Generative AI-Driven Human Digital Twin in IoT-Healthcare: A Comprehensive Survey

Jiayuan Chen, You Shi, Changyan Yi, *Member, IEEE*, Hongyang Du, Jiawen Kang, Dusit Niyato, *Fellow, IEEE*

***Abstract*—The Internet of things (IoT) can significantly enhance the quality of human life, specifically in healthcare, attracting extensive attentions to IoT-healthcare services. Meanwhile, the human digital twin (HDT) is proposed as an innovative paradigm that can comprehensively characterize the replication of the individual human body in the digital world and reflect its physical status in real time. Naturally, HDT is envisioned to empower IoT-healthcare beyond the application of healthcare monitoring by acting as a versatile and vivid human digital testbed, simulating the outcomes and guiding the practical treatments. However, successfully establishing HDT requires high-fidelity virtual modeling and strong information interactions but possibly with scarce, biased and noisy data. Fortunately, a recent popular technology called generative artificial intelligence (GAI) may be a promising solution because it can leverage advanced AI algorithms to automatically create, manipulate, and modify valuable while diverse data. This survey particularly focuses on the implementation of GAI-driven HDT in IoT-healthcare. We start by introducing the background of IoT-healthcare and the potential of GAI-driven HDT. Then, we delve into the fundamental techniques and present the overall framework of GAI-driven HDT. After that, we explore the realization of GAI-driven HDT in detail, including GAI-enabled data acquisition, communication, data management, digital modeling, and data analysis. Besides, we discuss typical IoT-healthcare applications that can be revolutionized by GAI-driven HDT, namely personalized health monitoring and diagnosis, personalized prescription, and personalized rehabilitation. Finally, we conclude this survey by highlighting some future research directions.**

***Index Terms*—IoT-healthcare, generative artificial intelligence, human digital twin, generative adversarial network, variational autoencoder, transformer, diffusion model**

I. Introduction

A. Background

THE advancements of Internet of the Things (IoT) in recent years is leading to a paradigm shift in the healthcare industry, which are termed IoT-healthcare [1], [2]. An Ireland healthcare organization estimates that roughly 17.5 million lives are lost annually due to inefficiencies in health data [3]. The main inefficiency is the delay access to and analyze health data, resulting in untimely interventions. Fortunately, the progression of IoT-healthcare presents a promising avenue to tackle these challenges. By enabling the real-time collection and transmission of pertinent health data to servers for in-depth healthcare analysis via personal IoT devices, it can prompt timely healthcare alerts and proactive interventions for saving lives. It has been widely applied in healthcare monitoring, including blood glucose monitoring, cardiac monitoring, respiration monitoring and blood pressure monitoring [4], [5].

Human digital twin (HDT), a promising technology and game changer for IoT-healthcare, is taking this field to another level. HDT can create a digital replica of a human body, comprehensively and precisely characterizing each individual in the digital space while reflecting its physical status in real time [6]–[10]. Besides, with visualization and interaction characteristics, HDT is envisioned as a versatile and vivid human digital testbed for revolutionizing the IoT-healthcare, beyond healthcare monitoring applications mentioned above. For instance, a patient's HDT can be implemented in silico treatment simulations and experiments, facilitating the development of finely-tailored, personalized treatment plans [11]. Additionally, a doctor's HDT with expert-level medical knowledge can be a personal 24/7 doctor to answer the patients' queries [12].

The successful establishment and implementation of HDT largely depend on the high-fidelity human modeling, supported by comprehensive individual-level data encompassing appearance, movement, and physiological data, acquired from multi-source, such as IoT devices. In addition, as an intelligent human digital testbed, HDT needs to generate various human-like feedback during immersive real-virtual interactions. These include, for example, providing intuitive feedback on drug-disease responses, and simulating haptic feedback to replicate the tactile sensations experienced by humans in real-world scenarios. All such requirements are difficult to meet due to several crucial reasons, e.g., data scarcity, bias, noise and intricate digital modeling. Fortunately, generative artificial intelligence (GAI) has been recognized as a promising technology that can effectively fulfill or assist the implementation of HDT for IoT-healthcare [13].

GAI can leverage advanced AI algorithms to automatically create, manipulate, and modify valuable while diverse data [14], [15]. Specifically, GAI models, such as generative adversarial network (GAN), variational autoencoder (VAE), transformer, and diffusion model, with their powerful creativities and data analysis abilities can generate ultra-realistic individual-level data and make informed decisions for HDT in IoT-healthcare, which will be elaborated in Section III. Thus, our survey focuses on how GAI enables HDT in IoT-healthcare, namely GAI-driven HDT in IoT-healthcare.

J. Chen Y. Shi and C. Yi are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 211106, China. (E-mail: {jiayuan.chen, shyou changyan.yi}@nuaa.edu.cn).

H. Du and D. Niyato are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. (Email: hongyang001@e.ntu.edu.sg and dniyato@ntu.edu.sg).

J. Kang is with the School of Automation, Guangdong University of Technology, China. (Email: kjwx886@163.com).

TABLE I
COMPARISON OF THE RELATED WORK WITH OUR SURVEY.

Reference	IoT-Healthcare	GAI	HDT	Remark
Habibzadeh et al., [4] 2020	√	×	×	This survey investigated enabling technologies in IoT-healthcare, focusing on communication, data analytics and acquisition. It also highlights various applications such as blood glucose monitoring, cardiac monitoring, etc.
Yang et al., [5] 2022	√	×	×	This survey analyzed IoT-healthcare applications, including activity recognition, fitness assistance and sleep monitoring, and surveyed implementation approaches using body-worn hardware or wireless signal transceivers.
Chen et al., [8] 2023	√	×	√	This survey focused on the networking architecture and key technologies of HDT in personalized healthcare applications.
Sun et al., [22] 2023	√	×	√	This survey highlighted the HDT can revolutionize IoT-healthcare by providing customized diagnosis and treatment.
AlAmir et al., [23] 2022	√	√	×	This survey discussed extension models of GANs, surveyed their applications in medical images, such as cross-modality synthesis, segmentation, etc.
Shokrollahi et al., [24] 2023	√	√	×	This paper highlight the critical role of GAI, such as transformers and diffusion models, in healthcare, which significantly enhance clinical diagnosis, data reconstruction, and drug synthesis.
Our Survey	√	√	√	This survey comprehensively explores the fundamentals of HDT and GAI, and discusses the implementation of GAI-driven HDT. Besides, it delves into the IoT-healthcare applications of GAI-driven HDT, including personalized health monitoring and diagnosis, personalized prescription, and personalized rehabilitation. It finally provides the future directions.

B. Related Work and Contributions

Given the increasing interest of GAI-driven HDT in IoT-healthcare, several surveys and tutorials have been recently published [16]–[21]. Table I presents a comparison of these related works compared with ours.

Specifically, Habibzadeh et al. in [4] surveyed the existing and emerging technologies that can enable IoT-healthcare. It presented the enabling technologies by investigating three of IoT-healthcare primary components: 1) sensing and data acquisition; 2) communication; and 3) data analytics and inference. Based on these, they also highlight some IoT-healthcare applications, including blood glucose monitoring, cardiac monitoring, respiration monitoring, blood pressure monitoring, among others. Yang et al. in [5] investigated the IoT-healthcare applications with high relevance to daily health routines, including activity recognition, fitness assistance, vital signs monitoring, daily dietary tracking, and sleep monitoring. Additionally, they surveyed the ways of implementing these applications based on leveraging of sensors, such as device-based paradigms using hardware on the body, and device-free paradigms using wireless signal transceivers. However, these surveys mainly focused on healthcare monitoring enabled by IoT, ignoring the power of integration of HDT, which can significantly enrich the application of IoT-healthcare.

With the prevalence of HDT, there are several surveys are liberating the power of HDT in IoT-healthcare. Chen et al. in [8] comprehensively explored the networking architecture and key supporting technologies for realizing HDT in personalized healthcare. Specifically, the networking architecture consisted of data acquisition, communication, computation, data management, data analysis and decision making layers. They surveyed the enabling technologies for each layer. Additionally, they delved into the application of HDT in personalized healthcare, including personalized diagnosis, prescription, surgery, and rehabilitation. Sun et al. in [22] highlighted that the HDT can revolutionize IoT-healthcare by providing customized diagnosis and treatment. They revealed that by using a patient's HDT, the medical system can predict the patient's immune response to infection or injury, which can help doctors diagnose diseases precisely. Additionally, they investigated that the patient's HDT can be used as a vivid digital testbed before the prescription or surgery, thereby supporting personalized treatment in a non-invasive manner. Furthermore, they surveyed more specific applications of HDT in IoT-healthcare, including cardiovascular disease, surgery, pharmacy, orthopaedics and COVID-19. However, none of them discuss the role of GAI for HDT in IoT-healthcare.

GAI, with its superior data generation and analysis capabilities, has attracted a myriad of researches recently, and many of them focused on its application in healthcare. AlAmir et al. in [23] discussed the recent advancements in GANs, particularly in the healthcare field. Specifically, they investigated the extension models of GANs by classifying and introducing them individually. Then, they surveyed the applications of these GAN models in medical images, including cross-modality synthesis, segmentation, augmentation, detection, classification, registration and reconstruction. Shokrollahi et al. in [24] delved into the critical role of GAI, such as transformers and diffusion models, in healthcare applications, including medical imaging, protein structure prediction, clinical documentation, diagnostic assistance, radiology interpretation, clinical decision support, drug design and molecular representation. Such applications have significantly enhanced clinical diagnosis, data reconstruction, and drug synthesis.

The surveys above are centered around IoT-healthcare, HDT and GAI. However, none of them explore the potential of GAI with remarkable data generation and analysis capabilities for enabling HDT in IoT-healthcare applications. This motivates us to compose this survey investigating the GAI-driven HDT in IoT applications. The contributions of this survey can be summarized as follows:

- We thoroughly review the HDT and GAI technique, including differences between HDT and the conventional DT and the framework of HDT, as well as the popular GAI models.
- We explore the implementation of GAI-driven HDT. We comprehensively explain how GAI enables each component of HDT's framework, including data acquisition, digital modeling, communication,data management and data analysis.
- We survey GAI-driven HDT in IoT-healthcare applications, including personalized health monitoring and diagnosis, prescription, and rehabilitation.
- We outline several open issues and future directions in GAI-driven HDT in IoT-healthcare, helping to drive the development of this field.

The rest of this paper is organized as follows: Section II presents the fundamentals of HDT and GAI, and gives an overview of the framework of GAI-driven HDT. In Section III, the implementation of GAI-driven HDT is analyzed in detail. Section IV surveyed the IoT-healthcare application of GAI-driven HDT, including personalized health monitoring and diagnosis, personalized prescription, and personalized rehabilitation. Section V explores several future research directions of GAI-driven HDT in IoT-healthcare. Section VI concludes this survey paper.

II. Fundamentals of HDT and GAI

A. Human Digital Twin

HDT as the versatile and vivid digital portrayal of individual, is breathing life into the digital world. According to Emergen Research, the global market for HDT will grow from 29.51$ billion in 2022 to about 530$ billion in 2032 [25]. With the continuous advancement of HDT, the IoT-healthcare is being revolutionized by HDT [8], [9], [13], [26].

HDT, focuses on digital replicas of human beings while the conventional DT limits the attention to non-living physical entities, e.g., machines [27] and networks [28]. Then, several distinguishing characteristics between HDT and the conventional DT are detailed below and summarized in Table II.

- Physiology and psychology: The most significant difference between HDT and conventional DT is physiology and psychology [29]–[31]. This includes attributes such as: i) physiological characteristics, e.g., brain electrophysiologic signals, blood oxygen level and heart rate; ii) perceptual abilities, e.g., visual sensitivity, pressure sensitivity and temperature sensitivity; iii) emotional state, e.g., happiness, depression and anxiety; iv) personality characteristics, e.g., personality type, propensity to trust, and propensity towards suspicion.
- Behavioral rule: Human beings' external behaviors highly depend on the individual subjective consciousness. Particularly, internal behaviors, such as the progression of diseases and emotional states, are generally influenced by multi-source and complex factors, including external environments. In contrast, machines generally follow predominantly model-based and predetermined behavioral rules [32], [33]. Therefore, humans are highly complex systems with greater uncertainty levels than machines. The abstract processes of human beings in HDT are significantly more challenging than those of machines in the conventional DT.
- Metrics for evaluation: The human-centered paradigm of HDT requires improving human beings' well-being by considering their roles, needs, talents and rights. Meanwhile, the conventional DT commonly takes the performance-centered interest first for improving production and economic benefits. Specifically, the conventional DT typically prioritize metrics such as efficiency, productivity, effectiveness, and profitability. However, HDT extends the scope of metrics to contain usability, user experience, etc [7].
- Data complexity: Human beings are more heterogeneous and unstructured than non-living machines. Consequently, unlike the conventional DT, building a high-fidelity digital representation model of any human entity in HDT requires diverse and complex data from multiple sources. In addition to physiological data, unstructured data from environmental factors and social media play a crucial role in abstracting human virtual twins. This is due to the significant correlation between human beings and such external data sources [8].
- Mobility pattern: Unlike the position-fixed machines in the conventional DT, the mobility patterns in HDT may be highly predictable. The mobility patterns of human beings can be categorized into human positional and postural mobility. Positional mobility, like a person moving from indoors to outdoors, may cause radio frequency (RF) propagation characteristics to change and even the service migrations. Additionally, the postural mobility of

TABLE II
DIFFERENCES BETWEEN HDT AND THE CONVENTIONAL DT.

Key Features	Conventional DT	HDT
Physiology and Psychology	Lack of physiology and psychology	Human beings are living entities, and have i) physiological characteristics; ii) perceptual abilities; iii) emotional state; and iv) personality characteristics.
Behavioral Rule	The behavioral rules of machines follow predominantly model-based and predetermined behavioral rules	Human external behaviors depend on individual subjective consciousness and internal behaviors, affected by multi-source and complex factors.
Metrics for Evaluation	Prioritizes performance-centered interests to achieve mass production and enhance economic benefits	It expands to human-centered metrics, aiming to enhance human well-being by acknowledging their roles, needs, talents, and rights.
Data Complexity	Mostly structured and homogeneous	Mostly data are heterogeneous and unstructured. Correlation exists between individuals and external data, such as environmental and social media data.
Mobility Pattern	Mostly fixed with no or limited mobility	The mobility patterns are highly complex, and are categorized as positional and postural mobility.

a human, like lying, sitting, walking, may cause signal strength to fluctuate due to the influence of human bodies on the path loss, known as body shadowing [8], [9], [13].

As the distinct features of HDT outlined above, the successful implementation of HDT relies on five essential components: data acquisition, communication, data management, digital modeling, and data analysis [7]–[9], [34]. These components form the framework of HDT. Note that, in HDT, the physical entity, i.e., the individual, in the physical world is called a physical twin (PT), while the corresponding virtual one in the digital world is called a virtual twin (VT). Section III delves into how GAI facilitates the actualization of HDT by bolstering each component within the framework of HDT. Before this, we introduce the GAI technique, unveiling its potential for enabling HDT.

