an optimal behavior in an interactive environment by using feedback from a series of previous actions [204]. Like in supervised learning, the RL maps between the inputs and outputs, but the feedback is a series of correct learning actions, as in unsupervised learning. Both RL and unsupervised learning perform learning in unlabeled data, but unsupervised learning discovers the similarities and differences between the data samples and RL learns an optimal behavior by achieving maximum rewards.

The following two terms are used in ML problems to define the data distribution and the purpose: the *domain* and *task* [206]. The domain *D* provides information about the inputs to an ML algorithm (data), and it is defined by two components, a **feature space *X*** and a **marginal probability distribution *P(X)*** [206]. The task *T* describes the purpose of the ML model, and two components are used to define the task *T*: a **label space *Y*** (outputs) and a **predictive function *f* (·)**. The predictive function is learned from the feature vector and label pairs $\{x_i, y_i\}$, where $x_i \in X$ and $y_i \in Y$ [206].

If a specific ML algorithm based on the CL approach is used to solve two problems (i.e., the source and target), the *domains* and *distributions* of the data, as well as the *tasks* of both the **source** and **target** problems, should be same. Additionally, the target data are usually a subset of the training data (source data). If either the *domains* and *distributions* or the *tasks* in both the source and target are dissimilar, the CL method is mostly unproper to develop accurate prediction models. In addition, there are four main challenges that arise when users attempt to develop accurate and reliable ML prediction models based on the CL approach [31]:

1. Appropriate modeling algorithms: there are many different types of ML algorithms, and choosing the right modeling algorithm for a particular task requires careful consideration of the data, the problem, and the desired outcome.

2. Hyperparameter tuning: each ML method has hyperparameters that must be set before training, such as the learning rate, regularization strength, number of layers, etc. Determining the optimal values for these hyperparameters can be time-consuming, as it often requires many attempts to attain the best configuration.

3. Data quality and privacy: preparing data to train ML models often requires extensive preprocessing of the raw data to enhance its quality and size. This involves techniques like normalization, scaling, transformation, feature selection, data augmentation, and data denoising, which demand careful considerations of the underlying data and the specific problem.

4. Significant hardware resources: DL algorithms particularly require significant computational resources, including powerful GPUs, high-speed storage, and large amounts of memory, to perform complex computations due to the deep architectures that consist of various types of numerous kernels and layers. Several challenges are associated with these requirements, such as cost, availability, scalability, energy consumption, maintenance, and upgrade requirements.

Addressing the above challenges requires careful consideration of the data, problem, and available resources, and often requires a combination of technical expertise, domain knowledge, and trial and error.

Within the domain of DH, the availability of insufficient data can present challenges to the development of efficient ML prediction models. These data challenges include various factors, such as limited data availability, data imbalance, concerns about data quality and consistency, and constraints on data access and sharing [31,207]. In the context of DH, dataset constraints related to limited samples, especially for rare diseases or conditions, can complicate the generalization of ML models based on CL. Additionally, imbalanced data are a common problem in DH, which leads to potential biases and poor performance on underrepresented classes [31,207]. Furthermore, DH datasets may be noisy, incomplete, or inconsistent, which can make it challenging to extract useful information and train accurate models [31]. The presence of sensitive information, such as patient health records, within these datasets further restricts data sharing, consequently limiting the availability of sufficient data for the development of ML models. Overall, these challenges collectively contribute to the complexity of developing accurate and scalable ML models based on CL within the realm of DH.

TL methods have been developed as a vital solution to address the above challenges associated with the CL approach in DH [35–54]. Figure 7 shows the general architecture of the TL approach. Many researchers describe **TL** as "the improvement of learning in a new task (the target task) through transferring knowledge from a related task (the source task) that has already been learned previously". The *source domain* and the *source task* are defined as *Ds* and *Ts*, respectively. The *target domain*, and *target task* are defined as *Dt* and *Tt*, respectively. The objective of TL is to transfer knowledge from the **source problem** to obtain a reliable solution in the **target problem**. Thus, the TL methods are ML optimization methods to speed up learning process by fast convergence, reducing the requirements of big data, decreasing the memory usage (to deal with complex computations), and improving the performance (in terms of the starting point and accuracy) [32]. The definition that focuses on *transferring a previous knowledge* can be related to fine-tuning and feature extraction methods only. In this paper, we expand the meaning of TL to cover any type of knowledge transfer from the source to the target, either previously learned or simultaneously learned, to include other types of TL, such as domain adaptation, multitask learning, and meta learning methods. In the next section, we will describe each of these approaches and their vital applications in DH.
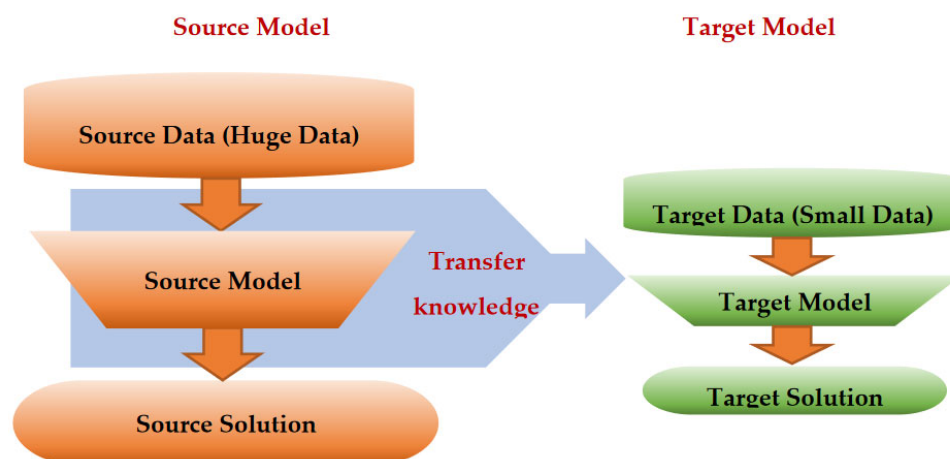
Figure 7. Transfer learning architecture.

3.2. Categories and Techniques of Transfer Learning

TL can be classified into three main branches based on the availability of the labeled data in the source and target task: inductive, transductive, and unsupervised [33,34], as illustrated in Figure 4. TL is also categorized into four main groups based on the knowledge transferred between domains [33]: instance transfer, feature-representation transfer, parameter transfer, and relational-knowledge transfer.

To use an effective TL method to obtain a reliable solution in the target problem, we need to answer the following three questions carefully: (1) What to transfer? (or what knowledge to transfer from the source to the target?); (2) How to transfer? (or how to develop a proper learning algorithm to transfer knowledge?); (3) When to transfer? (or when should the knowledge not be transferred?). There could be various possible answers for the above three questions based on the variations in the domain and task of the source and target models. However, the answer may lead to *negative transfer*, which requires a different strategy or method. The data labeling status can be used as a good sight to answer these questions, as shown in Figure 4. Another way to answer these questions can be the relation between the source and target domains. If the source and target domains are similar or closely related in terms of features and data distributions (i.e., $Xs = Xt$), the approach is defined as a homogeneous TL. If the source and target domains are dissimilar in terms of features or data distributions (i.e., $Xs \neq Xt$), the approach is defined as heterogeneous TL [206]. In heterogenous TL, the knowledge is transferred between different or unrelated source and target domains, which may require adaptation or alignment techniques to bridge the gap between these two domains. Thus, homogenous transfer can often be easier to implement due to the similarities between domains, while heterogeneous TL requires more sophisticated techniques to handle the dissimilarities and domain shifts between the source and target domains.

3.3. What to Transfer?

What is the type of knowledge needing to be transferred from a source model to a target model? The answer to this question is crucial to choosing the suitable strategy, and then the best algorithms, to develop accurate prediction models. Figure 4 presents the most common approaches to answer "*What to transfer?*" that are related to the three TL methods in Figure 4. These approaches are described as follows [33,34,206]:

- **Instance transfer:** The ideal solution in TL is to effectively reuse knowledge from one domain to enhance the performance in another domain. However, the direct reuse of data from the source domain in the target domain is typically not feasible. Instead, the focus is on specific data instances from the source domain that can be combined with target data to enhance the results. This process is known as inductive transfer.

This approach assumes that particular data portions from the source domain can be repurposed through techniques like instance reweighting and importance sampling.

- **Feature-representation transfer:** The goal of this approach is to decrease the differences between domains and improve the accuracy by finding valuable feature representations that can be shared from the source to the target domains. The choice between supervised and unsupervised methods for feature-based transfers depends on whether labeled data are accessible or not.

- **Parameter transfer:** This approach operates under the assumption that models for related tasks have certain shared parameters or a common distribution of hyperparameters. Multitask learning, where both the source and target tasks are learned simultaneously, is used in parameter-based TL.

- **Relational-knowledge transfer:** In contrast to the above three methods, relational-knowledge transfer aims to address non-independent and identically distributed data (non-IID), where each subsample exhibits significant variation and does not accurately represent the overall dataset distribution.

From Figure 8, we can conclude that **not all** the approaches mentioned above can be applied to all the three TL categories in Figure 4. For example, all the above approaches can be employed with inductive TL due to the availability of the labeled data for the target model. In contrast, the instance transfer and feature-representation transfer approaches are suitable for transductive TL, which is defined as suitable for situations involving similar source and target tasks, but without a requisite similarity in the source and target domains. This lack of similarity can appear as either variation in the feature space of the domains or variation in the marginal probability distribution of the domains (with a similar feature space) [33]. The transductive transferred knowledge attempts to solve these variations between the source and target domains, thus the absence of labeled data in the target is the case of this approach. This approach proves particularly valuable in addressing the challenges of a costly labeling process for target problems, such as medical image labeling [208–210]. In addition, the feature-representation transfer approach is utilized as an unsupervised TL method, requiring no labeled data to extract high-quality attributes from the raw data. As such, it is applicable to all three approaches and stands as the sole method for unsupervised TL [33].
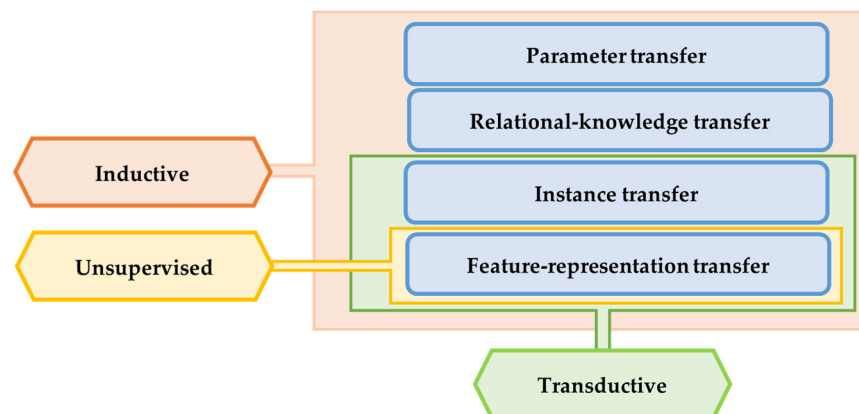

Figure 8. Approaches of knowledge transfer to answer *"What to transfer?"* for the three TL strategies: inductive, transductive, and unsupervised [206].

4. Applications of Transfer Learning on Digital Health Sensing Technologies

TL has emerged as a promising approach in digital healthcare, enabling the development of accurate and efficient ML models with limited data. Recent research has demonstrated the benefits of TL in a wide range of healthcare applications, including medical image analysis for disease diagnosis [210–212] and wearable sensor processing for patient monitoring [37,39,40]. For instance, TL has been used to improve the accuracy of the

automated diagnosis of lung cancer in CT scans [210], where the models were fine-tuned on large-scale image datasets. TL has also been used to develop personalized models for supporting decision-making by categorizing patients of Alzheimer's disease based on their MRI scans [213] into one of the following groups: Alzheimer's disease, late mild cognitive impairment, mild cognitive impairment, and normal cognition. Additionally, TL has shown great potential in remote patient monitoring, where it has been used to analyze wearable sensor data and predict the risk of falls in elderly people [155,214], the steep staging [39,40], and human activities [47–49]. It has been used to develop models for predicting blood glucose levels in patients with diabetes based on data collected from wearable sensors [215]. These studies have demonstrated the significant impact of TL on digital healthcare, highlighting its potential to improve patient outcomes and reduce healthcare costs by facilitating early diagnosis, personalized treatment, and remote monitoring.

Selecting the most suitable TL method and strategy is crucial to develop reliable prediction systems in digital healthcare. Factors such as the availability, size, and type of data, as well as the type of task and the relationship between source and target domains, must be considered when selecting a TL method. Additionally, privacy and data sharing must also be taken into account. In the following subsection, we present and explain several TL methods and approaches in digital healthcare to enhance diagnosis and prognosis outcomes, as well as digital healthcare services.

4.1. Methods, Strategies, and Applications of Transfer Learning in Digital Healthcare

In the realm of digital healthcare, a plethora of TL methods and strategies have been proposed and developed with the aim of bolstering the accuracy and training time of prediction models, mitigating the impact of data limitations, including issues with data quality, size, accurate labeling, bias, compatibility, and privacy, and reducing computation costs. In this context, we have outlined and summarized the most prevalent transfer learning strategies and methods employed for diverse applications in digital healthcare. Furthermore, we provide information regarding each study to inspire researchers to employ these approaches across a range of applications, improve their current systems through training or combining various techniques, and develop novel approaches. It is noteworthy that some of the studies cited herein rely on multi-TL approaches and strategies to address the manifold challenges and issues in ML and digital healthcare.

4.1.1. Feature Extraction

In the medical field, the availability of sufficient data for DL is crucial. When working with small medical datasets, traditional ML methods may be a suitable alternative to DL, which typically requires large amounts of data [162]. However, when working with medical images, traditional ML methods require a preprocessing step to extract, select, and/or combine meaningful features that can be challenging to implement effectively. To address this challenge, leveraging pretrained DL (source) models that have learned general features from large and diverse datasets can improve the model performance on downstream tasks (target tasks), reduce the need for extensive retraining on new data, and enable the effective transfer of knowledge between different tasks and domains.

With this method, users have the flexibility to employ the entirety of the pretrained network's layers, except the output layer [42], or select specific layers that yield meaningful features [43]. These chosen layers remain frozen to extract features; therefore, this method is considered as an unsupervised TL method, which is widely used in the preprocessing step to extract meaningful representations from the data without requiring existing labels [42]. These features are commonly called deep features, as they are extracted from pretrained DL models. Then, these features are either directed to a traditional ML algorithm (such as the SVN, KNN, etc.) or to a new output neural network prediction layer, as shown in Figure 9, to train a new ML model. An important consideration for this method is the need for consistent input vector dimensions in both the source and target models, coupled with the requirement for the relevance between the source and target domains. For example, most

of the available state-of-the-art pretrained models were developed based on image data for computer vision tasks (classification, detection, and segmentation), such as VGG 16, VGG 19, ResNet50, ResNet101, InceptionResNetV2, etc. Thus, these models cannot be used to extract features from non-imaging data, such as sensor and sound signals. Researchers have addressed these limitations by incorporating preprocessing techniques, such as resizing/cropping [41,43,44], domain transformation [42] (as shown in Figure 10), and feature fusions [43]. These steps are employed to harmonize the input data dimensions and establish domain relatedness prior to feeding the data into the network's input layers. The feature extraction method has been used widely to solve limited data samples, commonly within medical imaging data [41,43,44]. In [41], the authors employed three state-of-the-art pretrained DL classification models, namely, ResNet50, ResNet101, and InceptionResNetV2, to extract high-quality features from X-ray images. These features were utilized to train two different traditional ML classifiers, the SVM and KNN, using the 10-fold cross-validation method to classify patients' X-ray images into three categories: COVID-19, normal, and pneumonia. The authors concluded that a high classification accuracy of 99.86% was achieved using the SVM classifier. This suggests that the model could serve as a valuable decision support tool for radiologists. In [42], the authors explored the feasibility of using TL based on feature extraction to address the challenge of limited training data for the ECG signal classification. They used the pretrained DenseNet (the 161-layer deep CNN) to extract features from the ECG data to classify the ECG arrhythmia into four classes: normal sinus rhythm, ventricular fibrillation, atrial fibrillation and flutter, and ST segment change (ST). As the DenseNet model was trained on the image data, the authors applied a domain transformation to convert the signal representation (one-dimensional data(1D)) of the ECG arrhythmia to an image representation that was represented by the ECG spectrograms (two-dimensional (2D)). The extracted features from the ECG spectrograms were used to train an SVM classifier through 10-fold cross-validation. This model was based on deep features, and achieved an accuracy of 97.23% in classifying nearly 7000 instances, outperforming other models based on the CL approach using the SVM classifier using 1D and 2D ECG signals.
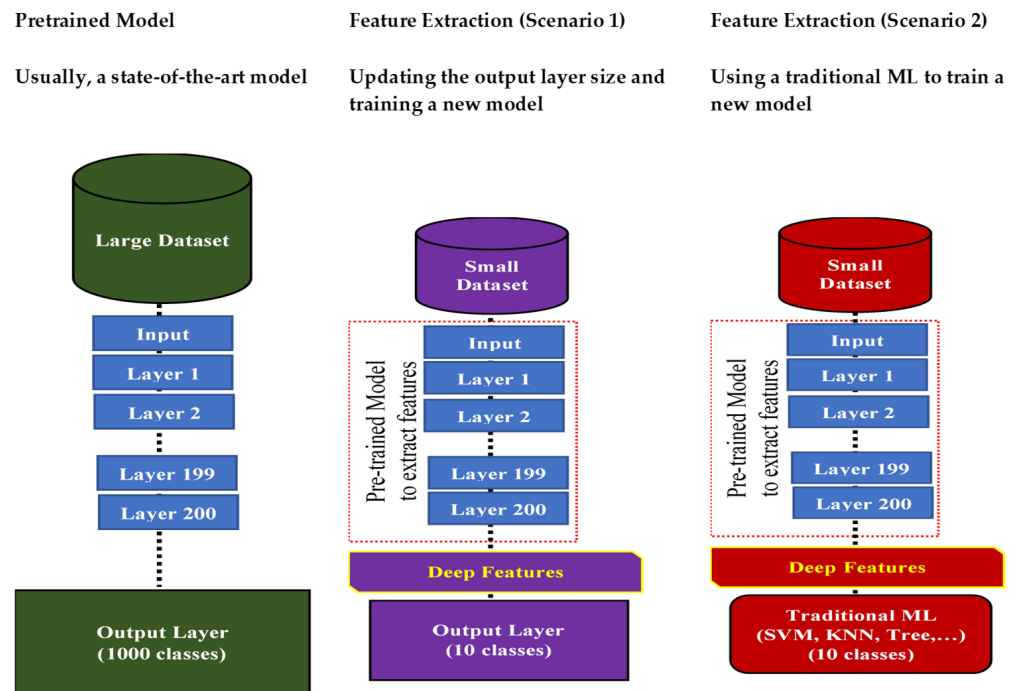

Figure 9. Transfer learning based on the feature extraction method.
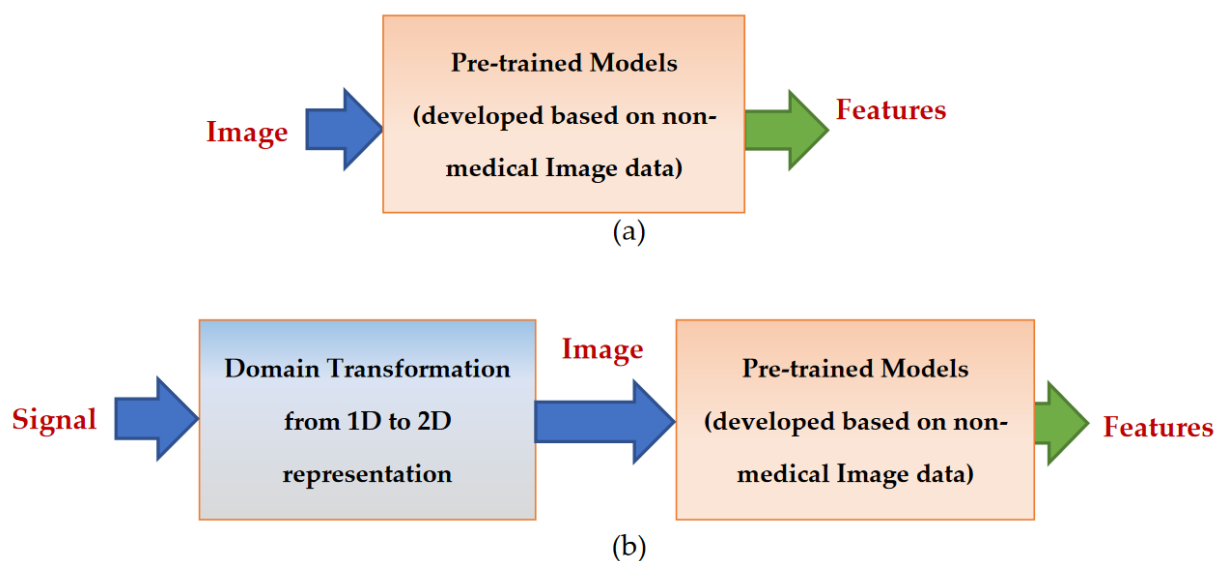
Figure 10. Feature extraction process using state-of-the-art image-based pretrained models that were developed on huge nonmedical imaging data for computer-vision tasks. (**a**) Medical imaging data can be directed to the input of these state-of-the-art image-based pretrained models after proper scaling and/or cropping; (**b**) One-dimensional data, such as sensor data and sound data, should be transformed to a two-dimensional shape (image representation) to be used with these models.

