Abstract—Artificial intelligence (AI) continues to transform data analysis in many domains. Progress in each domain is driven by a growing body of annotated data, increased computational resources, and technological innovations. In medicine, the sensitivity of the data, the complexity of the tasks, the potentially high stakes, and a requirement of accountability give rise to a particular set of challenges. In this review, we focus on three key methodological approaches that address some of the particular challenges in AI-driven medical decision making. 1) Explainable AI aims to produce a human-interpretable justification for each output. Such models increase confidence if the results appear plausible and match the clinicians expectations. However, the absence of a plausible explanation does not imply an inaccurate model. Especially in highly non-linear, complex models that are tuned to maximize accuracy, such interpretable representations only reflect a small portion of the justification. 2) Domain adaptation and transfer learning enable AI models to be trained and applied across multiple domains. For example, a classification task based on images acquired on different acquisition hardware. 3) Federated learning enables learning large-scale models without exposing sensitive personal health information. Unlike centralized AI learning, where the centralized learning machine has access to the entire training data, the federated learning process iteratively updates models across multiple sites by exchanging only parameter updates, not personal health data. This narrative review covers the basic concepts, highlights relevant corner-stone and state-of-the-art research in the field, and discusses perspectives.

Index Terms—Domain adaptation, explainable artificial intelligence, federated learning.

I. INTRODUCTION

Artificial intelligence (AI)—understood here as the capability of computers to transform input data to elicit an appropriate response [1], has a huge potential to transform many aspects of our lives. The methodological advances that drive the progress generally focus on core issues like solving specific tasks more accurately [2], creating models that generalize to unseen data [3], or addressing fundamentally new tasks [4]. Within these innovation cycles, the application of medical data faces particular challenges. Personal health related information is sensitive and therefore not readily sharable. When it is shared, the data itself is highly complex and heterogeneous. There is not only a large variation between individuals, but there is also measurement noise and systematic effects due to acquisition instruments and protocols. This noise, uncertainty, and heterogeneity in the data and labels create distinctly challenging sets of problems. The goal of AI in medicine is to support the clinical decision process. The decisions in clinical processes can have grave consequences, sometimes observed with a substantial delay and usually uncorrectable. Furthermore, state-of-the-art AI models tend to have a large number of parameters, which are necessary to learn complex interaction and generalize well. Thus, even when being deterministic, AI models do not behave intuitively and may fail or behave unexpectedly [5]. This perception of the black-box behavior leads to subjective mistrust and objectively limited reliability of an unknown extent.

With the growing availability of extensive medical data and the increase in computational power, AI has become more important in clinical practice for early disease detection, accurate diagnosis, predictions, and prognosis [6]. However, AI models—mostly instantiated as deep artificial neural networks, are sufficiently complicated to be non-intuitive. Thus, the computer models are perceived as a black boxes [7]. Then, the black-box nature leads to mistrust of predictive models by clinicians [8]. For this reason, explainable artificial intelligence (explainable AI) was proposed in order to make AI models more transparent and easier to understand and interpret [9]. Explainable AI can thus complement human perception of the naked data [10], [11] and will contribute to making AI models in medicine more trusted.

Computer-aided diagnosis is increasingly involved in clinical decision processes to advance the goal of personalized
medicine [12]. However, some challenges limit its application in clinical practice, such as domain transfer [13]. Specifically, it assumes that the training and test sets are independent and identically distributed (IID) using machine learning (ML) models. In reality, the distributions of the training and test sets may not be identical. This leads to domain shift when we transfer the model [14]. Despite advanced deep learning models [15], the error increases proportionately with different distributions between the training data set and the test data set [13]. Therefore, how to solve the domain transfer problem is a key to applying AI to clinical tasks. Domain adaptation (DA) breaks the traditional assumption that the distribution of the training and test data sets must be consistent with ML [16]. The data set is divided into source and target data in this context. The source data is related to the task to be solved, and the target data (with no labels or only a few labels) is related to the task. Note that there are always domain shifts between two domains [16]. In medical imaging, there are significant differences within data sets of the same type/modality due to various equipment, sites, protocols, etc [17]. Domain adaptation techniques aim to reduce these differences.

More and more attention has been paid to data privacy in recent years [18]. In clinical topics, personalized models require protection of personal health information [19]. Federated learning (FL) can integrate data from private sources without sharing private data itself [20]. Specifically, FL protects users’ privacy primarily by exchanging non-personal data such as updated models between agents (clients) of the learning federation [21].

This paper aims to provide a brief narrative review of explainable/interpretable, domain-adaptive, and federated AI for clinical applications. Fig. 1 illustrates a yearly growing number of publications in the field. These search results indicate a clear trend. However, it is worth noting that the search technique is not 100 percent sensitive because synonymous terms may be used for similar techniques; for instance, transfer learning is a type of DA. In addition, some research works may discuss one of the topics superficially yet being counted.

This review provides a brief overview of XAI, DA and FL and categorizes techniques within each field. For each topic, we motivate its use in medical research and healthcare applications and discuss challenges and potentials for feature research.

The remainder of this paper is structured as follows. Section II consists of the fundamental concepts for the three methodological approaches. We discuss in Section III the applications of XAI, DA and FL in medical research and healthcare. Finally, Section IV concludes by providing perspectives of challenges and untapped potential.

II. METHODOLOGICAL FOUNDATION AND BACKGROUND

A. Explainable Artificial Intelligence

Human-centric medical decision processes are designed to be trustworthy because of the high stakes that are involved. Isolated outputs from AI models derived from deterministic, yet highly non-linear networks that consist of artificial neurons organized in many–up to a thousand and more, layers

![Number of publications indexed on PubMed](https://pubmed.ncbi.nlm.nih.gov) that matched the search queries related to the topics in this survey. FL: "federated learning" AND ((medicine) OR (healthcare)), XAI: ((explainable AI") OR ("explainable artificial intelligence")) AND ((medicine) (healthcare)), DA: (domain adaptation") AND ((medicine) OR (healthcare)). The number for the year 2022 is a projection based on the assumption that publications appear at the same rate for the remaining two months of the year 2022.

[22] leave no room for interpretation and provide no explanation of the prediction. AI models are therefore casually considered as black boxes because the “How?” and “Why?” a machine arrives at its conclusion is in general not understandable by humans [9]. The field has a rich and nuanced taxonomy that is not always used consistently. Explainability and interpretability—sometimes used interchangeably, relate to human-focused understanding [23]–[25]. A machine is understandable if its decisions are based on descriptive and relevant information that is recognizable by humans. In this context, descriptive means how well the interpretation method captures the learned relationships and relevant means how useful the interpretation is [26]. Explainable AI models strive for high causability—the extent to which a human can casually understand the decision making process of artificial intelligence. Explainability and related research fields are elements of building reliable and trusted AI systems [10].

XAI elements are either injected into the model architecture or are external. The former leads to models that intrinsically increase transparency of the model, whereas the latter introduces explainability post-hoc [27]. Transparency is achieved by highlighting relevance of features, simplification, and visual data that is relevant and meaningful for the practitioners [28].

B. Domain-Adaptive Artificial Intelligence

Generalizability is the implicit ability to use a trained predictive model from one or more source domains in a different, but related, target domain. In contrast, domain adaptation is achieved through explicit modelling of domain differences. Assume we train a classifier model using data from a source domain \( D_s \) for some task \( T_s \), and to learn the same or similar task \( T_t \) using target data \( D_t \). If the two domains have the same distribution \( (D_s \sim D_t) \) and the tasks are identical, then the model trained for task \( T_s \) on \( D_s \) can also be used for the task \( T_t \) in \( D_t \). If the two domains are not from the same distribution \( (D_s \not\sim D_t) \), then the models trained on the source domain \( D_s \) are likely to perform worse on \( D_t \), DA is utilized to exploit similarities of tasks and learn differences in data domains to produce models for task \( T_t \) in \( D_t \) [29]. Fig. 2 shows an example of homogeneous domain adaptation, this is when the same
task is applied to the same features that have different distributions. Other types of DA methods are outlined in the taxonomy below.

1) Supervised, Semi-Supervised, and Unsupervised Domain Adaptation: Depending on the availability of annotation, DA applications can be classified as supervised [30], semisupervised [31], and unsupervised [32]. In supervised DA models, labels of all training instances are available. On the other hand, in semi-supervised learning, only a part of the training data is annotated. Learning with no labels is considered unsupervised. Specifically, most studies focus on unsupervised DA models due to insufficient numbers of labeled samples. For example, a novel unsupervised DA algorithm proposes to eliminate the domain differences for reducing the label noise [17]. While, a deep domain adaptation method is proposed to transfer domain knowledge from a well-labeled source domain to a partially labeled target domain [33]. This method minimizes the differences in the domain that have an important clinical impact for disease diagnosis.

2) Single- and Cross-Modality: Medical image derived from many types of scanner (e.g., magnetic resonance imaging, CT, X-ray) that are related to texture differences between the source and target domains [34]. This phenomenon is called the modality difference. When based on acquisition, DA methods are categorized as either single- or cross-modality DA. In single-modality DA, the adaptations happen within a single type of data [35]. In clinical practice, medical images are acquired on hardware from different vendors and at different centers [36]. Because different modalities provide complementary information and vary in acquisition cost and expertise, a diagnosis can be established with different modalities. Cross-modal DA helps to make models applicable across modalities. For example, some models can be transferred between CT and MRI images [37]. In addition, bidirectional cross-modality DA between MRI and CT images has been applied using the synergistic image and feature alignment framework (SIFA) [38]. Despite the practicability of DA, there are always remaining risks of inherent biases in the data [39] or unfairness in models [40].