B. GAI Techniques

In this subsection, we delve into the recent primary trends of GAI models, including generative adversarial network, variational autoencoder, transformer and diffusion model [14], [35]. These models find frequent applications in HDT for IoT-healthcare.

Generative adversarial network: As depicted in Fig. 1, the GAN comprises two neural networks engaged in a competitive process to create new samples resembling a specific distribution [36]. The first network, the generator, aims to produce synthetic samples by comprehending the underlying distribution of the training data. Meanwhile, the discriminator, the second network, distinguishes between real and synthetic data generated by the generator. Its task is to accurately differentiate real samples and provide feedback to enhance the quality of the generated samples. Throughout the GAN training process, these two networks iteratively refine their performance adversarially, engaging in a competitive interplay until they achieve a stable equilibrium. This continuous refinement allows the generator to create more realistic samples while enabling the discriminator to better distinguish between real and synthetic data.

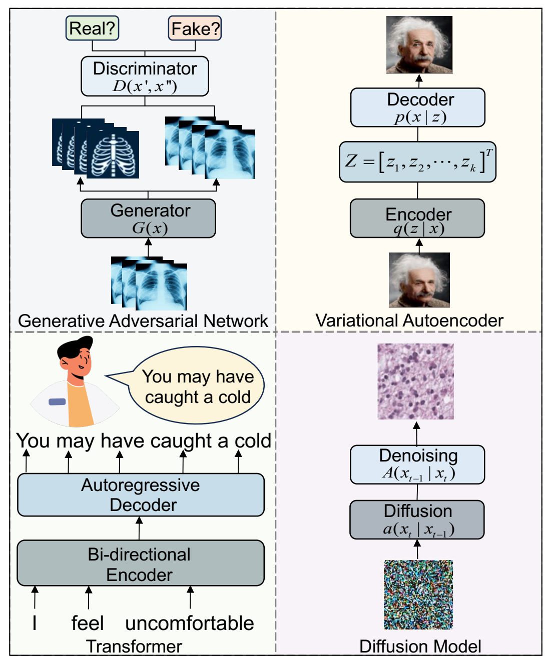

Fig. 1. The workflow of recent primary trends of GAI models, including generative adversarial network, variational autoencoder, transformer and diffusion model.

GAN has been successfully applied in HDT for IoT-healthcare. For instance, a heart DT model was integrated into optimizing a GAN to generate electrocardiogram (ECG) data [37], which can be used to solve the ECG data scarcity issue when digital modeling of cardiac activities. Additionally, a chained GAN-based approach, which connected multiple GAN models, was proposed to simulate the pathology of tissues in the human body. This approach can be used to model tissue digital twin in HDT, as well as help diagnose diseases and predict their progress [38].

Variational autoencoder: The core idea of VAE is to transform input data to a low-dimensional latent space representation [39]. Illustrated in Fig. 1, VAE comprises two neural networks. The first network, known as the encoder, maps input data to a latent space, often assumed to follow a Gaussian distribution characterized by learned mean and variance parameters. In contrast, the other network, the decoder, undertakes the task of reconstructing the original input data from a sample drawn from the latent space distribution. The decoder aims to generate a reconstructed sample closely resembling the input data. Throughout the training process, the encoder and decoder parameters are optimized to minimize the reconstruction error. Additionally, a regularization term, the Kullback-Leibler divergence, is introduced to ensure that the learned latent space distribution closely aligns with a standard Gaussian distribution. This regularization term contributes to the overall objective of refining the latent space representation in the VAE framework.

VAE has been successfully applied in HDT for IoT-healthcare. For instance, a VAE-based approach was introduced to reconstruct complete body movements using signals from the PT's head-mounted device (HMD) [40]. This method enabled the digital reconstruction of full-body motion based on signals from the head and hands. It proved beneficial for accurately modeling human motions in scenarios where motion data obtained from the PT may inadequately represent complete body movements, thereby aiding motion monitoring endeavors. Moreover, a VAE-based method was proposed to predict potential subsequent steps in the clinical measurement trajectory of a patient encountering an ischemic stroke [41]. This approach holds promise in simulating disease progression within HDT, thereby assisting in customizing treatment plans tailored to individual patients.

Transformer: Transformer excels at capturing contextual information and long-distance dependencies in text [42]. The architecture used in this model is an encoder-decoder structure, as illustrated in Fig. 1. The encoder employs a bidirectional information propagation process to comprehend the input text. The decoder, found in most transformer architectures, generates words sequentially. This type of decoder is commonly referred to as an autoregressive decoder. With this architecture, embedding with self-attention mechanisms, it can effectively process the relevant information in the input sequence, making the generated text more accurate, coherent, and able to consider more contextual information.

Transformer has been successfully applied in HDT for IoT-healthcare. For example, the state-of-the-art (SOTA) language generation model, generative pre-trained transformer (GPT), has been incorporated into electronic health records (EHR) workflows to autonomously respond to patients' healthcare inquiries [43]. This model can be visualized as an HDT representing a doctor, equipped with extensive medical knowledge, capable of addressing patient queries. Additionally, a pre-trained transformer-based approach has been proposed to concurrently learn gene and cell embeddings, enabling the capture of intricate gene-to-gene interactions at the single-cell level [44]. This approach holds promise in digitally modeling diverse facets of cellular processes within HDT, offering insights into personalized responses to treatments.

Diffusion model: Unlike GANs and VAEs, the diffusion model employs a series of sequential transformations on the input distribution [14], [45]. Specifically, as depicted in Fig. 1, this model constructs a Markov chain comprising diffusion steps where noise is incrementally introduced to the input data. Subsequently, a reverse process is implemented, gradually removing noise from the distribution to generate the desired data samples. This inverse method transforms noise distribution back to the original data distribution through a gradual denoising process.

Diffusion model has been successfully applied in HDT for IoT-healthcare [14], [46], [47]. For instance, a diffusion mode-based approach has been proposed to generate individual electroencephalogram (EEG) data [46]. It can help solve the EEG data scarcity issue in HDT, and enable HDT to digitally model brain activity patterns, supporting the neurological health monitoring. Additionally, by collaborating with an assistive modality embedding as prior information to diffusion model formulation, a diffusion model-based approach has been proposed for positron emission tomography (PET) denoising [47]. It can be used to denoise the acquired data with noise, providing cleaned data to HDT for better personalized healthcare monitoring and diagnostic accuracy.

In addition to the aforementioned models, other GAI models, such as normalizing flows and score-based generative models, have also been effectively implemented in HDT for IoT-healthcare [48], [49].

C. *Framework of GAI-driven HDT*

In this subsection, we give an overview of the framework of GAI-driven HDT, as shown in Fig. 2.

Data acquisition component is significantly crucial for HDT, which is the fuel of HDT. The required substantial health data, such as EEG, ECG, and magnetic resonance imaging (MRI) data, for HDT commonly collected from the IoT-enabled pervasive sensing and medical institutions [50]. However, these traditional data acquisitions methods are usually inefficient due to various factors. To this end, GAI can generate ultra-realistic health data based on the collected data for enriching the datasets.

The communication component plays a bridge role in HDT, which is responsible for bi-directional data transmissions between the physical and digital worlds, such as the transmission of the collected data and the feedback in the informal world. However, these data are usually large-scale, multi-modal, and time-sensitive, which are hard to fully met by traditional communication systems. To this end, GAI-enabled communication, such as GAI-enabled semantic communication [51]–[53] and cross-modal communication [54], can be applied to support communication in HDT by generating the transmitted data at the received sides, enhancing the data transmission performances.

Data management component is the core of HDT, where each component will interact with it for data access. Data from both physical and digital worlds, including collected, generated, and simulated ones, are large-scale, leakage-sensitive, and complex. To well manage these data, effective pre-processing, as well as security and privacy schemes are needed. To this end, GAI can be used in data imputation, data denoise, etc., for data pre-processing. Besides, GAI can

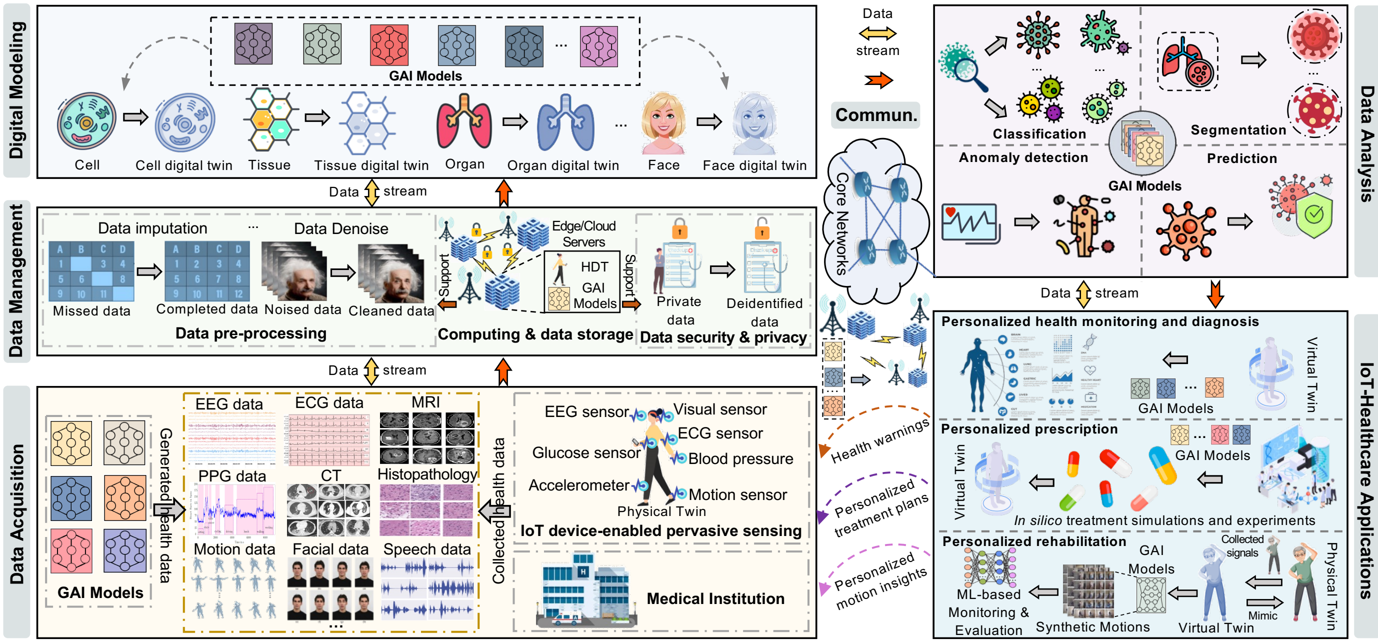

Fig. 2. The framework of GAI-driven HDT. It includes the GAI-enabled data acquisition, data management, data modeling and data analysis. With the implementation of them, the GAI-driven HDT can be applied in IoT-healthcare, including personalized health monitoring and diagnosis, personalized prescription, and personalized rehabilitation.

generate deidentified data while preserving all the patterns from the original data for protecting private data.

Digital modeling component is responsible for the human digitalization procedure in HDT. The digitalization procedure models the high-fidelity human body based on the collected data. However, traditional digital modeling methods highly rely on accurate simulation parameters, which is hard to achieve. To this end, GAI, with its robust generation capabilities, can be adapted to model humans digitally, including cells, tissues, organs, etc.

Data analysis component is essential in HDT, which is responsible for analyzing data in HDT for driving the IoT-healthcare applications. GAI with its strong analysis capabilities can be used in data classification for classifying diseases, data segmentation for obtaining key disease information, anomaly detection for identifying abnormal status, and prediction for predicting health status.

Based on the built GAI-driven HDT described above, it will significantly enhance the IoT-healthcare. It can be used in personalized health monitoring and diagnosis, prescription, and rehabilitation, acting as the intelligent and vivid human digital testbed.

In the following sections, we will delve into implementing GAI-driven HDT in IoT-healthcare.

III. GAI-driven HDT Implementation

A. GAI-enabled Data Acquisition

Data forms the cornerstone of HDT development and service provision [8]. Commonly employed data acquisition methods for HDT primarily stem from diverse sources, including medical institutions [55], as well as both non-invasive and invasive sensors equipped in PTs [34]. Nevertheless, these methods present challenges due to their time-consuming, costly, intrusive nature, and limited scalability. This restricts acquiring extensive individual-level data essential for robust HDT development and comprehensive service delivery. GAI can significantly assist the data acquisition in HDT, by offering diverse and highly realistic synthetic data. In the following, we introduce several common synthetic HDT data generated by GAI, including synthetic physiological, medical imaging and motion data, as summarized in Table III.

The physiological data, such as ECG, EEG, and photoplethysmogram (PPG), hold significant importance for HDT. By generating synthetic physiological data, GAI mitigates the limitations of traditional physiological data acquisition methods, reducing costs and time associated with collecting extensive real-world physiological data. For instance, Golany et al. in [37] proposed ECG simulator GAN (SimGAN) to create synthesized ECG data. Specifically, the authors used a system of ordinary differential equations (ODE) representing heart dynamics, and incorporated this ODE system into the GAN training to generate biologically plausible ECG data. Similarly, SynSigGAN proposed by Hazra et al. in [56] can generate synthesized PPG data, as shown in Fig. 3. In SynSigGAN, the bidirectional grid long short-term memory (LSTM) and the convolutional neural network (CNN) had been used for generator network and discriminator network, respectively, and it was trained on BIDMC PPG and respiration datasets [57]. In addition, diffusion model-based approaches also applied in physiological data generation. Tosato et al. in [46] proposed a diffusion probabilistic model (DDPM)-based approach for EEG data synthesis. The DDPM-based approach was trained on electrode-frequency distribution maps developed from a large emotion-labeled EEG datasets. Alcaraz

TABLE III
SUMMARY OF GAI-BASED APPROACH FOR DATA ACQUISITION IN HDT.