Vo et al. used ensemble deep pretrained convolutional neural networks to empower the meaning of the extracted deep features from multiscale images to grade breast cancer in histology images using a traditional ML (gradient boosting trees) [43]. The authors concluded that their method achieved a better performance compared to the state-of-the-art breast cancer classification systems in categorizing histological breast cancer images into four groups (normal, benign, in situ carcinoma, and invasive carcinoma) or two groups (noncarcinomas (combining the normal and benign classes) and carcinomas (combining the in situ and invasive classes)) due to the use of the ensemble deep convolutional neural networks (DCNNs) to combine various high-quality deep features; they reported the ensemble DCNNs model achieved an improved accuracy of at least 3.3%, 4.2%, 5.5%, and 3.6% for the images at the respective magnification factors of $40\times$, $100\times$, $200\times$, and $400\times$, respectively, compared to the other state-of-the-art approaches. In another breast cancer classification study [44], the authors utilized two TL methods to detect breast cancers in histopathological images: (1) fine-tuning, and (2) feature extraction. First, the authors fine-tuned two state-of-the-art imaging classifiers, the VGG16 and VGG19 networks, on histopathological breast images (this method will be discussed in the next subsection), and then they used these fine-tuned networks as pretrained models to extract the discriminated cancer features (deep features) from the histopathological images. To improve the performance, the authors used the GAN to increase the size of the data. The authors directed these extracted features (from the data, and augmented the data using the GAN) to a neural network to develop a reliable breast cancer detection system; they proposed three voting methods to calculate the accuracy for classifying malignant or benign patches, with method A relying on majority predictions, method B assigning correctness if two out of four patches are correct, and method C requiring at least one correct patch for the overall image to be classified as correct. The average attained accuracies of 94.9–99.2% were achieved by both methods B and C, and authors indicated the feasibility of using them in detecting the cancer when the patients have any potential signs before medical examinations.

In Table A3, we have summarized the methodology for the X-ray image classification using the TL feature extraction method to assist in diagnosing COVID-19 [41], thereby providing insights on how to implement TL feature extraction on DH sensing data.

4.1.2. Fine-Tuning

Fine-tuning is a TL method that involves taking a whole pretrained model, or part of it, and adapting it to a new downstream task with additional training on task-specific data [216]. The adaptation here includes model parameters as well as the model structure. Both feature extraction and fine-tuning utilize wellpretrained models that were developed on big data to assist in the development of a new task. The model parameters are frozen in the feature extraction method, but updatable in the fine-tuning method. In digital healthcare, fine-tuning can be particularly useful for tasks where specific features relevant to the task are not learned by the pretrained model. By fine-tuning the pretrained model on new task-specific data, the model can learn to adapt the features to the new task and improve the performance [216]. For example, in medical image analysis, fine-tuning can be used to train the pretrained convolutional neural network (that was already trained on a computer-vision classification task) on labeled medical image data for tasks such as tumor classification. The fine-tuning process requires three steps to adapt a pretrained model to a new task, as shown in Figure 11. Some researchers extracted part of a pretrained model (usually the top part that is close to the input) and modified it by adding new trainable layers in the output part to prepare it for a new task.
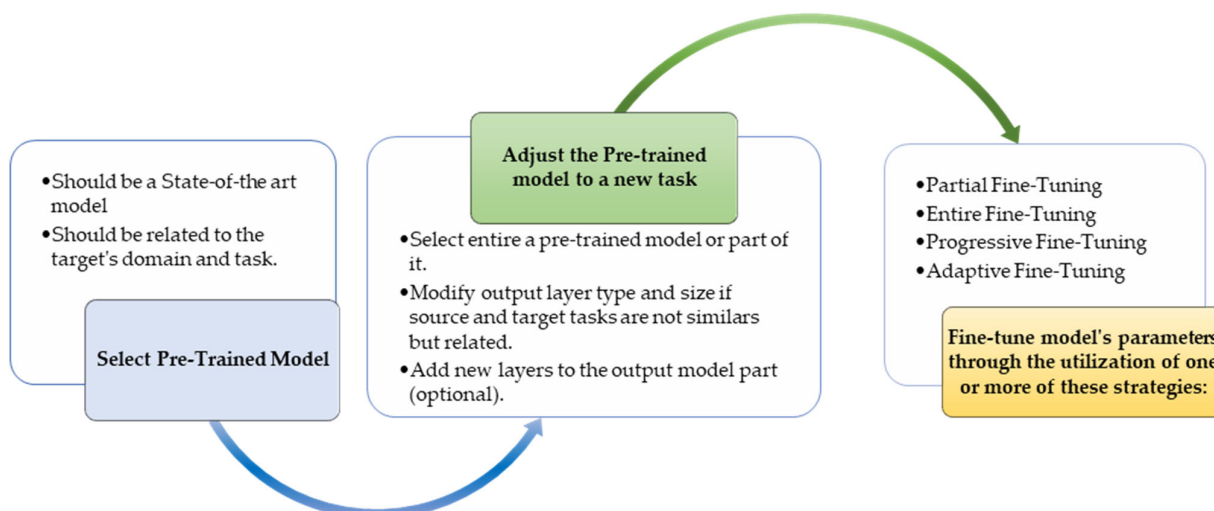

Figure 11. Steps to fine-tune a pretrained model.

Various strategies can be employed to refine a pretrained model through fine-tuning, as explained below. These strategies differ in terms of which of the parameters of a pretrained model are selected for updating, and how these parameters can be updated to develop an accurate model for a target task.

1. **Partial Fine-Tuning** (unfreezing some layers)

Here, users can selectively unfreeze and fine-tune only a subset of layers in the pretrained model [35]. Typically, researchers unfreeze the later layers (closer to the output) and keep the earlier layers frozen because the earlier layers develop low-level features (general features, such as lines, edges, and gradients), while the top layers develop high-level features (advanced descriptors, such as shape, type, and spatial and temporal information). This approach can be useful when the lower-level features are universally applicable and only the task-specific high-level features need adjustment.

2. **Fully Fine-Tuning (unfreezing entire extracted layers)**

In this approach, all layers of the pretrained model are unfrozen, and both the lower-level features and higher-level representations are adjusted to the new task's data [217,218]. This approach can be effective when the new (target) task's dataset is significantly different from the dataset on which the original (source) model was trained. This technique is

especially useful when additional data have been amassed after training a model on an extensive dataset. Instead of retraining the model using both old and new data, fine-tuning can be exclusively applied to the new data. This approach does not necessitate any layer removal, modification, or addition, given that the target task aligns with the source task. However, if the source and target tasks are not identical, but interconnected, certain modifications become imperative for the successful application of this approach.

3. **Progressive Fine-Tuning (partially unfreezing the layers and training them on a multistage)**

In the first stage, the initial layers of a pretrained model are frozen and the remaining layers are fine-tuned. The second phase involves gradually unfreezing the frozen layers in the first stage and fine-tuning the unfrozen layers.

4. **Adaptive Fine-Tuning (differentiating the learning rates for layer groups)**

This method adjusts the learning rate for the different layers during the fine-tuning of a pretrained model. The layers closer to the input capture the general features, which are valuable for a new task, and thus have smaller learning rates to preserve these features [219]. On the other hand, the layers closer to the output learn features that are more related to decision making on the task's specifics, and may require larger learning rates for efficient adaptation. By customizing the learning rates across the layers, adaptive fine-tuning enhances the convergence, the stability, and the model's ability to transfer knowledge to new tasks. Here, it is worth highlighting that adjusting the learning rates for the layer groups is a distinct approach from the more general concept of adapting the learning rates within the optimizer.

Figure 12 clarifies the idea of the entire fine-tuning and partial fine-tuning approaches. If the target task is not similar to the source task, modifying a pretrained model is essential for developing a proper model for the target task; modifications may include updating the output layer only to make it compatible with the target task purpose or increasing the network capacity by adding new layers on the model's output part to develop more robust task-specific features. Although increasing the model capacity also increases the computation costs, it is useful for improving the performance [220].

After describing the common fine-tuning approaches, the subsequent discussion explores the relevant applications for a better clarification and a source of inspiration. To develop the dental caries detection system, researchers have implemented modifications to the pretrained VGG16 model by adding specific layers after removing the output layer from the VGG16 [217]. They called their model the ConvNet, and trained end-to-end on oral photographs captured using consumer-grade cameras. Then, they fine-tuned the ConvNet with true positives against only false positives as a second training stage to decrease the false-positive predictions and boost the performance by achieving 85.65%, 81.9%, and 64.6% of the area under the curve (AUC), image-wise sensitivity, and box-wise sensitivity, respectively. The authors reported that their approach successfully classified the presence of dental caries in the provided images and accurately identified the localization of the bounding boxes. This outcome underscores the potential of their methodology as a valuable tool for cost-effective and efficient dental caries screening among large populations. Koike et al. investigated the effectiveness of using TL for heart-sound classification [218]. They fine-tuned a pretrained model that was trained on large-scale audio data, the PhysioNet CinC Challenge dataset, to classify heart sounds into normal and abnormal sounds. The authors reported that their method demonstrated superior performance compared to other models that were pretrained on images, achieving the highest unweighted average recall of 89.7%. This highlights the potential of using TL methods as a noninvasive way to monitor a person's health status by automatically extracting higher representations from heart sounds without requiring human domain knowledge. In [36], researchers addressed two key challenges in human activity recognition (HAR) using TL. First, conventional training struggles with new users' diverse activity patterns. Second, training from scratch being impractical for mobile apps due to computation and time constraints. Their innovative approach in-

volved a thorough analysis to identify common and user-specific features. By transferring the reusable parts of an offline classifier to new users and fine-tuning for uniqueness, they achieved remarkable results—up to a 43% accuracy boost and 66% training time reduction. Additionally, hardware assessments indicated a 43% drop in power consumption and a 68% drop in energy consumption. In another study [35], partial fine-tuning was employed to address the cross-sensor challenge that arises when sensor variations between the source and target dataset are presented to train a human activity recognition (HAR) system; the authors utilized nine pretrained state-of-the-art convolutional neural network (CNN) models on the KU-HAR dataset as the source. As all these pretrained CNNs were developed using image data, the researchers generated the scalogram from the accelerometer and gyroscope data of smartphones as a virtual image representation by using different configurations of mother wavelets. The most superior performance from the source KU-HAR dataset was achieved by partially fine-tuning the DenseNet121 architecture using the Morlet wavelet (with a classification accuracy of 97.48% and an F1 score of 97.52%), thereby outperforming the state-of-the-art achievement; they found that freezing the first 308 layers of the pretrained model resulted in faster training and a smoother gradient descent on a small dataset. This model also achieved improvements in both the accuracy and F1 score by 0.21% and 0.33%, respectively, on the entire UCI-HAPT target dataset. In addition, they reported that the fine-tuned larger datasets led to the negative transfer causing a drop in the accuracy. Another study involved the transfer of knowledge from in-hospital multi-sensor data, which are generally more comprehensive and dependable, to wearable devices to benefit from its size and cost [37]; this approach sought to improve the accuracy of the models trained on the wearable device data, which are typically derived from a limited set of sensors. The authors used in-hospital recordings to boost the performance via TL on a sleep-staging task using a single channel of the EEG captured from an in-home commercial system [37]. They used two pretrained neural networks, bespoke (their own network) and the DeepSleepNET, to be trained on six publicly available in-hospital datasets based on PSG sensors as the source data, and then retrained these models on a wearable head device EEG-based sensor as a target dataset. They tested several transferability measures (such as the log expected empirical prediction (LEEP), H-score, hypothesis margin, silhouette score, and the target density around the source) to determine the most effective one for assessing the performance on unseen target data. They used two CNN structures: their own structure and the DeepSleepNET structure. They used several TL approaches to retrain bespoke (their own network) and the DeepSleepNET, aimed at developing the best model, such as Head Re-train, subspace alignment, Per-Class CORAL, CORAL, and deep domain confusion (DDC). They found that retraining the head layers (the closest to the output) was the most effective TL method, and the transferability measures provided useful indicators of the TL effectiveness. In [40], the authors proposed a method to improve the performance of a sleep-staging system based on small PPG data through a combined domain and decision TL. They used a pretrained RNN model based on large ECG data from an unwearable device to enhance the performance of the wearable-PPG-based data. The authors compared different training strategies, including CL and three TL approaches (domain transfer, decision transfer, and combined domain and decision transfer). The authors reported that the models developed using any of the three TL approaches achieved a better accuracy than those based on CL (trained from scratch). While the accuracy of each domain transfer and decision transfer was similar, the combined domain and decision method outperformed the other two TL methods by achieving 0.65 ± 0.11 and $76.36 \pm 7.57\%$ for Cohen's kappa and accuracy, respectively. The authors concluded that training a successful structure from scratch is not a good strategy, although it can be considered as transferring knowledge based on the model structure.
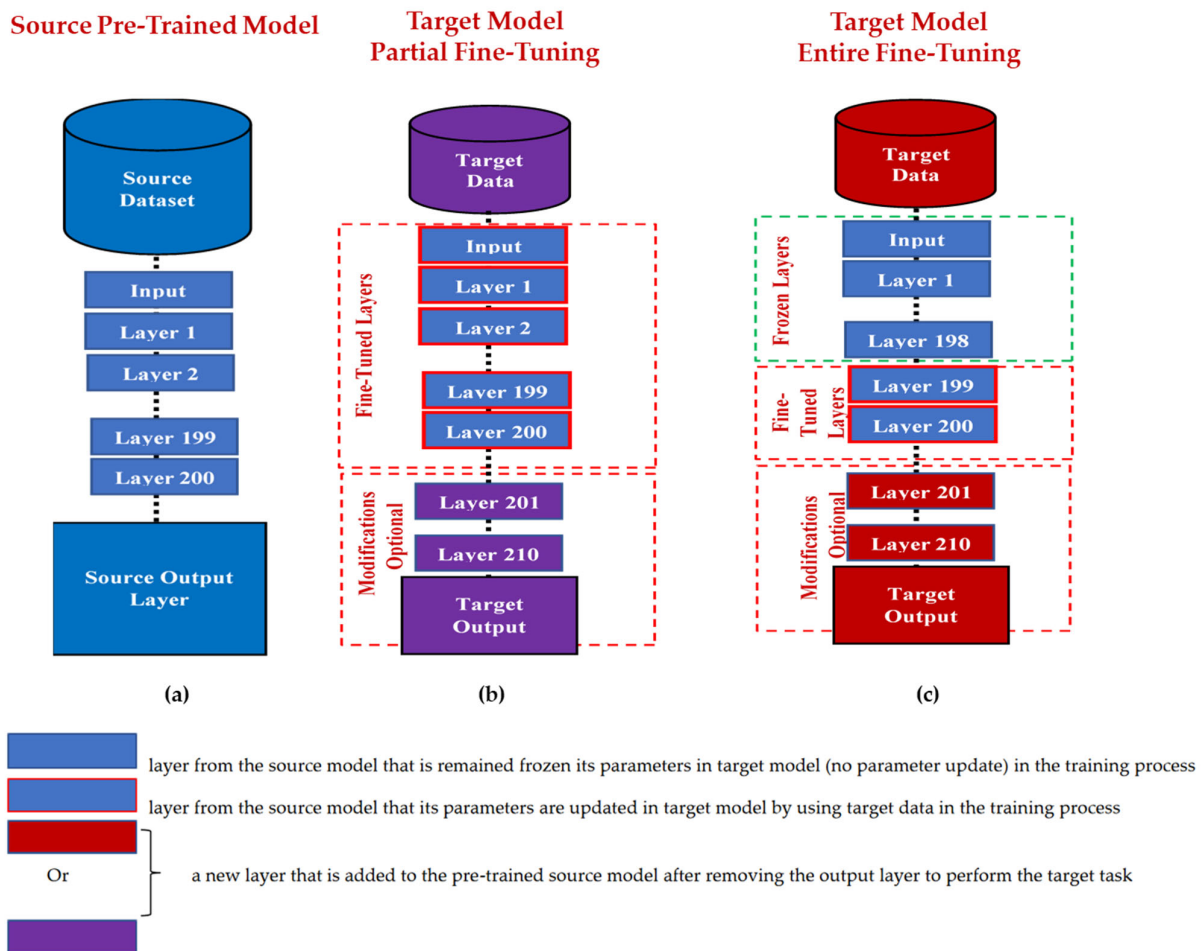
Figure 12. Fine-tuning approaches: (**a**) The source pretrained model is a DL model that consists of 200 layers and developed on big source data to perform a specific task (the source task) by using a specific output layer; (**b**) The entire fine-tuning of the extracted 200 layers from the source (pretrained) model to develop a target task that can either be the same as the source task or different and requiring a modification in the output part (the model base); (**c**) Partially fine-tuning by freezing the first 189 layers, and the fine-tuning of only the last 2 layers (layers 199 and 200) to develop a target task that can either be the same as source task or different and requiring a modification in the output part (the model base).

The following two studies demonstrated the effectiveness of TL in creating accurate sleep-staging systems using a variety of physiological signals. In [38], the authors proposed a TL method to develop an automatic and high-performing sleep-staging system based on long-term scalp EEG recordings. They trained a hybrid DL network of a CNN and RNN, called the CRNN, on large clinical PSG data from over 6000 patients and fine-tuned it on long-term ambulatory scalp EEG recordings. The pretrained and fine-tuned CRNN models achieved a Cohen's kappa close to the expert level (0.75 ± 0.11), with the fine-tuned CRNN increasing the cross-validated Cohen's kappa to 0.78. In [39], the authors used a pretrained CNN model based on derived information from large ECG data to develop a sleep-staging model based on small PPG data recorded from wearable devices. The authors reported improving the accuracy and Cohen's kappa coefficient of the fine-tuned model on the PPG data to 1–9% and 0.01–0.13, respectively, compared to training without TL. The advantages of using fine-tuned TL in wearable devices for digital healthcare can be summarized as follows: First, by leveraging pretrained models, wearable devices can achieve a high accuracy even with limited amounts of individual data. Second, fine-tuning the pretrained model on individual data enables personalized health insights and recommendations,

which can improve the health outcomes and promote patient engagement. Finally, fine-tuning allows for the continuous learning and improvement of the model, making it more robust and adaptable to individual wearers over time.

In Table A4 we have summarized the methodology for human activity recognition based on wearable sensors using the TL fine-tuning method, providing insights on how to address the cross-sensor challenges that arise with sensor variations between the source and target datasets using TL methods [35].