3) Single- and Multi-Source Domain Adaptation: Depending on the number of source domains, it is possible to classify DA into single-source DA (SDA), multi-source DA (MDA) and source-free DA [41]. For example, trained models using single-source domains lack generalization. Multi-source DA collects labeled data from multiple sources with different distributions [42] and implicitly learn to be robust. For example, MDA learning can alleviate the difficulties caused by aligning multimodal characteristics between multiple domains [43]. Although MDA is advanced, MDA applications are challenging due to the heterogeneity of data between domains. For this reason, most existing studies are based on SDA [44]. Another issue could be related to the source domains, cause to data privacy or lack of storage space for small devices [45]. A generalized source-free domain adaptation (G-SFDA) was proposed to possess source domain features via transfer learning [46].

4) One- and Multi-Step DA: On the basis of the distance between the source and the target domain, domain adaptation methods can be roughly categorized into two types: one-step and multistep domain adaptation methods. The one-step DA methods often simply align the feature distributions of the two domains by minimizing their distances. But it is difficult for the network to minimize the huge domain discrepancy. Specifically, in remote domain transfer learning, when the distance between the source domain and the target domain is too long [47], it is impossible to achieve and can even affect the performance of the learning model in the target domain. This situation is called negative transfer [48]. In order to avoid this situation, it is necessary to find an “intermediate bridge” to connect the source and the target domain. That is called multistep DA. It can be transformed into one-step domain adaptation by choosing an appropriate intermediate domain. For example, through efficient instance selection, transfer features are successfully applied between remote domains and effectively solve the remote domain problem [49].

5) Shallow and Deep DA: According to the type of model, we can divide the DA method into shallow and deep DA. Usually, shallow DA leverages feature engineering and is used with traditional machine learning. But few timely reviews the emerging deep learning based methods. Unlike the shallow DA, deep DA methods can leverage deep networks to learn more transferable representations by embedding DA in a deep learning pipeline [50]. In addition, deep DA integrates a feature into the learning model to improve performance metrics [13].

These methods may reduce the differences in the domain that can have a positive impact in clinical research. Specifically, more effort is required to fully apply such methods in clinical topics to avoid bias.

C. Federated Learning

Federated learning (FL) is a way of distributed collaborative machine learning in which the training data is not shared. This concept of data-private collaborative learning [51], that
initially was motivated in the context of mobile devices in which each one has only a small fraction of the data, equally applies to medical application in which the data is distributed across hospitals [52].

This framework works by collaborating multiple devices to update a machine learning model without sharing their private data under the supervision of a central server. An example in medicine is the COVID-19 chest scan engine, which was built by many researchers and medical practitioners from different parts of the world. In the COVID-19 pandemic, along with increased privacy concerns, FL is mainly useful to support the diagnosis and detection of COVID-19. This was done by coordinating massive hospitals to build a common AI model [53].

We now give a general formal definition of FL. Let \( N \) be the number of data clients for example hospitals, \( \{H_1, H_2, \ldots, H_N\} \), all of whom wish to contribute to the training of a predictive model by combining their respective data \( \{D_1, D_2, \ldots, D_N\} \). A traditional method is to combine all data and use \( D = D_1 \cup D_2 \cup \ldots \cup D_N \) to train a model \( M_{all} \). FL is a learning process in which the data clients collaboratively train a model \( M_{FL} \), in which process any data client \( H_i \) does not share its data \( D_i \) to others. In addition, the accuracy of \( M_{FL} \), denoted as \( V_{FL} \), should approach the performance of \( M_{all} \), \( V_{all} \). Formally, let \( \delta \) be a non-negative real number, if
\[
|V_{FL} - V_{all}| < \delta. \tag{1}
\]
Then, the federated learning algorithm has an accuracy loss of \( \delta \).

FL is particularly attractive for smart healthcare, by coordinating multiple hospitals to perform an AI training without sharing raw data [54]. However, FL is an emerging research area and has not yet gained much trust in the medical community; therefore, it has required more research, particularly with respect to security and privacy aspects. Although this framework sounds ideal in theory, it is not immune to attacks. Moreover, its current development is not mature enough to be expected to solve all privacy issues by default. The most specific security threats facing the FL framework are communication bottlenecks, poisoning, backdoor attacks, and inference-based attacks, which can be considered the most critical to FL privacy [20]. In this context, FL is represented in three main groups a) privacy protection [55], b) effective communication [56] and c) data heterogeneity [57].

1) **Privacy Protection**: For generalization of the predictive model in medical data analysis, recent work focuses on considering multisite samples with multidimensionality. However, data sharing between institutions has been controlled by laws such as the United States Health Insurance Circulation and Accountability Act (HIPAA) [58] and the European General Data Protection Regulation (GDPR) [59]). FL demonstrates its feasibility in sharing and protecting the privacy of 20 medical institutions and centers [60]. FL is represented by models that are trained simultaneously at each site. These models are periodically aggregated and redistributed, requiring only the weight transfer of learned model between institutions. This step eliminates the requirement of sharing data directly [52]. Although FL can efficiently protect the data privacy, their application still bares some security-related risks [61].

2) **Effective Communication**: Communication is a critical bottleneck in federated networks [62]. FL network can consist of thousands of agents contributing model updates. To obtain high-performance models, the multi-participant situation is inevitable. In this context, network communication is slower than local computation by many orders of magnitude due to limited resources such as bandwidth, energy, and power. To solve this problem, two key aspects can be used: a) reducing the total number of communication rounds and b) reducing the size of the messages transmitted in each round [63]. It can perform client-side filtering, reduce the frequency of updates, and P2P (peer-to-peer) learning (e.g., compression) [64].

3) **Data Heterogeneity**: Currently, data samples in hospitals are not independent and distributed identically (IID). It leads to an increase in the bias of the model [65]. A proposed initial shared model used for hospitals that can easily adapt it to the local dataset [66]. Through intelligently choosing the participant, it can reduce the times of communication in FL and counterbalance the bias introduced by IID data. However, the data are not always IID type. To systematically and intuitively understand the current FL method in different data distributions, new studies have been conducted a comprehensive experimental comparison between the new algorithms [67]. This means that FL can solve the problem of multisite data, such as in multiple clinical institutions [68].

Regardless of these advanced FL techniques, more efforts are needed to consider these models in healthcare field.

III. METHODS

Deep learning models are widely applied in medicine. However, these models require more interpretability to simplify the steps of the prediction process. Despite advances in XAI models, there is a huge need to consider XAI with DA and FL when the topics are related to medicine to achieve the goal of personalized medicine. For example, DA-based XAI models have the ability to improve the performance in computer aided diagnosis [69]. But the same model will not be able to provide a feasible performance using multi-institutional data. For this reason, FL is a good option for data sharing under the condition of protecting data privacy [70]. However, there is an additional issue related to intra- and inter-site difference of distributions. This leads to a domain shift between sites. Recent work proposes to integrate unsupervised domain adaptation (UDA) into the FL framework [71]. For example, UDA guides the model to learn domain-agnostic features through adversarial learning or a specific type of batch normalization [47]. Furthermore, DA techniques such as mixture of experts (MoE) with adversarial domain alignment can be used to improve the accuracy of different sites in the FL learning setup [72]. These techniques will be presented below in three subsections related to DA, XAI and FL.

A. **Current Methods of Explainable Artificial Intelligence**

With high-performance deep learning models of increasing complexity, the way specific results were obtained is not accessible to humans. A number of methods were proposed to
increase explainability in recent years, 17 of which were discussed in a recent survey [73] and a selection of which is listed in Table I and discussed in following paragraphs. These techniques are mainly divided into two categories, i) explainability in modeling and ii) explainability after modeling (see Fig. 3).

1) Explainability in Modeling: Explainability is derived from modifications of the model architecture. This process can be represented in simplification (and/or quantification) layers and loss function, or by adding an explainable network and visualization module. For example, CNN architectures consist of convolutions, pooling, and fully connected layers; To increase the explainability of CNN, layers could be quantified, modified, and combined with a visual model. Replacement of convolutional layers with max-pooling layers provides better visualization without losing performance metrics [74]. Another scenario related to CNN is represented by adding attention network and modifying the loss function [75]. Recent studies have been quantified the convolution layers by Gaussian mixture model and/or texture features for clinical classifications [76]–[81]. However, there are only a few studies focused on explainable hidden layers.

2) Explainability After Modeling: It requires many experiments to balance model accuracy based on post-hoc interpretability (explainability after modeling). It is grouped into the following categories:

a) Feature relevance analysis: It considers the importance of features in the prediction model (e.g., layer-wise relevance back propagation (LRP)). LRP is an explainable AI model that explores the relation between output and input data by generating separate correlation maps for each individual image. LRP with explainability concept is also used to prune CNN. It finds the most relevant units (i.e., weights or filters) with their relevance scores [82]. Deep-LIFT is also a method to improve the effectiveness of feature learning and improve the interpretability of the model itself. It works by comparing the activation of each neuron to its reference activation and assigns scores according to the difference between the two [83], [84].