Ref.	Focus	Limitation of Traditional Methods	GAI-based Solution
[37]	Physiological data acquisition	Time-consuming and costly	Proposed a GAN-based approach, ECG simulator GAN, to generate synthesized ECG data
[56]			Proposed a GAN-based approach, SynSigGAN, to generate synthesized PPG data
[46]			Proposed a diffusion model-based approach to generate synthesized EEG data.
[58]			Proposed a diffusion model-based approach, structured state space diffusion-ECG, to generate ECG data
[61]	Medical imaging data acquisition	Costly, as well as patient discomfort and safety concerns	Proposed a diffusion model-based approach to generate histopathology image data
[63]			Proposed a diffusion model-based approach, sequence-aware diffusion model, to generate longitudinal cardiac and brain MRI
[49]			Proposed a normalizing flows-based approach to generate chest x-ray and skin cancer images
[40]	Motion data acquisition	Limited number of motion sensors equipped in the human, and the poor transmission comprises the collected motion data	Proposed a VAE-based approach to generate full-body motion using collected impoverished motion data
[67]			Proposed a diffusion model-based approach to generate the full-body poses conditioned on the sparse tracking motion data

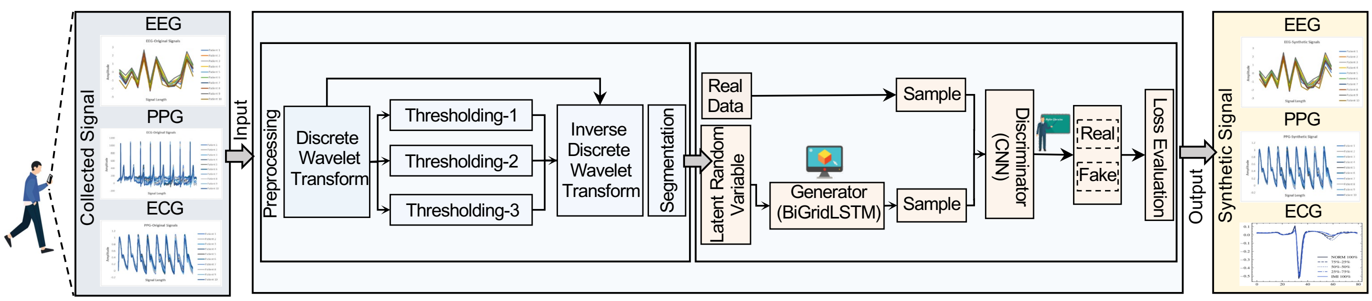

Fig. 3. Overview of the GAN-based biomedical signal synthesis, SynsigGAN, proposed in [56]. The collected signals proceed through the preprocessing stage, eliminating noise and refining the signals using discrete wavelet transform, thresholding, and inverse discrete wavelet transform. After preprocessing, the signals are forwarded to the segmentation stage that uses the Z-score to solve the amplitude scaling problem and eliminate offset. Next is the GAN, which takes in the segmented signals and generates synthetic biomedical signals using bidirectional grid long short-term memory for generator network and convolutional neural network for the discriminator. Finally, SynsigGAN outputs the synthesized biomedical signals.

et al. in [58] designed a diffusion model-based approach, called structured state space diffusion-ECG (SSSD-ECG), for generating synthetic 12-lead ECG data. SSSD-ECG was built on the SSSD^{S4} model architecture [59] and trained on PTB-XL dataset [60], which was a publicly available collection of clinical 12-lead ECG data comprising 21,837 records from 18,885 patients. These highly realistic synthetic physiological data can significantly amplify the physiological datasets that required by HDT, to enhance the performance.

The medical imaging data hold significant importance for HDT. Medical imaging provides detailed insights into a PT's anatomy, allowing for a precise digital representation. This information aids in creating accurate simulations of physiological structures and functions within the VT. Besides, by integrating this data, the VT can simulate and predict the progression of diseases. However, existing medical imaging acquisition methods rely on medical imaging devices, such as MRI or computed tomography (CT) scanners, which encounter high costs, patient discomfort, and safety concerns. To this end, GAI can generate synthetic medical images, augmenting existing datasets without additional patient scans. For instance, Moghadam et al. in [61] proposed a diffusion model-based approach for the synthesis of histopathology images, as shown in Fig. 4. Specifically, the authors used color normalization to force the diffusion model-based approach to learn morphological patterns, and used perception prioritized weighting, aiming to prioritize focusing on diffusion stages with more important structural histopathology contents. Experimental results showed that the proposed approach outperformed the GAN-based approach proposed in [62] by generating high quality histopathology images of brain cancer. Similarly, Yoon et al. in [63] proposed a sequence-aware diffusion model (SADM) for the generation of longitudinal medical images. SADM used a sequence-aware transformer as the conditional module in the diffusion model. This enabled learning longitudinal dependency even with missing data during training and allowed autoregressive generation of a sequence of medical images during inference. Experimental results showed that SADM can generate high quality longitudinal cardiac and brain MRI. While GAN and diffusion model made remarkable progress

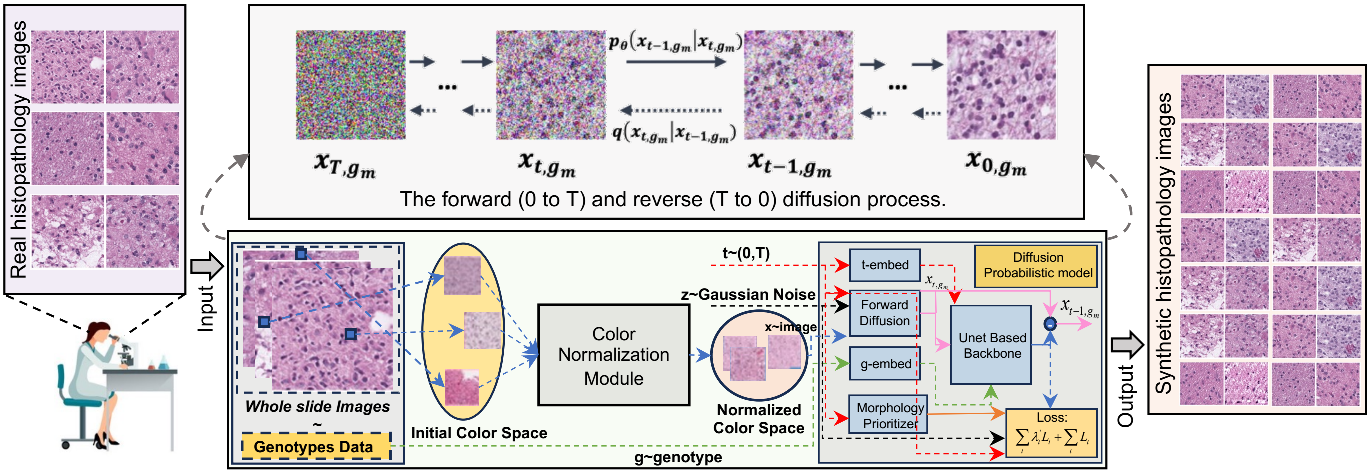

Fig. 4. Overview of the diffusion model-based histopathology image synthesis approach proposed in [61]. The real histopathology images are extracted the genotype information firstly, then it is used as a conditional input to the diffusion probabilistic model, which generates synthetic histopathology images that are tailored to specific genotypes.

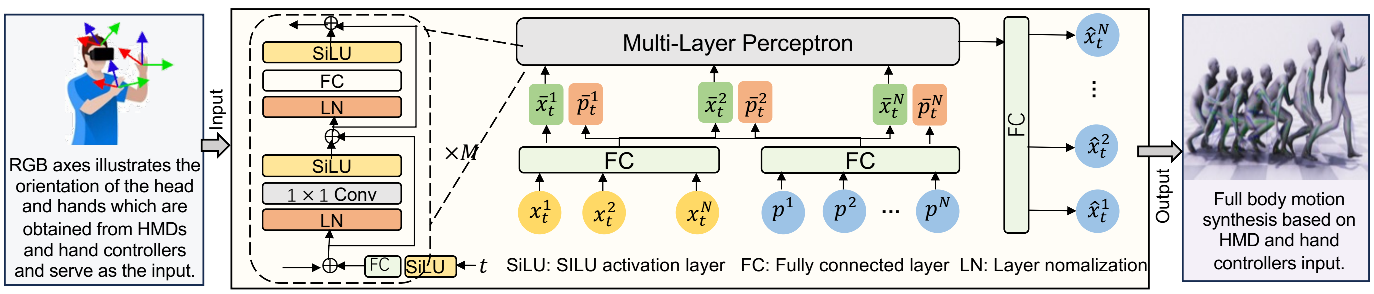

Fig. 5. Overview of AGRoL proposed in [67]. It takes the orientations of the head and hands from HMDs and hand controllers as the input. These input processed by AGRoL. The architecture of AGRoL is presented in the middle of this figure, where t is the noising step. $x_t^{1:N}$ denotes the motion sequence of length N at step t, which is pure Gaussian noises when $t = 0$. $p^{1:N}$ denotes the sparse upper body signals of length N. $\hat{x}_t^{1:N}$ denotes the denoised motion sequence at step t. The output is the synthesized full body motion.

in medical image generation, they cannot explicitly learn the probability density function of the input data and are highly sensitive to the hyperparameter selections. To mitigate these issues, Hajij et al. in [49] investigated normalizing flows (NFs) based approach as an alternative for synthesizing medical images. Particularly, the authors trained a RealNVP [64], a popular NF model for medical image synthesis, on two medical imaging datasets: chest X-ray [65] and skin cancer [66]. The experimental results showed that the NF-based medical image synthesis approach is an attractive alternative to GAN-based and diffusion model-based approaches.

The motion data is significant for HDT, which can be used to model and simulate a PT's entire body posture and movements in the digital space. Motion modeling in HDT can enable critical HDT services, such as motion monitoring during rehabilitation and injury prediction during exercise. Existing motion data acquisition mainly through wearable devices [8], which has several drawbacks hindering the complete and accurate motion data acquisition. For example, humans may not always be equipped with a large number of motion sensors for acquiring comprehensive motion data, which results in the acquired data can only characterize partial motion. Moreover, issues related to signal transmission, such as poor connectivity, signal interference, or obstructions caused by complex human mobility, can result in data loss or corruption during transmission. In this regard, GAI has been successfully applied in motion data synthesis, which can be a promising solution. For instance, Dittadi et al. in [40] proposed a VAE-based approach to generate full-body motion based on an impoverished control signal coming from HMDs. Specifically, to reconstruct the articulated poses of a human skeleton from noisy streams of head and hand pose, the authors proposed VAE-based approach decomposed the problem into a generative model of human pose, with an inference model that mapped input signals into the learned latent embedding. Experiment results showed that the proposed approach can faithfully reconstruct the walking motion of the person wearing an HMD. Additionally, to accelerate the motion generation rate to meet online application requirements, Du et al. in [67] proposed a diffusion model-based approach, called avatars grow legs (AGRoL), to generate the full-body poses conditioned on the sparse tracking signals from HMDs, as shown in Fig. 5. To enable gradual denoising and produce smooth motion sequences, the authors proposed a block-wise injection scheme that added diffusion timestep embedding before every intermediate neural network block. With this timestep embedding strategy, AGRoL achieved SOTA performance on the full-body motion synthesis task without any extra losses that were commonly used in other

TABLE IV
SUMMARY OF GAI-BASED APPROACH FOR COMMUNICATION IN HDT.

Ref.	Focus	Limitation of Traditional Methods	GAI-based Solution
[70]	Massive data transmission	High bandwidth consumption, more latency, and experience with bad QoS	Proposed a GAN-based semantic communication framework to transmit images or videos, where the proposed pix-to-pix GAN was used to reconstruct and denoise the received frames
[72]			Proposed a diffusion model-based semantic communication, where a semantic diffusion model was designed to reconstruct photorealistic images at the receiver side
[73]	Cross-modal information transmission	Missing signals and distortion	Proposed a GAN-based approach to transform the images into the corresponding haptic signals
[74]			Proposed the cross-domain GAN, DiscoGAN, to generate the desired haptic spectrograms belonging to corresponding category
[75]			Proposed a VAE-based approach for bidirectional mapping between visual and haptic signals

motion prediction methods. In addition, due to the lightweight architecture, AGRoL can generate realistic, smooth motions while achieving real-time inference speed, making it suitable for online applications.

In summary, Generative AI offers promising solutions to address the limitations of current data acquisition methods in HDT by generating diverse and high-realistic datasets.

B. GAI-enabled Communication

HDT relies on real-time data transmission to keep synchronization between any PT-VT pair to ensure high-fidelity of VT [8]. This synchronization is, however, data-driven and delay-sensitive. Furthermore, data acquired in the physical world is often massive and complex. In addition to real-time synchronization, interacting with VTs involves more complex information that needs to be transmitted between the PT and users in the physical world. Multi-modal information, such as 3D virtual items, text, images, haptic signals, among others, needs to be transmitted in HDT under various applications, such as the virtual surgery, to enhance the immersive experience. These specific characteristics place a significant burden on current communication networks. To address these issues, this subsection delves into the GAI-aid semantic communication [52] and cross-modal communication [54] for communication in HDT, as summarized in Table IV.

Semantic communication is expected to enable the data transmission between the PT and VT pair in HDT, tackling the challenges of unnecessary transmission of vast amounts that cause high bandwidth consumption, more latency, and experience with bad quality of service (QoS) by only transmitting meaningful and task-oriented information extracted from the original information [68], [69]. Generally, semantic communication extracts the "meaning" of any transmitted information at the transmitter and encodes the extracted features. Then, this semantic information is transmitted to the intending receiver and is "interrupted " and decoded by the receiver. GAI with its creativity is applied in receiver side to reconstruct the original information from the received semantic information. For instance, Raha et al. in [70] proposed a GAN-aid semantic communication framework for transmitting images or videos. Specifically, by considering resource limitations inherent in edge devices, such as HMDs, and the need for low-latency transmission in HDT applications, such as the real-time PT-VT synchronization task, the authors utilized a lightweight mobile segment anything model [71] for essential semantic information extraction from the images or videos. Then, on the receiver side, the authors proposed a pix-to-pix GAN approach to reconstruct and denoise the received semantic frames. The simulation results showed that the proposed framework can reduce up to 93.45% of the communication cost while maintaining the original information. In addition to GAN-aid semantic communication, the diffusion model with strong synthesizing multimedia content abilities has also been applied in this field. For example, Grassucci et al. in [72] introduced a semantic diffusion model designed to reconstruct photorealistic images at the receiver side. This model was trained using noisy semantics and incorporated a fast denoising semantic block to enhance the quality of inferred images. Consequently, the receiver can reconstruct semantically-consistent samples from the compressed semantic images transmitted by the sender over a noisy channel. These methods can be applied in HDT for rehabilitation training to reduce bandwidth usage and latency. The cameras are utilized to capture the patient's motion, and from the captured extensive images or videos, semantic features are extracted. Subsequently, the highly-compressed semantic information is transmitted to the digital world, where GAI is employed to regenerate the images or videos, and then, synchronizing the motion of the corresponding VT. This approach enables real-time monitoring and analysis of rehabilitation conditions [13], [69].

The interactions in HDT applications, such as virtual surgery, commonly involve with audio, visual and haptic signals to provide users with human interactive and immersive experiences. To support such interactions within HDT, cross-modal communication [54], which involves collaborative audio-visual and haptic interactions, presents a promising solution. It adeptly resolves the distinct requirements among these modalities when they coexist. However, long transmission distance between the physical and digital world or possible poor network conditions may result in missing signals and distortion, negatively impacting the users' experiences. To this end, researchers have explored using GAI to enable cross-modal signal reconstruction within the cross-modal communication to compensate for defective signals and achieve high-fidelity communication. For instance, by fully using of the correlation between image and haptic modalities, Liu et al. in [73] proposed a GAN-based approach to transform the images into the corresponding haptic signals. Specifically, the authors extract the image features firstly to obtain the required category information. Then, they adopted the cross-domain GAN, DiscoGAN [74], to generate the desired haptic spectrograms belonging to that category. In addition to cross-modal unidirectional mapping illustrated above, the cross-modal bidirectional mapping is investigated to enhance the cross-modal signal reconstruction. For example, Fang et al. in [75] proposed a VAE-based approach for bidirectional mapping between visual and haptic signals. Specifically, the authors firstly adopted the visual VAE model and the haptic VAE for compressing visual and haptic data, respectively. Then, they employed a conditional flow model to connect the latent feature spaces of these two VAE models. The forward process of the flow model was the mapping from the haptic to the visual latent feature space, while the reverse process was the mapping from the visual to the haptic latent feature space. Based on this, the cross-modal bidirectional mapping between visual and haptic signals can be successfully implemented on one model. These GAI-aid cross-modal communication approaches are crucial for cross-modal interactions in HDT. For example, during virtual acupuncture training, when a doctor interacts with a patient's VT, haptic signals may be missed or distorted due to the degraded channel conditions, while the visual signals are successfully received by the doctor's VR device [76]. To provide the immersive experience, the GAI-aid haptic signal reconstruction approach can be utilized to reconstruct the haptic signals from the visual signals.