4.1.3. Domain Adaptation

Domain adaptation (DA) is a subfield of TL that focuses on developing effective techniques to address the issue of heterogeneous data distributions [47,48]. Transductive TL methods, including domain adaptation, have demonstrated significant success in addressing domain differences between the source and target domain distributions, thereby eliminating the need for expensive target domain data annotations [48]. The source and target domains are different ($Ds \neq Dt$), but they are related in this approach, as we can adapt one to the other. In digital healthcare, domain adaptation can be particularly useful when working with data from different sources. By adapting the model to the target domain, domain adaptation can improve the model performance and reduce the need for retraining the model on new data. For example, in medical image analysis, domain adaptation can be used to adapt pretrained models on medical images from one hospital to medical images from another hospital with different imaging protocols. Similarly, in wearable and unwearable devices, domain adaptation can be used to adapt pretrained prediction models from two different sources of sensor data that measure related signals [40]. In [47], the authors proposed a novel cross-domain learning framework called stratified transfer learning (STL) to improve the classification accuracy for activity recognition by exploiting the intra-affinity of the classes. STL first obtains pseudo labels for the target domain and performs an iterative intraclass knowledge transfer to transform both domains into the same subspaces. The labels of the target domain are obtained via the second annotation. Comprehensive experiments on three large public activity recognition datasets (all data have different sensors that were mounted on different body locations) show that STL significantly outperformed the other state-of-the-art methods, with an improvement of 7.68% in the classification accuracy. The authors also investigated the performance of STL across different degrees of similarities and activity levels between the domains, and discussed its potential in other pervasive computing applications. In [48], the authors proposed a deep multi-source adversarial domain adaptation (MSADA) framework for human activity recognition from wearable sensor data in heterogeneous environments, where multiple distinct source domains are present. The proposed framework selected the most relevant feature representations from multiple source domains and established mappings to the target domain by learning the perplexity scores. The authors demonstrated that the learned mappings reflected prior knowledge on the semantic relationships between domains, making the MSADA a powerful tool for exploratory activity data analysis; the proposed multisource domain adaptation approach achieved a 2% and a 13% improvement in accuracy on both the OPPORTUNITY dataset and the DSADS dataset, respectively. In [221], the authors proposed a new method called joint probability domain adaptation with a bi-projection matrix algorithm (JPDA-BPM) to overcome the challenge of collecting enough labeled data for emotion recognition based on physiological signals, which is time-consuming and expensive. The proposed method considered the differences in feature distributions between the source and target domains, which improved the algorithm's performance. The authors also proposed a substructure-based joint probability domain adaptation algorithm (SSJPDA) to overcome the effect of physiological signal noise. The proposed algorithm was tested on the DEAP dataset and the results showed that the average recognition accuracy of the proposed SSJPDA-BPM algorithm in the multimodal fusion physiological data from the DEAP dataset was 63.6% and 64.4% in valence and arousal,

respectively. Compared with joint probability domain adaptation (JPDA), the performance of the valence and arousal recognition accuracy increased by 17.6% and 13.4%, respectively.

4.1.4. Multitask Learning

Multitask learning (MTL) is inspired by the way humans can learn multiple related tasks simultaneously, which is often faster and more efficient than focusing on each task separately for extended periods. This parallels how children typically learn to read and write concurrently. Multitask learning is a technique that involves training a single ML model to perform multiple related tasks simultaneously [222,223] rather than training multiple models (one for each specific task), as shown in Figure 13. This approach is an inductive transfer mechanism that aims to improve the generalization performance by utilizing domain-specific information in the training of jointly related tasks [46]. In this method, the transferred knowledge is represented by sharing the feature learning simultaneously among different (but related) tasks, leveraging task similarities to enhance the performance and generalization [224,225]. This method has been used as a powerful tool to reduce computation costs and the need for big training/validation data, as well as to overcome expensive data labeling requirements [222,223]. It does not require transferring knowledge from previously learned tasks (source models) to leverage a new task (the target model), as in fine-tuning method; instead, it focuses on sharing joint feature adaptation among various related tasks during the learning process. In this matter, each task can be considered as a target task, while the other tasks are source tasks that jointly participate in developing each other. Thus, to develop a multitask learning model, there is no need for two separate datasets (the source and the target); a single dataset is sufficient to learn multiple tasks.
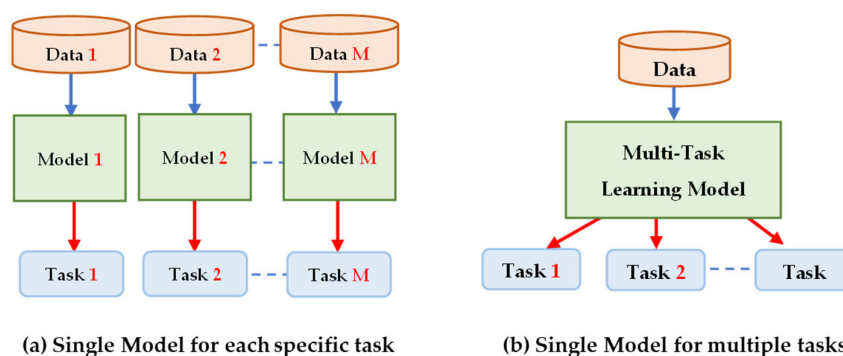

(a) Single Model for each specific task   **(b) Single Model for multiple tasks**

Figure 13. Single task learning vs. multitask learning.

In multitask learning, as in any learning procedure (CL or TL), data can be collected from a multisource or a single source, as shown in Figure 14. In digital healthcare, multitask learning based on multisource data can be particularly useful when working with data from multiple sensors (i.e., wearable and attachable devices) that predict different aspects of a patient's health, such as wearable-based multi sensor data to segment and recognize activities and cycle phases simultaneously [224], or data from a specific single sensor (i.e., a medical imaging scanner, such as an MRI) to provide multiple automatic diagnoses and prognoses (such as simultaneously detecting, segmenting, and classifying different parts of the human spine, including the discs, vertebrae, and neural foramen) using MRI images in categories such as normal, slight, marked, and severe [225]. Two challenges that are taken into consideration for developing a multitask learning model are: a network sharing architecture that focuses on answering the questions *What to share?* and *Where to share?*, and loss-balancing methods that relate to the answer for *How to balance the learning of multitasks?*. For *What to share?* and *Where to share?*, the multitask DL model can be represented in two ways based on the parameter-sharing scheme: hard-parameter sharing and soft-parameter sharing, as shown in Figure 15. The hard-parameter sharing is the most common scheme, generalized around sharing the early hidden layers of the model among

all tasks while using a few different task-specific output layers. So, the low-level features are more general and common, while the high-level features are more task-specific features. This scheme assists in reducing overfitting. For example, four wearable-based inertial sensors placed at the wrists and on the shoes to segment and recognize activities and cycle phases simultaneously using the hard-parameter sharing scheme [224]; the overall achieved F1-score and the phase-detection F1-score were 92.6% and 98.2%, respectively. Additionally, the achieved mean stride time error and swing duration error were 5.3 ± 51.9 ms and $0.0 \pm 5.9\%$, respectively, in the gait analysis. In another study, the hard-parameter sharing multitask learning architecture was developed based on wearable technology to collaboratively learn two correlated tasks, rhythm-classification and signal-quality classification (excellent, acceptable, and noise) [223]. The idea behind jointly learning these two tasks was to overcome the low performance in the rhythm-classification task that occurs due to noisy signals measured by wrist-worn wearable devices. The authors adopted unsupervised transfer learning through the convolutional denoising autoencoder architecture in their classification model to enhance the quality of the noisy signals and improve the performance of atrial fibrillation detection, achieving an F1 score of 0.96 compared to 0.54 when the single task was performed.
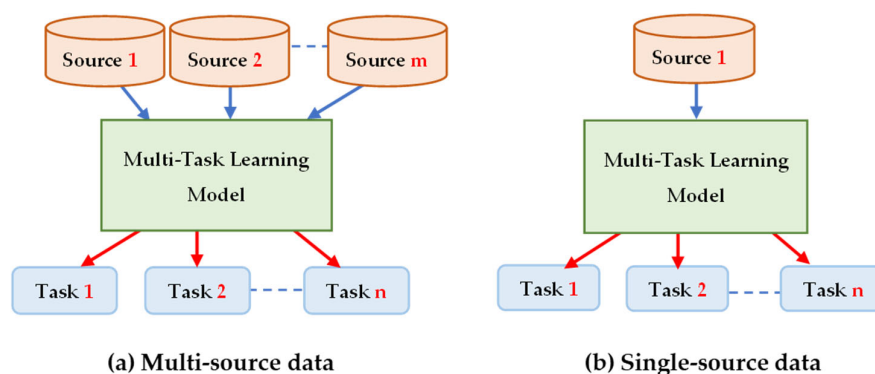

(a) Multi-source data   **(b) Single-source data**

Figure 14. Multitask learning approaches: (**a**) Multisource data can be used to develop multitask learning; for example, data from multiple/different sensors (such as MRI images and X-ray images) can be used to develop a multitask model to identify the samples (scans) that have tumors (task 1), localize and identify the tumor region in the scans (task 2), and classify the tumor into benign or malignant (task 3); (**b**) Single-source data, such as MRI scans, can be used to develop a multitask model to identify the samples (scans) that have tumors (task 1), localize and identify the tumor region in the scans (task 2), and classify the tumor into benign or malignant (task 3).

While, in the soft parameter sharing scheme, each task has its own parameters, this approach focuses on regularizing the distances between these task-specific parameters to encourage the similarity among the related tasks. Although this approach shows superiority in reducing the dissimilarity between different tasks, it suffers from huge memory requirements and expensive computations due to a large number of parameters [226]. Thus, a hybrid approach that combines both soft- and hard-parameter sharing schemes was developed to reduce the computation costs and memory requirements. To combine the benefits of these two approaches and reduce their drawbacks, various schemes of the hybrid approach have been developed, such as the select-or-skip policy learning scheme to choose which layers to be trained, shared, and skipped for each task [226], and the attention module scheme to learn task-related features by applying a soft-attention mask to the features in the shared network [227]. For example, the attention-based U-Net model was modified to implement two brain tumor diagnosis tasks: classification (meningioma, glioma, and pituitary) and segmentation in the MRI images [228]; the authors developed their model based on the attention U-Net, and modified the encoder part to not only develop low-level segmentation features, but also to classify an MRI image into one of the following brain tumor categories: glioma, meningioma, and pituitary; the segmentation

of these brain tumors in the MRI images was performed in the decoder part. This scheme improved both the segmentation performance by achieving a Dice coefficient of 0.74 (a 5% increase) and a Jaccard index of 0.6 (a 9% increase) compared to the U-Net segmentation model, and the classification performance achieved an accuracy of 98.04% (increasing by at least 4%) compared to other pretrained classification models (such as VGG 16, VGG19, and ResNet50).
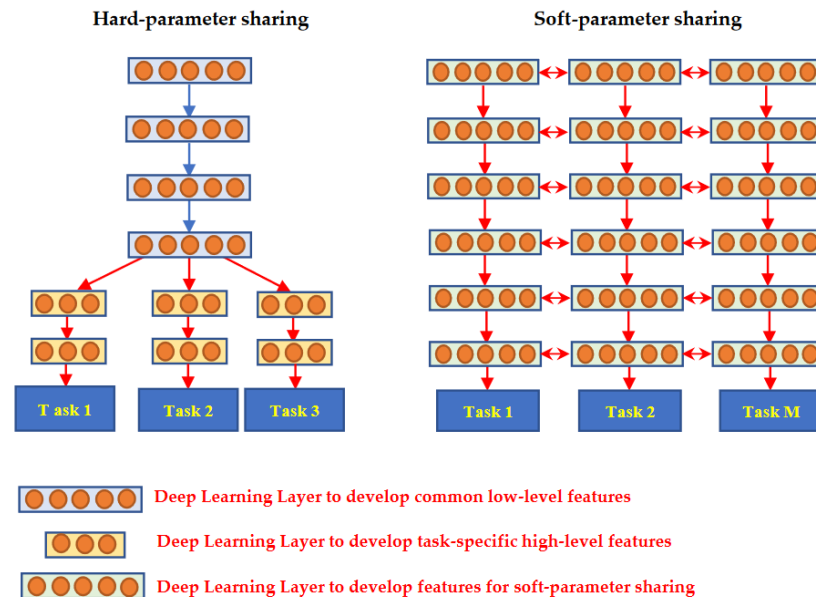

Figure 15. Multitask learning approaches based on the parameter-sharing scheme.

To balance the learning of the multitask (when each task has a specific type of loss), various loss optimization procedures have been developed and applied [222,229]. Let a multitask model contain t tasks. The objective loss (L_{obj}) can be explicated as $L_{obj} = \sum_{i}^{t} \lambda_i L_i$. The hyperparameter λ_i represents the weight assigned to each task-specific loss. In the simplest case, for the *equal weighting method*, $\lambda_i = 1/t$. However, this approach may not be effective when there is significant variation in the difficulty of learning across tasks [222]. Thus, various weighting strategies have been proposed and developed, such as *uncertainty weights* [230], *gradient normalization* [231], the *dynamic weight average* [227], the *projecting conflicting gradient* [232], *impartial multitask learning* [233], and *random loss weighting* [229].

In [45], the authors proposed a multitask multi-kernel based on a logistic regression (MTMKLR) model for emotion recognition. The authors used the multitask learning as an inductive TL approach to improve the generalization by using the domain information contained in the training signals of related tasks as an inductive bias [45,46], and the multi-kernel method to address the challenge of fusing different physiological signals for emotion recognition, as different types of physiological signals may carry different information related to emotions [45]. The authors concluded that, by treating different physiological signals as multiple kernels, the proposed method was able to combine them into a single model that could capture more complex relationships between signals and improve the recognition accuracy. The authors used the multitask method to simultaneously address the challenge of revealing the importance of different signals for recognition. By considering the classification of low/high valence and low/high arousal as multiple tasks, the proposed method was able to learn multiple decision boundaries that corresponded to different emotions and capture the relevance of different signals to different emotions. Therefore, the proposed multitask multi-kernel logistic regression (MTMKLR) approach solves both the problems of fusing different signals and identifying their importance for emotion recognition; this MTMKLR approach increased the accuracy to more than 10% compared to conventional kernel logistic regression.

In Table A5 we have summarized the methodology for implementing brain tumor classification and segmentation in MRI scans jointly in a single model using the multitask method [228], providing insights on how to improve the accuracy of two different but related tasks using limited data using these methods.

4.1.5. Zero-Shot, One-Shot, and Few-Shot Learning

One-shot, few-shot, and zero-shot learning are meta-learning methods. Meta-learning methods are designed around the idea of "*learn how to learn*", and they are related to TL in that they all involve leveraging existing knowledge or structures to improve the generalization performance on new tasks or domains [234–236]. These methods can be considered a specialized form of TL, where the model learns how to learn and adapt across different tasks, rather than just transferring fixed knowledge from one specific task to another. In other words, they focus on enabling a model to quickly adapt to new tasks, which aligns with the concept of TL, where knowledge acquired from one domain or task is applied to improve performance on a related but different domain or task. These methods do not require a vast amount of labeled data to predict a new category; just a few to zero examples (i.e., samples) are sufficient. To train and test a prediction model using few-shot learning, one requires both training and test datasets, each comprising multiple tasks. In the training dataset, each task consists of a labeled support set (training data), used to train the model, and a labeled query set (validation data), used for model validation. The notation 'N-way K-shot' is typically employed in few-shot learning, with 'N' denoting the number of distinct classes within each task and 'K' representing the number of samples (i.e., examples) within each class. During the training stage, it is important to ensure that identical tasks are not used. However, some tasks may share classes with other tasks. For instance, Task 2 in the training set shown in Figure 16 might contain the same 'rectangular' class as Task 1 (though Task 1 and Task 2 are distinct and should not be considered identical). Within the query set, tasks consist of samples that have not been seen before, but they belong to the same classes as those in the support set. The test dataset contains tasks that have not been encountered during training, and the query set within the test dataset is typically unlabeled, requiring predictions from the developed model. Usually, a large dataset containing numerous distinct training tasks is used to develop a pretrained model based on few-shot learning. Subsequently, this pretrained model is fine-tuned on a limited dataset to implement a new few-shot learning task. In the context of this paper, the training dataset could consist of images from a computer-vision classification task for pretrained model development, with the pretrained model then being fine-tuned on medical imaging data to implement a medical image classification task. One-shot learning is a learning paradigm where the model is trained to recognize new objects or categories based on a single example (i.e., $K = 1$) rather than a few examples. This is typically achieved by leveraging the prior knowledge or structure about the underlying data to generalize to new examples. Zero-shot learning is another learning paradigm where the model is trained to recognize new objects or categories that were not present in the training dataset. This is typically achieved by leveraging additional information, such as attributes or relationships, to generalize to new examples. In [54], the authors proposed a method to address the limitations of existing CNNs in semantic segmentation based on medical images due to high manual annotation costs and privacy issues. The authors combined domain adaptation and meta learning based on few-shot learning to adjust an optimization algorithm so that the segmentation model could learn with only a few examples instead of big annotated data. They used an optimization based on the meta-learning method to align the source and target data in a domain-invariant discriminative feature space. They proposed augmenting model–agnostic meta-learning (MAML) and Reptile algorithms (meta-learning benchmarks) to learn from diverse segmentation tasks across the entire task distribution. The proposed method focused on learning from the diversity of image features that characterize a specific tissue type while showing diverse signal intensities. The advantages of the proposed method include an

improved learning capability, avoidance of overfitting, and fast domain adaptation with a small dataset. To examine their proposed method, the authors used Medical Segmentation Decathlon (MSD) data that contained several medical image segmentation tasks; they chose six segmentation tasks randomly as source tasks (the heart/King's College London, the liver/IRCAD, the prostate/Nijmegen Medical Centre, and the pancreas/Memorial Sloan Kettering Cancer Center, spleen/Memorial Sloan Kettering Cancer Center, and colon/Memorial Sloan Kettering Cancer Center), and they involved the remaining two tasks (liver and colon) as target tasks. The authors concluded that their proposed method outperformed existing benchmarks (for the MAML and Reptile, 2% and 2.4% in terms of the Dice similarity coefficient (DSC), respectively), and improved the generalization of the model on tasks with few examples.
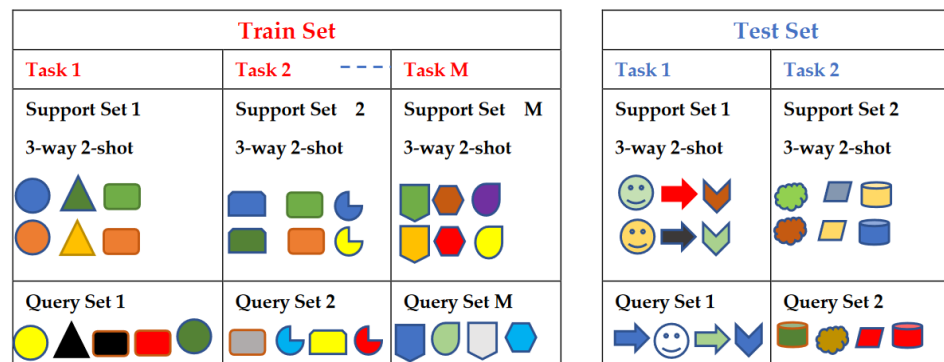

Figure 16. Example of the train and test datasets for the few-shot learning. Notes: (1) Both support and query sets are labeled in the training dataset, and (2) the support set is labeled in the test dataset, but query set is unlabeled in the test dataset and it is used to test the model performance.