b) Visual explanation: To further improve model visualization, heat maps, saliency maps, or class activation methods are

| Application | XAI / Explainability level | Prediction model | Reference |
|-------------|---------------------------|-----------------|-----------|
| Preclinical relevance assessment | Model combination / G | GNN, FNN | [89] |
| Detection and prediction for Alzheimer’s disease | SHAP / L & G | RF | [90] |
| Deterioration risk prediction of hepatitis | SHAP, LIME, PDP / L & G | RF | [91] |
| Non-communicable diseases prediction | DeepSHAP / L & G | DNN | [92] |
| Spinal posture classification | LIME / L | SVM | [93] |
| Classification of estrogen receptor status | SmoothGrad / L | DCNN | [94] |
| Differential diagnosis of COVID-19 | Grad-CAM / G | CNN | [95] |
| Prediction of mortality | Shapley additive / L | RNN | [96] |
| Medical image segmentation | Combination / L & G | CNN, Attention Mechanism | [97] |
| Early detection of Parkinson’s disease | LIME / L | VGG-16 | [98] |
| COVID-19 diagnosis | Grad-CAM ++, LRP / L | DNN | [99] |
| Evaluation of cancer in Barrett’s esophagus | 5 XAI methods / L | AlexNet, SqueezeNet, ResNet50, VGG16 | [100] |
| Heart anomaly detection | SHAP and Occlusion maps / L | CNN, MLP | [101] |
| Distinguish VTE patients | DeepLIFT / L | ANN | [102] |
| Detection of ductal carcinoma in situ | LRP / L | ResNet-50 CNN | [103] |
| Lesion prediction | Activation Maps, Saliency Maps / L | U-net | [104] |
| Early detection of COVID-19 | Saliency Maps / L | Fuzzy-enhanced CNN | [105] |
| EEG emotion recognition | SmoothGrad, saliency maps / L | CNN | [106] |
| Cerebral hemorrhage detection and localization | Grad-CAM / L | ResNet | [107] |
| Macular disease classification | t-SNE, model simplification / L | CNN | [108] |
| Automated diagnosis and grading of ulcerative colitis | Grad-CAM / L | CNN | [109] |
| Diagnostic for breast cancer | Causal-TabNet / G | GNN | [110] |
| Prediction of depressive symptoms | LIME / L | RNN | [111] |
| Early prediction of antimicrobial multidrug resistance | SHAP / L & G | LSTM | [112] |
| Predictions of clinical time series | CAM / Local | FCN | [113] |

L: Local; G: Global; ANN: Artificial Neural Network; GNN: Graph Neural Networks; FNN: Fuzzy Neural Network; DNN: Deep Neural Networks; RNN: Recurrent Neural Network; FCN: Fully Convolutional Network; DCNN: Deep Convolutional Neural Network; VGG-16: GG-Very-Deep-16 CNN; VGG-19: GG-Very-Deep-19 CNN; Fuzzy-enhanced CNN: Fuzzy-enhanced Convolutional Neural Network; RF: Random forest; SVM: Support Vector Machine; MLP: Multi-Layer Perceptron; LSTM: Long Short-Term Memory network; PDP: Partial Dependence Plot; t-SNE: t-distributed Stochastic Neighbor Embedding. Seven Network: SqueezeNet, Inception, ResNet, ResNeXt, Xception, ShuffleNet, DenseNet; Five XAI methods: Saliency, guided backpropagation, integrated gradients, input gradients, and DeepLIFT.
Fig. 3. Example of flowchart for the explainable artificial intelligence model, 1) data acquisition consists of images and clinical report for each of patients. 2) Example of four methods during modeling: it can modify layer (Layer modification) to simplify the structure or combined (Model combination) with other model to build a full explainable model. In addition, an attention network could be used for visualization and even to modify the loss function. 3) The two most used methods after modeling: feature relevance analysis and visual explanation.

3) Explainability Level: The explainability scale is divided into different categories in different situations. For example, XAI is divided into five scales according to ethology in explaining animal intentionality [88]. According to the transparency of the model, XAI is divided into simulatability, decomposability, and algorithmic transparency [25]. Furthermore, XAI with deep learning models could be grouped into a) local and b) global explanations. For local explanation, the model considers a local approximation of how the black box works. It mainly focuses on explanation of data instances. The global explanation describes the whole mechanism of the model. Generally, a group of data instances is needed to generate one or more explanation maps.

Table I summarizes XAI in deep learning models applied to clinical tasks. We note that CNNs are the most widely used deep learning models and gradient-based explainability is the most applied principle. The level of explainability is mostly local.

Agnostic model In addition to the previous XAI methods, shapley additive explanations (SHAP) and local explainable model-agnostic explanation (LIME) are agnostic to the deep learning models they are applied to. Specifically, LIME uses explainability models (such as linear models and decision trees) to approximate the target predictions from the black-box model. It detects the relationships between the output of the black-box model by slightly perturbing the input in the XAI model. For example, LIME demonstrated feasible explanations in artificial neural networks by altering the input feature and fitting it to a linear model [114]. The SHAP explains how each feature affects the predicted value (known as Shaply). It is computed based on the average of the marginal contributions of all permutations (permutations of each feature). It provides different explanations for a particular model (e.g., feature importance visualization, linear formulation). With SHAP [115], it illustrates how the design parameters affect performance using DNN features.

B. Domain Adaptation With Deep Learning

We present the main steps in DA model by summarizing the existing literature. We grouped the existing methods into 1) DA-related data, 2) DA-related features, and 3) building and adversarial domain adaptation. In addition, we report these models in Table II. We note that the problem of domain bias occurs when the target domain is different from the source domain. This bias type limits the performance of transfer deep learning model that leads to disable the generalization of model.

1) Domain Adaptation Related Data: It assumes that DA can solve the problem of domain shift. However, there is a bias due to different hospital equipments. Instance Re-weighting Adaptation is a method to solve this issue when the domain bias is small. It removes samples from the source domain that mismatch the similar distribution in the target domain. For example, samples with higher correlation with target samples are assigned larger weights [116]. The assigned
data are then re-entered into the training model to reduce the problems arising from domain shifts [117]. This idea has been developed with a task of re-weighting mechanism that considers loss function with dynamically update weights [118]. Regardless of the advances algorithms and methods used in DA related data, these models are able to only solve the problem of small domain bias.

2) Domain Adaptation Related Features: It is based on the feature similarity between the source and the target domains. For example, the feature transformation strategy can solve the bias problem by transferring the source and target samples from the original feature space to the new feature representation space [119]. A novel feature-based transfer learning algorithm for remote domains requires only a small portion of labeled target samples from a different domain [120]. However, transfer features can only reduce the distribution variance from a single perspective, and it is a challenge to find an optimal transfer learning method for a given dataset. For this reason, the distribution variance was reduced from multiple perspectives by applying multi-feature-based transfer learning methods [121].

3) Building an Adversarial Domain Adaptation: Data- and feature-based DA aims to align a labeled source domain and an unlabeled target domain, which requires accessing the source data. It will raise concerns about data privacy, portability, and transmission efficiency. Therefore, a method that can adapt source-trained models towards target distributions without accessing source data will be promising. To adapt a model, samples from the source and target domains can share similar representations to provide similar predictions [122]. These models need to be pre-trained using the source domain data and re-trained by the target domain [122], which will increase the processing time. Therefore, these types of predictive models are not preferred. Another scenario based on adversarial domain adaptation is currently one of the mainstream methods. It is known that deep models (e.g., adversarial domain adaptation) can use the transformed features for domain adaptation [123]. For example, an unsupervised method based DA used for predicting glaucoma fundus [124]. Specifically, they designed a loss function that learns the source region features and possesses the original labels unchanged. An improved adversarial domain adaptation network has been considered for tumor image diagnosis with few and/or no labels [125].

Table II reports the DA methods used for clinical applications. We find that the dominant research approaches in recent years are related to adversarial networks, with a lower dominant for DA methods based on features and data. It observes that the most of these approaches based on unsupervised DA. This is due to the limitation of medical data. These contributions provide a solid foundation for future DA methods applied in medicine.

C. Federated Learning Modeling

FL updates the models to achieve high performance by considering co-modeling step between different data sources (or client). FL is divided into the following categories (see Fig. 4):

1) Horizontal Federate Learning: Horizontal federated learning (HFL) is a system in which all the parties share the same feature space. It does the training process as follows: a) Participants first download the latest model from the server individually and train the model based on the local data, b) The trained model with its parameters (i.e., cryptographic gradients) is uploaded to the server, and c) The server aggregates the gradients of each user and returns the updated model to each participant to update local models [154]. With this advantage, HFL is widely used in clinical topics [155].

2) Vertical Federated Learning: Vertical federated learning (VFL) lets multiple parties that possess different attributes (e.g., features and/or labels) of the same data entity (e.g., a person) to jointly train a model. Its learning process consists of a) encrypted sample alignment and b) an encryption training model. a) Encrypted sample alignment: Using encryption-based user ID alignment technology to align common users. These common users data are then considered for b) encryption training model. However, in learning process, VFL requires multiple communications between all participants that need a high processing time. For this reason, many studies have developed an asynchronous training algorithm based on vertically partitioned data [156]. We note that many existing methods focus on HFL when all feature sets and labels are available.
3) **Federated Transfer Learning**: Federated transfer learning (FTL) generalizes the concept of federated learning and emphasizes that collaborative modeling learning can be performed on any data distribution and entity [157]. For example, participants first compute and encrypt their intermediate results with their gradients and losses. The gradients and losses are then collected and decrypted by a collaborator. Finally, the participants receive the aggregated results that are used to update the respective models. We note that FTL uses homomorphic encryption for security and the polynomial difference approximation for a) privacy and b) prevention of data leakage [154]. Unlike HFL that possesses the original data and models locally, FTL makes the participants encrypt the gradients before data transmission, and adding random masks to avoid the participants from getting information of each other's. In healthcare applications, FTL can not only provide a standard framework, but also achieve efficient classifications and avoid the problem of unbalanced classes [158].

4) **Federated Optimization**: Compared with traditional machine learning, FL focuses on significant variability of system features on each device in the network (systems heterogeneity), and non-identically distributed data across the network (statistical heterogeneity) [159]. Furthermore, aggregation FL requires defining an aggregation strategy, such as a method to combine the local models coming from the clients into a global one. For example, the standard and simplest aggregation strategy is federated averaging (known as FedAvg [160]). Another strategy is called FedProx [161] is a generalization of FedAvg with some modifications to address data and the heterogeneity of the system. Furthermore, FedProx is considered to explore the interactions between statistical and systematic heterogeneity [159]. In addition, there is a new version of FedAvg called FedAvg+ based on personalization and cluster FL [162]. Also, FedEM model is based on the mixed distribution assumption with the expectation maximum algorithm [163]. Unfortunately, these optimization algorithms are still rarely applied in healthcare topics.