C. GAI-enabled Data Management

Data management is a crucial step in the successful implementation of HDT. First, the collected multi-source data in HDT may have the characteristics of heterogeneity, multi-scale and high noises. Hence, as the critical step in data management, data pre-processing is indispensable in HDT, such that issues like missing data and noise data can be properly handled, and thereby providing high-quality data for downstream tasks in HDT [8]. Moreover, the data in HDT are highly sensitive, especially for individual-level data, and any leakage may result in serious ethical and moral concerns. Therefore, effective security and privacy schemes are imperative to protect data in HDT. In the following, we will discuss the application of GAI in data pre-processing, followed by its application in security and privacy schemes, as summarized in Table V.

Traditional data imputation methods, such as K-nearest neighbors-based [77] and deep learning-based imputation [78], often rely on existing observed data for modeling and prediction, and cannot directly generate new data samples. This will make it hard for traditional data imputation methods to handle missing severe data issues, such as data modality missing. To this end, GAI, with its powerful generative capabilities, can effectively handle the missing data issues in HDT. For instance, with the potential of substituting missing data accurately and efficiently, Dong et al. in [79] used GAN to impute missing values in large clinical datasets collected from PTs with mixed-type variables. Specifically, the method adopted by authors was generative adversarial imputation nets (GAIN), where the generator observed some components of a real clinical data vector, imputed the missing components conditioned on what was actually observed, and outputs a complete clinical data vector. The discriminator then took a completed clinical data vector and attempted to determine which components were observed and which were imputed. Additionally, to ensure that discriminator forced generator to learn the desired data distribution, the authors provided discriminator with some additional information in the form of a hint vector. The experimental results showed that GAIN outperformed the traditional imputation models, MICE [80] and missForest [81], in terms of accuracy in the imputation of missing data in clinical datasets, particularly for imbalanced and skewed data, and when the missingness rate was high (50%). Acquiring multi-modality data is crucial for the implementation of HDT. However, the missing modalities may happen in some conditions, such as failure in data transmission. Fortunately, diffusion models have shown favorable results for generating missing modalities utilizing cross-modalities and producing ones using other modality types. For instance, Lyu et al. in [82] proposed a GAI framework, called diffusion and score-matching models, which took advantage of the recently introduced denoising diffusion probabilistic models (DDPMs) [83] and score-based diffusion models [84], for translating MRI to CT. Specifically, they presented conditional DDPM and conditional stochastic differential equation (SDE) [85], where their reverse process was conditioned on T2w MRI images. The authors adopted the DDPM and SDE with three different sampling methods, namely Euler-Maruyama, Prediction-Corrector, and probability flow ordinary differential equation. Their extensive experiments on the Gold Atlas male pelvis dataset [86] demonstrated that the proposed diffusion models outperformed both GAN and CNN-based methods [87] regarding structural similarly index measure and peak signal-to-noise ratio. Similarly, to cope with the missing modality issue, Meng et al. in [48] proposed a unified multi-modal conditional score-based generative approach (UMM-CSGM), which synthesized the missing modality based on all remaining modalities as conditions. The proposed model was a conditional SDE format, employing only a score-based network to learn different cross-modal conditional distributions. The experimental results showed that the UMM-CSGM could generate missing-modality images with higher fidelity and structural information of the brain tissue compared to GAN-based methods [88]–[92].

Noise data is a common issue during the data acquisition in HDT. Noise reduces the data quality and is especially significant when the point of interests are minor and have relatively low contrast, which hinders the downstream tasks in HDT [93]. Traditional data denoise methods, such as CNN-based [94] and U-Net-based methods [95], highly rely on a significant amount of labeled training data to build the denoising models. However, obtaining such data is challeng-

TABLE V
SUMMARY OF GAI-BASED APPROACH FOR DATA MANAGEMENT IN HDT.

Ref.	Focus	Limitation of Traditional Methods	GAI-based Solution
[79]	Data imputation	Lack of generative capability	Proposed the generative adversarial imputation nets to impute missing values in large clinical datasets collected from PTs with mixed-type variables
[82]			Proposed a GAI framework, called diffusion and score-matching models, to translate MRI to CT
[48]			Proposed a unified multi-modal conditional score-based generative approach, to synthesize the missing modality based on all remaining modalities as conditions
[47]	Data denoise	Require massive labeled training data	Proposed a diffusion model-based framework for PET denoising, called PET-DDPM
[97]			Proposed a diffusion model-based denoising method, called denoising diffusion models, to denoise diffusion MRI
[100]	Data security and privacy	Difficult to balance between data privacy and usability	Proposed a GAN-based framework to generate synthetic data for anonymization of EHR data while closely approximating the distribution of original HER data
[104]			Proposed a GAN-based framework, EHR-Safe, to generate EHR data while guaranteeing fidelity and privacy

ing, especially at the individual level. To this end, with the strong generative abilities of GAI, it is convenient for diverse denoising problems in HDT data. For instance, PET is a medical imaging technique used to detect metabolic activities within the human body. The information provided by PET can be used to update or calibrate the VT models. However, due to various physical degradation factors, PET often suffers from low signal-to-noise ratio (SNR) and resolution. To denoise the PET images, a DDPM-based method has been proposed for PET image denoising. Gong et al. in [47] proposed the DDPM-based framework for PET denoising, called PET-DDPM, which collaborated with an assistive modality embedding as prior information to DDPM formulation. Quantitative results demonstrated that PET-DDPM outperformed U-Net based denoising networks [96] in peak SNR and structural similarity index measure, showcasing its superior performance in PET denoising. For denoising MRI that with severely SNR-limited, Xiang et al. in [97] proposed a self-supervised denoising method, called denoising diffusion models for denoising diffusion MRI (DDM^2). Specifically, their approach consisted of three stages, which integrated statistic-based denoising theory into diffusion models and performed denoising through conditional generation. During inference, they represented input noisy measurements as a sample from an intermediate posterior distribution within the diffusion Markov chain. Ultimately, their trained diffusion model can produce clean MRI images unsupervised.

HDT environments may suffer from various security threats, such as eavesdropping attacks in data transmission between PTs and VTs causing privacy leakage. A typical privacy protection mechanism for HDT data is anonymization [98], [99]. However, this method can easily compromise the data distribution, and it is hard to balance data privacy and usability. To this end, GAI can generate synthesized HDT data, which maximally preserves original data distribution while guaranteeing data security and privacy. For instance, to lower the risk of breaching individual confidentiality during data

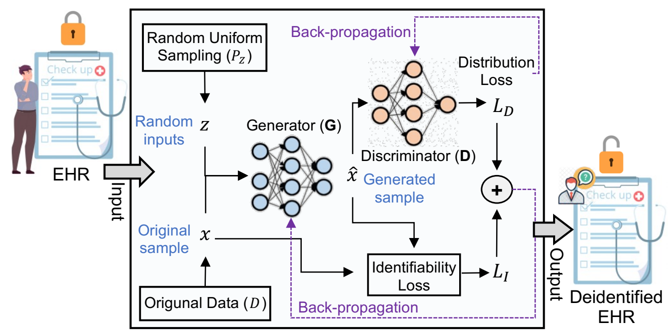

Fig. 6. Block diagram of ADS-GAN for individual EHR anonymization [100]. The generator uses original sample (x) and random vector (z) to generate sample ($\hat{x}$). The summation of distribution loss (LD) and identifiability loss (LI) is back-propagated to the generator. Both the generator and discriminator are implemented with multi-layer perceptrons.

sharing in HDT, Yoon et al. in [100] proposed a GAN-based framework, called anonymization through data synthesis using GAN (ADS-GAN), for generating synthetic data that closely approximates the distribution of variables in an original EHR dataset, achieving anonymization of EHR, as shown in Fig. 6. Specifically, ADS-GAN generated the synthetic data conditioned on the original data, and different from traditional conditional GAN framework, conditioning variables in ADS-GAN were not pre-determined but instead were optimized from real EHR. Therefore, the ADS-GAN can generate better quality synthetic data than the traditional conditional GAN frameworks [87], [101]–[103] while minimizing individual identifiability. However, such methods have limitations regarding the fundamental aspects of real-world EHR data, such as dealing with missing features, varying feature length, categorical features and static features (beyond time series). These fundamental challenges require a holistic re-design in GAN-based synthetic data generation systems. To this end, Yoon et al. in [104] further designed, EHR-Safe, that can jointly represent these diverse data modalities while preserving

TABLE VI
SUMMARY OF GAI-BASED APPROACH FOR DIGITAL MODELING IN HDT.

Ref.	Focus	Limitation of Traditional Methods	GAI-based Solution
[107]	Digital modeling of human cells	Require massive and accurate simulation parameters, which are particularly challenging to acquire from biological entities in the human body	Proposed a VAE-based cell DT, scGen, for predicting transcriptional response to drug perturbations
[109]			Proposed a VAE-based cell DT to predict cell morphology by generating cell images
[44]			Proposed a transformer-based cell DT, scGPT, to facilitated the modeling of various aspects of cellular processes
[111]	Digital modeling of human tissues		Proposed a GAN-based tissue DT, PathologyGANs, for simulating the morphological characteristics of cancer tissue
[38]			Proposed a GAN-based tissue DT, ScarGAN, to simulate the pathology of scar tissue on healthy myocardium
[115]	Digital modeling of human organs		Proposed a GAN-based organ DT, ReconGAN, to predict the risk of vertebral fracture
[117]			Proposed a GAN-based cardiac DT to synthesize and simulate myocardial velocity

the privacy of EHR data. Specifically, EHR-Safe was based on a two-stage model that consisted of sequential encoder-decoder networks and GANs. To circumvent the heterogeneity of EHR data mentioned above, the authors used the sequential encoder-decoder networks to learn the mapping from the raw EHR data to low-dimensional representations and vice versa, and the mapped information is used for GAN training. The experimental results showed that the fidelity of synthetic data generated by EHR-Safe is almost-identical with real data, while yielding ideal performance in practical privacy metrics.

In summary, GAI is a promising technique for data management in HDT, handling data pre-processing and security and privacy issues.

D. GAI-enabled Digital Modeling

In HDT, digital modeling refers to digitally model the PT and virtualized in the digital world based on the acquired data and various digital modeling technologies, forming the VT.

The classic HDT digital modeling technology, mechanistic modeling, has pioneered the integration of biology and physiology domain knowledge to allow robust and accurate modeling [105]. For instance, a human heart DT model has been developed, which consisting of approximately 100 million virtual heart cell patches, with each patch modeled by around 50 equations [106]. The human heart DT accurately represented the interconnected cardiac muscle cells, effectively simulating the transmission of electrical currents through these cells and the subsequent initiation of the heartbeat. However, as the human heart DT, mechanistic modeling requires massive and accurate simulation parameters, which are particularly challenging to acquire from biological entities in the human body, i.e., from molecular to organ level, and are typically limited to only a subset of all available biomolecular [105]. GAI techniques can overcome these challenges, through learning the underlying distribution and sequential or temporal relations of data for modeling. In the following, we provide a survey of the application of GAI in HDT digital modeling, including digital modeling of human cells, tissues and organs, as summarized in Table VI.

First, GAI techniques have shown their strong abilities in digitally modeling human cells. For instance, Lotfollahi et al. in [107] built a VAE-based cell DT, scGen, for predicting transcriptional response to drug perturbations. Specifically, scGen combined VAE and latent space vector arithmetics for high-dimensional sing-cell gene expression data. scGen can accurately model perturbation and infection response of cells across cell types. The results of the simulations showcased that scGen successfully learned cell-type responses, indicating its ability to capture distinguishing features between responding and non-responding genes and cells. Furthermore, the simulation results provided evidence for the enhanced generalization capabilities of scGen compared to mechanistic modeling approaches, which are typically tailored to specific cellular settings. In addition, the authors further developed another VAE-based cell DT, compositional perturbation autoencoder (CPA), to predict cellular response to unseen drugs, drug combinations and dosages in high-throughput screens [108]. CPA combined the interpretability of linear models with the flexibility of deep learning approaches for single-cell response modeling. The authors envisioned that with the accurate modeling of the cell, CPA will accelerate therapeutic applications using single-cell technologies. Similarly, Donovan-Maiye et al. in [109] developed a VAE-based cell DT to predict cell morphology by generating cell images. Specifically, they employed stacked conditional β-VAE to first learn a latent representation of cell morphology, and then learn a latent representation of subcellular structure localization which is conditioned on the learned cell morphology under treatment. Inspired by parallels between linguistic constructs and cellular biology, where text comprises words, similarly, cells are defined by genes, pre-trained transformers are envisioned to model the cell DT. Cui et al. in [44] developed a transformer-based cell DT, scGPT, a generative pre-trained foundation model that harnessed the power of pre-trained transformers on a vast amount of single-cell sequencing data, as shown in Fig. 7. The use of transformers in scGPT enabled the simultaneous learning of gene and cell embeddings, which facilitated the modeling of various aspects of cellular processes. In addition, by leveraging

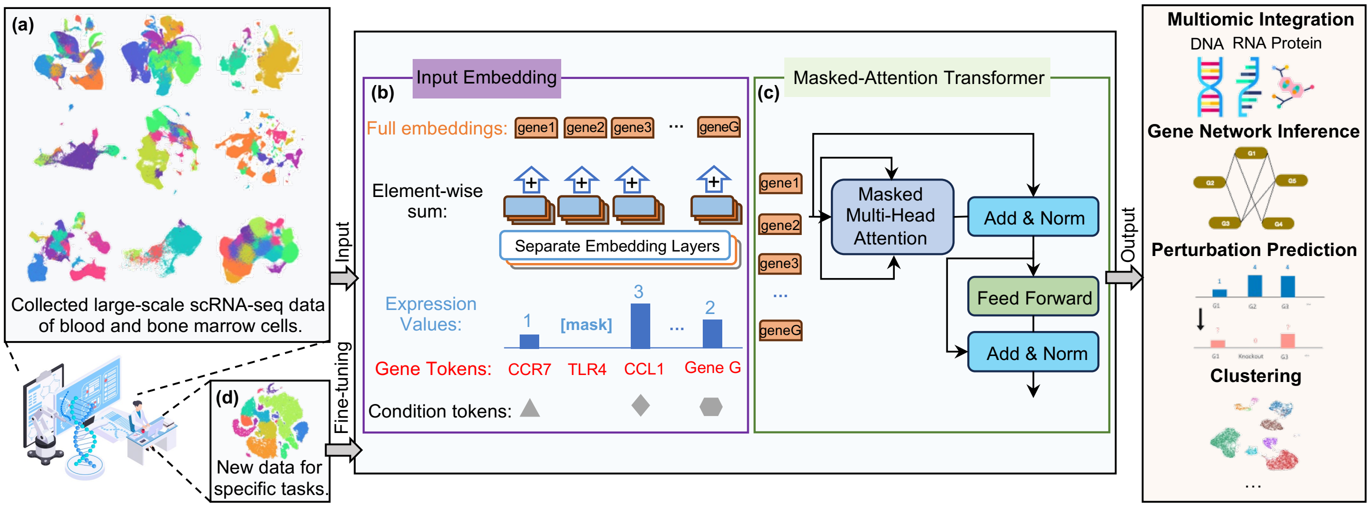

Fig. 7. Overview of the transformer-based cell DT, scGPT, proposed in [44]. (a) Large-scale scRNA-seq datasets from cell atlas are obtained from medical institutions and input to the input embedding module. (b) The input contains three layers of information, the gene token, expression value, and condition tokens (modality, batch, perturbation condition, et al.). (c) A specially designed mask-attention transformer is proposed to conduct generative pre-training on single-cell sequencing data. (d) For downstream applications (e.g., multiomic integration, gene network inference, perturbation prediction, and clustering), the pre-trained model weights can be finetuned on the new data.