In [53], the authors proposed a TL approach based on meta-transfer and few-shot for automatic arrhythmia detection based on wearable device data. The source and target tasks were related, but different, and the domains were the same (ECG data) but from different devices and datasets. The researchers used few-shot to overcome the pretraining in big data for a new class. The proposed method involved transforming the ECG signals into time–frequency spectrograms and using a 2D-CNN model for feature extraction. The feature extractor was pretrained on an auxiliary dataset to improve the efficiency and alleviate the training sample requirements. The proposed meta-transfer scheme was used to recognize the unseen target samples and improve the generalization to new arrhythmia types, even with limited-sized datasets. They used the meta-transfer strategy to leverage the learned knowledge to mitigate overfitting, a challenge commonly faced by traditional few-shot learning methods. They conducted comprehensive experiments to evaluate the effectiveness of their proposed approach (2-way and 4-way with few-shots (1–5)), which outperformed other developed meta learning methods in terms of the accuracy, with improvements of 3–12%, 1–11.5%, and 1.8–4%, in the 1-shot 2-way, the 10-shot 2-way, and the 10-shot 4-way, respectively. In addition, this model achieved competitive accuracies when compared to large-scale training, especially when the 10-shot was applied.

In [52], the authors proposed a novel zero-shot image-retrieval model for medical images using meta-learning and ensemble learning techniques for medical image retrieval to improve the generalizability of the model for new and emerging infectious diseases where historical data were not available. To conduct experiments, they randomly sampled 5% of the images from the NIH Chest X-Ray dataset and created a smaller dataset that contained 5606 images classified into 15 classes. To simulate the situation of new diseases, they randomly selected one disease as the new disease and used the other 14 types of diseases as the training set. The goal of the experiment was to train a retrieval model on the 14 diseases and achieve a good retrieval performance on the new disease without using any data from the new disease. The triplet loss was used to optimize the model

during the training process by decreasing the distance between the image's hash codes of the same class and increasing the distance between the image's hash codes of different classes. The authors used the pretrained model Alexnet to extract the image features and the mean average precision (mAP) based on Hamming ranking as the evaluation metric. The proposed method achieved a 3% to 5% improvement when four distinct hash codes were applied (8 bits, 16 bits, 32 bits, 48 bits) in the retrieval mAP compared to the traditional method (baseline), which can aid doctors in making more accurate diagnoses of new diseases.

4.1.6. Federated Learning

The sharing of sensitive health data in DH is a major concern, as it poses a significant risk to patient privacy and security. To address this problem, federated learning has emerged as a promising approach, in which ML models can be trained on distributed data from various resources while preserving data privacy [237]. Federated learning is an ML technique that allows multiple healthcare centers to collaboratively train a shared model without sharing their data, as shown in Figure 17. Instead, each healthcare center trains a local model on their own data and then shares the model updates with a central server, which aggregates them to create a new, improved model. In this manner, the principles of federated learning align with those of TL. The core of knowledge transfer in federated learning primarily centers around the exchange of the architecture details of the local models among multiple parties; subsequently, these parties adapt and enhance the model parameters using their private data [237]. This methodology not only enhances model accuracy and generalization, but also mitigates the risks associated with centralized data storage. Furthermore, it enables the development of personalized health monitoring and disease detection systems that can be used by individuals in real-time. In this approach of TL, there is no a singular source (pretrained) model to be involved in generating an accurate target (new) model. Instead, the approach involves the aggregation of multiple refined local (source) models to produce a robust global (target) model.
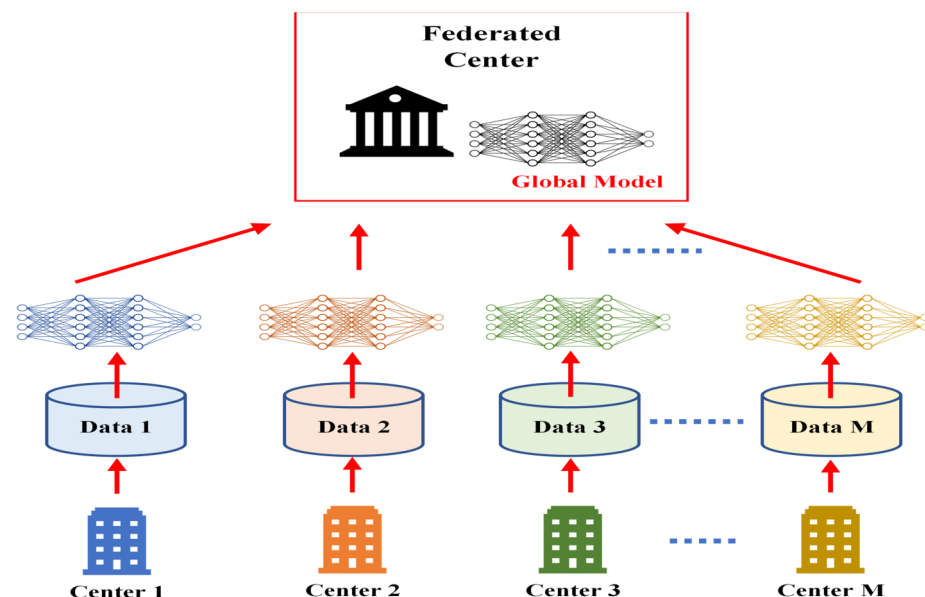

Figure 17. Federate learning architecture. The different colors are assigned for different centers to clarify that each center updated the parameters of the global model with its own data. For example, the first center 1 (blue) used the global model and updated its parameters using its own data (blue), and after finishing, sent the blue model to center 2 (orang); center 2 updated the parameters of the blue model using its own data to generate the orange model, and continued the same process to end with the yellow model as a final developed global model.

In [49], the authors developed a framework called the FedHAR for human activities recognition (HAR) based on multimodality sensor data. The framework was designed to address challenges related to privacy preservation, label scarcity, real-time processing, and heterogeneity. The proposed FedHAR model uses a hierarchical attention architecture for the alignment of different level features and a semi-supervised online learning strategy for online HAR tasks. The proposed FedHAR framework utilizes a semi-supervised online learning strategy that aggregates gradients from both labeled and unlabeled clients. This approach is related to TL, as it involves leveraging information from multiple sources (in this case, both labeled and unlabeled clients) to improve the performance of the model on a given task [49]. The authors concluded that their FedHAR model outperformed the state-of-the-art baselines on two public datasets; in the fine-tuning of each unlabeled client, PerFedHAR achieved an average of a +10% improvement across all metrics on the two datasets. In another study [50], the authors proposed a new framework called FedHealth, which is the first federated TL framework for wearable healthcare devices. FedHealth aggregates data from different organizations while ensuring privacy and security. The framework achieves personalized model learning through knowledge transfer, resulting in excellent performance in smartphone-based human activity recognition. Compared to traditional learning approaches, FedHealth significantly improves the recognition accuracy. The authors believe that FedHealth is an extensible framework that can be used in many healthcare applications. As a proof-of-concept, the authors designed a FedHealth system and applied it in Parkinson's disease auxiliary diagnosis. The results showed that the system achieved good performance while preserving users' privacy in real-world scenarios. In terms of the classification accuracy, FedHealth demonstrated a substantial enhancement of 21.6% and 16.8% in two datasets compared to the non-federated systems. In [51], the authors proposed a method for COVID-19 diagnosis using two-stage federated learning as a TL technique based on convolutional neural networks (CNNs) to classify lung CT scans. The authors utilized the LeNet pretrained model to train a classification model to categorize CT scans into healthy and pneumonia in the first stage, then categorize pneumonia into COVID-19 pneumonia and non-COVID-19 pneumonia in the second stage. In this work, the authors highlighted the two main challenges in medical image classification—the difficulty in acquiring enough samples and privacy concerns. Thus, they used federated learning to ensure privacy by decentralizing the model training on different devices without sharing data, while they utilized TL based on the fine-tuning approach to deal with the limited data. The authors explored the impact of the dataset distribution and training epochs on the model performance. The authors achieved a high performance with their proposed method, attaining a high area under the curve (AUC) of 0.99 for diagnosing COVID-19 while preserving privacy.

While federated learning has achieved notable milestones in digital healthcare, obstacles linked to data heterogeneity and the demanding nature of the system complexity are present [238].

To provide a comprehensive understanding of the above methods, Table 2; Table 3 outline the key features and challenges relevant to each of the TL methods that have been discussed above. This analysis serves as a guide to help researchers and practitioners select the most appropriate TL method for their specific needs, based on the characteristics of their data and the intended use of their prediction system.

Some of the discussed literature in this section mentioned the DL framework they used in developing their research without explaining the specific reasons for their choice; these DL open-source frameworks are: Keras on top of the TensorFlow [38,54,223,224], TensorFlow [36,43,44,51], and PyTorch [53], as shown in Figure 18. We observed that certain references provide open-access code on GitHub, such as [53,226,227]. However, the authors did not mention the availability of the source code in their papers, so we did not cite the GitHub sources, and we did not include them in Figure 18.

Table 2. Characteristics and challenges of the feature extraction, fine-tuning, and domain adaption transfer learning methods.

| Transfer Learning Method	Feature Extraction	Fine-Tuning	Domain Adaptation
Source and Target Domains	Similar	Similar or related	Related but not different
Source and target tasks	Similar/related/different	Similar or related	Similar or related
Model complexity	Low/moderate	Moderate/high	High
Features	1. Cannot develop a final prediction model, but can provide features; thus, it is used as a preprocessing method. 2. Relies on a well-generalized pretrained model to extract high-quality representation from the data. 3. Does not need any annotation information to extract features. 4. Reduces the need for big data.	1. Relies on a well-generalized pretrained model. 2. Good for reducing computation costs if the freezing layers are considered. 3. A good method to improve the performance and generalization when new data are collected, which reduces the need to retrain on both previous and new data.	1. Handles shifts in data distribution, allowing better performance. 2. Reduces data requirements 3. Better generalization.
Challenges	1. May not be able to capture complex relationships between domains and tasks. 2. Needs a second phase focused on modeling to develop a reliable prediction model.	1. May require large amounts of labeled data for the target task. 2. May be overfit to the source task. 3. Limited generalization to new tasks or domains. 4. May require various experiments to come up with the best configuration. 5. May increase computation costs due to the increasing model capacity when the model modification is considered.	1. Assumes similarity between domains and tasks. 2. Lack of data availability or data heterogeneity negatively effects the learning process.

■ **Keras** ■ **Tensorflow** ■ **PyTorch**
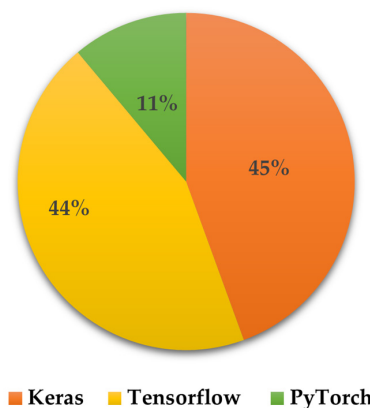
Figure 18. DL open-source frameworks that were used in the studies to develop and test prediction models based on TL for the applications of the DH sensing data.

Table 3. Characteristics and challenges of the multitask, federated, and few-shot, one-shot, and zero-shot transfer learning methods.

Transfer Learning Method	Multitask	Federated	Few-/One-/Zero-Shot
Source and Target Domains	Similar	Multiple distributed Similar	Similar/related/different
Source and target tasks	Multiple and related	Similar	Similar or related
Model complexity	High	High	High
Features	1. Develops multiple related tasks simultaneously. 2. No previously pretrained model is required. 3. Reduces the model development time, as the single model descripts multiple tasks.	1. Maintains data privacy and shares an updated global model. 2. Better generalization, as multi centers or parties collaboratively develop models. 3. Reduces the requirement for large data per institution.	1. Reduces the dependency on large annotated datasets. 2. Quick adaptation to new tasks or concepts, even with very limited data.
Challenges	1. Requires sufficient data to describe all tasks. 2. Increasing complexity when the number of tasks is increased. 3. Requires task relatedness. 4. Hard to generalize and difficult in interpretation.	1. Data heterogeneity. 2. Limited data. 3. Bias and fairness due to data heterogeneity that may lead to negative learning. 4. Complex system architecture that may require secure encryption methods to prevent data disclosure, or the creation of suitable node schedulers to optimize the utilization of distributed computational devices and minimize idle time.	1. May require significant computational resources. 2. May be limited by the size and diversity of the support set. 3. May not generalize well to unseen domain or tasks.

4.2. Advantages and Disadvantages of Transfer Learning

From the explained and discussed applications of TL in DH in the previous section, we can summarize the following advantages of using the TL techniques:

- **Improved performance**: TL can help improve the performance of ML models, especially in cases where the training data are limited.
- **Reduced training time**: TL can reduce the amount of time and resources required to train an ML model, as the pretrained model can provide a starting point for learning.
- **Reduced need for large datasets**: TL can help mitigate the need for large datasets, as the pretrained model can provide a starting point for learning on smaller datasets.
- **Increased generalization**: TL can help improve the generalization of ML models, as the pretrained model has already learned the general features that can be applied to new datasets.
- **Maintain data privacy**: Multiple centers can collaboratively develop a global model without the need to share data to protect data sharing privacy.

Despite the above advantages of TL, challenges are present too. They are summarized as follows:

- **Domain-specific knowledge** [239,240]: TL requires domain-specific knowledge to be effective. For example, if the Ds is image data, while the Dt is sound data, it is obvious

that their features and distributions are dissimilar. Without finding a way to connect these two different domains, TL cannot be feasible.

- **Limited flexibility:** If the source task and target task are different and not related, it may not be easy to adapt the source task to a new task.
- **Risk of negative transfer:** TL can lead to a negative transfer for various reasons: distinct domains, conflicting assumptions, incompatible features, unbalanced transfer (if the source domain dominates the target domain, the model might be overfit to the source domain's characteristics, leading to poor generalization on the target task), and model complexity. Additionally, transferring knowledge from noisy and limited source data cannot lead to positive outcomes in the target data.
- **Limited interpretability:** TL can make it more challenging to interpret the features learned by the model, as they may be influenced by the source model and may not necessarily be relevant to the target domain.

The mentioned disadvantages of TL should not be interpreted as invalidating its effectiveness. These disadvantages highlight the need for careful consideration and understanding of the source and target domains, as well as the potential challenges that might arise during the transfer process. Proper adaptation techniques, domain-alignment methods, and thoughtful model selection can help mitigate these issues and make TL more effective.

5. Conclusions and Future Work

TL has emerged as a powerful technique for improving the performance and efficiency of ML models within the perspective of data quality and size, as well as computation complexity and time development. From the above studies, it has been shown to be effective in addressing several challenges and limitations in deploying ML on DH sensing technologies, such as the availability of sufficient data, variations in the domains, variations and complexities in the tasks, as well as issues related to data privacy and sharing. To overcome the challenge of insufficient data, TL is instrumental in leveraging pretrained models that have already been trained on large amounts of available data. By fine-tuning these models on smaller labeled datasets, TL improves the accuracy and efficiency of ML models, even when limited labeled data is available. Additionally, it can help to adapt models to new domains or datasets, reduce the need for extensive feature engineering by using pretrained models to extract high-level features from the raw data, and reduce the size and computational complexity of models by fine-tuning them on smaller, more targeted datasets. In addition, the meta-TL methods (zero-shot, single-shot, and few-shot) contributed positively to the ability to achieve a high prediction accuracy, even when few- or zero-labeled samples were available. Further, federate learning successfully addresses the data privacy and sharing limitations in medical data by providing the ability to improve the prediction model generalizations and update the model parameters without the need to share data among multiple healthcare centers. In addition, using multiple TL strategies is crucial to improving the performance of a specific problem, as they can address more challenges and limitations.

Although TL offers many potential benefits in DH sensor data, challenges are still present. One major challenge for complex new tasks is the need for large, diverse datasets to train pretrained models to be fine-tuned for specific applications. In some cases, such as with rare diseases or specific patient populations, there may not be enough data available to train effective models. Additionally, transferring models from one domain to another can be difficult due to variations in data distributions and feature representations, which may potentially lead to reduced performance and accuracy in some cases. Another challenge is the need for the interpretability and transparency in TL models used in healthcare, as decisions made by these models can have significant impacts on patient outcomes. Moreover, ethical considerations must be taken into account, such as ensuring that models are not biased or discriminating against certain patient groups. These challenges and limitations need further research to be carefully considered and addressed. We also highlight specific

open research topics in the concept of TL on DH that need further development, improvement, and experimentation to enhance the contributions of TL on the DH domain. These questions are the focus of our ongoing research.

1.   Adaptive Learning for real-time DH sensing Data:

Research Challenge: DH sensing data are continuously generated in real-time, presenting a dynamic and evolving landscape. TL models need to adapt to handle incremental learning, ensuring they stay up to date with the latest data.

Research Direction: How can TL models be developed to effectively adapt to real-time or near-real-time data streams, facilitating continuous learning and timely decision-making?

2.   Enabling TL on Edge Devices (EDs) for timely healthcare applications:

Research Challenge: To enhance in-time diagnosis and personal healthcare monitoring, there is a need to enable TL approaches on edge devices.

Research Direction: How can we simplify the embedding of TL models on portable devices, ensuring the efficient, real-time analysis of DH sensing data? This includes optimizing the model size, energy efficiency, and deployment strategies for edge computing.

Appendix A

In this Appendix, we present tables that summarize the features and applications of common wearable/attachable devices, as well as sensing technologies in DH.

Table A1. Features of common wearable and attachable sensing technologies in digital health.

Technology	Worn/Attached Location	Features and Applications
Smart watches and fitness trackers (**Wearable**)	Wrist, upper arm, waist, and ankle	Track physical activity, heart rate, sleep patterns, and other health metrics.
Smart lenses (**Wearable**)	Head/eye	Embedded with sensors to monitor glucose levels and other parameters in the tears, and send the data to a connected device.
Mouthguards (**Wearable**)	Head/mouth	Monitor various health metrics, such as the heart rate, breathing rate, and oxygen saturation, by measuring changes in the saliva and oral fluids.
Continuous glucose monitoring and insulin pumps (**Wearable/attached to the skin using adhesive patches**)	Abdomen (belly), upper buttocks, upper arm, and thigh	Help people with diabetes manage their blood sugar levels by continuously monitoring glucose levels and automatically delivering insulin as needed.
Headbands and hats (**Wearable**)	Head/around the forehead, and over the ears	Measure the brain activity, heart rate, and other vital signs.
Chest straps and attached bands (**Wearable**)	Around the chest or the wrist	Measure the heart rate variability, respiratory rate, and other health metrics by sensing changes in the skin conductivity or other physiological parameters used in the field of fitness and wellness.

Table A1. *Cont.*

Technology	Worn/Attached Location	Features and Applications
ECG patches (**Attached to the skin using a medical-grade adhesive**)	Chest, upper back, or upper arm	It can monitor the heart rate, rhythm, and other cardiac metrics, and are often used to diagnose arrhythmias and other heart conditions.
Blood pressure cuffs (**Wearable**)	Wrist or upper arm	Measure blood pressure and can help diagnose and manage hypertension and other cardiovascular conditions. Some of these devices contain memory to store measurements and send information wirelessly to healthcare providers.
Smart clothing shirts, pants, and sports bras (**Wearable**)	Human body	Embedded with sensors to monitor vital signs, physical activity, and other health parameters.
Wearable cameras		Capture images and video of a patient's environment, which can be used for telemedicine and remote monitoring purposes.
Smart jewelry rings, bracelets, and necklaces (**Wearable**)	Different body locations	Equipped with sensors to track various health metrics.
Smart shoes (**Wearable**)	Foot	Detect gait patterns, track steps, and monitor posture.
Skin patches (**Attachable**)	Skin	Attached to the skin to monitor various physiological parameters, such as the heart rate, blood glucose levels, temperature, and hydration.
Smart helmets (**Wearable**)	Head	Enhance safety and provide connectivity. Equipped with sensors to monitor head impact forces and detect signs of concussion in athletes.

Table A2. Sensor technologies in digital health that have been used with ML/DL techniques.