Table III reports FL techniques used recently for healthcare data. It observed that the accuracy of FL model under the optimization algorithm is close to the central model, which demonstrates the potential of federated learning. However, the number of clients studied is still limited to 100 [148]. Moreover, most of the data from these clients are derived from private datasets or multi-site datasets, which will limit the development of FL in practical medical applications [164]. So far, FL for healthcare is still in its early stages. The preliminary investigations provide a good foundation for future work, regardless of whether they involved replicating FL environments or doing limited experiments across hospitals [165].

**D. Overview of Methods and Datasets Used for XAI, DA, and FL**

Table IV compares methods of XAI, DA and FL based on key strengths and weaknesses and lists code and datasets within each topic. To overcome the limited availability of medical datasets, many of the XAI, DA and FL based on deep learning techniques employ transfer learning from a large dataset such as ImageNet [166]. Thus, we include the original datasets and code used for the recent and popular techniques. For example, Grad-CAM is one of the common XAI methods used for multitask in medical topics like grading of ulcerative colitis [109], detecting cerebral hemorrhage [167], and predicting COVID-19 [67], [168] with a high degree of image interpretation. LIME also shows a feasible rate in multitask such as spinal posture classification [93], detection of Parkinson’s disease [98], and predicting depressive symptoms [111]. Many other techniques used recently (e.g., Shapley Additive Explanations, Graph LIME, etc.) are almost based on deep learning models that need high power and performance computers.

DA in high-dimensional complex data like medical imaging relies on deep learning to implicitly model nonlinear transformations [169]. For example, the difference of the feature distribution between source and target domains is narrowed by the deep domain confusion model for cross subject recognition [170]. Recent algorithms like deep CORAL [171], deep adaptation networks [172], deep subdomain associate adapta-
TABLE III
SUMMARY OF FEDERATED LEARNING MODELS USED FOR CLINICAL TOPICS

| Application                        | Model / Learning / optimization | n   | Performance (Local / Central / FL models) | Reference |
|------------------------------------|---------------------------------|-----|------------------------------------------|-----------|
| Predicting clinical outcomes       | EXAM / HFL / ———                | 20  | AUC: 0.79 / — / 0.92                     | [139]     |
| Precision medicine                 | RETAIN / HFL / FedAvg            | 42  | AUC: — / 0.73 / 0.62                     | [140]     |
| Spinal cord gray matter segmentation | DNN / HFL / FedAvg / FedProx     | 4   | Sensitivity: — / 0.76 / 0.78             | [141]     |
| Prostate cancer diagnosis          | 3D AH-Net / HFL / ———           | 3   | Dice: 0.812 / 0.872 / — / 0.89          | [52]      |
| Detection diabetic retinopathy     | AlexNet / FTL / FedAvg / FedProx | 5   | Acc: 0.81-0.92 / — / 0.65-0.90          | [142]     |
| Prediction in COVID-19 patients    | LASSO / MLP / HFL / ———         | 5   | AUC: — / 0.719 / 0.822 / 0.786-0.836    | [143]     |
| Image reconstruction               | U-net / HFL / FedAvg             | 4   | SSIM: 0.94 / — / 0.93                    | [144]     |
| Computational pathology            | CNN / HFL / ———                | 3   | AUC: — / 0.985±0.004 / 0.976±0.007       | [145]     |
| Predicting outcomes in patients    | ResNet-34 / HFL / FedAvg         | 20  | AUC: 0.79 / — / 0.92                     | [146]     |
| Multimodal melanoma detection      | EfficientNet / HFL / FedAvg      | 5   | Acc: — / 0.83 / 0.83                     | [147]     |
| Arrhythmia detection of Non-IID ECG| ——— / HFL / EWC                 | 100 | F1-score: — / — / 0.80                   | [148]     |
| Joint training                     | ——— / HFL / FedACS              | 3   | Acc: — / — / 0.70                       | [149]     |
| MultiView training                 | CNN / VFL / ———                | 2   | Acc: 0.62 / — / 0.71                     | [150]     |
| Prognostic prediction              | ELM / HFL / ———                | 4   | Acc: — / 0.89 / 0.8899                   | [151]     |
| Designing ECG monitoring healthcare system | CNN / FTL / ———        | 3   | Acc: — / 0.94 / 0.92                     | [70]      |
| Multi-center imaging diagnostics   | 3D-CNN, ResNet-18 / HFL / ———   | 4   | AUC: — / 0.85 / 0.86                     | [152]     |
| COVID-19 detection                 | ResNet / HFL / FedMoCo           | 3   | Acc: — / 0.96 / 91.56                    | [153]     |

Local model: training alone in client; Central model: use all client data for training; FL model: model under FL training methods; n: number of institutions; EXAM: Electronic Medical Record (EMR) chest X-ray AI model; HFL: Horizontally Federated Learning; VFL: Vertically Federated Learning; FTL: Federated Transfer Learning; CNN: Deep Neural Networks; CNN: Convolutional Neural Network; 3D-CNN: three-Dimensional Convolutional Neural Network; AlexNet, EfficientNetB7, ResNet, ResNet-34, ResNet-18, U-net: a class in CNN; AUC: Area Under Curve; Acc: Accuracy; 3D AH-Net: 3D Anisotropic Hybrid Network; LASSO: Least Absolute Shrinkage and Selection Operator; MLP: Multi-layer Perceptron; Dice: Dice coefficient range; Kaapa score: is a measure of classification accuracy; SSIM: Structural Similarity Index Measure; PSNR: Peak-Signal-to-Noise Ratio; ELM: Extreme Learning Machine; RETAIN: has two parallel RNN branches merged at a final logistic layer; EWC: elastic weight consolidation; FedAvg: Federated Averaging; FedACS: FedAvg with adaptive client sampling; FedMoCo: a robust federated contrastive learning framework; FedProx: A optimal method to tackle heterogeneity in federated networks.

Recent studies of FL provide promising results in healthcare applications. For example, data from 20 institutes across the globe to train a model (EXAM: electronic medical record (EMR) chest X-ray AI model), for predicting COVID-19 [139]. This model is based on HFL that considers differential privacy to mitigate risk of data “interception” during site-server communication. In [142], three main models using standard transfer learning, FedAvg, and Federated Proximal frameworks, respectively, are considered to diagnose diabetic retinopathy. In [176], a new study proposes federated disentangled representation learning for unsupervised brain anomaly detection using MR scans from four different institutions. To solve the problems of security and trustworthiness, densely connected CNN based on FedAvg is proposed and used for COVID-19 [177]. Due to the limitations of sharing the data across hospitals and countries, more collaborations will improve the feasibility of FL in medical applications.

IV. CONCLUSION AND FUTURE PERSPECTIVE

The rate at which the complexity and volume of digital patient data are acquired, far exceeds the rate at which expert clinicians can be trained to analyze it. Therefore, it is necessary to employ artificial intelligence to keep up with the growth of the data. Deep learning based image analysis reached the performance of expert clinicians in making referral recommendations based on widespread available optical coherence tomography [194]. In many scenarios, however, artificial intelligence is not yet readily able to replace expert clinicians because the applicable AI methods are not plug-in replacements for human decision-making. In neuroimaging, for example, the transition to the clinics did not take place despite many successes in research studies [195]. Applications of AI with a human in the loop relieve the clinician from some—typically tedious work. Atri-U, for example, is an application of AI in the clinical routine in which the AI back-end proposes a segmentation, two landmarks, and a time point to compute the volume of the left atrium [196]. The clinician that bares the responsibility can either accept, modify, or ignore the proposals. The latter scenario introduces a non-disruptive innovation into an existing diagnostic process with a substantially reduced average process time without compromising quality. Further developments of such AI tools for clinical practice will depend on the key technologies discussed previously, and for each of which we will conclude with a summary.

1) Explainable AI: We found that the most XAI with deep learning addresses post-hoc interpretability to increase face value, like the CNN-CAM. We also found that the explainability levels are limited local explainability. This means that...
TABLE IV
THE STRENGTHS AND WEAKNESSES OF XAI, DA, AND FL TECHNIQUES

| Technique | Strengths | Weaknesses | Code | Data | Reference |
|-----------|-----------|-----------|------|------|-----------|
| Gradient-weighted Class Activation Mapping | - Simple to use  
- No need to change the network structure  
- Efficient to compute  
- High degree of image explanation | - The weight of saliency map depends on the model and data set  
- Sometimes prominent areas cannot be correctly labeled  
- Poor robustness | C1 | D1, D2 | [67], [86], [109], [167], [168] |
| Local Explainable Model-agnostic Explanation | - Uses a simple model for local explanation | - Poor robustness  
- Inefficient calculation | C2 | D3 | [93], [98], [111], [178] |
| Graph LIME | - Simple to use  
- Trustworthy | - Inefficient calculation  
- Instable | *C3 | D4 | [91], [179] |
| Shapley Additive Explanations | - Useful for interpretation  
- Repeatability | - Inefficient computation | C4 | D5, D6 | [96], [180] |

Domain Adaptation

| Technique | Strengths | Weaknesses | Code | Data | Reference |
|-----------|-----------|-----------|------|------|-----------|
| Deep Domain Confusion | - Reduce the distribution difference between the source and target domains | - Only adapts to one layer of network  
- A single fixed kernel may not be the optimal kernel | C5 | D7 | [170], [181] |
| Deep Adaptation Networks | - Reduce domain differences | - Domain difference problem cannot be solved by semi-supervised learning | C5 | D7 | [182], [183] |
| Deep Subdomain Adaptation Network | - The weight could be defined flexible | - The weight value of network maybe could not be the best | C5 | D8 | [184], [185] |
| Dynamic Adversarial Adaptation Networks | - Solve the dynamic distribution adaptation problem in adversarial networks | - Can not handle regression problems | C5 | D7 | [186], [187] |

Federated Learning

| Technique | Strengths | Weaknesses | Code | Data | Reference |
|-----------|-----------|-----------|------|------|-----------|
| EXAM / HFL | - Can use large, and heterogeneous data-sets  
- Robust and generalizable | - Bias may occur due to limitations in data quality. | C6 | D9 | [139] |
| AlexNet / FTL / FedAvg & FedProx | - Prediction rate is high | - Sub-optimal security  
- Data accuracy pitfalls | – | D10–D12 | [142] |
| CNN / HFL / FedDis | - Mitigate the statistical heterogeneity  
- Good generalizability with clinical applications  
- Flexibility | - Privacy concerns | C7 | D13–D17 | [176] |
| 3D-CNN / HFL / FedAvg | - High performance metrics  
- Provided visual explanations  
- Suitable for transfer learning | - Weak federated training process for the unstable internet connection.  
- Inefficient computation | C8 | D18 | [177] |

FedDi: Federated Disentred representation learning for unsupervised brain operatory segmentation; EXAM: Electronic Medical Record (EMR) chest X-ray AI model, "-" not available, "*" not official, C: code, D: data.