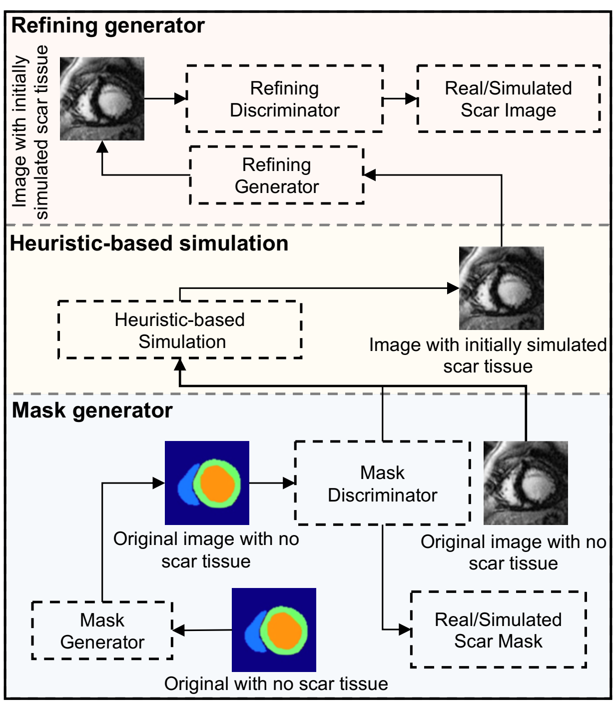

Fig. 8. Overview of the GAN-based tissue DT, ScarGAN, propoesd in [38]. A mask generator simulates the shape of scar tissuesegmentation mask (left ventricular endo is light blue; left ventricular myo is green; left ventricular endo is orange; scartissue is red); a heuristic-based method provides an initial simulated scar tissue using the simulated shape; a refining generator add details of scar tissue to the image.

the attention mechanism of transformers, scGPT captures gene-to-gene interactions at the single-cell level, providing an additional layer of interpretability.

GAI has been successfully applied in the digital modeling of human tissues for digital pathology [110]. For instance, Quiros et al. in [111] developed a GAN-based tissue DT, PathologyGANs, for simulating the morphological characteristics of cancer tissue. Specifically, PathologyGANs combined BigGAN [112], StyleGAN [113], and relativistic average discriminator [114] to learn representations of entire tissue architecture. Then, they used these characteristics to structure PathologyGANs' latent space (e.g., color, texture, spatial features of cancer and normal cells, and their interaction). Thus, PathologyGANs can generate high-fidelity cancer tissue images from the structured latent space. The simulation showed that the quality of the generated cancer tissue images did not allow pathologists to reliably find differences between real and generated images. It indicated that the proposed GAN-based tissue DT can accurately characterize the features of cancer tissues. To reduce the frequent need to collect scans from patients, Lau et al. in [38] built a GAN-based tissue DT, ScarGAN, to simulate the pathology of scar tissue on healthy myocardium, as shown in Fig. 8. Specifically, ScarGAN was based on chained GAN, and the simulation process included 3 steps: i) a mask generator to simulate the shape of the scar tissue; ii) a domain-specific heuristic to produce the initial simulated scar tissue from the mask; iii) a refining generator to add details to the simulated scar tissue. The experimental results conducted by the authors showed that ScarGAN can high realistically simulate scar tissue on normal scans, such that experienced radiologists could not distinguish between real and simulated scar tissue.

GAI has been used in the digital modeling of human organs for computationally reproducing normal and pathological organ function and treatment effect. For instance, Ahmadian et al. in [115] built a GAN-based organ DT, coined ReconGAN, to predict the risk of vertebral fracture (VF). Specifically, ReconGAN consisted of a deep convolutional GAN (DCGAN) [116], image-processing steps, and finite element (FE) based shape optimization to reconstruct the vertebra model. The synthetic trabecular microstructural models generated by DCGAN were infused into the vertebra cortical shell extracted from the patient's diagnostic CT scans using an FE-based shape optimization approach to achieve a smooth transition between trabecular to cortical regions. The final geometrical model of the vertebra was converted into a high-fidelity FE model to simulate the VF response using a continuum damage model under compression and flexion loading conditions. The experiments implemented by the authors demonstrated that the built GAN-based vertebra DT can accurately simulate and predict the risk of VF in a cancer patient with spinal metastasis. As the core organ of human bodies, the GAI aided human heart digital modeling has also been investigated. For instance, to overcome the long acquisition time and complex acquisition of cardiac data, Xing et al. in [117] developed a GAN-based cardiac DT, hybrid deep learning (HDL), to synthesize and simulate myocardial velocity maps from real-world three-directional CINE multi-slice myocardial velocity mapping (3Dir MVM) data. Specifically, the HDL was featured by a hybrid UNet and a GAN with a foreground-background generation scheme. The experimental results demonstrated that the synthetic 3Dir MVM data generated from the HDL algorithm can accurately and quantitatively assess the cardiac motion in three orthogonal directions of the left ventricle.

In summary, the data-driven GAI based digital modeling can effectively overcome the shortcomings of mechanistic modeling, including requiring accurate simulation parameters and sufficient prior domain knowledge. Therefore, GAI is a promising technique for digital modeling of HDT.

E. GAI-enabled Data Analysis

Data analysis is the crucial component in HDT, which analyzes the data collected from the physical world and the data generated in the digital world. The results from data analysis are essential information in providing HDT services, such as diagnosis and prescription. Traditional data analysis methods, such as CNN-based [118] and U-Net-based methods [119], highly rely on a large amount of labeled data and difficult in handling high-dimensional HDT data. To this end, GAI has showcased remarkable potential in data analysis, including classification, segmentation, anomaly detection and prediction, as summarized in Table VII.

The data classification is crucial for HDT services, such as early prevention and diagnosis of diseases. For example, when doctors remotely diagnose a patient through the patient's VT, classification models can quickly and automatically identify and classify the patient's diseases. Recent studies have shown the potential of the GAI-based methods in this field. For instance, Dhinagar et al. in [120] investigated vision transformer (ViT) architecture for high-stakes neuroimaging downstream tasks, focusing on Alzheimer's disease classification based on 3D brain MRI. The authors evaluated the effects of different training strategies including pre-training, data augmentation and learning rate warm-ups followed by annealing, and emphasized the importance of these strategies in neuroimaging applications. Similarly, Dong et al. in [121] proposed a ViT-based end-to-end multi-label arrhythmia classification model, called CNN-DVIT, for the 12-lead ECG with varied-length

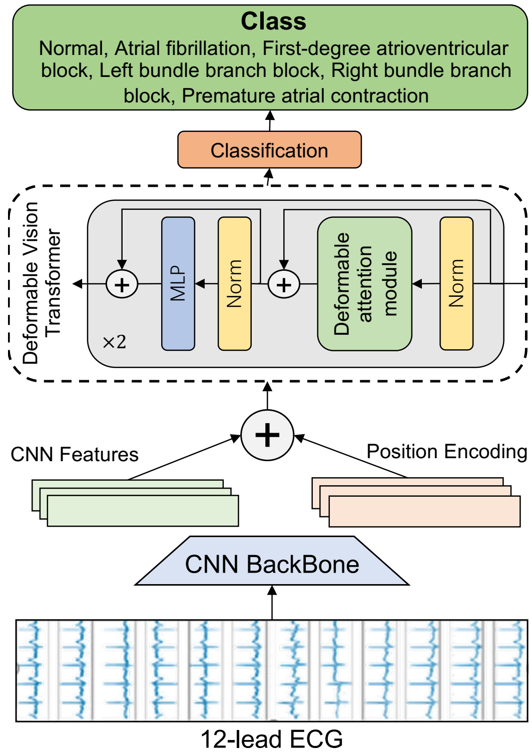

Fig. 9. Overview of the GAI-based classification model, CNN-DVIT, proposed in [121]. It is able to take continuous 12-lead ECG signals as input and output the arrhythmia diagnosis result in an end-to-end manner. Specifically, the model is composed of three main components: (a) a CNN-based backbone for feature extraction from each lead; (b) a deformable attention transformer encoder module to combine the CNN-extracted features and the positional encoding; and (3) the classification layer to obtain the probability that each patient may have for each type of heart disease.

recordings, as shown in Fig. 9. Specifically, CNN-DVIT was based on a combination of CNN with depthwise separable convolution, and a ViT architecture with deformable attention. Besides, the authors introduced the spatial pyramid pooling layer to accept varied-length ECG signals. Experimental results showed that CNN-DVIT outperformed the most recent transformer-based ECG classification methods [122].

Data segmentation play a crucial role in data analysis within the realm of HDT. Segmentation allows complex HDT data to be subdivided and analyzed to extract critical information. For instance, through image segmentation, the different regions and structures within an image can be separated, such as organs and tumors. This facilitates quantitative analysis and measurements, such as measuring organ volumes and calculating tumor growth rates. This important information can be provided for HDT's subsequent services, such as surgical planning, treatment plans and diagnosis. GAI has been successfully implemented in this field. It is worth note that , for data analysis in HDT, analyzing organs or other human structures from medical images is not a deterministic pixel-wise process, but underlies the assessment of the whole image or, on smaller scale, assessing the neighboring pixels' diversity. In this regard, Rahman et al. in [123] leveraged the stochastic sampling step in the diffusion model to produce diverse and multiple masks. Specifically, the au-

TABLE VII
SUMMARY OF GAI-BASED APPROACH FOR DATA ANALYSIS IN HDT.

Ref.	Focus	Limitation of Traditional Methods	GAI-based Solution
[120]	Data classification	Require a large amount of labeled data and difficult in handling high-dimensional HDT data	Proposed a vision transformer architecture for Alzheimer's disease classification based on 3D brain MRI
[121]			Proposed a vision transformer-based end-to-end multi-label arrhythmia classification model, CNN-DVIT, for the 12-lead ECG with varied-length recordings
[123]	Data segmentation		Proposed a diffusion model-based ambiguous segmentation network, collectively intelligent medical diffusion, that can generate multiple plausible annotations from a single input image
[124]			Proposed a diffusion adversarial representation learning model for self-supervised vessel segmentation
[125]	Anomaly detection		Proposed a cascade VAE-based anomaly detector for outlier detection in medical images
[126]			Proposed an auto-encoding GAN framework for anomaly detection in chest radiographs
[127]			Proposed a weakly supervised anomaly detection method based on denoising diffusion implicit models
[41]	Prediction		Proposed a β-VAE based HDT to forecast clinical measurement trajectory in patients who went on to experience an ischemic stroke
[130]			Applied BERT-based models to ophthalmology clinical text from EHR, predicting the progression of glaucoma

thors introduced a single diffusion model-based ambiguous segmentation network, called collectively intelligent medical diffusion (CIMD), that can generate multiple plausible annotations from a single input image. The CIMD utilized the noisy segmentation ground-truth masks concatenated to the original image to prevent the conventional diffusion process usage in the segmentation task from producing more resilient results, rather than arbitrary masks. The authors validated the CIMD in three different medical image modalities, namely CT, ultrasound and MRI. The experimental results showed that CIMD outperformed existing SOTA ambiguous segmentation networks in terms of accuracy while preserving naturally occurring variation. Similarly, Kim et al. in [124] proposed a diffusion adversarial representation learning (DARL) model, which leveraged a denoising diffusion probabilistic mode with adversarial learning, for self-supervised vessel segmentation, aiming to diagnose vascular diseases and treatment planning, as shown in Fig. 10. Specifically, DARL model consisted of two main modules, where the diffusion module learned background image distribution, and the generation module generated vessel segmentation masks or synthetic angiograms using the proposed switchable spatially-adaptive denormalization algorithm. Experimental results showed that DARL model significantly outperforms existing unsupervised and self-supervised vessel segmentation methods.

In the context of HDT, anomaly detection for HDT data plays a critical role. Anomaly detection helps identify abnormal patterns or conditions in HDT data early, enabling proactive health alerts from HDT and timely intervention before a condition deteriorates. GAI has been successfully implemented in this field. For instance, Guo et al. in [125] designed a cascade VAE-based anomaly detector (CVAD) for outlier detection in medical images. Specifically, with a focus on the generalizability of anomaly detector, CVAD combined latent representation at multiple scales, before being fed to a discriminator to distinguish the out-of-distribution (OOD) data from the in-distribution data. The reconstruction error and the OOD probability predicted by the binary discriminator were used to determine the anomalies. Extensive experiments on multiple intra- and inter-class OOD medical imaging datasets showed CVAD's effectiveness and generalizability. The GAN-based approaches are also successfully applied in anomaly detection. For instance, Nakao et al. in [126] designed an auto-encoding GAN (α-GAN) framework, which was a combination of a GAN and a VAE, for anomaly detection in chest radiographs. The experimental results showed that α-GAN can correctly visualize various lesions including a lung mass, cardiomegaly, pleural effusion, bilateral hilar lymphadenopathy, and even dextrocardia. However, the VAE-based and GAN-based anomaly detection methods are often complicated to train or have difficulties to preserve fine details in the medical images [127]. To this end, Wolleb et al. in [127] proposed a weakly supervised anomaly detection method based on denoising diffusion implicit models (DDIMs) [45]. Specifically, the authors combined the deterministic iterative noising and denoising scheme with classifier guidance for image-to-image translation between diseased and healthy subjects. The authors applied this method on two different medical datasets, namely BRATS2020 brain tumor datasets and the CheXpert datasets [128]. Experimental results showed that this method can preserves details of the input image unaffected by the disease while realistically representing the diseased part.