Sensor (Type of Data)	Features	Applications
Electrocardiogram (ECG) (**Time series**)	Measures electrical signals of the heart over time	Detecting arrhythmias and predicting heart disease [144,145]
Blood glucose monitoring (**Time series**)	Measures glucose levels over time	Predicting blood glucose levels [146,147]
Pulse oximeter (**Time series**)	Measures oxygen saturation and heart rate over time	Monitoring patients with respiratory (chronic obstructive pulmonary disease [148], COVID-19 [149], cardiac conditions [150])
Electroencephalogram (EEG) (**Time series**)	Measures electrical activity in the brain over time	Predicting epilepsy [151], seizure risk [152], and diagnosing neurological disorders [153,154]
Accelerometer (**Time series**)	Measures movement over time	Monitoring physical activity [35,37,49] and predicting falls [155,156]
Blood pressure monitor (MEMS) (**Time series**)	Measures blood pressure over time	Cardiovascular monitoring [157]
X-ray (**Image**)	Images of internal structures, such as bones or organs	Diagnosing internal injuries or diseases (i.e., coronavirus [158,159], heart diseases [160])

Table A2. *Cont.*

Sensor (Type of Data)	Features	Applications
MRI (**Image**)	Images of internal structures, such as the brain or joints	Diagnosing internal injuries or diseases (i.e., cardiovascular diseases [161], cancers [162,163], knee injuries [164])
CT scan (**Image**)	Images of internal structures, such as the brain or abdomen	Diagnosing internal injuries and diseases (i.e., cancers [165,166], cerebral aneurysm [167], lung diseases [168], and brain injuries [169])
Ultrasound (**Image**)	Images of internal structures, such as the fetus or organs	Diagnosing internal injuries or diseases (i.e., carpal tunnel [170], liver diseases [171,172], and kidney injuries [173])
Spirometer (**Time series**)	Measures lung function, including volume and flow rates	Predicting respiratory disease progression and monitoring the response to treatment [174,175]
Photoplethysmography (PPG) (**Time series**)	Measures various physiological parameters (heart rate, blood oxygen saturation, blood pressure, glucose levels, and emotional state)	Predicting glucose levels in patients with diabetes [176,177] and monitoring the emotional state or stress levels [178,179]
Electro-oculogram (EOG) (**Time series**)	Measures electrical signals from eye muscles and movements	Monitoring sleep patterns [180,181] and predicting eye disorder [182]
Infrared thermometer (**Time series**)	Measures body temperature from a distance	Monitoring patients with fever or hypothermia [183,184]
Optical coherence tomography (OCT) (**Image**)	Images of internal structures, such as the retina or cornea	Diagnosing eye diseases [185,186]
Capsule endoscope (**Video**)	Images and videos of the digestive tract	Diagnosing gastrointestinal disorders [187,188]
Acoustic (**Video**)	Measures acoustic features of the voice, e.g., the pitch, volume, and tone	Predicting Parkinson's disease [189,190], diagnosing voice disorders [191], and detecting cardiac diseases [192]
Electrodermal activity sensor (EDA) (**Time series**)	Measures the electrical activity of sweat glands	Predicting emotional or psychological states [193,194] and monitoring stress levels [195]
Magnetometer (**Time series**)	Measures magnetic fields in the body	Monitoring cardiac function [196] and detecting locomotion and daily activities [197]
Photoacoustic imaging (**Image**)	Combines optical and ultrasound imaging for high-resolution images	Diagnosing cancer [198,199] and brain diseases [200]
Smart clothing (**Time series**)	Monitors vital signs and activity levels through sensors woven into clothing	Monitoring sleep [85], human motion [88], and detecting cardiovascular diseases [86].
Pulse oximeter (**Time series**)	Measures oxygen saturation in the blood through a sensor on a finger or earlobe	Monitoring patients with respiratory [201] or cardiac conditions [202]
Multi-sensors (Multimodal signals data) (**Time series**)	Selects a few data features for better performance and higher accuracy	Multitask emotion recognition (valence, arousal, dominance, and liking) after watching videos [45]
Multi-sensors (Multimodal imaging data) (**Image**)	Provides information about tissues and internal organs, and functional information about metapolicy activities	Early detections of COVID-19 to assign appropriate treatment plans [203]

Appendix B

In this Appendix, we describe data characteristics and methodologies for some examples that we illustrated in Section 4.1.

Table A3. Methodologies to implement feature extraction TL for automatically patients' X-ray images classification into one of three categories: COVID-19, normal, and pneumonia to assist in the diagnosis COVID-19 [41].

Task, Goal, and ML/DL Software to Develop the Model	Data Characteristics	Development Procedure	Achievements
Task: Automatically classify patients' X-ray images into one of three categories: COVID-19, normal, and pneumonia. **Goal:** Overcome the difficulty in selecting the optimal engineering features to develop a reliable prediction model. Reduce the high dimensionality of the raw data, and improve its meaning. **Software:** Not specified.	**Source:** COVID-19 radiography database (open access provided by Kaggle). **This database consists of 4 datasets:** • COVID-19 Database: The Italian Medical and Interventional Radiology Society. • Novel Corona Virus 2019 Dataset: Cohen Morrison and Dao on GitHub. • COVID-19 (+) Chest X-ray Images collected from various scientific articles. • Kaggle chest X-ray database. **Number of samples**: 219 X-ray images from patients with COVID-19, 1341 normal individuals, and 1345 pneumonia patients.	• Extract features from X-ray data by using the three state-of-the-art pretrained models (ResNet50, ResNet101, and InceptionResNetV2). • Combine the three sets of deep features. • Apply the feature selection method using particle swarm optimization and the ant colony algorithm to select the optimal features. • Feed the selected features to either the SVM or KNN classifier for model training. • Compare the results of the various configurations, such as models without feature selection methods using the SVM classifier, models with feature selection methods using the SVM classifier, models without feature selection methods using the KNN classifier, or models with feature selection methods using the KNN classifier	• They found the SVM classifier achieved better performance in terms of the F1 score and accuracy compared with the KNN classifier in all four configurations: only ResNet50 features, only ResNet101 features, only IncResNetV2 features, and combined features (presented in Tables 4 and 5 in the study [41]). • The combined features achieved the best performance in both the SVM and KNN classifiers. In addition, the feature selection methods improved slightly the performance of the SVM classifier, while they slightly reduced the performance of the KNN classifier in the combined features configuration (see Tables 4–10 in the study [41]). • The highest performance was achieved by the configuration (combined features, particle swarm optimization, and linear SVM), with an accuracy of 99.86%

Table A4. Methodologies to implement the fine-tuning TL for human activity recognition based on wearable sensors [35].

Task, Goal, and ML/DL Software to Develop the Model	Data Characteristics	Development Procedure	Achievements
Task: Classify human activities based on smartphone sensor data. **Goal**: Use fine-tuning to speed up training processes, overcome overfitting, and achieve a high classification accuracy in a new target task. **Software:** Not specified.	**Source**: Two state-of-the-art datasets were used: the "Khulna University Human Activity Recognition (KU-HAR)" and "the University of California Irvine Human Activities and Postural Transitions (UCI-HAPT)" **Data usage**: • The KU-HAR (source dataset) to develop the pretraining model and to select the best model. • The UCI-HAPT (target dataset) to test and fine-tune the pretrained model on a new task. **KU-HAR dataset (2021):** Contains 20,750 samples of 18 different activities (stand, sit, talk–sit, talk–stand, stand–sit, lay, lay–stand, pick, jump, push-up, sit-up, walk, walk backwards, walk–circle, run, stair–up, stair–down, and table tennis). Each sample lasted 3 s. The data were collected using a smartphone's accelerometer and gyroscope sensors, worn at the waist. The data were gathered from 90 people aged 18 to 34. The data were not cleaned or filtered in any way so to consider a realistic dataset of real-world conditions. The dataset is unbalanced, and no data samples overlap with each other. **UCI-HAPT (2014):** Contains 10,929 samples collected from 30 volunteers (aged 19–48) using a waist-mounted smartphone triaxial accelerometer and gyroscope at the sampling rate of 50 Hz. It contains 12 activities (walking, walking upstairs, walking downstairs, sitting, standing, laying, stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, and lie-to-stand).	• Extract frequency- and time-domain features (scalograms for the visual representation of a signal) from the KU-HAR data samples (i.e., the signals) using the continuous wavelet transform. • To achieve the best performance, use various scalogram representations to train various pretrained models (DenseNet121, DenseNet169, DenseNet201, ResNet50, ResNet101, ResNet152, Xception, InceptionV3, and InceptionResNetV2). • Choose the best model based on the highest achieved accuracy. • To achieve an identical sample form of the KU-HAR data, preprocess the UCI-HAPT data (target data) by increasing the sampling rate of the raw data from 50 HZ to 100 Hz, and then use a nonoverlapping 3 s windowing technique to sample the data. This step produced 4847 six-channel time-domain samples. • Select 30% of the data samples randomly for validation, and the remaining samples can be used to train and fine-tune the model. • Extract the time- and frequency-domain features (scalograms) from the training samples. • Feed these training samples to the best pretrained model. • Apply partial fine-tuning on the UCI-HAPT dataset by freezing the layers close to the input and unfreezing the layers that are close to the output. To obtain best performance, the authors gradually unfroze the output layers and fine-tuned the model. This step was important to reduce the computations and overcome overfitting.	• Partial fine-tuning of the DenseNet121 architecture using Morlet wavelet achieved the best performance on the source KU-HAR dataset (classification accuracy of 97.48% and an F1 score of 97.52%), outperforming the state-of-the-art achievements. • Freezing the first 308 layers of the pretrained model resulted in faster training and a smoother gradient descent on a small dataset. • This model also achieved improvements in both the accuracy and F1 score by 0.21% and 0.33%, respectively, on entire the UCI-HAPT target dataset.

Table A5. Methodologies to implement multitask TL brain tumor classification and segmentation in MRI scans [228].

Task, Goal, and ML/DL Software to Develop the Model	Data Characteristics	Development Procedure	Achievements
Task: Automatically classify patients' MRI scans into one of three brain tumors: meningioma, glioma, and pituitary tumors, and segment the tumor regions from the MRI scans. **Goal**: Reduce development processes and improve the performance by jointly training two distinct but related tasks. **Software**: Not specified.	**Source**: Figshare MRI dataset. **This dataset consists of:** 3064 2D T1-weighted contrast-enhanced modalities (coronal, axial, and sagittal) collected from 233 patients. The classification and segmentation labels are included. **Datadistribution**: • 23% samples of meningioma (708 slices) • 46.5% samples of glioma (1426 slices) • 30% samples of pituitary (930 slices) • 80% samples for training and 20% for validation	• Develop a multitask DL classification and segmentation model based on the modified U-Net to successfully predict two distinct diagnosis tasks, but related simultaneously. The state-of-the-art DL segmentation architecture for medical imaging data analysis, called the U-Net, was modified by the authors by adding a classification layer to the end of the encoder branch to implement the classification task along with the segmentation task. • Feed the training samples to the developed mode. • Train and validate the model to justify its effectiveness.	• The authors developed a multitask DL classification and segmentation model based on the modified U-Net, called the attention-guided encoder–decoder network (MAG-Net). The authors added a classification layer to the end of the encoder part of the U0-net segmentation model to perform the classification task in addition to the segmentation. • Using multitask model improved both the segmentation and classification results. • The segmentation task in the multitask model achieved a Dice coefficient of 0.74 (a 5% increase) and a Jaccard index of 0.6 (a 9% increase) compared to a U-Net segmentation model. • The classification task in the multitask model achieved an accuracy of 98.04% (increasing by at least 4%) compared to the other state-of-the-art pretrained classification models, such as VGG 16, VGG19, and ResNet50.

References

1. Gentili, A.; Failla, G.; Melnyk, A.; Puleo, V.; Tanna, G.L.D.; Ricciardi, W.; Cascini, F. The Cost-Effectiveness of Digital Health Interventions: A Systematic Review of The Literature. *Front. Public Health* **2022**, *10*, 787135. [CrossRef] [PubMed]
2. Georgiou, A.; Li, J.; Hardie, R.A.; Wabe, N.; Horvath, A.R.; Post, J.J.; Eigenstetter, A.; Lindeman, R.; Lam, Q.; Badrick, T.; et al. Diagnostic Informatics-The Role of Digital Health in Diagnostic Stewardship and the Achievement of Excellence, Safety, and Value. *Front. Digit. Health* **2021**, *3*, 659652. [CrossRef]
3. Jagadeeswari, V.; Subramaniyaswamy, V.; Logesh, R.; Vijayakumar, V. A Study on Medical Internet of Things and Big Data in Personalized Healthcare System. *Health Inf. Sci. Syst.* **2018**, *6*, 14. [CrossRef] [PubMed]
4. Mbunge, E.; Muchemwa, B.; Jiyane, S.; Batani, J. Sensors and Healthcare 5.0: Transformative Shift in Virtual Care Through Emerging Digital Health Technologies. *Glob. Health J.* **2021**, *5*, 169–177. [CrossRef]
5. Butcher, C.J.; Hussain, W. Digital Healthcare: The Future. *Future Healthc. J.* **2022**, *9*, 113–117. [CrossRef] [PubMed]
6. Liao, Y.; Thompson, C.A.; Peterson, S.K.; Mandrola, J.; Beg, M.S. The Future of Wearable Technologies and Remote Monitoring in Health Care. *Am. Soc. Clin. Oncol. Educ. Book* **2019**, *39*, 115–121. [CrossRef] [PubMed]

7. Vesnic-Alujevic, L.; Breitegger, M.; Guimarães Pereira, Â. 'Do-It-Yourself' Healthcare? Quality of Health and Healthcare Through Wearable Sensors. *Sci. Eng. Ethics* **2018**, *24*, 887–904. [CrossRef]

8. Anikwe, C.V.; Nweke, H.F.; Ikegwu, A.C.; Egwuonwu, C.A.; Onu, F.U.; Alo, U.R.; Teh, Y.W. Mobile and Wearable Sensors for Data-Driven Health Monitoring System: State-of-The-Art and Future Prospect. *Expert Syst. Appl.* **2022**, *202*, 117362. [CrossRef]

9. Verma, P.S.; Sood, S.K. Cloud-Centric IoT based Disease Diagnosis Healthcare Framework. *J. Parallel Distrib. Comput.* **2018**, *116*, 27–38. [CrossRef]

10. Rutledge, C.M.; Kott, K.; Schweickert, P.A.; Poston, R.; Fowler, C.; Haney, T.S. Telehealth and eHealth in Nurse Practitioner Training: Current Perspectives. *Adv. Med. Educ. Pract.* **2017**, *8*, 399–409. [CrossRef]

11. Guk, K.; Han, G.; Lim, J.; Jeong, K.; Kang, T.; Lim, E.K.; Jung, J. Evolution of Wearable Devices with Real-Time Disease Monitoring for Personalized Healthcare. *Nanomaterials* **2019**, *9*, 813. [CrossRef] [PubMed]

12. Kim, J.; Imani, S.; de Araujo, W.R.; Warchall, J.; Valdés-Ramírez, G.; Paixão, T.R.L.C.; Mercier, P.P.; Wang, J. Wearable Salivary Uric Acid Mouthguard Biosensor with Integrated Wireless Electronics. *Biosens. Bioelectron.* **2015**, *74*, 1061–1068. [CrossRef] [PubMed]

13. Kim, J.; Campbell, A.S.; Wang, J. Wearable Non-Invasive Epidermal Glucose Sensors: A Review. *Talanta* **2018**, *177*, 163–170. [CrossRef] [PubMed]

14. Viswanath, B.; Choi, C.S.; Lee, K.; Kim, S. Recent Trends in the Development of Diagnostic Tools for Diabetes Mellitus using Patient Saliva. *Trends Anal. Chem.* **2017**, *89*, 60–67. [CrossRef]

15. Arakawa, T.; Tomoto, K.; Nitta, H.; Toma, K.; Takeuchi, S.; Sekita, T.; Minakuchi, S.; Mitsubayashi, K. A Wearable Cellulose Acetate-Coated Mouthguard Biosensor for In Vivo Salivary Glucose Measurement. *Anal. Chem.* **2020**, *92*, 12201–12207. [CrossRef] [PubMed]

16. Zheng, Y.L.; Ding, X.R.; Poon, C.C.Y.; Lo, B.P.L.; Zhang, H.; Zhou, X.L.; Yang, G.Z.; Zhao, N.; Zhang, Y.T. Unobtrusive Sensing and Wearable Devices for Health Informatics. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1538–1554. [CrossRef] [PubMed]

17. Sempionatto, J.R.; Brazaca, L.C.; García-Carmona, L.; Bolat, G.; Campbell, A.S.; Martin, A.; Tang, G.; Shah, R.; Mishra, R.K.; Kim, J.; et al. Eyeglasses-Based Tear Biosensing System: Non-Invasive Detection of Alcohol, Vitamins and Glucose. *Biosens. Bioelectron.* **2019**, *137*, 161–170. [CrossRef]

18. Constant, N.; Douglas-Prawl, O.; Johnson, S.; Mankodiya, K. Pulse-Glasses: An Unobtrusive, Wearable HR Monitor with Internet-of-Things Functionality. In Proceedings of the 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Cambridge, MA, USA, 9–12 June 2015; pp. 1–5. [CrossRef]

19. Liu, X.; Lillehoj, P.B. Embroidered Electrochemical Sensors for Biomolecular Detection. *Lab A Chip* **2016**, *16*, 2093–2098. [CrossRef]

20. Jung, P.G.; Lim, G.; Kong, K. A Mobile Motion Capture System based on Inertial Sensors and Smart Shoes. *J. Dyn. Syst. Meas. Control* **2014**, *136*, 011002. [CrossRef]

21. Kim, J.H.; Roberge, R.; Powell, J.B.; Shafer, A.B.; Williams, W.J. Measurement Accuracy of Heart Rate and Respiratory Rate During Graded Exercise and Sustained Exercise in the Heat Using the Zephyr BioHarnessTM. *Int. J. Sports Med.* **2013**, *34*, 497–501. [CrossRef]

22. Mihai, D.A.; Stefan, D.S.; Stegaru, D.; Bernea, G.E.; Vacaroiu, I.A.; Papacocea, T.; Lupușoru, M.O.D.; Nica, A.E.; Stiru, O.; Dragos, D.; et al. Continuous Glucose Monitoring Devices: A Brief Presentation (Review). *Exp. Ther. Med.* **2022**, *23*, 174. [CrossRef] [PubMed]

23. Gamessa, T.W.; Suman, D.; Tadesse, Z.K. Blood Glucose Monitoring Techniques: Recent Advances, Challenges and Future Perspectives. *Int. J. Adv. Technol. Eng. Explor.* **2018**, *5*, 335–344. [CrossRef]

24. Javaid, M.; Haleem, A.; Singh, R.P.; Suman, R.; Rab, S. Significance of Machine Learning in Healthcare: Features, Pillars and Applications. *Int. J. Intell. Netw.* **2022**, *3*, 58–73. [CrossRef]

25. Qayyum, A.; Qadir, J.; Bilal, M.; Al-Fuqaha, A. Secure and Robust Machine Learning for Healthcare: A Survey. *IEEE Rev. Biomed. Eng.* **2021**, *14*, 156–180. [CrossRef]

26. Islam, M.S.; Hasan, M.M.; Wang, X.; Germack, H.D.; Noor-E-Alam, M. A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. *Healthcare* **2018**, *6*, 54. [CrossRef] [PubMed]

27. Bohr, A.; Memarzadeh, K. The Rise of Artificial Intelligence in Healthcare Applications. *Artif. Intell. Healthc.* **2020**, 25–60. [CrossRef]

28. Love-Koh, J.; Peel, A.; Rejon-Parrilla, J.C.; Ennis, K.; Lovett, R.; Manca, A.; Taylor, M. The Future of Precision Medicine: Potential Impacts for Health Technology Assessment. *Pharmacoeconomics* **2018**, *36*, 1439–1451. [CrossRef] [PubMed]

29. Berger, J.S.; Haskell, L.; Ting, W.; Lurie, F.; Chang, S.C.; Mueller, L.A.; Elder, K.; Rich, K.; Crivera, C.; Schein, J.R.; et al. Evaluation of Machine Learning Methodology for the Prediction of Healthcare Resource Utilization and Healthcare Costs in Patients With Critical Limb Ischemia-Is Preventive and Personalized Approach on the Horizon? *EPMA J.* **2020**, *11*, 53–64. [CrossRef]