C1 https://github.com/zhoubolei/CAM, https://github.com/Cloud-CV/Grad-CAM
C2 https://github.com/marcotcr/lime
C3 https://github.com/WilliamCCHuang/GraphLIME
C4 https://github.com/slundberg/shap
C5 https://github.com/ting2696/Deep-Symmetric-Adaptation-Network.
C6 https://github.com/jindongwang/transferlearning/tree/master/code/DeepDA/, https://github.com/Cloud-CV/GraphLIME
C7 https://github.com/HUST-EIC-AI-LAB/UCAIDI
D1 http://host.robots.ox.ac.uk/pascal/VOC/voc2007
D2 ImageNet Object Localization Challenge [166]
D3 https://github.com/marcotcr/lime-experiments
D4 Datasets are from Cora and Pubmed.
D5 [188]
D6 http://yann.lecun.com/exdb/mnist
D7 [189]–[193], ultrasoundcases.info
D8 https://github.com/jindongwang/transferlearning/blob/master/data/dataset.md, [186]
D9 https://www.kaggle.com/datasets/mariaherrerot/eyepacspreprocess
D10 https://www.adni.loni.usc.edu/data-553samples/access-data/
D11 https://ciee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-rid
D12 https://www.kaggle.com/datasets/mariaherrerot/aptos2019
D13 https://www.oasis-brains.org
D14 http://adni.loni.usc.edu/data-553samples/access-data/
D15 http://lit.fci.uj.i-si.tools.php?lang=en
D16 https://smart-stats-tools.org/lesion-challenge-5562015
D17 https://www.med.upenn.edu/sbia/brats2018/data.html
D18 https://www.nhsx.nhs.uk/covid-19-response/data-and-covid-19/national-covid-19-chest-imaging-database-nccid/

the model understanding can only explain the model output rather than the internal decision-making. Global explainability is desirable in clinical tasks to achieve trust. More specifically, the practical XAI application is inevitable to consider people who do not have the relevant AI background. Additional explanations such as manual instructions will be greatly recommended.

2) Domain Adaptation: Domain adaptation can effectively promote the transfer of models between different domains. However, data heterogeneity and label validity are still a critical challenge. Many DA methods can eliminate the domain bias and the difference in data distribution, such as data-based methods and feature-based methods. Unfortunately, these methods are too rigid and require a lot of processing time.
Unlike adversarial DA, which shows strong performance, it can effectively reduce domain differences and improve generalization performance.

3) Federated Learning: Federate learning is not well applied to healthcare data, and most studies focus on horizontal federated learning, which does not fully exploit the potential of federated learning. Only few optimization algorithms can achieve performance similar to the central model. We note that FL faces the challenges of poor data quality and insufficient model accuracy. This is due to the absence of standard and uniform data collection among institutions. Therefore, it mentions that the standard unified data with model security protection mechanism is the key to improving the deep learning model for clinical tasks.

4) Outlook to Further Challenges: In the past decade, many methods originating in computer vision and other non-medical fields found applications in medical research healthcare with some delay and minor adaptations. If this trend continues, we expect to see more of the emerging XAI, DA, and FL methods in medicine for tasks that are similarly structured as in non-medical fields. We note, however, that healthcare processes often incorporate irregularly sampled, longitudinal, multi-modal data for decision-making. The imputation of missing data can lead to inaccurate prediction that may result in biased estimations [197], [198]. With strong assumptions and simplified models, clinical decision processes can be modeled and optimized using mixed linear models that naturally handle missing data, as long as the input data is low-dimensional and in a homogeneous domain [199]. To train deep learning models with similar flexibility but more complex data, much larger training data sets are required. Combined, FA may contribute to leverage massive pools of private data that will be transformed into a common reference domain with DA to train interpretable and thus more trustworthy AI models using XAI.

REFERENCES

[1] S. C. Shapiro, Encyclopedia of Artificial Intelligence. 2nd ed. New York, USA: Wiley, 1992.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet large scale visual recognition challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, Dec. 2015.

[3] H. D. Li, J. C. Li, X. M. Guan, B. H. Liang, Y. T. Lai, and X. L. Luo, “Research on overfitting of deep learning,” in Proc. 15th Int. Conf. Computational Intelligence and Security, Macao, China, 2019, pp. 78–81.

[4] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danilhika, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsik, J. Molloy, T. L. Paine, C. Gulcehre, Z. Y. Wang, T. Pfaff, Y. H. Wu, R. Ring, D. Yokogawa, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver, “Grandmaster level in StarCraft II using multi-agent reinforcement learning,” Nature, vol. 575, no. 7782, pp. 350–354, Nov. 2019.

[5] A. Nguyen, J. Yosinski, and J. Clune, “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Boston, USA, 2015, pp. 427–436.

[6] G. Varoquaux and V. Cheplygina, “Machine learning for medical imaging: Methodological failures and recommendations for the future,” NPJ Digit. Med., vol. 5, no. 1, p. 48, Apr. 2022.

[7] M. Gaur, K. Faldu, and A. Sheth, “Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable?” IEEE Internet Comput., vol. 25, no. 1, pp. 51–59, Jan.-Feb. 2021.

[8] A. Adadi and M. Berrada, “Explainable AI for healthcare: From black box to interpretable models,” in Embedded Systems and Artificial Intelligence, V. Bhateja, S. C. Satapathy, and H. Satori, Eds. Singapore, Singapore: Springer, 2020, pp. 327–337.

[9] G. Yang, Q. H. Ye, and J. Xia, “Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond,” Inf. Fusion, vol. 77, pp. 29–52, Jan. 2022.

[10] A. Holzinger, “The next frontier: AI we can really trust,” in Proc. Joint European Conf. Machine Learning and Knowledge Discovery in Databases, 2021, pp. 427–440.

[11] A. Holzinger, M. Dehmer, F. Emmert-Streib, R. Cucchiara, I. Augenstein, J. Del Ser, W. Samek, I. Jurisica, and N. Díaz-Rodríguez, “Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence,” Inf. Fusion, vol. 79, pp. 263–278, Mar. 2022.

[12] H. P. Chan, L. M. Hadijsi, and R. K. Samala, “Computer-aided diagnosis in the era of deep learning,” Med. Phys., vol.47, no.5, pp. e218–e227, Jun. 2020.

[13] H. Guan and M. X. Liu, “Domain adaptation for medical image analysis: A survey,” IEEE Trans. Biomed. Eng., vol. 69, no. 3, pp. 1113–1185, Mar. 2022.

[14] G. Bode, S. Thul, M. Baranski, and D. Müller, “Real-world application of machine-learning-based fault detection trained with experimental data,” Energ, vol.198, p. 117323, May 2020.

[15] Y. G. Lei, B. Yang, X. W. Jiang, F. Jia, N. P. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: A review and roadmap,” Mech. Syst. Signal Process., vol.138, p. 106587, Apr. 2020.

[16] C. C. Zhang and Q. J. Zhao, “Deep discriminative domain adaptation,” Inf. Sci., vol.575, pp. 599–610, Oct. 2021.

[17] Y. F. Zhang, Y. Wei, Q. Y. Wu, P. L. Zhao, S. C. Niu, J. Z. Huang, and M. K. Tan, “Collaborative unsupervised domain adaptation for medical image diagnosis,” IEEE Trans. Image Process., vol.29, pp. 7834–7844, Jul. 2020.

[18] P. Yang, X. N. Xiong, and J. L. Ren, “Data security and privacy protection for cloud storage: A survey,” IEEE Access, vol.8, pp.131723–131740, Jul. 2020.

[19] C. Zhang, Y. Xie, H. Bai, B. Yu, W. H. Li, and Y. Gao, “A survey on federated learning,” Knowl.-Based Syst., vol.216, p. 106775, Mar. 2021.

[20] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantahana, and G. Srivastava, “A survey on security and privacy of federated learning,” Future Gener. Comput. Syst., vol.115, pp.619–640, Feb. 2021.

[21] Y. G. Lei, B. Yang, X. W. Jiang, F. Jia, N. P. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: A review and roadmap,” Mech. Syst. Signal Process., vol.138, p. 106587, Apr. 2020.

[22] C. Zhang, Y. Xie, H. Bai, B. Yu, W. H. Li, and Y. Gao, “A survey on federated learning,” Knowl.-Based Syst., vol.216, p. 106775, Mar. 2021.

[23] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantahana, and G. Srivastava, “A survey on security and privacy of federated learning,” Future Gener. Comput. Syst., vol.115, pp.619–640, Feb. 2021.

[24] Y. G. Lei, B. Yang, X. W. Jiang, F. Jia, N. P. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: A review and roadmap,” Mech. Syst. Signal Process., vol.138, p. 106587, Apr. 2020.

[25] C. Zhang, Y. Xie, H. Bai, B. Yu, W. H. Li, and Y. Gao, “A survey on federated learning,” Knowl.-Based Syst., vol.216, p. 106775, Mar. 2021.

[26] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantahana, and G. Srivastava, “A survey on security and privacy of federated learning,” Future Gener. Comput. Syst., vol.115, pp.619–640, Feb. 2021.