Prediction plays a critical role in data analysis, which can provide informed information for HDT services. For instance, the disease progression and drug response can be predicted in patient' VTs based on HDT data. With this, doctors can

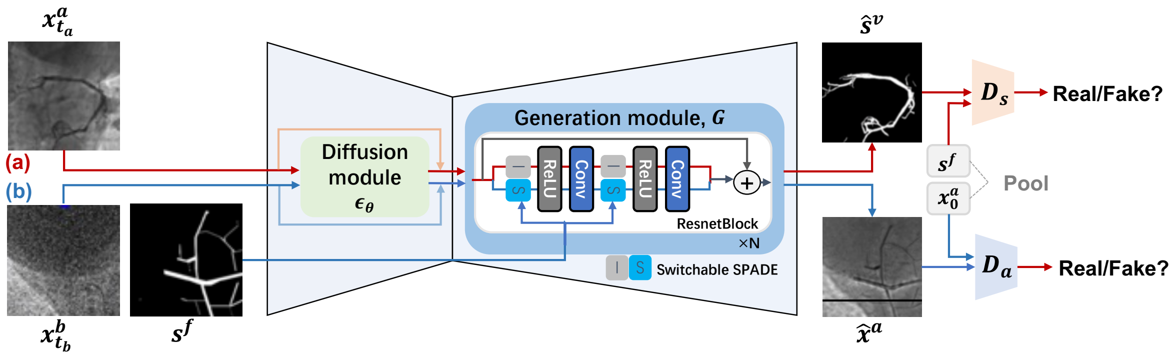

Fig. 10. Overview of the diffusion model and GAN-based vessel segmentation method, DARL, proposed in [124]. In path (a), given a real noisy angiography image $x^a_{t_a}$, DARL estimates vessel segmentation mask $\hat{s}^v$. In path (b), given a real noisy background image $x^b_{t_b}$ and a vessel-like fractal mask s^f, DARL generates a synthetic angiography image $\hat{x}^a$.

optimize the intervention and design the customized treatment plans for patients. GAI has been successfully implemented in this field. For instance, Allen et al. in [41] proposed a β-VAE based HDT to forecast clinical measurement trajectory in patients who went on to experience an ischemic stroke. Specifically, the β-VAE based HDT model was used to generate possible next steps in a patient's disease progression. The model was trained on data extracted from the medical information mart for intensive care-IV database [129]. The database contains 1216 patients with useable trajectories that experienced ischemic stroke. Experimental results demonstrated that the model can accurately forecast the progression of relevant clinical measurements in patients at risk of ischemic stroke, which was virtually indistinguishable from real patient data. Additionally, transformer-based approaches have also been used in prediction of disease progression. For instance, Hu et al. in [130] applied four BERT-based models, including the original BERT [131], BioBERT [132], RoBERTa [133], and DistilBERT [134], to ophthalmology clinical text from EHR, predicting the progression of glaucoma. Experiment results showed that these four BERT-based models outperformed clinical predictions by an ophthalmologist's review of the same clinical information. The authors qualitatively evaluated the BERT-based models by performing explainability studies based on the self-attention mechanisms of the BERT-models. Based on this, the authors can evaluate what types of words were most important for predictions.

In summary, GAI is a key enabling technology for data analysis in HDT, analyzing the HDT data for classification, segmentation, anomaly detection and prediction.

IV. GAI-driven HDT in IoT-Healthcare Applications

The remarkable capabilities of GAI-driven HDT in a wide range of IoT-healthcare applications, including personalized health monitoring and diagnosis, prescription, and rehabilitation. In this section, we will delve into the details of these IoT-healthcare applications, exploring how GAI-driven HDT is revolutionizing each of them.

A. Personalized Health Monitoring and Diagnosis

The GAI for HDT can be applied in personalized health monitoring and diagnosis. By leveraging the robust data analysis capabilities, GAI can analyze collected patient data in HDT, enabling anomaly detection for personalized health monitoring and diagnosis, as shown in Fig. 11. For example, a VT continuously receives data streams from its corresponding PT, namely the patient, encompassing vital metrics like heart rate, blood pressure, and ECG signals, etc. Utilizing GAI, the VT can analyze this data, identifying deviations from expected norms, such as cardiac irregularities or other health concerns. Detected anomalies trigger alerts to the patient and healthcare providers, signaling potential irregularities or cardiac issues. Consequently, GAI for HDT is a personalized health monitoring system that delivers timely health warnings.

GAI has been successfully applied in anomaly detection from the collected patient data in HDT. For example, Wang et al. in [135] proposed a GAN-based approach for ECG signal analysis, achieving automatic cardiac diagnosis, as shown in Fig. 11. Specifically, the proposed approach involved two-level hierarchical deep learning framework with GAN. The first-level model was composed of a memory-augmented deep autoencoder with GAN (MadeGAN), which aimed to differentiate abnormal signals from normal ECGs for anomaly detection. The second-level learning aimed at robust multi-class classification for different arrhythmia identification, which was achieved by integrating the transfer learning technique to transfer knowledge from the first-level learning with the multi-branching architecture to handle the data-lacking and imbalanced data issues. Experimental results showed that the proposed approach can effectively capture the disease-altered feature patterns from ECG signals, yielding better performance in predicting heart disease compared to other existing methods. Moreover, this hierarchical deep learning framework can be broadly implemented to study other patient data, such as EEG and PPG, for smart anomaly detection. Following this, Another GAI technique, VAE, has been applied in this field. Staffini et al. in [136] proposed a disentangled VAE with a bidirectional LSTM (BiLSTM) backend to detect in an unsupervised manner anomalies in heart rate data collected during sleep time with a wearable device. Specifically, the

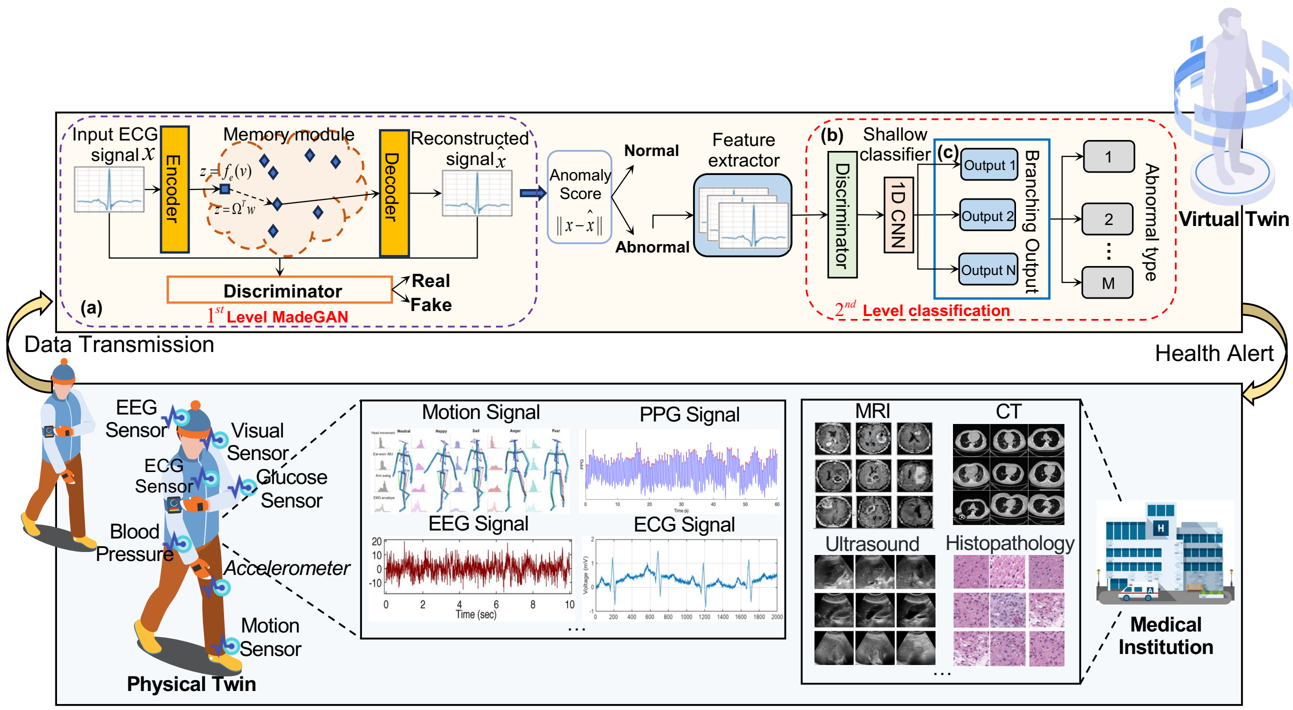

Fig. 11. The application of GAI-driven HDT in the personalized health monitoring and diagnosis, where the GAN-based approach, proposed in [135] is used as an example. The collected ECG signals from the physical twin are transmitted to the GAI-driven HDT for ECG signal analysis. Specifically, the GAN-based approach is a two-level hierarchical deep learning framework: (a) first-level MadeGAN for anomaly detection; (b) second-level classification for arrhythmia type identification; (c) Multi-branching output. If there are abnormities in collected data, the virtual twin will initiate health alerts to either the physical twin or medical institution for interventions.

added BiLSTM backend to the VAE model allowed it to capture contextual relationships in VAE-processed heart rate sequences by analyzing the information flow's forward and backward directions. The proposed approach can better model the considered time series and learn more accurate patterns. Experimental results showed that the proposed approach outperformed five well-known algorithms in terms of anomaly detection from the heart rate data.

Beyond cardiac diseases, GAI has demonstrated successful applications in anomaly detection for monitoring and diagnosing other conditions like breast cancer [137] and Alzheimer's [138]. These advancements utilize diverse patient data collected within HDT. Thus, GAI-driven HDT holds the promise of delivering precise, timely, and proactive personalized health monitoring and diagnosis, significantly enhancing patient care and overall well-being.

B. *Personalized Prescription*

The GAI-driven HDT can be applied in the personalized prescription. GAI empowers the HDT to predict and simulate treatment outcomes, enabling personalized prescription, as shown in Fig. 12. For instance, considering an Alzheimer's patient's VT, doctors can virtually test various candidate drugs on the VT. Then, leveraging GAI within the VT, the interactions of each drug with the patient's Alzheimer's disease can be predicted and simulated. Upon deriving insights from the GAI-driven HDT, doctors can pinpoint the most effective drug tailored to the patient's specific health profile.

GAI has been successfully applied in treatment outcome predictions and simulations inside the HDT for personalized prescriptions. For instance, Jarada et al. in [139] proposed a VAE-based approach, called similarity network fusion-collective VAE (SNF-CVAE), for predicting drug-disease interactions, as shown in Fig. 12. Specifically, SNF-CVAE integrated similarity measures, similarity selection, SNF, and CVAE to conduct a non-linear analysis and improve the drug-disease interaction prediction accuracy. Additionally, to further demonstrate the reliability and robustness of SNF-CVAE, the authors conducted two case studies on the top predicted drug candidates for potentially treating Alzheimer's disease and Juvenile rheumatoid arthritis, which were successfully validated against clinical trials and published studies. In addition, GAN has also been applied in this field. Xu et al. in [140] introduced a GAN-based method for predicting the short-term therapeutic outcomes of anti-vascular endothelial growth factor (VEGF) therapy in treating diabetic macular edema (DME). They employed the state-of-the-art GAN-style algorithm, pix2pixHD [141], renowned for progressively synthesizing high-resolution and realistic images. This approach was utilized to generate post-therapeutic optical coherence tomography (OCT) images. Their experimental results showcased that the established pix2pixHD model effectively produced post-therapeutic OCT images close to the ground truth, providing credible insights to aid ophthalmologists in predicting and assessing the response to anti-VEGF treatment.

The continuous advancement of GAI-driven HDT can offer the potential for more diverse and accurate predictions and

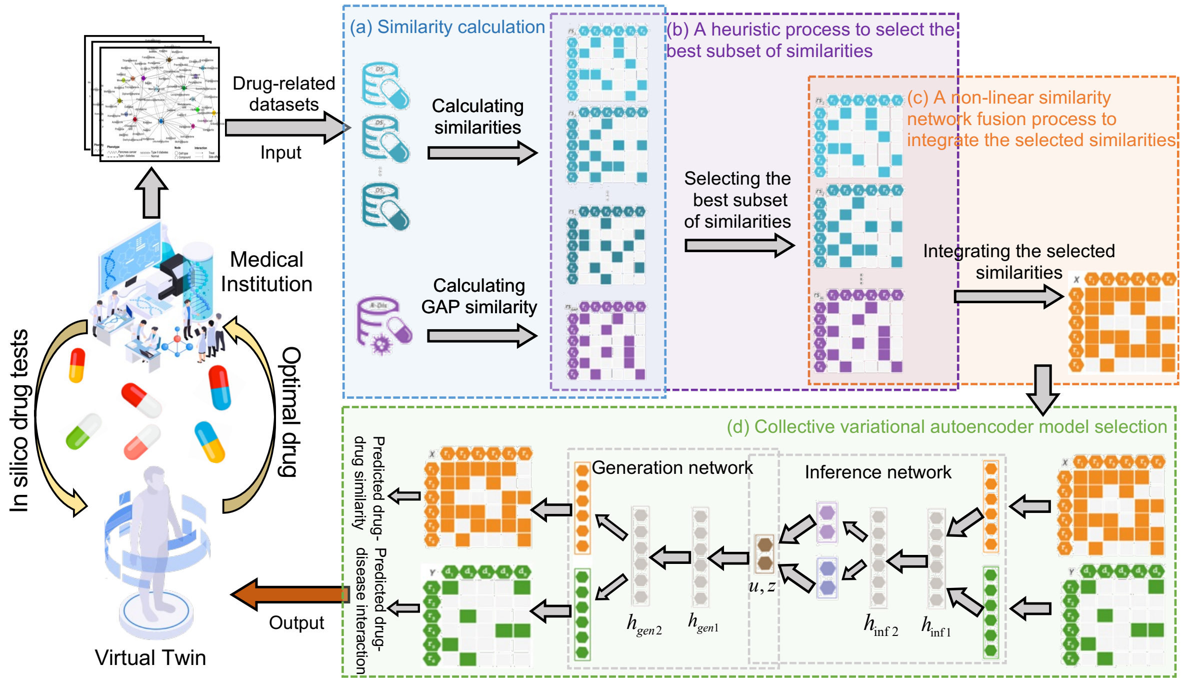

Fig. 12. The application of GAI-driven HDT in the personalized prescription, where the GAI-based approach, SNF–CVAE method, proposed in [139] is used as an example: (a) Calculating drug similarity matrices using the drug-related datasets obtaining from the medical institution. (b) Applying a heuristic process to select the most informative, least redundant subset of drug similarity matrices. (c) Applying a non-linear similarity network fusion process to integrate the selected similarity matrices into a comprehensive drug similarity matrix. (d) Training a collective VAE model with drug similarity information and drug–disease interactions to predict novel drug–disease interactions. Finally, the output is used to choose the optimal drug for the patient.