30. Kelly, C.J.; Karthikesalingam, A.; Suleyman, M.; Corrado, G.; King, D. Key Challenges for Delivering Clinical Impact with Artificial Intelligence. *BMC Med.* **2019**, *17*, 195. [CrossRef]

31. Ellis, R.J.; Sander, R.; Limon, A. Twelve Key Challenges in Medical Machine Learning and Solutions. *Intell.-Based Med.* **2022**, *6*, 100068. [CrossRef]

32. Hosna, A.; Merry, E.; Gyalmo, J.; Alom, Z.; Aung, Z.; Azim, M.A. Transfer Learning: A Friendly Introduction. *J. Big Data* **2022**, *9*, 102. [CrossRef] [PubMed]

33. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]

34. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [CrossRef]

35. Pavliuk, O.; Mishchuk, M.; Strauss, C. Transfer Learning Approach for Human Activity Recognition Based on Continuous Wavelet Transform. *Algorithms* **2023**, *16*, 77. [CrossRef]

36. An, S.; Bhat, G.; Gümüşsoy, S.; Ogras, Ü.Y. Transfer Learning for Human Activity Recognition Using Representational Analysis of Neural Networks. *ACM Trans. Comput. Healthc.* **2023**, *4*, 1–21. [CrossRef]

37. Waters, S.H.; Clifford, G.D. Comparison of Deep Transfer Learning Algorithms and Transferability Measures for Wearable Sleep Staging. *Biomed. Eng. Online* **2022**, *21*, 66. [CrossRef] [PubMed]

38. Abou Jaoude, M.; Sun, H.; Pellerin, K.R.; Pavlova, M.; Sarkis, R.A.; Cash, S.S.; Westover, M.B.; Lam, A.D. Expert-Level Automated Sleep Staging of Long-Term Scalp Electroencephalography Recordings using Deep Learning. *Sleep* **2020**, *43*, zsaa112. [CrossRef]

39. Li, Q.; Li, Q.; Cakmak, A.S.; Da Poian, G.; Bliwise, D.L.; Vaccarino, V.; Shah, A.J.; Clifford, G.D. Transfer learning from ECG to PPG for improved sleep staging from wrist-worn wearables. *Physiol. Meas.* **2021**, *42*, 044004. [CrossRef]

40. Radha, M.; Fonseca, P.; Moreau, A.; Ross, M.; Cerny, A.; Anderer, P.; Long, X.; Aarts, R.M. A Deep Transfer Learning Approach for Wearable Sleep Stage Classification with Photoplethysmography. *npj Digit. Med.* **2021**, *4*, 135. [CrossRef]

41. Narin, A. Accurate detection of COVID-19 Using Deep Features based on X-Ray Images and Feature Selection Methods. *Comput. Biol. Med.* **2021**, *137*, 104771. [CrossRef]

42. Salem, M.A.; Taheri, S.; Yuan, J. ECG Arrhythmia Classification Using Transfer Learning from 2-Dimensional Deep CNN Features. In Proceedings of the 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), Cleveland, OH, USA, 17–19 October 2018; pp. 1–4.

43. Vo, D.M.; Nguyen, N.Q.; Lee, S.W. Classification of Breast Cancer Histology Images using Incremental Boosting Convolution Networks. *Inf. Sci.* **2019**, *482*, 123–138. [CrossRef]

44. Thuy, M.B.H.; Hoang, V.T. Fusing of Deep Learning, Transfer Learning and GAN for Breast Cancer Histopathological Image Classification. In *Advanced Computational Methods for Knowledge Engineering: Proceedings of the 6th International Conference on Computer Science, Applied Mathematics and Applications, ICCSAMA 2019*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; Volume 6, pp. 255–266. [CrossRef]

45. He, X.; Huang, J.; Zeng, Z. Logistic Regression Based Multi-task, Multi-kernel Learning for Emotion Recognition. In Proceedings of the 2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM), Chongqing, China, 3–5 July 2021; pp. 572–577. [CrossRef]

46. Caruana, R. Multitask Learning. In *Learning to Learn*; Thrun, S., Pratt, L., Eds.; Springer: Boston, MA, USA, 1998. [CrossRef]

47. Wang, J.; Chen, Y.; Hu, L.; Peng, X.; Philip, S.Y. Stratified Transfer Learning for Cross-Domain Activity Recognition. In Proceedings of the 2018 IEEE international conference on pervasive computing and communications (PerCom), Athens, Greece, 19–23 March 2018; pp. 1–10. [CrossRef]

48. Chakma, A.; Faridee, A.Z.M.; Khan, M.A.A.H.; Roy, N. Activity Recognition in Wearables Using Adversarial Multi-Source Domain Adaptation. *Smart Health* **2021**, *19*, 100174. [CrossRef]

49. Presotto, R.; Civitarese, G.; Bettini, C. Semi-Supervised and Personalized Federated Activity Recognition based on Active Learning and Label Propagation. *Pers. Ubiquitous Comput.* **2022**, *26*, 1281–1298. [CrossRef]

50. Chen, Y.; Qin, X.; Wang, J.; Yu, C.; Gao, W. FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare. *IEEE Intell. Syst.* **2020**, *35*, 83–93. [CrossRef]

51. Zhang, A.S.; Li, N.F. When Accuracy Meets Privacy: Two-Stage Federated Transfer Learning Framework in Classification of Medical Images on Limited Data: A COVID-19 Case Study. *arXiv* **2022**, arXiv:2203.12803. [CrossRef]

52. Mishra, N.; Rohaninejad, M.; Chen, X.; Abbeel, P. A Simple Neural Attentive Meta-Learner. *arXiv* **2018**, arXiv:1707.03141.

53. Liu, T.; Yang, Y.; Fan, W.; Wu, C. Reprint of: Few-Shot Learning for Cardiac Arrhythmia Detection based on Electrocardiogram Data from Wearable Devices. *Digit. Signal Process.* **2022**, *125*, 103574. [CrossRef]

54. Zhang, P.; Li, J.; Wang, Y.; Pan, J. Domain Adaptation for Medical Image Segmentation: A Meta-Learning Method. *J. Imaging* **2021**, *7*, 31. [CrossRef]

55. Xu, S.; Kim, J.; Walter, J.R.; Ghaffari, R.; Rogers, J.A. Translational Gaps and Opportunities for Medical Wearables in Digital Health. *Sci. Transl. Med.* **2022**, *14*, eabn6036. [CrossRef]

56. Ullah, F.; Haq, H.; Khan, J.; Safeer, A.A.; Asif, U.; Lee, S. Wearable IoTs and Geo-Fencing Based Framework for COVID-19 Remote Patient Health Monitoring and Quarantine Management to Control the Pandemic. *Electronics* **2021**, *10*, 2035. [CrossRef]

57. Sardar, A.W.; Ullah, F.; Bacha, J.; Khan, J.; Ali, F.; Lee, S. Mobile Sensors based Platform of Human Physical Activities Recognition for COVID-19 spread minimization. *Comput. Biol. Med.* **2022**, *146*, 105662. [CrossRef] [PubMed]

58. Papini, G.B.; Fonseca, P.; Van Gilst, M.M.; Bergmans, J.W.M.; Vullings, R.; Overeem, S. Wearable Monitoring of Sleep-Disordered Breathing: Estimation of the Apnea–Hypopnea index using Wrist-Worn Reflective Photoplethysmography. *Sci. Rep.* **2020**, *10*, 13512. [CrossRef] [PubMed]

59. Wang, X.; Ellul, J.; Azzopardi, G. Elderly Fall Detection Systems: A Literature Survey. *Front. Robot. AI* **2020**, *7*, 71. [CrossRef] [PubMed]

60. Thilo, F.J.S.; Hahn, S.; Halfens, R.J.G.; Schols, J.M.G.A. Usability of a Wearable Fall Detection Prototype from the Perspective of Older People–A Real Field Testing Approach. *J. Clin. Nurs.* **2018**, *28*, 310–320. [CrossRef] [PubMed]

61. Konstantinidis, D.; Iliakis, P.; Tatakis, F.; Thomopoulos, K.; Dimitriadis, K.; Tousoulis, D.; Tsioufis, K. Wearable Blood Pressure Measurement Devices and New Approaches in Hypertension Management: The Digital Era. *J. Hum. Hypertens.* **2022**, *36*, 945–951. [CrossRef]

62. Tran, A.; Zhang, X.; Zhu, B. Mechanical Structural Design of a Piezoresistive Pressure Sensor for Low-Pressure Measurement: A Computational Analysis by Increases in the Sensor Sensitivity. *Sensors* **2018**, *18*, 2023. [CrossRef] [PubMed]

63. Chandrasekhar, A.; Kim, C.S.; Naji, M.; Natarajan, K.; Hahn, J.O.; Mukkamala, R. Smartphone-Based Blood Pressure Monitoring Via the Oscillometric Finger-Pressing Method. *Sci. Transl. Med.* **2018**, *10*, eaap8674. [CrossRef] [PubMed]

64. Jafarzadeh, F.; Rahmani, F.; Azadmehr, F.; Falaki, M.; Nazari, M. Different Applications of Telemedicine—Assessing the Challenges, Barriers, And Opportunities- A Narrative Review. *J. Fam. Med. Prim. Care* **2022**, *11*, 879–886. [CrossRef]

65. Bouabida, K.; Lebouché, B.; Pomey, M.-P. Telehealth and COVID-19 Pandemic: An Overview of the Telehealth Use, Advantages, Challenges, and Opportunities during COVID-19 Pandemic. *Healthcare* **2022**, *10*, 2293. [CrossRef]

66. Doniec, R.J.; Piaseczna, N.J.; Szymczyk, K.A.; Jacennik, B.; Sieciński, S.; Mocny-Pachońska, K.; Duraj, K.; Cedro, T.; Tkacz, E.J.; Glinkowski, W.M. Experiences of the Telemedicine and eHealth Conferences in Poland—A Cross-National Overview of Progress in Telemedicine. *Appl. Sci.* **2023**, *13*, 587. [CrossRef]

67. Serhani, M.A.; EL Kassabi, H.T.; Ismail, H.; Nujum Navaz, A. ECG Monitoring Systems: Review, Architecture, Processes, and Key Challenges. *Sensors* **2020**, *20*, 1796. [CrossRef] [PubMed]

68. Turakhia, M.P.; Hoang, D.D.; Zimetbaum, P.; Miller, J.D.; Froelicher, V.F.; Kumar, U.N.; Xu, X.; Yang, F.; Heidenreich, P.A. Heidenreich, Diagnostic Utility of a Novel Leadless Arrhythmia Monitoring Device. *Am. J. Cardiol.* **2013**, *112*, 520–524. [CrossRef] [PubMed]

69. Braunstein, E.D.; Reynbakh, O.; Krumerman, A.; Di Biase, L.; Ferrick, K.J. Inpatient Cardiac Monitoring Using a Patch-Based Mobile Cardiac Telemetry System During The COVID-19 pandemic. *J. Cardiovasc. Electrophysiol.* **2020**, *31*, 2803–2811. [CrossRef] [PubMed]

70. Gomes, N.; Pato, M.; Lourenço, A.R.; Datia, N. A Survey on Wearable Sensors for Mental Health Monitoring. *Sensors* **2023**, *23*, 1330. [CrossRef]

71. Perna, G.; Riva, A.; Defillo, A.; Sangiorgio, E.; Nobile, M.; Caldirola, D. Heart Rate Variability: Can It Serve as a Marker of Mental Health Resilience? *J. Affect. Disord.* **2020**, *263*, 754–761. [CrossRef]

72. Dobson, R.; Li, L.L.; Garner, K.; Tane, T.; McCool, J.; Whittaker, R. The Use of Sensors to Detect Anxiety for In-the-Moment Intervention: Scoping Review. *JMIR Ment. Health* **2023**, *10*, e42611. [CrossRef]

73. Crupi, R.; Faetti, T.; Paradiso, R. Preliminary Evaluation of Wearable Wellness System for Obstructive Sleep Apnea Detection. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 4142–4145. [CrossRef]

74. Chung, A.H.; Gevirtz, R.N.; Gharbo, R.S.; Thiam, M.A.; Ginsberg, J.P. Pilot Study on Reducing Symptoms of Anxiety with a Heart Rate Variability Biofeedback Wearable and Remote Stress Management Coach. *Appl. Psychophysiol. Biofeedback* **2021**, *46*, 347–358. [CrossRef]

75. Morin, C.M.; Bjorvatn, B.; Chung, F.; Holzinger, B.; Partinen, M.; Penzel, T.; Ivers, H.; Wing, Y.K.; Chan, N.Y.; Merikanto, I.; et al. Insomnia, Anxiety, and Depression during the COVID-19 Pandemic: An International Collaborative Study. *Sleep Med.* **2021**, *87*, 38–45. [CrossRef]

76. Can, Y.S.; Arnrich, B.; Ersoy, C. Stress Detection in Daily Life Scenarios Using Smart Phones and Wearable Sensors: A Survey. *J. Biomed. Inform.* **2019**, *92*, 103139. [CrossRef]

77. Bunn, J.A.; Navalta, J.W.; Fountaine, C.J.; Reece, J.D. Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015–2017. *Int. J. Exerc. Sci.* **2018**, *11*, 503–515.

78. McGinnis, R.S.; Mcginnis, E.W.; Petrillo, C.J.; Price, M. Mobile Biofeedback Therapy for the Treatment of Panic Attacks: A Pilot Feasibility Study. In Proceedings of the 2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Chicago, IL, USA, 19–22 May 2019; pp. 1–4. [CrossRef]

79. Hilty, D.M.; Armstrong, C.M.; Luxton, D.D.; Gentry, M.T.; Krupinski, E.A. A Scoping Review of Sensors, Wearables, and Remote Monitoring for Behavioral Health: Uses, Outcomes, Clinical Competencies, and Research Directions. *J. Technol. Behav. Sci.* **2021**, *6*, 278–313. [CrossRef]

80. Kwon, S.; Kim, H.; Yeo, W.-H. Recent Advances in Wearable Sensors and Portable Electronics for Sleep Monitoring. *iScience* **2021**, *24*, 102461. [CrossRef]

81. Liao, L.D.; Wang, Y.; Tsao, Y.C.; Wang, I.J.; Jhang, D.F.; Chuang, C.C.; Chen, S.F. Design and Implementation of a Multifunction Wearable Device to Monitor Sleep Physiological Signals. *Micromachines* **2020**, *11*, 672. [CrossRef] [PubMed]

82. Pham, N.T.; Dinh, T.A.; Raghebi, Z.; Kim, T.; Bui, N.; Nguyen, P.; Truong, H.; Banaei-Kashani, F.; Halbower, A.C.; Dinh, T.N.; et al. WAKE: A Behind-The-Ear Wearable System for Microsleep Detection. In Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services (MobiSys '20), Association for Computing Machinery, New York, NY, USA, 15–19 June 2020; pp. 404–418. [CrossRef]

83. Meng, K.; Zhao, S.; Zhou, Y.; Wu, Y.; Zhang, S.; He, Q.; Wang, X.; Zhou, Z.; Fan, W.; Tan, X.; et al. A Wireless Textile-Based Sensor System for Self-Powered Personalized Health Care. *Matter* **2020**, *2*, 896–907. [CrossRef]

84. Fang, Y.; Zou, Y.; Xu, J.; Chen, G.; Zhou, Y.; Deng, W.; Zhao, X.; Roustaei, M.; Hsiai, T.K.; Chen, J. Ambulatory Cardiovascular Monitoring Via a Machine-Learning-Assisted Textile Triboelectric Sensor. *Adv. Mater.* **2021**, *33*, 2104178. [CrossRef] [PubMed]

85. Yang, Z.; Pang, Y.; Han, X.; Yang, Y.; Ling, J.; Jian, M.; Zhang, Y.; Yang, Y.; Ren, T.L. Graphene Textile Strain Sensor with Negative Resistance Variation for Human Motion Detection. *ACS Nano* **2018**, *12*, 9134–9141. [CrossRef] [PubMed]

86. Di Tocco, J.; Lo Presti, D.; Rainer, A.; Schena, E.; Massaroni, C. Silicone-Textile Composite Resistive Strain Sensors for Human Motion-Related Parameters. *Sensors* **2022**, *22*, 3954. [CrossRef] [PubMed]

87. Vaddiraju, S.; Burgess, D.J.; Tomazos, I.; Jain, F.C.; Papadimitrakopoulos, F. Technologies for Continuous Glucose Monitoring: Current Problems and Future Promises. *J. Diabetes Sci. Technol.* **2010**, *4*, 1540–1562. [CrossRef]

88. Dungan, K.; Verma, N. Monitoring Technologies—Continuous Glucose Monitoring, Mobile Technology, Biomarkers of Glycemic Control. In *Endotext*; MDText.com, Inc.: South Dartmouth, MA, USA, 2000. Available online: https://www.ncbi.nlm.nih.gov/books/NBK279046/ (accessed on 8 July 2023).