[27] Y. G. Lei, B. Yang, X. W. Jiang, F. Jia, N. P. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: A review and roadmap,” Mech. Syst. Signal Process., vol.138, p. 106587, Apr. 2020.
features from MRI scans predict survival outcome of recurrent A. Chaddad, M. L. Zhang, L. Hassan, and T. Niazi, “Modeling of textures to predict immune cell status and survival of brain tumour patients,” in Proc. IEEE 18th Int. Symp. Biomedical Imaging, Nice, France, 2021, pp. 1067–1071.

[80] A. Chaddad, L. Hassan, and C. Desrosiers, “Deep Radiomic analysis for predicting coronavirus disease 2019 in computerized tomography and X-ray images,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 1, pp. 3–11, Jan. 2022.

[81] A. Chaddad, P. Daniel, M. L. Zhang, S. Rathore, P. Sargos, C. Desrosiers, and T. Niazi, “Deep radiomic signature with immune cell markers predicts the survival of glioma patients,” Neurocomputing, vol. 469, pp. 366–375, Jan. 2022.

[82] S. K. Yeom, P. Seegerer, S. Lapuschkin, A. Binder, S. Wiedemann, K. R. Müller, and W. Samek, “Pruning by explaining: A novel criterion for deep neural network pruning,” Pattern Recognit., vol. 115, p. 107899, Jul. 2021.

[83] J. B. Li, C. Q. Zhang, J. T. Zhou, H. Z. Fu, S. Y. Xia, and Q. H. Hu, “Deep-LIFT: Deep label-specific feature learning for image annotation,” IEEE Trans. Cybern., vol. 52, no. 8, pp. 7732–7741, Aug. 2022.

[84] A. Shrikumar, P. Greenside, and A. Kundaje, “Learning important features through propagating activation differences,” in Proc. 34th Int. Conf. Machine Learning, Sydney, Australia, 2017, pp. 3145–3153.

[85] M. Liu, X. Chen, and S. C. Yan, “Network in network,” arXiv preprint arXiv: 1213:4400, 2013.

[86] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Computer Vision, Venice, Italy, 2017, pp. 618–626.

[87] H. F. Wang, Z. F. Wang, M. N. Du, F. Yang, Z. J. Zhang, S. R. Ding, P. Mardziel, and X. Hu, “Score-CAM: Score-weighted visual explanations for convolutional neural networks,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops, Seattle, USA, 2020, pp. 111–119.

[88] R. Dazeley, P. Vamplew, C. Foale, C. Young, S. Aryan, and F. Cruz, “Levels of explainable artificial intelligence for human-aligned conversational explanations,” Artif. Intell., vol. 299, p. 103525, Oct. 2021.

[89] J. Jiménez-Luna, M. Skalic, N. Weskamp, and G. Schneider, “Coloring molecules with explainable artificial intelligence for preclinical relevance assessment,” J. Chem. Inf. Model., vol. 61, no. 3, pp. 1083–1094, Feb. 2021.

[90] S. El-Sappagh, J. M. Alonso, S. M. R. Islam, A. M. Sultan, and K. S. Kwak, “A multitlayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer’s disease,” Sci. Rep., vol. 11, no. 1, p. 2660, Jan. 2021.

[91] J. F. Peng, K. Q. Zou, M. Zhou, Y. Teng, X. Y. Zhu, F. F. Zhang, and J. Xu, “An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients,” J. Med. Syst., vol. 45, no. 5, p. 61, Apr. 2021.

[92] K. Davadgol, J. W. Bae, V. H. Pham, N. Theera-Umpon, and K. H. Rya, “Explainable artificial intelligence based framework for non-communicable diseases prediction,” IEEE Access, vol. 9, pp. 123672–123688, Sept. 2021.

[93] C. Dindorf, J. Konradi, C. Wolf, B. Taetz, G. Bleser, J. Huthwelker, F. Werthmann, E. Bartaguz, J. Kniepert, P. Drees, U. Betz, and M. Fröhlich, “Classification and automated interpretation of spinal posture data using a pathology-independent classifier and explainable artificial intelligence (XAI),” Sensors, vol. 21, no. 18, p. 6323, Sept. 2021.

[94] Z. Papanastasopoulos, R. K. Samala, H. P. Chan, L. Hadjiiski, C. Paramagul, M. A. Helvie, and C. H. Neal, “Explainable AI for medical imaging: Deep-learning CNN ensemble for classification of estrogen receptor status from breast MRI,” in Proc. SPIE 11314, Medical Imaging 2020: Computer-Aided Diagnosis, Houston, Texas, United States, pp. 113140P.

[95] R. K. Singh, R. Pandey, and R. N. Babu, “COVIDScreen: Explainable deep learning framework for differential diagnosis of COVID-19 using chest X-rays,” Neural Comput. Appl., vol. 33, no. 14, pp. 8871–8892, Jan. 2021.


[96] H. C. Thorsen-Meyer, A. B. Nielsen, A. P. Nielsen, B. S. Kaas-Hansen, P. Toft, J. Schierbeck, T. Strom, P. J. Chimura, M. Heimann, L. Dybdahl, L. Svangsgaard, P. Huslen, K. Belling, S. Brunak, and A. Perner, “Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: A retrospective study of high-frequency data in electronic patient records,” Lancet Digit. Health, vol. 2, no. 4, pp. e179–e191, Apr. 2020.

[97] R. Gu, G. T. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. T. Zhang, “CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation,” IEEE Trans. Med. Imaging, vol. 40, no. 2, pp. 699–711, Feb. 2021.

[98] P. R. Magesh, R. D. Mylooth, and R. J. Tom, “An explainable machine learning model for early detection of Parkinson’s disease using LIME on DeTSCAN imagery,” Comput. Biol. Med., vol. 126, p. 104041, Nov. 2020.

[99] R. Karim, T. Döhmern, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, and S. Decker, “DeepCOVIDExplainer: Explainable COVID-19 diagnosis from chest X-ray images,” in Proc. IEEE Int. Conf. Bioinformatics and Biomedicine, Seoul, Korea (South), 2020, pp. 1034–1037.

[100] L. A. Jr. de Souza, R. Mendel, S. Strasser, A. Ebibgo, A. Probst, H. Messmann, J. P. Papa, and C. Palm, “Convolutional neural networks for the evaluation of cancer in Barrett’s esophagus: Explainable AI to lighten up the black-box,” Comput. Med. Biol., vol. 135, p. 104578, Aug. 2021.

[101] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, “A robust interpretable deep learning classifier for heart anomaly detection without segmentation,” IEEE J. Biomed. Health Inform., vol. 25, no. 6, pp. 2162–2171, Jun. 2021.

[102] M. Razzaq, L. Gounidi, M. J. Iglesias, G. Munsch, M. Bruzelius, M. Ibrahim-Kosta, L. Butler, J. Odeberg, P. E. Morange, and D. A. Pomberger, “Explainable artificial neural network for recurrent venous thromboembolism based on plasma proteomics,” in Proc. 19th Int. Conf. Computational Methods in Systems Biology, Bordeaux, France, 2021, pp. 108–121.

[103] M. La Ferla, M. Montebello, and D. Seychell, “An XAI approach to deep learning models in the detection of ductal carcinoma in situ,” arXiv preprint arXiv:2106.14186, 2021.

[104] C. Gillmann, L. Peter, C. Schmidt, D. Saur, and G. Scheuermann, “Visualizing multimodal deep learning for lesion prediction,” IEEE Comput. Graph. Appl., vol. 41, no. 5, pp. 90–98, Sep.–Oct. 2021.

[105] C. Iercitano, N. Mammonte, M. Versaci, G. Varone, A. R. Ali, A. Armentano, G. Calabrese, A. Ferrarelli, L. Turano, C. Tebala, Z. Hussain, Z. Sheikh, A. Sheikh, G. Sceni, A. Hussain, and F. C. Morabito, “A fuzzy-enhanced deep learning approach for early detection of COVID-19 pneumonia from portable chest X-ray images,” Neurocomputing, vol. 481, pp. 202–215, Apr. 2022.

[106] J. M. Mayor-Torres, S. Medina-DeVilliers, T. Clarkson, M. D. Lerner, and G. Riccardi, “Evaluation of interpretability for deep learning algorithms in EEG emotion recognition: A case study in autism,” arXiv preprint arXiv:2111.13208, 2021.

[107] K. H. Kim, H. W. Koo, B. J. Lee, S. W. Yoon, and M. J. Sohn, “Cerebral hemorrhage detection and localization with medical imaging for cerebrovascular disease diagnosis and treatment using explainable deep learning,” J. Korean Phys. Soc., vol. 79, no. 3, pp. 321–327, May 2021.

[108] R. Arefin, M. D. Samad, F. A. Akylken, and A. Davanian, “Non-transfer deep learning of optical coherence tomography for post-hoc explanation of macular disease classification,” in Proc. IEEE 9th Int. Conf. Healthcare Informatics, Victoria, Canada, 2021, pp. 48–52.

[109] R. T. Sutton, O. R. Zaiâne, R. Goebel, and D. C. Barillot, “Artificial intelligence enabled automated diagnosis and grading of ulcerative colitis endoscopy images,” Sci. Rep., vol. 12, no. 1, p. 2748, Feb. 2022.

[110] D. H. Chen, H. J. Zhao, J. R. He, Q. Pan, and W. L. Zhao, “An causal XAI diagnostic model for breast cancer based on mammography reports,” in Proc. IEEE Int. Conf. Bioinformatics and Biomedicine, Houston, USA, 2021, pp. 3341–3349.

[111] Z. Uddin, K. K. Dysthe, A. Folstad, and P. B. Brandtzæg, “Deep learning for prediction of depressive symptoms in a large textual dataset,” Neural Comput. Appl., vol. 34, no. 1, pp. 721–744, Jan. 2022.
[129] Q. Dou, C. Ouyang, C. Chen, H. Chen, B. Glocker, X. H. Zhaung, and P. A. Heng, “PnP-AdaNet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation,” IEEE Access, vol. 7, pp. 99065–99076, Jul. 2019.