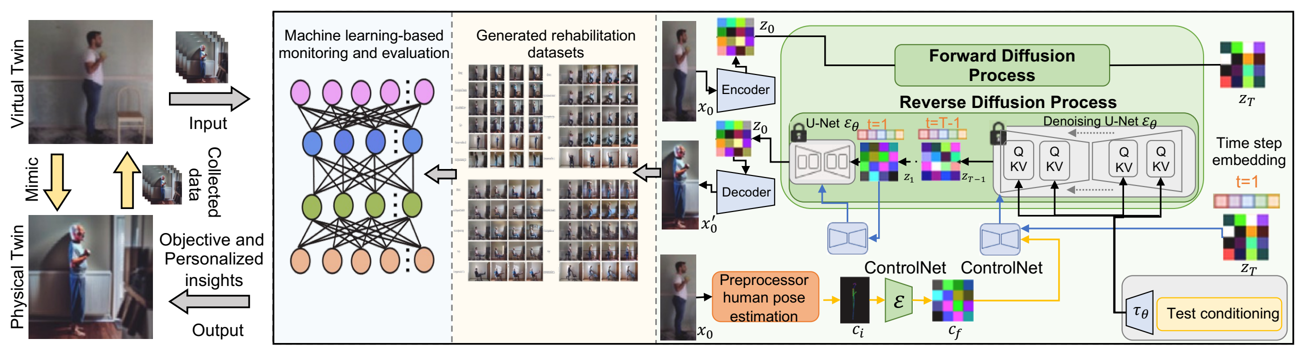

Fig. 13. The application of GAI-driven HDT in the personalized rehabilitation, where the GAI-based approach, pose-guided condition diffusion model, proposed in [142] is used as an example. The physical twin mimics the motions of virtual twin, and the motions of the physical twin are collected and input into the machine learning model for obtaining the physical twin's objective and personalized insights. The training datasets of the machine learning model are enriched by the pose-guided condition diffusion model, which is based on a latent diffusion model combined with the ControlNet.

simulations of treatment outcomes. This capability facilitates more versatile personalized prescriptions, ultimately optimizing treatment outcomes.

C. Personalized Rehabilitation

The GAI for HDT can be applied in the personalized rehabilitation training. Based on the built patient's VT, the doctor or physical therapist can simulate different rehabilitation training scenarios by controlling the VT's target posture. For example, if the patient needs balance training, the doctor can adjust the VT's posture to create a personalized unstable state that requires balance subject to the patient's profile. Then, the patient, namely the PT, observes the VT's movements and attempts to mimic its posture to improve their balance. Then, the IoT devices are used to collect the patient's motion data, and transmit to the VT for monitoring and evaluation. Then, the VT serves as virtual "coach" providing real-time feedback and guidance to help the patient posture correction and actively progress through their rehabilitation training.

During this personalized rehabilitation training, machine learning-based approaches are usually utilized to enhance the monitoring and evaluation process in the VT, providing objective and personalized insights. This method relies on substantial high-quality data to obtain a robust and accurate machine learning model. This data should encompass a diverse range of exercises and rehabilitation contexts, as well as holistic patient profiles. However, acquiring such data poses significant obstacles, such as data availability and privacy concerns, especially in individual-level. To this end, GAI can be used to generate synthetic data in the context of personalized rehabilitation training for enhancing the performance of the virtual "coach" role of VT, as shown in Fig. 13. For instance, Mennella et al. in [142] proposed a diffusion model-based approach, pose-guided condition diffusion model, to generate synthetic data that mimicked realistic-looking human movements in a rehabilitation context. Specifically, the data generation framework used the latent diffusion model in combination with ControlNet [143], and was trained on a prelabeled dataset of 6 rehabilitation exercises. The total 22,352 synthetic images can accurately capture the spatial consistency of human joint relationships for predefined exercise movements. Additionally, for enhancing the machine learning-based post-stroke rehabilitation assessment, Boukhennoufa et al. in [144] proposed a GAN-based approach to generate synthetic data for amplifying two post-stroke rehabilitation datasets. Specifically, the proposed approach incorporated a Siamese network and an additional discriminator to address original GAN mode collapse issue. Experimental results showed that the proposed approach could generate more diverse and realistic data, and improving classification models' accuracy in post-stroke rehabilitation assessment.

These synthetic datasets can be tailored to target various aspects of rehabilitation, including postural correction, joint mobility, balance training, and functional movements. Additionally, it can also be customized according to patient profiles, achieving personalized rehabilitation training. In summary, these GAI-based data augmentation methods are significant for HDT in rehabilitation training by enhancing the monitoring and evaluation process.

V. Future Research Directions

The GAI-driven HDT in IoT-healthcare is revolutionizing the healthcare industry, as surveyed above. This exciting field is nascent, and therefore, some critical issues remain unexplored and are of great importance.

A. Deriving Energy-Efficient Scheme for GAI-driven HDT

Implementing such a large model, GAI-driven HDT, requires substantial requirements in computation, communication, and storage resources due to the high parametric complexity of GAI-driven HDT and the necessity for vast datasets. For instance, 10,000 graphical processing unit (GPU) cards were employed to run an HDT brain [145], [146]. It will certainly result in large energy consumption, and generate significant carbon dioxide emissions, jeopardizing sustainability. Thus, the energy-efficient scheme for the implementation of GAI-driven HDT is imperative. Techniques such as pruning, knowledge distillation, quantization, and green learning, have been proposed to address this issue. However, these solutions come at the cost of sacrificing the accuracy of GAI-driven HDT, which is a critical issue, especially in IoT-healthcare. Therefore, it should carefully balance accuracy and sustainability in energy-efficient GAI-driven HDT.

B. Designing Human-centric Evaluation Metrics

Human-centric metric design for evaluating the performance of GAI-driven HDT in IoT-healthcare is inherently challenging. It should encompass the accuracy and reliability of the GAI generated content compared to the PTs' real health data. This involves defining parameters to evaluate the fidelity of the VTs' behavior, responses, and predictive capabilities in the context of the IoT-healthcare enabled by GAI. One of potential solution is the integration of GAI evaluation metrics, such as inception score and Frechet inception distance, and healthcare knowledge. Based on these, the continual evaluation and refinement of the GAI-driven HDT can ultimately enhance the performance in IoT-healthcare.

C. Accelerating Real-time Responsiveness of GAI-driven HDT

The slow generation process of GAI models poses significant challenges for developing GAI-driven HDT in IoT-healthcare. This delay hampers real-time applications, such as immersive interactions with VTs, and limits the responsiveness necessary for instant healthcare interventions, such as timely diagnostics or immediate treatment recommendations in critical conditions. Thus, accelerating real-time responsiveness of GAI-driven HDT is imperative to ensure timely and effective healthcare outcomes. One potential solution is refining and optimizing GAI algorithms to expedite the generation process. Advanced parallel computing techniques, such as leveraging high-performance computing or distributed computing frameworks, can significantly accelerate model training and generation, mitigating the latency and enhancing agility in responding to critical healthcare needs.

VI. Conclusions

In this survey, we have shed light on implementing GAI-driven HDT in IoT-healthcare. Specifically, we have introduced IoT-healthcare, and envisioned the potential of utilizing GAI-driven HDT. Then, we have reviewed the fundamentals of HDT and GAI and illustrated the framework of GAI-driven HDT. Furthermore, we have delved into the implementation of GAI-driven HDT, including GAI enabled data acquisition, communication, data management, digital modeling, and data analysis. Based on this, we presented the IoT-healthcare applications of GAI-driven HDT, including personalized health monitoring and diagnosis, personalized prescription, and personalized rehabilitation. Finally, we have outlined some future research directions.

References

[1] S. Selvaraj and S. Sundaravaradhan, "Challenges and opportunities in IoT healthcare systems: a systematic review," *SN Appl. Sci.*, vol. 2, no. 1, p. 139, 2020.
[2] M. H. Kashani, M. Madanipour, M. Nikravan *et al.*, "A systematic review of IoT in healthcare: Applications, techniques, and trends," *J. Netw. Comput. Appl.*, vol. 192, p. 103164, 2021.
[3] G. H. Hub, "Global heart hub," https://globalhearthub.org, 2023.
[4] H. Habibzadeh, K. Dinesh, O. R. Shishvan *et al.*, "A survey of healthcare internet of things (HIoT): A clinical perspective," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 53–71, 2020.
[5] Y. Yang, H. Wang, R. Jiang *et al.*, "A review of IoT-enabled mobile healthcare: technologies, challenges, and future trends," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9478–9502, 2022.

[6] the Computational Science Lab (Alfons Hoekstra) of the Informatics Institute, "Ecosystem for digital twins in healthcare," https://www.edith-csa.eu/, 2023.

[7] B. Wang, H. Zhou, X. Li *et al.*, "Human digital twin in the context of Industry 5.0," *Robotics and Computer-Integrated Manufacturing*, vol. 85, p. 102626, 2024.

[8] J. Chen, C. Yi, S. D. Okegbile *et al.*, "Networking architecture and key supporting technologies for human digital twin in personalized healthcare: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, pp. 1–1, 2023.

[9] S. D. Okegbile, J. Cai, C. Yi *et al.*, "Human digital twin for personalized healthcare: Vision, architecture and future directions," *IEEE Netw.*, 2022.

[10] A. E. Coşgun, "Digital twin and human digital twin for practical implementation in industry 5.0," in *Global Perspectives on Robotics and Autonomous Systems: Development and Applications*. IGI Global, 2023, pp. 168–183.

[11] B. Björnsson, C. Borrebaeck, N. Elander *et al.*, "Digital twins to personalize medicine," *Genome Med.*, vol. 12, pp. 1–4, 2020.

[12] Google, "Med-palm," https://sites.research.google/med-palm/, 2023.

[13] J. Chen, C. Yi, H. Du *et al.*, "A revolution of personalized healthcare: Enabling human digital twin with mobile AIGC," *arXiv preprint arXiv:2307.12115*, 2023.

[14] H. Du, R. Zhang, Y. Liu, J. Wang, Y. Lin, Z. Li, D. Niyato, J. Kang, Z. Xiong, S. Cui *et al.*, "Beyond deep reinforcement learning: A tutorial on generative diffusion models in network optimization," *arXiv preprint arXiv:2308.05384*, 2023.

[15] S. Bond-Taylor, A. Leach, Y. Long *et al.*, "Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7327–7347, 2022.

[16] Y. Lin, L. Chen, A. Ali *et al.*, "Human digital twin: A survey," *arXiv preprint arXiv:2212.05937*, 2022.

[17] M. Xu, H. Du, D. Niyato *et al.*, "Unleashing the power of edge-cloud generative AI in mobile networks: A survey of aigc services," *arXiv preprint arXiv:2303.16129*, 2023.

[18] B. Wang, H. Zhou, X. Li *et al.*, "Human digital twin in the context of industry 5.0," *Robot. Comput.-Integr. Manuf.*, vol. 85, p. 102626, 2024.

[19] J. Chen, C. Yi, S. D. Okegbile *et al.*, "Networking architecture and key supporting technologies for human digital twin in personalized healthcare: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, 2023.

[20] P. Pataranutaporn, V. Danry, J. Leong *et al.*, "AI-generated characters for supporting personalized learning and well-being," *Nat. Mach. Intell.*, vol. 3, no. 12, pp. 1013–1022, 2021.

[21] H. Pascual, X. M. Bruin, A. Alonso *et al.*, "A systematic review on human modeling: Digging into human digital twin implementations," *arXiv preprint arXiv:2302.03593*, 2023.

[22] T. Sun, X. He, and Z. Li, "Digital twin in healthcare: Recent updates and challenges," *Digit. Health*, vol. 9, p. 20552076221149651, 2023.

[23] M. AlAmir and M. AlGhamdi, "The role of generative adversarial network in medical image analysis: An in-depth survey," *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–36, 2022.

[24] Y. Shokrollahi, S. Yarmohammadtoosky, M. M. Nikahd *et al.*, "A comprehensive review of generative ai in healthcare," *arXiv preprint arXiv:2310.00795*, 2023.

[25] E. Research, "Digital human avatar market, by product type (interactive digital human avatar and non-interactive digital human avatar), by industry verticals (gaming, retail, it & telecommunications, education, and others), and by region forecast to 2032," *Emergen Research*, 2023. [Online]. Available: https://www.emergenresearch.com/industry-report/digital-human-avatar-market

[26] H. Lonsdale, G. M. Gray, L. M. Ahumada *et al.*, "The perioperative human digital twin," *Anesth. Analg.*, vol. 134, no. 4, pp. 885–892, 2022.

[27] F. Tao, H. Zhang, A. Liu *et al.*, "Digital twin in industry: State-of-the-art," *IEEE Trans. Industr. Inform.*, vol. 15, no. 4, pp. 2405–2415, 2019.

[28] S. Jang, J. Jeong, J. Lee *et al.*, "Digital twin for intelligent network: Data lifecycle, digital replication, and ai-based optimizations," *IEEE Commun.*, pp. 1–7, 2023.

[29] Z. Hu, S. Lou, Y. Xing *et al.*, "Review and perspectives on driver digital twin and its enabling technologies for intelligent vehicles," *IEEE Trans. Intell. Veh.*, vol. 7, no. 3, pp. 417–440, 2022.

[30] A. Bilberg and A. A. Malik, "Digital twin driven human–robot collaborative assembly," *CIRP annals*, vol. 68, no. 1, pp. 499–502, 2019.

[31] B. M. Rosa and G. Z. Yang, "A flexible wearable device for measurement of cardiac, electrodermal, and motion parameters in mental healthcare applications," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2276–2285, 2019.

[32] R. Ferdousi, F. Laamarti, M. A. Hossain *et al.*, "Digital twins for well-being: an overview," *Digital Twin*, vol. 1, p. 7, 2022.

[33] Y. Lin, L. Chen, A. Ali *et al.*, "Human digital twin: A survey," *arXiv preprint arXiv:2212.05937*, 2022.

[34] H. Y. Zhu, N. Q. Hieu, D. T. Hoang *et al.*, "A human-centric metaverse enabled by brain-computer interface: A survey," *arXiv preprint arXiv:2309.01848*, 2023.

[35] J. Wu, W. Gan, Z. Chen *et al.*, "AI-generated content (AIGC): A survey," *arXiv preprint arXiv:2304.06632*, 2023.

[36] A. Creswell, T. White, V. Dumoulin *et al.*, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag*, vol. 35, no. 1, pp. 53–65, 2018.

[37] T. Golany, K. Radinsky, and D. Freedman, "SimGANs: Simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification," in *Proc. ICML*. PMLR, 2020, pp. 3597–3606.

[38] F. Lau, T. Hendriks, J. Lieman-Sifry *et al.*, "ScarGAN: chained generative adversarial networks to simulate pathological tissue on cardiovascular MR scans," in *Proc. DLMIA*. Springer, 2018, pp. 343–350.

[39] D. P. Kingma, M. Welling *et al.*, "An introduction to variational autoencoders," *Foundations and Trends in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.

[40] A. Dittadi, S. Dziadzio, D. Cosker, and ohters, "Full-body motion from a single head-mounted device: Generating SMPL poses from partial observations," in *Proc. IEEE/CVF ICCV*, 2021, pp. 11 687–11 697.