89. Ma, X.; Ahadian, S.; Liu, S.; Zhang, J.; Liu, S.; Cao, T.; Lin, W.; Wu, D.; de Barros, N.R.; Zare, M.R.; et al. Smart Contact Lenses for Biosensing Applications. *Adv. Intell. Syst.* **2021**, *3*, 2000263. [CrossRef]

90. Chen, Y.; Zhang, S.; Cui, Q.; Ni, J.; Wang, X.; Cheng, X.; Alem, H.; Tebon, P.; Xu, C.; Guo, C.; et al. Microengineered Poly (HEMA) Hydrogels for wearable Contact Lens Biosensing. *Lab Chip* **2020**, *20*, 4205–4214. [CrossRef]

91. Gao, W.; Emaminejad, S.; Nyein, H.Y.Y.; Challa, S.; Chen, K.; Peck, A.; Fahad, H.M.; Ota, H.; Shiraki, H.; Kiriya, D.; et al. Fully Integrated Wearable Sensor Arrays for Multiplexed in Situ Perspiration Analysis. *Nature* **2016**, *529*, 509–514. [CrossRef] [PubMed]

92. Farandos, N.M.; Yetisen, A.K.; Monteiro, M.J.; Lowe, C.R.; Yun, S.H. Contact Lens Sensors in Ocular Diagnostics. *Adv. Healthc. Mater.* **2015**, *4*, 792–810. [CrossRef] [PubMed]

93. Lee, H.; Choi, T.K.; Lee, Y.B.; Cho, H.R.; Ghaffari, R.; Wang, L.; Choi, H.J.; Chung, T.D.; Lu, N.; Hyeon, T.; et al. A Graphene-Based Electrochemical Device with Thermoresponsive Microneedles for Diabetes Monitoring and Therapy. *Nat. Nanotechnol.* **2016**, *11*, 566–572. [CrossRef] [PubMed]

94. Lee, H.; Song, C.; Hong, Y.S.; Kim, M.; Cho, H.R.; Kang, T.; Shin, K.; Choi, S.H.; Hyeon, T.; Kim, D.H. Wearable/Disposable Sweat-Based Glucose Monitoring Device with Multistage Transdermal Drug Delivery Module. *Sci. Adv.* **2017**, *3*, e1601314. [CrossRef] [PubMed]

95. Kim, S.; Lee, G.; Jeon, C.; Han, H.H.; Kim, S.; Mok, J.W.; Joo, C.; Shin, S.; Sim, J.; Myung, D.; et al. Bimetallic Nanocatalysts Immobilized in Nanoporous Hydrogels for Long-Term Robust Continuous Glucose Monitoring of Smart Contact Lens. *Adv. Mater.* **2022**, *34*, 2110536. [CrossRef] [PubMed]

96. Jalloul, N. Wearable Sensors for The Monitoring of Movement Disorders. *Biomed. J.* **2018**, *41*, 249–253. [CrossRef] [PubMed]

97. Uddin, M.Z.; Soylu, A. Human Activity Recognition Using Wearable Sensors, Discriminant Analysis, And Long Short-Term Memory-Based Neural Structured Learning. *Sci. Rep.* **2021**, *11*, 16455. [CrossRef]

98. Csizmadia, G.; Liszkai-Peres, K.; Ferdinandy, B.; Miklósi, Á.; Konok, V. Human Activity Recognition of Children with Wearable Devices Using Lightgbm Machine Learning. *Sci. Rep.* **2022**, *12*, 5472. [CrossRef]

99. Pacchierotti, C.; Salvietti, G.; Hussain, I.; Meli, L.; Prattichizzo, D. The hRing: A Wearable Haptic Device to Avoid Occlusions in Hand Tracking. In Proceedings of the 2016 IEEE Haptics Symposium (HAPTICS), Philadelphia, PA, USA, 8–11 April 2016; pp. 134–139. [CrossRef]

100. Fan, C.; Gao, F. Enhanced Human Activity Recognition Using Wearable Sensors via a Hybrid Feature Selection Method. *Sensors* **2021**, *21*, 6434. [CrossRef]

101. Li, Y.; Tang, H.; Liu, Y.; Qiao, Y.; Xia, H.; Zhou, J. Oral Wearable Sensors: Health Management Based on the Oral Cavity. *Biosens. Bioelectron. X* **2022**, *10*, 100135. [CrossRef]

102. Li, X.; Luo, C.; Fu, Q.; Zhou, C.; Ruelas, M.; Wang, Y.; He, J.; Wang, Y.; Zhang, Y.S.; Zhou, J. A Transparent, Wearable Fluorescent Mouthguard for High-Sensitive Visualization and Accurate Localization of Hidden Dental Lesion Sites. *Adv. Mater.* **2020**, *32*, e2000060. [CrossRef]

103. Quadir, N.A.; Albasha, L.; Taghadosi, M.; Qaddoumi, N.; Hatahet, B. Low-Power Implanted Sensor for Orthodontic Bond Failure Diagnosis and Detection. *IEEE Sens. J.* **2018**, *18*, 3003–3009. [CrossRef]

104. Bodini, A.; Borghetti, M.; Paganelli, C.; Sardini, E.; Serpelloni, M. Low-Power Wireless System to Monitor Tongue Strength Against the Palate. *IEEE Sens. J.* **2021**, *21*, 5467–5475. [CrossRef]

105. Gawande, P.; Deshmukh, A.; Dhangar, R.; Gare, K.; More, S. A Smart Footwear System for Healthcare and Fitness Application—A Review. *J. Res. Eng. Appl. Sci.* **2020**, *5*, 10–14. [CrossRef]

106. Mehendale, N.; Gokalgandhi, D.; Shah, N.; Kamdar, L. A Review of Smart Technologies Embedded in Shoes. *SSRN Electron. J.* **2020**, *44*, 150. [CrossRef]

107. Bae, C.W.; Toi, P.T.; Kim, B.Y.; Lee, W.I.; Lee, H.B.; Hanif, A.; Lee, E.H.; Lee, N.-E. Fully Stretchable Capillary Microfluidics-Integrated Nanoporous Gold Electrochemical Sensor for Wearable Continuous Glucose Monitoring. *ACS Appl. Mater. Interfaces* **2019**, *11*, 14567–14575. [CrossRef]

108. Lin, S.; Wang, B.; Zhao, Y.; Shih, R.; Cheng, X.; Yu, W.; Hojaiji, H.; Lin, H.; Hoffman, C.; Ly, D.; et al. Natural Perspiration Sampling and in Situ Electrochemical Analysis with Hydrogel Micropatches for User-Identifiable and Wireless Chemo/Biosensing. *ACS Sens.* **2019**, *5*, 93–102. [CrossRef]

109. Darwish, A.; Hassanien, A.E. Wearable and Implantable Wireless Sensor Network Solutions for Healthcare Monitoring. *Sensors* **2011**, *11*, 5561–5595. [CrossRef]

110. Dinis, H.; Mendes, P.M. Recent Advances on Implantable Wireless Sensor Networks. *Wirel. Sens. Netw. Insights Innov.* **2017**. [CrossRef]

111. Klosterhoff, B.S.; Tsang, M.; She, D.; Ong, K.G.; Allen, M.G.; Willett, N.J.; Guldberg, R.E. Implantable Sensors for Regenerative Medicine. *J. Biomech. Eng.* **2017**, *139*, 021009. [CrossRef]

112. McShane, M.J.; Zavareh, A.T.; Jeevarathinam, A.S. Implantable Sensors. *Encycl. Sens. Biosens.* **2023**, *4*, 115–132. [CrossRef]

113. Bisignani, A.; De Bonis, S.; Mancuso, L.; Ceravolo, G.; Bisignani, G. Implantable Loop Recorder in Clinical Practice. *J. Arrhythmia* **2018**, *35*, 25–32. [CrossRef]

114. Talebian, S.; Foroughi, J.; Wade, S.J.; Vine, K.L.; Dolatshahi-Pirouz, A.; Mehrali, M.; Conde, J.; Wallace, G.G. Biopolymers for Antitumor Implantable Drug Delivery Systems: Recent Advances and Future Outlook. *Adv. Mater.* **2018**, *30*, e1706665. [CrossRef]

115. Pial, M.M.H.; Tomitaka, A.; Pala, N.; Roy, U. Implantable Devices for the Treatment of Breast Cancer. *J. Nanotheranostics* **2022**, *3*, 19–38. [CrossRef]

116. Chong, K.P.; Woo, B.K. Emerging wearable technology applications in gastroenterology: A Review of the Literature. *World J. Gastroenterol.* **2021**, *27*, 1149–1160. [CrossRef]

117. Beardslee, L.A.; Banis, G.E.; Chu, S.; Liu, S.; Chapin, A.A.; Stine, J.M.; Pasricha, P.J.; Ghodssi, R. Ingestible Sensors and Sensing Systems for Minimally Invasive Diagnosis and Monitoring: The Next Frontier in Minimally Invasive Screening. *ACS Sens.* **2020**, *5*, 891–910. [CrossRef]

118. Dagdeviren, C.; Javid, F.; Joe, P.; von Erlach, T.; Bensel, T.; Wei, Z.; Saxton, S.; Cleveland, C.; Booth, L.; McDonnell, S.; et al. Flexible Piezoelectric Devices for Gastrointestinal Motility Sensing. *Nat. Biomed. Eng.* **2017**, *1*, 807–817. [CrossRef]

119. Mimee, M.; Nadeau, P.; Hayward, A.; Carim, S.; Flanagan, S.; Jerger, L.; Collins, J.; McDonnell, S.; Swartwout, R.; Citorik, R.J.; et al. An Ingestible Bacterial-Electronic System to Monitor Gastrointestinal Health. *Science* **2018**, *360*, 915–918. [CrossRef]

120. Wang, J.; Coleman, D.C.; Kanter, J.; Ummer, B.; Siminerio, L. Connecting Smartphone and Wearable Fitness Tracker Data with a Nationally Used Electronic Health Record System for Diabetes Education to Facilitate Behavioral Goal Monitoring in Diabetes Care: Protocol for a Pragmatic Multi-Site Randomized Trial. *JMIR Res. Protoc.* **2018**, *7*, e10009. [CrossRef]

121. Eades, M.T.; Tsanas, A.; Juraschek, S.P.; Kramer, D.B.; Gervino, E.V.; Mukamal, K.J. Smartphone-Recorded Physical Activity for Estimating Cardiorespiratory Fitness. *Sci. Rep.* **2021**, *11*, 14851. [CrossRef]

122. Seifert, A.; Schlomann, A.; Rietz, C.; Schelling, H.R. The Use of Mobile Devices for Physical Activity Tracking in Older Adults' Everyday Life. *Digit. Health* **2017**, *3*, 205520761774008. [CrossRef]

123. De Ridder, B.; Van Rompaey, B.; Kampen, J.K.; Haine, S.; Dilles, T. Smartphone Apps Using Photoplethysmography for Heart Rate Monitoring: Meta-Analysis. *JMIR Cardio* **2018**, *2*, e4. [CrossRef]

124. Pipitprapat, W.; Harnchoowong, S.; Suchonwanit, P.; Sriphrapradang, C. The Validation of Smartphone Applications for Heart Rate Measurement. *Ann. Med.* **2018**, *50*, 721–727. [CrossRef]

125. Chan, A.H.Y.; Pleasants, R.A.; Dhand, R.; Tilley, S.L.; Schworer, S.A.; Costello, R.W.; Merchant, R. Digital Inhalers for Asthma or Chronic Obstructive Pulmonary Disease: A Scientific Perspective. *Pulm. Ther.* **2021**, *7*, 345–376. [CrossRef]

126. Zabczyk, C.; Blakey, J.D. The Effect of Connected Smart Inhalers on Medication Adherence. *Front. Med. Technol.* **2021**, *3*, 657321. [CrossRef]

127. Le, M.-P.T.; Voigt, L.; Nathanson, R.; Maw, A.M.; Johnson, G.; Dancel, R.; Mathews, B.; Moreira, A.; Sauthoff, H.; Gelabert, C.; et al. Comparison of Four Handheld Point-Of-Care Ultrasound Devices by Expert Users. *Ultrasound J.* **2022**, *14*, 27. [CrossRef]

128. Kwon, J.M.; Kim, K.H.; Jeon, K.H.; Lee, S.Y.; Park, J.; Oh, B.H. Artificial Intelligence Algorithm for Predicting Cardiac Arrest using Electrocardiography. *Scand. J. Trauma Resusc. Emerg. Med.* **2020**, *28*, 98. [CrossRef]

129. Biondi, A.; Santoro, V.; Viana, P.F.; Laiou, P.; Pal, D.K.; Bruno, E.; Richardson, M.P. Noninvasive mobile EEG as a Tool for Seizure Monitoring and Management: A Systematic Review. *Epilepsia* **2022**, *63*, 1041–1063. [CrossRef]

130. Kviesis-Kipge, E.; Rubins, U. Portable Remote Photoplethysmography Device for Monitoring of Blood Volume Changes with High Temporal Resolution. In Proceedings of the 2016 15th Biennial Baltic Electronics Conference (BEC), Tallinn, Estonia, 3–5 October 2016. [CrossRef]

131. Zhou, J.; Li, X.; Wang, X.; Yu, N.; Wang, W. Accuracy of Portable Spirometers in The Diagnosis of Chronic Obstructive Pulmonary Disease a Meta-Analysis. *npj Prim. Care Respir. Med.* **2022**, *32*, 15. [CrossRef]

132. Papadea, C.; Foster, J.; Grant, S.; Ballard, S.A.; Cate IV, J.C.; Michael Southgate, W.; Purohit, D.M. Evaluation of the I-STAT Portable Clinical Analyzer for Point-Of-Care Blood Testing in The Intensive Care Units of a University Children's Hospital. *Ann. Clin. Lab. Sci.* **2002**, *32*, 231–243.

133. Schrading, W.A.; McCafferty, B.; Grove, J.; Page, D.B. Portable, Consumer-Grade Pulse Oximeters Are Accurate for Home and Medical Use: Implications for Use in the COVID-19 Pandemic and Other Resource-Limited Environments. *J. Am. Coll. Emerg. Physicians Open* **2020**, *1*, 1450–1458. [CrossRef]

134. Zijp, T.R.; Touw, D.J.; van Boven, J.F.M. User Acceptability and Technical Robustness Evaluation of a Novel Smart Pill Bottle Prototype Designed to Support Medication Adherence. *Patient Prefer. Adherence* **2020**, *14*, 625–634. [CrossRef] [PubMed]

135. Hussain, S.; Mubeen, I.; Ullah, N.; Shah, S.S.U.D.; Khan, B.A.; Zahoor, M.; Ullah, R.; Khan, F.A.; Sultan, M.A. Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review. *BioMed Res. Int.* **2022**, *2022*, 5164970. [CrossRef] [PubMed]

136. Bi, W.L.; Hosny, A.; Schabath, M.B.; Giger, M.L.; Birkbak, N.J.; Mehrtash, A.; Allison, T.; Arnaout, O.; Abbosh, C.; Dunn, I.F.; et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA Cancer J. Clin.* **2019**, *69*, 127–157. [CrossRef]

137. Gulum, M.A.; Trombley, C.M.; Kantardzic, M. A Review of Explainable Deep Learning Cancer Detection Models in Medical Imaging. *Appl. Sci.* **2021**, *11*, 4573. [CrossRef]

138. Rana, M.; Bhushan, M. Machine Learning and Deep Learning Approach for Medical Image Analysis: Diagnosis to Detection. *Multimed. Tools Appl.* **2022**, *82*, 26731–26769. [CrossRef] [PubMed]

139. Iyawa, G.E.; Herselman, M.; Botha, A. Digital Health Innovation Ecosystems: From Systematic Literature Review to Conceptual Framework. *Procedia Comput. Sci.* **2016**, *100*, 244–252. [CrossRef]

140. Vadillo Moreno, L.; Martín Ruiz, M.L.; Malagón Hernández, J.; Valero Duboy, M.Á.; Lindén, M. The Role of Smart Homes in Intelligent Homecare and Healthcare Environments. *Ambient Assist. Living Enhanc. Living Environ.* **2017**, 345–394. [CrossRef]

141. Chang, K.D.; Raheem, A.A.; Rha, K.H. Novel Robotic Systems and Future Directions. *Indian J. Urol.* **2018**, *34*, 110–114. [CrossRef]

142. St Mart, J.P.; Goh, E.L. The Current State of Robotics in Total Knee Arthroplasty. *EFORT Open Rev.* **2021**, *6*, 270–279. [CrossRef]

143. Acker, G.; Hashemi, S.M.; Fuellhase, J.; Kluge, A.; Conti, A.; Kufeld, M.; Kreimeier, A.; Loebel, F.; Kord, M.; Sladek, D.; et al. Efficacy and safety of CyberKnife Radiosurgery in Elderly Patients with Brain Metastases: A Retrospective Clinical Evaluation. *Radiat. Oncol.* **2020**, *15*, 225. [CrossRef]

144. Ayano, Y.M.; Schwenker, F.; Dufera, B.D.; Debelee, T.G. Interpretable Machine Learning Techniques in ECG-Based Heart Disease Classification: A Systematic Review. *Diagnostics* **2022**, *13*, 111. [CrossRef] [PubMed]

145. Ahmed, A.A.; Ali, W.; Abdullah, T.A.A.; Malebary, S.J. Classifying Cardiac Arrhythmia from ECG Signal Using 1D CNN Deep Learning Model. *Mathematics* **2023**, *11*, 562. [CrossRef]

146. El Idrissi, T.; Idri, A. Deep Learning for Blood Glucose Prediction: CNN vs. LSTM. In *Computational Science and Its Applications—ICCSA 2020. ICCSA 2020*; Gervasi, O., Murgante, B., Misra, S., Garau, C., Blečić, I., Taniar, D., Apduhan, B.O., Rocha, A.M.A.C., Tarantino, E., Torre, C.M., et al., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12250, pp. 379–393. [CrossRef]

147. Sun, Q.; Jankovic, M.V.; Bally, L.; Mougiakakou, S.G. Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network. In Proceedings of the 2018 14th Symposium on Neural Networks and applications (NEUREL), Belgrade, Serbia, 20–21 November 2018; pp. 1–5. [CrossRef]

148. Abineza, C.; Balas, V.E.; Nsengiyumva, P. A Machine-Learning-based Prediction Method for Easy COPD Classification based on Pulse Oximetry clinical Use. *J. Intell. Fuzzy Syst.* **2022**, *43*, 1683–1695. [CrossRef]

149. Rohmetra, H.; Raghunath, N.; Narang, P.; Chamola, V.; Guizani, M.; Lakkaniga, N. AI-Enabled Remote Monitoring of Vital Signs for COVID-19: Methods, Prospects and Challenges. *Computing* **2021**, *105*, 783–809. [CrossRef]

150. Lai, Z.; Vadlaputi, P.; Tancredi, D.J.; Garg, M.; Koppel, R.I.; Goodman, M.; Hogan, W.; Cresalia, N.; Juergensen, S.; Manalo, E.; et al. Enhanced Critical Congenital Cardiac Disease Screening by Combining Interpretable Machine Learning Algorithms. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2021**, *2021*, 1403–1406. [CrossRef] [PubMed]

151. Rajaguru, H.; Prabhakar, S.K. Sparse PCA and Soft Decision Tree Classifiers for Epilepsy Classification from EEG Signals. In Proceedings of the 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 20–22 April 2017; Volume 1, pp. 581–584. [CrossRef]

152. Sharma, A.; Rai, J.K.; Tewari, R.P. Epileptic Seizure Anticipation and Localisation of Epileptogenic Region Using EEG Signals. *J. Med. Eng. Technol.* **2018**, *42*, 203–216. [CrossRef]

153. Alturki, F.A.; AlSharabi, K.; Abdurraqeeb, A.M.; Aljalal, M. EEG Signal Analysis for Diagnosing Neurological Disorders Using Discrete Wavelet Transform and Intelligent Techniques. *Sensors* **2020**, *20*, 2505. [CrossRef]

154. Tawhid, M.N.A.; Siuly, S.; Wang, H. Diagnosis of Autism Spectrum Disorder from EEG using a Time–Frequency Spectrogram Image-based Approach. *Electron. Lett.* **2020**, *56*, 1372–1375. [CrossRef]

155. Maray, N.; Ngu, A.H.; Ni, J.; Debnath, M.; Wang, L. Transfer Learning on Small Datasets for Improved Fall Detection. *Sensors* **2023**, *23*, 1105. [CrossRef]

156. Butt, A.; Narejo, S.; Anjum, M.R.; Yonus, M.U.; Memon, M.; Samejo, A.A. Fall Detection Using LSTM and Transfer Learning. *Wirel. Pers. Commun.* **2022**, *126*, 1733–1750. [CrossRef]

157. Kaisti, M.; Panula, T.; Leppänen, J.; Punkkinen, R.; Jafari Tadi, M.; Vasankari, T.; Jaakkola, S.; Kiviniemi, T.; Airaksinen, J.; Kostiainen, P.; et al. Clinical Assessment of a Non-Invasive Wearable MEMS Pressure Sensor Array for Monitoring of Arterial Pulse Waveform, Heart Rate and Detection of Atrial Fibrillation. *npj Digit. Med.* **2019**, *2*, 39. [CrossRef] [PubMed]

158. El Asnaoui, K.; Chawki, Y. Using X-Ray Images and Deep Learning for Automated Detection of Coronavirus Disease. *J. Biomol. Struct. Dyn.* **2021**, *39*, 3615–3626. [CrossRef] [PubMed]

159. Erdaw, Y.; Tachbele, E. Machine Learning Model Applied on Chest X-ray Images Enables Automatic Detection of COVID-19 Cases with High Accuracy. *Int. J. Gen. Med.* **2021**, *14*, 4923–4931. [CrossRef] [PubMed]

160. Matsumoto, T.; Kodera, S.; Shinohara, H.; Ieki, H.; Yamaguchi, T.; Higashikuni, Y.; Kiyosue, A.; Ito, K.; Ando, J.; Takimoto, E.; et al. Diagnosing Heart Failure from Chest X-Ray Images Using Deep Learning. *Int. Heart J.* **2020**, *61*, 781–786. [CrossRef] [PubMed]

161. Leiner, T.; Rueckert, D.; Suinesiaputra, A.; Baeßler, B.; Nezafat, R.; Išgum, I.; Young, A.A. Machine learning in cardiovascular magnetic resonance: Basic concepts and applications. *J. Cardiovasc. Magn. Reson.* **2019**, *21*, 61. [CrossRef] [PubMed]

162. Chato, L.; Latifi, S. Machine Learning and Radiomic Features to Predict Overall Survival Time for Glioblastoma Patients. *J. Pers. Med.* **2021**, *11*, 1336. [CrossRef]