[130] H. Guan, L. Wang, and M. X. Liu, “Multi-source domain adaptation via optimal transport for brain dementia identification,” in Proc. IEEE 18th Int. Symp. Biomedical Imaging, Nice, France, 2021, pp. 1514–1517.

[131] H. Guan, Y. B. Liu, E. K. Yang, P. T. Yap, D. G. Shen, and M. X. Liu, “Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification,” Med. Image Anal., vol. 71, p. 102076, Jul. 2021.

[132] J. Wang, L. C. Zhang, Q. Wang, L. Chen, J. Shi, X. B. Chen, Z. Y. Li, and D. G. Shen, “Multi-class ASL classification based on functional connectivity and functional correlation tensor via multi-source domain adaptation and multi-view sparse representation,” IEEE Trans. Med. Imaging, vol. 39, no. 10, pp. 3137–3147, Oct. 2020.

[133] H. F. Cui, C. Yuwen, J. Lei, Y. Xia, and Y. N. Zhang, “Bidirectional cross-modality unsupervised domain adaptation using generative adversarial networks for cardiac image segmentation,” Comput. Biol. Med., vol. 136, p. 104726, Sept. 2021.

[134] R. Xu, Z. Cong, X. C. Ye, S. Kido, and N. Tomiyama, “Unsupervised content-preserved adaptation network for classification of pulmonary textures from different CT scanners,” in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, Barcelona, Spain, 2020, pp. 1060–1064.

[135] P. Wang, P. F. Li, Y. M. Li, J. Xu, and M. F. Jiang, “Classification of histopathological whole slide images based on multiple weighted semi-supervised domain adaptation,” Biomed. Signal Process. Control, vol. 73, p. 103400, Mar. 2022.

[136] Y. Wang, Y. Q. Feng, L. Zhang, Z. Z. Wang, Q. Lv, and Z. Yi, “Deep adversarial domain adaptation for breast cancer screening from mammograms,” Med. Image Anal., vol. 73, p. 102147, Oct. 2021.

[137] J. Li, C. L. Feng, X. Z. Lin, and X. H. Qian, “Utilizing GCN and meta-learning strategy in unsupervised domain adaptation for pancreatic cancer segmentation,” IEEE J. Biomed. Health Inform., vol. 26, no. 1, pp. 79–89, Jan. 2022.

[138] M. Saiz-Vivó, A. Colomé, C. Fonfría, L. Martí-Bonmatí, and V. Naranjo, “Supervised domain adaptation for automated semantic segmentation of the atrial cavity,” Entropy, vol. 23, no. 7, p. 898, Jul. 2021.

[139] I. Dayan, H. R. Roth, A. X. Zhong, A. Harouni, A. Gentili, A. Z. Abidin, A. Liu, A. B. Costa, B. Wood, C. S. Tsai, C. H. Wang, C. N. Hsu, C. K. Lee, C. Ruan, D. G. Xu, D. F. Wu, E. Huang, F. Kitamura, G. Lacey, G. C. de Antônio Corrada, H. H. Shin, H. Obinata, H. Ren, J. Crane, J. Tetraut, J. H. Guan, J. Garrett, J. G. Park, K. Dreyer, K. Juru, K. Kersten, M. A. B. C. Rockenbach, M. Linguraru, M. Haider, M. Abdelmaseeh, N. Rieke, P. Damasceno, P. C. e Silva, P. Wang, S. X., S. C. Kawano, S. Sriswasdi, S. Y. Park, T. Grist, V. Buch, W. Jantaranab jakul, W. Wang, W. Y. Tak, X. Li, X. H. Lin, F. Kwon, F. Gilbert, J. Kaggsie, J. G. Park, K. Dreyer, K. Juru, K. Kersten, M. A. B. C. Rockenbach, M. G. Linguraru, M. A. Haider, M. Abdelmaseeh, N. Rieke, P. F. Damasceno, P. M. C. e Silva, P. C. Wang, S. Xu, S. Kawano, S. Sriswasdi, S. Y. Park, T. M. Grist, V. Buch, W. Jantaranab jakul, W. Wang, W. Y. Tak, X. Li, X. H. Lin, Y. J. Kwon, A. Qurain, A. Feng, A. N. Priest, B. Turkby, B. Glicksberg, B. Bizzo, B. S. Kim, C. Tor-Dieze, C. C. Lee, C. J. Hsu, C. N. Lin, C. L. Lai, C. P. Hess, C. Compas, D. Bhatia, E. K. Oermann, E. Leibovitz, K. Marshall, H. M. I. Wang, J. H. Y. Koh, K. K. N. Murthy, L. C. Fu, M. R. F. de Mendonça, M. Fralick, M. K. Kang, M. Adil, N. Gangai, P. Vateekul, P. Elnajjar, P. Elnajjar, S. Hickmann, S. Majumdar, S. McLeod, S. Reed, S. Graf, S. Harmon, T. Kodama, T. Puthanakit, T. Mazzulli, V. de Lima Lavor, Y. Rakvongthai, Y. R. Lee, Y. H. Wen, “Federated learning used for predicting outcomes in SARS-COV-2 patients,” Res. Sq., DOI: 10.21203/rs.3.rs.126892/v1.

[140] B. L. Y. Agbule, J. P. Li, A. U. Haq, E. K. Bankas, S. Ahmad, I. O. Agyemang, D. Kulevome, W. D. Ndiaye, B. Cobbina, and S. Latipova, “Multimodal melanoma detection with federated learning,” in Proc. 18th Int. Computer Conf. Wavelet Active Media Technology and Information Processing, Chengdu, China, 2021, pp. 238–244.

[141] M. F. Zhang, Y. N. Wang, and T. Luo, “Federated learning for arrhythmia detection of Non-ID ECG,” in Proc. IEEE 6th Int. Conf. Computer and Communications, Chengdu, China, 2020, pp. 1176–1180.

[142] Y. C. Gu, Q. Q. Hu, X. L. Wang, Z. Zhou, and S. X. Lu, “FedACS: An efficient federated learning method among multiple medical institutions with adaptive client sampling,” in Proc. 14th Int. Congr. Image and Signal Processing, BioMedical Engineering and Informatics, Shanghai, China, 2021, pp. 1–6.

[143] Y. Kang, Y. Liu, and T. J. Chen, “FedMVT: Semi-supervised vertical federated learning with MultiView training,” arXiv preprint arXiv: 2008.10838, 2020.

[144] W. H. Sun, Y. Q. Chen, D. X. Yang, J. B. Cao, and Y. X. Song, “FedIO: Bridge inner-and outer-hospital information for perioperative complications prognostic prediction via federated learning,” in Proc. IEEE Int. Conf. Bioinformatics and Biomedicine, Houston, USA, 2021, pp. 3215–3221.

[145] A. Linardos, K. Kushibar, S. Walsh, P. Gkontra, and K. Lekadir, “Federated learning for multi-center imaging diagnostics: A simulation study in cardiovascular disease,” Sci. Rep., vol. 12, no. 1, p. 3551, Mar. 2022.

[146] N. Q. Dong and I. Voiculescu, “Federated contrastive learning for decentralized unlabeled medical images,” in Proc. 24th Int. Conf. Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 2021, pp. 378–387.

[147] Q. Yang, Y. Liu, T. J. Chen, and Y. X. Tong, “Federated machine learning: Concept and applications,” ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, p. 12, Mar. 2019.

[148] J. Lo, T. T. Yu, D. Ma, P. X. Zang, J. P. Owen, Q. Q. Zhang, R. K.
Wang, M. F. Beg, A. Y. Lee, Y. L. Jia, and M. V. Sarunic, “Federated learning for microvasculature segmentation and diabetic retinopathy classification of OCT data,” *Ophthalmol. Sci.*, vol. 1, no. 4, p. 100069, Dec. 2021.

[156] B. Gu, A. Xu, Z. Y. Huo, C. Deng, and H. Huang, “Privacy-preserving asynchronous vertical federated learning algorithms for multiparty collaborative learning,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6103–6115, Nov. 2022.

[157] Y. Liu, Y. Kang, C. P. Xing, T. J. Chen, and Q. Yang, “A secure federated transfer learning framework,” *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 70–82, Jul.–Aug. 2020.

[158] S. Li, T. X. Cai, and R. Duan, “Targeting underrepresented populations in precision medicine: A federated transfer learning approach,” arXiv preprint arXiv: 2108.12112, 2021.

[159] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” in *Proc. Machine Learning and Systems*, Austin, USA, 2020, pp. 429–450.

[160] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Proc. 20th Int. Conf. Artificial Intelligence and Statistics, Fort Lauderdale, USA, 2017, pp. 1273–1282.

[161] S. U. Stich, “Local SGD converges fast and communicates little,” in *Proc. 7th Int. Conf. Learning Representations*, New Orleans, LA, USA, May 2019.

[162] F. Sattler, K. R. Müller, and W. Samek, “Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3710–3722, Aug. 2021.

[163] O. Marfoq, G. Neglia, A. Bellet, L. Kameni, and R. Vidal, “Federated multi-task learning under a mixture of distributions,” in *Proc. 35th Advances in Neural Information Processing Systems*, 2021, pp. 15434–15447.

[164] A. C. Hauschild, M. Lemanyczk, J. Matschinske, T. Frisch, O. Zolotareva, A. Holzinger, J. Baumbach, and D. Heider, “Federated random forests can improve local performance of predictive models for various healthcare applications,” *Bioinformatics*, vol. 38, no. 8, pp. 2278–2286, Apr. 2022.

[165] A. Chowdhury, H. Kassem, N. Padroy, R. Umetone, and A. Karargyris, “A review of medical federated learning: Applications in oncology and cancer research,” in *Proc. 7th Int. MICCAI Brainlesions Workshop*, 2022, pp. 3–24.

[166] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami, USA, 2009, pp. 248–255.