163. Li, H.; Lee, C.H.; Chia, D.; Lin, Z.; Huang, W.; Tan, C.H. Machine Learning in Prostate MRI for Prostate Cancer: Current Status and Future Opportunities. *Diagnostics* **2022**, *12*, 289. [CrossRef]

164. Siouras, A.; Moustakidis, S.; Giannakidis, A.; Chalatsis, G.; Liampas, I.; Vlychou, M.; Hantes, M.; Tasoulis, S.; Tsaopoulos, D. Knee Injury Detection Using Deep Learning on MRI Studies: A Systematic Review. *Diagnostics* **2022**, *12*, 537. [CrossRef]

165. Said, Y.; Alsheikhy, A.A.; Shawly, T.; Lahza, H. Medical Images Segmentation for Lung Cancer Diagnosis Based on Deep Learning Architectures. *Diagnostics* **2023**, *13*, 546. [CrossRef]

166. Sreenivasu, S.V.N.; Gomathi, S.; Kumar, M.J.; Prathap, L.; Madduri, A.; Almutairi, K.M.A.; Alonazi, W.B.; Kali, D.; Jayadhas, S.A. Dense Convolutional Neural Network for Detection of Cancer from CT Images. *BioMed Res. Int.* **2022**, *2022*, 1293548. [CrossRef]

167. Dai, X.; Huang, L.; Qian, Y.; Xia, S.; Chong, W.; Liu, J.; Di Ieva, A.; Hou, X.; Ou, C. Deep Learning for Automated Cerebral Aneurysm Detection on Computed Tomography Images. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 715–723. [CrossRef] [PubMed]

168. Abraham, G.K.; Bhaskaran, P.; Jayanthi, V.S. Lung Nodule Classification in CT Images Using Convolutional Neural Network. In Proceedings of the 2019 9th International Conference on Advances in Computing and Communication (ICACC), Kochi, India, 6–8 November 2019; pp. 199–203. [CrossRef]

169. Vidhya, V.; Gudigar, A.; Raghavendra, U.; Hegde, A.; Menon, G.R.; Molinari, F.; Ciaccio, E.J.; Acharya, U.R. Automated Detection and Screening of Traumatic Brain Injury (TBI) Using Computed Tomography Images: A Comprehensive Review and Future Perspectives. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6499. [CrossRef]

170. Shinohara, I.; Inui, A.; Mifune, Y.; Nishimoto, H.; Yamaura, K.; Mukohara, S.; Yoshikawa, T.; Kato, T.; Furukawa, T.; Hoshino, Y.; et al. Using Deep Learning for Ultrasound Images to Diagnose Carpal Tunnel Syndrome with High Accuracy. *Ultrasound Med. Biol.* **2022**, *48*, 2052–2059. [CrossRef] [PubMed]

171. Mohammad, U.F.; Almekkawy, M. Automated Detection of Liver Steatosis in Ultrasound Images Using Convolutional Neural Networks. In Proceedings of the 2021 IEEE International Ultrasonics Symposium (IUS) 2021, Xi'an, China, 11–16 September 2021; pp. 1–4. [CrossRef]

172. Xu, S.S.-D.; Chang, C.-C.; Su, C.-T.; Phu, P.Q. Classification of Liver Diseases Based on Ultrasound Image Texture Features. *Appl. Sci.* **2019**, *9*, 342. [CrossRef]

173. Lv, Y.; Huang, Z. Account of Deep Learning-Based Ultrasonic Image Feature in the Diagnosis of Severe Sepsis Complicated with Acute Kidney Injury. *Comput. Math. Methods Med.* **2022**, *2022*, 8158634. [CrossRef] [PubMed]

174. Russell, A.M.; Adamali, H.; Molyneaux, P.L.; Lukey, P.T.; Marshall, R.P.; Renzoni, E.A.; Wells, A.U.; Maher, T.M. Daily Home Spirometry: An Effective Tool for Detecting Progression in Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* **2016**, *194*, 989–997. [CrossRef]

175. Jung, T.; Vij, N. Early Diagnosis and Real-Time Monitoring of Regional Lung Function Changes to Prevent Chronic Obstructive Pulmonary Disease Progression to Severe Emphysema. *J. Clin. Med.* **2021**, *10*, 5811. [CrossRef]

176. Islam, T.T.; Ahmed, M.S.; Hassanuzzaman, M.; Bin Amir, S.A.; Rahman, T. Blood Glucose Level Regression for Smartphone PPG Signals Using Machine Learning. *Appl. Sci.* **2021**, *11*, 618. [CrossRef]

177. Susana, E.; Ramli, K.; Murfi, H.; Apriantoro, N.H. Non-Invasive Classification of Blood Glucose Level for Early Detection Diabetes Based on Photoplethysmography Signal. *Information* **2022**, *13*, 59. [CrossRef]

178. Celka, P.; Charlton, P.H.; Farukh, B.; Chowienczyk, P.; Alastruey, J. Influence of Mental Stress on The Pulse Wave Features of Photoplethysmograms. *Healthc. Technol. Lett.* **2019**, *7*, 7–12. [CrossRef] [PubMed]

179. Přibil, J.; Přibilová, A.; Frollo, I. Stress Level Detection and Evaluation from Phonation and PPG Signals Recorded in an Open-Air MRI Device. *Appl. Sci.* **2021**, *11*, 11748. [CrossRef]

180. Fan, J.; Sun, C.; Long, M.; Chen, C.; Chen, W. EOGNET: A Novel Deep Learning Model for Sleep Stage Classification Based on Single-Channel EOG Signal. *Front. Neurosci.* **2021**, *15*, 573194. [CrossRef] [PubMed]

181. Weinhouse, G.L.; Kimchi, E.; Watson, P.; Devlin, J.W. Sleep Assessment in Critically Ill Adults: Established Methods and Emerging Strategies. *Crit. Care Explor.* **2022**, *4*, e0628. [CrossRef]

182. Tag, B.; Vargo, A.W.; Gupta, A.; Chernyshov, G.; Kunze, K.; Dingler, T. Continuous Alertness Assessments. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019. [CrossRef]

183. Hausfater, P.; Zhao, Y.; Defrenne, S.; Bonnet, P.; Riou, B. Cutaneous Infrared Thermometry for Detecting Febrile Patients. *Emerg. Infect. Dis.* **2008**, *14*, 1255–1258. [CrossRef] [PubMed]

184. Sessler, D.I. Temperature Monitoring and Perioperative Thermoregulation. *Anesthesiology* **2008**, *109*, 318–338. [CrossRef] [PubMed]

185. Jaffe, G.J.; Caprioli, J. Optical Coherence Tomography to Detect and Manage Retinal Disease and Glaucoma. *Am. J. Ophthalmol.* **2004**, *137*, 156–169. [CrossRef]

186. Le, D.; Son, T.; Yao, X. Machine Learning in Optical Coherence Tomography Angiography. *Exp. Biol. Med.* **2021**, *246*, 2170–2183. [CrossRef]

187. Ahmed, M. Video Capsule Endoscopy in Gastroenterology. *Gastroenterol. Res.* **2022**, *15*, 47–55. [CrossRef]

188. Akpunonu, B.; Hummell, J.; Akpunonu, J.D.; Ud Din, S. Capsule Endoscopy in Gastrointestinal Disease: Evaluation, Diagnosis, And Treatment. *Clevel. Clin. J. Med.* **2022**, *89*, 200–211. [CrossRef]

189. Berus, L.; Klancnik, S.; Brezocnik, M.; Ficko, M. Classifying Parkinson's Disease Based on Acoustic Measures Using Artificial Neural Networks. *Sensors* **2018**, *19*, 16. [CrossRef] [PubMed]

190. Lahmiri, S.; Dawson, D.A.; Shmuel, A. Performance of Machine Learning Methods in Diagnosing Parkinson's Disease Based on Dysphonia Measures. *Biomed. Eng. Lett.* **2017**, *8*, 29–39. [CrossRef] [PubMed]

191. Reid, J.; Parmar, P.; Lund, T.; Aalto, D.K.; Jeffery, C.C. Development of a machine-learning based voice disorder screening tool. *Am. J. Otolaryngol.* **2022**, *43*, 103327. [CrossRef] [PubMed]

192. Brunese, L.; Martinelli, F.; Mercaldo, F.; Santone, A. Deep Learning for Heart Disease Detection Through Cardiac Sounds. *Procedia Comput. Sci.* **2020**, *176*, 2202–2211. [CrossRef]

193. Joudeh, I.O.; Cretu, A.-M.; Guimond, S.; Bouchard, S. Prediction of Emotional Measures via Electrodermal Activity (EDA) and Electrocardiogram (ECG). *Eng. Proc.* **2022**, *27*, 47. [CrossRef]

194. Gorson, J.; Cunningham, K.; Worsley, M.; O'Rourke, E. Using Electrodermal Activity Measurements to Understand Student Emotions While Programming. In Proceedings of the 2022 ACM Conference on International Computing Education Research, 7–11 August 2022; Association for Computing Machinery: New York, NY, USA; Volume 1, pp. 105–119. [CrossRef]

195. Rahma, O.N.; Putra, A.P.; Rahmatillah, A.; Putri, Y.S.K.A.; Fajriaty, N.D.; Ain, K.; Chai, R. Electrodermal Activity for Measuring Cognitive and Emotional Stress Level. *J. Med. Signals Sens.* **2022**, *12*, 155–162. [CrossRef]

196. Ghasemi-Roudsari, S.; Al-Shimary, A.; Varcoe, B.; Byrom, R.; Kearney, L.; Kearney, M. A Portable Prototype Magnetometer to Differentiate Ischemic and Non-Ischemic Heart Disease in Patients with Chest Pain. *PLoS ONE* **2018**, *13*, e0191241. [CrossRef]

197. Ha, S.; Choi, S. Convolutional Neural Networks for Human Activity Recognition Using Multiple Accelerometer and Gyroscope Sensors. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 381–388. [CrossRef]

198. Mehrmohammadi, M.; Yoon, S.J.; Yeager, D.; Emelianov, S.Y. Photoacoustic Imaging for Cancer Detection and Staging. *Curr. Mol. Imaging* **2013**, *2*, 89–105. [CrossRef]

199. Gharieb, R.R. Photoacoustic Imaging for Cancer Diagnosis: A Breast Tumor Example. In *Photoacoustic Imaging*; IntechOpen: London, UK, 2020. [CrossRef]

200. Qiu, T.; Lan, Y.; Gao, W.; Zhou, M.; Liu, S.; Huang, W.; Zeng, S.; Pathak, J.L.; Yang, B.; Zhang, J. Photoacoustic Imaging as A Highly Efficient and Precise Imaging Strategy for The Evaluation of Brain Diseases. *Quant. Imaging Med. Surg.* **2021**, *11*, 2169–2186. [CrossRef]

201. Wick, K.D.; Matthay, M.A.; Ware, L.B. Pulse Oximetry for The Diagnosis and Management of Acute Respiratory Distress Syndrome. *Lancet Respir. Med.* **2022**, *10*, 1086–1098. [CrossRef]

202. Jawin, V.; Ang, H.-L.; Omar, A.; Thong, M.-K. Beyond Critical Congenital Heart Disease: Newborn Screening Using Pulse Oximetry for Neonatal Sepsis and Respiratory Diseases in a Middle-Income Country. *PLoS ONE* **2015**, *10*, e0137580. [CrossRef] [PubMed]

203. Horry, M.J.; Chakraborty, S.; Paul, M.; Ulhaq, A.; Pradhan, B.; Saha, M.; Shukla, N. COVID-19 Detection through Transfer Learning Using Multimodal Imaging Data. *IEEE Access* **2020**, *8*, 149808–149824. [CrossRef] [PubMed]

204. Nian, R.; Liu, J.; Huang, B. A review on Reinforcement Learning: Introduction and Applications in Industrial Process Control. *Comput. Chem. Eng.* **2020**, *139*, 106886. [CrossRef]

205. van Engelen, J.E.; Hoos, H.H. A Survey on Semi-Supervised Learning. *Mach. Learn.* **2019**, *109*, 373–440. [CrossRef]

206. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1345–1459. [CrossRef]

207. Zhang, A.; Xing, L.; Zou, J.; Wu, J.C. Shifting Machine Learning for Healthcare from Development to Deployment and From Models to Data. *Nat. Biomed. Eng.* **2022**, *6*, 1330–1345. [CrossRef] [PubMed]

208. Sami, S.M.; Nasrabadi, N.M.; Rao, R. Deep Transductive Transfer Learning for Automatic Target Recognition. In Proceedings of the Automatic Target Recognition XXXIII, Orlando, FL, USA, 30 April–5 May 2023; SPIE: Bellingham, DC, USA, 2023; Volume 12521, pp. 31–40. [CrossRef]

209. Kushibar, K.; Salem, M.; Valverde, S.; Rovira, À.; Salvi, J.; Oliver, A.; Lladó, X. Transductive Transfer Learning for Domain Adaptation in Brain Magnetic Resonance Image Segmentation. *Front. Neurosci.* **2021**, *15*, 608808. [CrossRef]

210. Nishio, M.; Fujimoto, K.; Matsuo, H.; Muramatsu, C.; Sakamoto, R.; Fujita, H. Lung Cancer Segmentation with Transfer Learning: Usefulness of a Pretrained Model Constructed from an Artificial Dataset Generated Using a Generative Adversarial Network. *Front. Artif. Intell.* **2021**, *4*, 694815. [CrossRef]

211. Tschandl, P.; Sinz, C.; Kittler, H. Domain-Specific Classification-Pretrained Fully Convolutional Network Encoders for Skin Lesion Segmentation. *Comput. Biol. Med.* **2019**, *104*, 111–116. [CrossRef]

212. Oh, K.; Chung, Y.C.; Kim, K.W.; Kim, W.S.; Oh, I.S. Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning. *Sci. Rep.* **2019**, *9*, 18150. [CrossRef]

213. Raza, N.; Naseer, A.; Tamoor, M.; Zafar, K. Alzheimer Disease Classification through Transfer Learning Approach. *Diagnostics* **2023**, *13*, 801. [CrossRef]

214. Martinez, M.; De Leon, P.L. Falls Risk Classification of Older Adults Using Deep Neural Networks and Transfer Learning. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 144–150. [CrossRef] [PubMed]

215. De Bois, M.; El Yacoubi, M.A.; Ammi, M. Adversarial Multi-Source Transfer Learning in Healthcare: Application to Glucose Prediction for Diabetic People. *Comput. Methods Programs Biomed.* **2021**, *199*, 105874. [CrossRef] [PubMed]

216. Iman, M.; Arabnia, H.R.; Rasheed, K. A Review of Deep Transfer Learning and Recent Advancements. *Technologies* **2023**, *11*, 40. [CrossRef]

217. Zhang, X.; Liang, Y.; Li, W.; Liu, C.; Gu, D.; Sun, W.; Miao, L. Development and Evaluation of Deep Learning for Screening Dental Caries from Oral Photographs. *Oral Dis.* **2020**, *28*, 173–181. [CrossRef]

218. Koike, T.; Qian, K.; Kong, Q.; Plumbley, M.D.; Schuller, B.W.; Yamamoto, Y. Audio for Audio is Better? An Investigation on Transfer Learning Models for Heart Sound Classification. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 74–77. [CrossRef]

219. Chakraborty, A.; Anitescu, C.; Zhuang, X.; Rabczuk, T. Domain Adaptation based Transfer Learning Approach for Solving PDEs on Complex Geometries. *Eng. Comput.* **2022**, *38*, 4569–4588. [CrossRef]

220. Wang, Y.-X.; Ramanan, D.; Hebert, M. Growing a Brain: Fine-Tuning by Increasing Model Capacity. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3029–3038. [CrossRef]

221. Fu, Z.; Zhang, B.; He, X.; Li, Y.; Wang, H.; Huang, J. Emotion Recognition based on Multi-Modal Physiological Signals and Transfer Learning. *Front. Neurosci.* **2022**, *16*, 1000716. [CrossRef] [PubMed]

222. Zhao, Y.; Wang, X.; Che, T.; Bao, G.; Li, S. Multi-Task Deep Learning for Medical Image Computing and Analysis: A Review. *Comput. Biol. Med.* **2023**, *153*, 106496. [CrossRef]

223. Torres-Soto, J.; Ashley, E.A. Multi-Task Deep Learning for Cardiac Rhythm Detection in Wearable Devices. *npj Digit. Med.* **2020**, *3*, 116. [CrossRef]

224. Martindale, C.F.; Christlein, V.; Klumpp, P.; Eskofier, B.M. Wearables-based Multi-Task Gait and Activity Segmentation Using Recurrent Neural Networks. *Neurocomputing* **2021**, *432*, 250–261. [CrossRef]

225. Hong, Y.; Wei, B.; Han, Z.; Li, X.; Zheng, Y.; Li, S. MMCL-Net: Spinal Disease Diagnosis in Global Mode Using Progressive Multi-Task Joint Learning. *Neurocomputing* **2020**, *399*, 307–316. [CrossRef]

226. Sun, X.; Panda, R.; Feris, R.; Saenko, K. Adashare: Learning What to Share for Efficient Deep Multi-Task Learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 8728–8740.

227. Liu, S.; Johns, E.; Davison, A.J. End-to-End Multi-Task Learning with Attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1871–1880.

228. Gupta, S.; Punn, N.S.; Sonbhadra, S.K.; Agarwal, S. MAG-Net: Multi-Task Attention Guided Network for Brain Tumor Segmentation and Classification. In *Big Data Analytics*; Srirama, S.N., Lin, J.C.W., Bhatnagar, R., Agarwal, S., Reddy, P.K., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; p. 13147. [CrossRef]

229. Lin, B.; Ye, F.; Zhang, Y. A Closer Look at Loss Weighting in Multi-Task Learning. *arXiv* **2021**, arXiv:2111.10603.

230. Kendall, A.; Gal, Y.; Cipolla, R. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7482–7491.

231. Chen, Z.; Badrinarayanan, V.; Lee, C.Y.; Rabinovich, A. Gradnorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks. *Int. Conf. Mach. Learn.* **2018**, *80*, 794–803.

232. Yu, T.; Kumar, S.; Gupta, A.; Levine, S.; Hausman, K.; Finn, C. Gradient Surgery for Multi-Task Learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 5824–5836.

233. Liu, L.; Li, Y.; Kuang, Z.; Xue, J.; Chen, Y.; Yang, W.; Liao, Q.; Zhang, W.; Towards Impartial Multi-Task Learning. Int. *Conf. Learn.* 2021. Available online: https://openreview.net/forum?id=IMPnRXEWpvr (accessed on 7 December 2023).

234. Snell, J.; Swersky, K.; Zemel, R. Prototypical Networks for Few-Shot Learning. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: New York, NY, USA, 2017; pp. 4080–4090.

235. Chen, Y.; Liu, Z.; Xu, H.; Darrell, T.; Wang, X. Meta-Baseline: Exploring Simple Meta-Learning for Few-Shot Learning. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021. [CrossRef]

236. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a Few Examples. *ACM Comput. Surv.* **2020**, *53*, 1–34. [CrossRef]

237. Sheller, M.J.; Edwards, B.; Reina, G.A.; Martin, J.; Pati, S.; Kotrotsou, A.; Milchenko, M.; Xu, W.; Marcus, D.; Colen, R.R.; et al. Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data. *Sci. Rep.* **2020**, *10*, 12598. [CrossRef]

238. Rieke, N.; Hancox, J.; Li, W.; Milletarì, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K.; et al. The Future of Digital Health with Federated Learning. *npj Digit. Med.* **2020**, *3*, 119. [CrossRef]

239. Dash, T.; Chitlangia, S.; Ahuja, A.; Srinivasan, A. A Review of Some Techniques for Inclusion of Domain-Knowledge into Deep Neural Networks. *Sci. Rep.* **2022**, *12*, 1040. [CrossRef]

240. Williams, J.; Tadesse, A.; Sam, T.; Sun, H.; Montañez, G.D. Limits of Transfer Learning. In *Machine Learning, Optimization, and Data Science*; Springer: Cham, Switzerland, 2021. [CrossRef]