[167] Y. Otaki, A. Singh, P. Kavanagh, R. J. H. Miller, T. Parekh, B. K. Tamarappoo, T. Sharir, A. J. Einstein, M. B. Fish, T. D. Ruddy, P. A. Heng, C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, “Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, pp. 865–872, Jul. 2019.

[168] M. Meng, J. H. Hu, Y. Y. Gao, W. Z. Kong, and Z. Z. Luo, “A deep subdomain associate adaptation network for cross-session and cross-subject EEG emotion recognition,” *Biomed. Signal Process. Control*, vol. 78, p. 103873, Sept. 2022.

[169] Z. K. Tian, D. H. Li, Y. Yang, F. Z. Hou, Z. Y. Yang, Y. Song, and Q. Gao, “A novel domain adversarial networks based on 3D-LSTM and local-domain discriminator for hearing-impaired emotion recognition,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 1, pp. 363–373, Jan. 2023.

[170] V. M. Bashyam, J. Doshi, G. Erus, D. Srinivasan, A. Abdulkadir, A. Singh, M. Habes, Y. Fan, C. L. Masters, P. Maruff, C. J. Zhao, H. Völzke, S. C. Johnson, J. Fripp, N. Koutsouris, T. D. Satterthwaite, D. H. Wolf, R. E. Gur, R. C. Gur, J. C. Morris, M. S. Albert, H. J. Grabe, S. M. Resnick, N. R. Bryan, K. Wittfeld, R. Bülow, D. A. Wolk, H. C. Shou, I. M. Nasrallah, C. Davatzikos, and The iSTAGING and PHENOM Consortia,” “Deep generative medical image harmonization for improving cross-site generalization in deep learning predictors,” *J. Magn. Reson. Imaging*, vol. 55, no. 3, pp. 908–916, Mar. 2022.

[171] C. I. Berecza, B. Wuestler, D. Rueckert, and S. Albarqouni, “Federated disentangled representation learning for unsupervised brain scan detection,” *Nat. Mach. Intell.*, vol. 4, no. 8, pp. 685–695, Aug. 2022.

[172] X. Bai, H. C. Wang, L. Y. Ma, Y. C. Xu, J. F. Gan, Z. W. Fan, F. Yang, K. Ma, J. H. Yang, S. Bai, C. Shu, Y. X. Zou, R. H. Huang, C. Z. Zhang, X. W. Liu, D. D. Tu, C. O. Xu, W. Q. Zhang, X. Wang, A. G. Chen, Y. Zeng, D. H. Yang, M. W. Wang, N. Holalikere, N. J. Halin, I. R. Kamel, J. Wu, X. H. Peng, X. Wang, J. B. Shao, P. Mongkolwat, J. J. Zhang, W. Y. Liu, M. Roberts, Z. Z. Teng, L. Beer, L. L. Sanchez, E. Sala, D. L. Rubin, A. Weller, J. Lasenby, C. S. Zheng, J. M. Wang, Z. Li, C. Schönlieb, and T. Xia, “Advancing COVID-19 diagnosis with privacy-preserving collaboration in artificial intelligence,” *Nat. Mach. Intell.*, vol. 3, no. 12, pp. 1081–1089, Dec. 2021.

[173] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why should I trust you?’: Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, USA, 2016, pp. 1135–1144.

[174] Q. Huang, M. Yamada, Y. Tian, D. Singh, and Y. Chang, “GraphLIME: Local interpretable model explanations for graph neural networks,” *IEEE Trans. Knowl. Data Eng.*, DOI: 10.1109/TKDE.2022.3187455

[175] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” in *Proc. 31st Int. Conf. Neural Information Processing Systems*, Long Beach, USA, 2017, pp. 4768–4777.

[176] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, “Deep domain confusion: Maximizing for domain invariance,” arXiv preprint arXiv: 1412.3474, 2014.

[177] M. S. Long, Y. Cao, J. M. Wang, and M. I. Jordan, “Learning transferable features with deep adaptation networks,” in *Proc. 32nd Int. Conf. Machine Learning*, Lille, France, 2015, pp. 97–105.

[178] X. T. Han, L. Qi, Q. Yu, Z. Q. Zhou, Y. F. Zheng, Y. H. Shi, and Y. Gao, “Deep symmetric adaptation network for cross-modality medical image segmentation,” *IEEE Trans. Med. Imaging*, vol. 41, no. 1, pp. 121–132, Jan. 2022.

[179] Y. C. Zhu, F. Z. Zhuang, J. D. Wang, G. L. Ke, J. W. Chen, J. Bian, H. Xiong, and Q. He, “Deep subdomain adaptation network for image classification,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021.

[180] Y. R. Jin, Z. Y. Li, Y. Q. Liu, J. L. Liu, C. J. Qin, L. Q. Zhao, and C. L. Liu, “Multi-class 12-lead ECG automatic diagnosis based on a novel subdomain adaptive deep network,” *Sci. China Technol. Sci.*, vol. 64, no. 11, pp. 2617–2630, Sept. 2022.

[181] J. D. Wang, Y. Q. Chen, W. J. Feng, H. Yu, M. Y. Huang, and Q. Yang, “Transfer learning with dynamic distribution adaptation,” *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 1, p. 6, Feb. 2020.

[182] B. Xu, K. W. Wu, Y. Wu, J. He, and C. Y. Chen, “Dynamic adversarial domain adaptation based on multitoken maximum mean discrepancy for breast ultrasound image classification,” *Expert Syst. Appl.*, vol. 207, p. 117978, Nov. 2022.

[183] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms,” arXiv preprint
Ahmad Chaddad received the master degree (2008) from Université de Technologie de Compiègne and the Ph.D. degree (2012) from the University of Lorraine, France. He worked for seven years at McGill University, Canada, École de Technologie Supérieure (ETS), Montreal, The University of Texas MD Anderson Cancer Center, USA, and Villanova University, USA. In 2020, he joined the School of Artificial Intelligence, Guilian University of Electronic Technology, as a Professor. He is an Associate Member at LIVIA, ETS (Canada) and LCOMS, University of Lorraine (France). He has authored more than 80 research papers (60 papers first co-author). His current research interests include AI and radiomics analysis to improve personalized medicine strategies, by allowing clinicians to monitor disease in real-time as patients move through treatment. Dr. Chaddad is a member of several international technical and organizational committees.

Qizong Lu is undergraduate student in School of Artificial Intelligence, Guilian University of Electronic Technology. From 2021 to 2022, he works on the current problems faced the clinical practice. His current research interests include machine learning and biomedical engineering. He has authored a research paper about radiomics analysis for Autism spectrum disorder.

Jiali Li is undergraduate student in School of Artificial Intelligence, Guilian University of Electronic Technology. She works on artificial intelligence for clinical diagnosis. Her current research interests include machine learning and biomedical engineering. She has authored a research paper about radiomics analysis for prostate cancers.

Reem Katesh is an Assistant Professor at Taibah University in the Department of Information Systems, Saudi Arabia. She received the M.A.Sc. in information system security and Ph.D. in information and systems engineering at Concordia University, Canada. Her research interests are cybersecurity, physical system, IoT, and Secure AI.

Camel Tanougast received the Ph.D. degree in microelectronic and electronic instrumentation from the Henri Poincaré University of Nancy, France, in 2001, and the authorization/accreditation to supervise research (HDR - Habilitation degree) from the University of Metz, in 2009. He is currently a Professor in electronics and embedded systems with the University of Lorraine, France. He joined the Electronic Architectures Group, Electronic Instrumentation Laboratory of Nancy, in 2003, and the Micro-electronic and Sensor Interface Laboratory of Metz (LICM), in 2008. He has
been the Head Research of the networked adaptive and self-organized systems at LICM. He joined the Laboratory of Design, Optimization and Modeling Systems (LCOMS), in 2013. He is currently the Deputy Director and the ASEC Team Leader of LCOMS for research in microelectronics, communicating embedded systems, and smart sensors. He has authored or coauthored more than 120 publications and several research books, participated in several editorial boards of books and international journals, and has been the Supervisor of 21 Ph.D. thesis. His research interests include radio communication, reconfigurable systems and NoCs, design and implementation of real time processing architectures, optimization, and implementation design, System-on-Chip development, FPGA design, computing vision, radiomic analysis, image processing, cryptography, and the digital television broadcast (DVB).

Ahmed Bouridane (Senior Member, IEEE) received an "Ingenieur d’Etat" degree in electronics from "Ecole Nationale Polytechnique" of Algiers (ENPA), Algeria, in 1982, an M.Phil. degree in electrical engineering (VLSI design for signal processing) from the University of Newcastle-Upon-Tyne, U.K., in 1988, and a Ph.D. degree in electrical engineering (computer vision) from the University of Nottingham, U.K., in 1992. From 1992 to 1994, he worked as a research developer in telesurveillance and access control applications. In 1994, he joined Queen’s University Belfast, U.K., initially as Lecturer in computer architecture and image processing and later on he was promoted to reader in computer science. In 2009, he joined Northumbria University at Newcastle leading the Computational Intelligence and Visual Computing Lab. His is now a professor of machine intelligence and director of the Cybersecurity and Data Analytics Research Center at the University of Sharjah, UAE. His research interests are in machine learning with applications to imaging for forensics and security, quantitative pathology and biomedical engineering, homeland security and video analytics. He has authored and co-authored more than 350 publications and three research books on imaging for forensics and security and biometric security and privacy.

Ahmed Abdulkadir received the master degree (2012) in life sciences from the École Polytechnique Fédérale de Lausanne, Switzerland and the Ph.D. (2018) in computer engineering from the Albert-Ludwigs University Freiburg, Freiburg-im-Breisgau, Germany. He worked as a Researcher at the University of Bern, Switzerland, and the University of Pennsylvania, USA. He is currently completing a three-year mobility fellowship granted by the Swiss National Science Foundation at the Lausanne University Hospital, Switzerland. His work focuses on the use of statistical methods and AI to study aging and dementia in order to improve the diagnosis of brain disorders.