[41] A. Allen, A. Siefkas, E. Pellegrini *et al.*, "A digital twins machine learning model for forecasting disease progression in stroke patients," *Appl. Sci.*, vol. 11, no. 12, p. 5576, 2021.

[42] A. Vaswani, N. Shazeer, N. Parmar *et al.*, "Attention is all you need," *Proc. NeurIPS*, vol. 30, 2017.

[43] Microsoft and Epic, "Alphasense," https://research.alpha-sense.com, 2023.

[44] H. Cui, C. Wang, H. Maan *et al.*, "scGPT: Towards building a foundation model for single-cell multi-omics using generative AI," *bioRxiv*, pp. 2023–04, 2023.

[45] H. Du, Z. Li, D. Niyato, J. Kang, Z. Xiong, H. Huang, and S. Mao, "Diffusion-based reinforcement learning for edge-enabled AI-generated content services," *arXiv preprint arXiv:2303.13052*, 2023.

[46] G. Tosato, C. M. Dalbagno, and F. Fumagalli, "EEG synthetic data generation using probabilistic diffusion models," *arXiv preprint arXiv:2303.06068*, 2023.

[47] K. Gong, K. Johnson, G. El Fakhri *et al.*, "PET image denoising based on denoising diffusion probabilistic model," *Eur. J. Nucl. Med. Mol. Imaging*, pp. 1–11, 2023.

[48] X. Meng, Y. Gu, Y. Pan *et al.*, "A novel unified conditional score-based generative framework for multi-modal medical image completion," *arXiv preprint arXiv:2207.03430*, 2022.

[49] M. Hajij, G. Zamzmi, R. Paul *et al.*, "Normalizing flow for synthetic medical images generation," in *Proc. HI-POCT*, 2022, pp. 46–49.

[50] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, J. Zhang, and X. Shen, "Semantic communications for wireless sensing: RIS-aided encoding and self-supervised decoding," *IEEE J. Sel. Areas Commun.*, to appear, 2023.

[51] H. Du, J. Wang, D. Niyato *et al.*, "AI-generated incentive mechanism and full-duplex semantic communications for information sharing," *IEEE J. Sel. Areas in Commun.*, vol. 41, no. 9, pp. 2981–2997, 2023.

[52] C. Liang, H. Du, Y. Sun *et al.*, "Generative AI-driven semantic communication networks: Architecture, technologies and applications," *arXiv preprint arXiv:2401.00124*, 2023.

[53] H. Du, G. Liu, D. Niyato *et al.*, "Generative AI-aided joint training-free secure semantic communications via multi-modal prompts," *arXiv preprint arXiv:2309.02616*, 2023.

[54] X. Wei, D. Wu, L. Zhou *et al.*, "Cross-modal communication technology: A survey," *Fundam. Res.*, 2023.

[55] J. Corral-Acero, F. Margara, M. Marciniak *et al.*, "The 'digital twin' to enable the vision of precision cardiology," *Eur. Heart J.*, vol. 41, no. 48, pp. 4556–4564, 2020.

[56] D. Hazra and Y.-C. Byun, "SynSigGAN: Generative adversarial networks for synthetic biomedical signal generation," *Biology*, vol. 9, no. 12, p. 441, 2020.

[57] M. A. Pimentel, A. E. Johnson, P. H. Charlton *et al.*, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1914–1923, 2016.

[58] J. M. L. Alcaraz and N. Strodthoff, "Diffusion-based conditional ECG generation with structured state space models," *Comput. Biol. Med.*, p. 107115, 2023.

[59] ——, "Diffusion-based time series imputation and forecasting with structured state space models," *Trans. Mach. Learn. Res.*, 2022.

[60] P. Wagner, N. Strodthoff, R.-D. Bousseljot *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Sci. data*, vol. 7, no. 1, p. 154, 2020.

[61] P. A. Moghadam, S. Van Dalen, K. C. Martin *et al.*, "A morphology focused diffusion probabilistic model for synthesis of histopathology images," in *Proc. IEEE/CVF CVPR*, 2023, pp. 2000–2009.

[62] A. B. Levine, J. Peng, D. Farnell *et al.*, "Synthesis of diagnostic quality cancer pathology images by generative adversarial networks," *J. pathol.*, vol. 252, no. 2, pp. 178–188, 2020.

[63] J. S. Yoon, C. Zhang, H.-I. Suk *et al.*, "SADM: Sequence-aware diffusion model for longitudinal medical image generation," in *Proc. IPMI*. Springer, 2023, pp. 388–400.

[64] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," *arXiv preprint arXiv:1605.08803*, 2016.

[65] R. Summers, "Nih chest x-ray dataset of 14 common thorax disease categories," *NIH Clinical Center: Bethesda, MD, USA*, 2019.

[66] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. data*, vol. 5, no. 1, pp. 1–9, 2018.

[67] Y. Du, R. Kips, A. Pumarola *et al.*, "Avatars grow legs: Generating smooth human motion from sparse tracking inputs with diffusion model," in *Proc. IEEE/CVF CVPR*, 2023, pp. 481–490.

[68] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. S. Shen, and C. Miao, "Semantic communications for future Internet: Fundamentals, applications, and challenges," *IEEE Communications Surveys & Tutorials*, 2023.

[69] H. Du, R. Zhang, D. Niyato, J. Kang, Z. Xiong, S. Cui, X. Shen, and D. I. Kim, "User-centric interactive AI for distributed diffusion model-based AI-generated content," *arXiv preprint arXiv:2311.11094*, 2023.

[70] A. D. Raha, M. S. Munir, A. Adhikary *et al.*, "Generative AI-driven semantic communication framework for NextG wireless network," *arXiv preprint arXiv:2310.09021*, 2023.

[71] C. Zhang, D. Han, Y. Qiao *et al.*, "Faster segment anything: Towards lightweight SAM for mobile applications," *arXiv preprint arXiv:2306.14289*, 2023.

[72] E. Grassucci, S. Barbarossa, and D. Comminiello, "Generative semantic communication: Diffusion models beyond bit recovery," *arXiv preprint arXiv:2306.04321*, 2023.

[73] H. Liu, D. Guo, X. Zhang *et al.*, "Toward image-to-tactile cross-modal perception for visually impaired people," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 521–529, 2021.

[74] T. Kim, M. Cha, H. Kim *et al.*, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. ICML*. PMLR, 2017, pp. 1857–1865.

[75] Y. Fang, X. Zhang, W. Xu *et al.*, "Bidirectional visual-tactile cross-modal generation using latent feature space flow model," *Available at SSRN 4593113*.

[76] A. Li, Y. Chen, S. Ni *et al.*, "Haptic signal reconstruction in ehealth internet of things," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17 047–17 057, 2021.

[77] B. R. Barricelli, E. Casiraghi, J. Gliozzo *et al.*, "Human digital twin for fitness management," *IEEE Access*, vol. 8, pp. 26 637–26 664, 2020.

[78] S. Phung, A. Kumar, and J. Kim, "A deep learning technique for imputing missing healthcare data," in *Proc. IEEE EMBC*. IEEE, 2019, pp. 6513–6516.

[79] W. Dong, D. Y. T. Fong, J.-s. Yoon *et al.*, "Generative adversarial networks for imputing missing data for big data clinical research," *BMC medical res. methodol.*, vol. 21, pp. 1–10, 2021.

[80] P. Hayati Rezvan, K. J. Lee, and J. A. Simpson, "The rise of multiple imputation: a review of the reporting and implementation of the method in medical research," *BMC med. res. methodol.*, vol. 15, pp. 1–14, 2015.

[81] D. J. Stekhoven and P. Bühlmann, "MissForest—non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2012.

[82] Q. Lyu and G. Wang, "Conversion between CT and mri images using diffusion and score-matching models," *arXiv preprint arXiv:2209.12104*, 2022.

[83] C. Saharia, J. Ho, W. Chan *et al.*, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, 2022.

[84] Y. Song, J. Sohl-Dickstein, D. P. Kingma *et al.*, "Score-based generative modeling through stochastic differential equations," *Proc. ICLR*, 2021.

[85] ——, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.

[86] T. Nyholm, S. Svensson, S. Andersson *et al.*, "MR and CT data with multiobserver delineations of organs in the pelvic area—part of the gold atlas project," *Med. phys.*, vol. 45, no. 3, pp. 1295–1300, 2018.

[87] I. Gulrajani, F. Ahmed, M. Arjovsky *et al.*, "Improved training of wasserstein GANs," *Proc. NeurIPS*, vol. 30, 2017.

[88] P. Huang, D. Li, Z. Jiao *et al.*, "CoCa-GAN: common-feature-learning-based context-aware generative adversarial network for glioma grading," in *Proc. MICCAI*. Springer, 2019, pp. 155–163.

[89] T. Zhou, H. Fu, G. Chen *et al.*, "Hi-Net: hybrid-fusion network for multi-modal mr image synthesis," *IEEE trans. med. imaging*, vol. 39, no. 9, pp. 2772–2781, 2020.

[90] A. Sharma and G. Hamarneh, "Missing MRI pulse sequence synthesis using multi-modal generative adversarial network," *IEEE trans. med. imaging*, vol. 39, no. 4, pp. 1170–1183, 2019.

[91] X. Liu, F. Xing, G. El Fakhri *et al.*, "A unified conditional disentanglement framework for multimodal brain mr image translation," in *Proc. ISBI*. IEEE, 2021, pp. 10–14.

[92] A. Chartsias, T. Joyce, M. V. Giuffrida *et al.*, "Multimodal MR synthesis via modality-invariant latent representation," *IEEE trans. med. imaging*, vol. 37, no. 3, pp. 803–814, 2017.

[93] S. Kazeminia, C. Baur, A. Kuijper *et al.*, "GANs for medical image analysis," *Artif. Intell. Med.*, vol. 109, p. 101938, 2020.

[94] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu *et al.*, "Noise reduction in ECG signals using fully convolutional denoising autoencoders," *IEEE Access*, vol. 7, pp. 60 806–60 813, 2019.

[95] S. Nasrin, M. Z. Alom, R. Burada *et al.*, "Medical image denoising with recurrent residual U-Net (R2U-Net) base auto-encoder," in *Proc. IEEE NAECON*. IEEE, 2019, pp. 345–350.

[96] P. Isola, J.-Y. Zhu, T. Zhou *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017, pp. 1125–1134.

[97] T. Xiang, M. Yurt, A. B. Syed *et al.*, "DDM^2 : Self-supervised diffusion mri denoising with generative diffusion models," *Proc. ICLR*, 2023.

[98] A. Majeed, "Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 31, no. 4, pp. 426–435, 2019.

[99] A. Aminifar, F. Rabbi, V. K. I. Pun *et al.*, "Diversity-aware anonymization for structured health data," in *Proc. IEEE EMBC*. IEEE, 2021, pp. 2148–2154.

[100] J. Yoon, L. N. Drumright, and M. van der Schaar, "Anonymization through data synthesis using generative adversarial networks (ADS-GAN)," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 8, pp. 2378–2388, 2020.

[101] E. Choi, S. Biswal, B. Malin *et al.*, "Generating multi-label discrete patient records using generative adversarial networks," in *Proc. ML4H*. PMLR, 2017, pp. 286–305.

[102] L. Xie, K. Lin, S. Wang, F. Wang, and J. Zhou, "Differentially private generative adversarial network," *arXiv preprint arXiv:1802.06739*, 2018.

[103] J. Jordon, J. Yoon, and M. Van Der Schaar, "PATE-GAN: Generating synthetic data with differential privacy guarantees," in *Proc. ICLR*, 2018.

[104] J. Yoon, M. Mizrahi, N. F. Ghalaty *et al.*, "EHR-Safe: generating high-fidelity and privacy-preserving synthetic electronic health records," *NPJ Digit. Med.*, vol. 6, no. 1, p. 141, 2023.

[105] M. Bordukova, N. Makarov, R. Rodriguez-Esteban *et al.*, "Generative artificial intelligence empowers digital twins in drug discovery and clinical trials," *Expert Opin. Drug Discov.*, pp. 1–10, 2023.

[106] B. S. Center, "Alya Red: A computational heart," https://www.bsc.es/news/bsc-news/alya-red-computational-heart, 2023.

[107] M. Lotfollahi, F. A. Wolf, and F. J. Theis, "scGen predicts single-cell perturbation responses," *Nat. Methods*, vol. 16, no. 8, pp. 715–721, 2019.

[108] M. Lotfollahi, A. Klimovskaia Susmelj, C. De Donno *et al.*, "Predicting cellular responses to complex perturbations in high-throughput screens," *Mol. Syst. Biol.*, p. e11517, 2023.

[109] R. M. Donovan-Maiye, J. M. Brown, C. K. Chan *et al.*, "A deep generative model of 3D single-cell organization," *PLOS Comput. Biol.*, vol. 18, no. 1, p. e1009155, 2022.

[110] L. Jose, S. Liu, C. Russo *et al.*, "Generative adversarial networks in digital pathology and histopathological image processing: A review," *J. Pathol. Inform.*, vol. 12, no. 1, p. 43, 2021.
[111] A. C. Quiros, R. Murray-Smith, and K. Yuan, "PathologyGAN: Learning deep representations of cancer tissue," in *Proc. PMLR*, vol. 121. PMLR, 06–08 Jul 2020, pp. 669–695.
[112] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.
[113] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF CVPR*, 2019, pp. 4401–4410.
[114] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard gan," *arXiv preprint arXiv:1807.00734*, 2018.
[115] H. Ahmadian, P. Mageswaran, B. A. Walter *et al.*, "Toward an artificial intelligence-assisted framework for reconstructing the digital twin of vertebra and predicting its fracture response," *Int. J. Numer. Method Biomed. Eng.*, vol. 38, no. 6, p. e3601, 2022.
[116] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
[117] X. Xing, J. Del Ser, Y. Wu *et al.*, "HDL: Hybrid deep learning for the synthesis of myocardial velocity maps in digital twins for cardiac analysis," *IEEE J. Biomed. Health. Inform.*, 2022.
[118] R. Martinez-Velazquez, R. Gamez, and A. El Saddik, "Cardio twin: A digital twin of the human heart running on the edge," in *Proc. IEEE MeMeA*. IEEE, 2019, pp. 1–6.
[119] X. Li, Z. Gong, H. Yin *et al.*, "A 3D deep supervised densely network for small organs of human temporal bone segmentation in CT images," *Neural Netw.*, vol. 124, pp. 75–85, 2020.
[120] N. J. Dhinagar, S. I. Thomopoulos, E. Laltoo *et al.*, "Efficiently training vision transformers on structural MRI scans for alzheimer's disease detection," *arXiv preprint arXiv:2303.08216*, 2023.
[121] Y. Dong, M. Zhang, L. Qiu *et al.*, "An arrhythmia classification model based on vision transformer with deformable attention," *Micromachines*, vol. 14, no. 6, p. 1155, 2023.
[122] C. Che, P. Zhang, M. Zhu *et al.*, "Constrained transformer network for ECG signal processing and arrhythmia classification," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 1–13, 2021.
[123] A. Rahman, J. M. J. Valanarasu, I. Hacihaliloglu *et al.*, "Ambiguous medical image segmentation using diffusion models," in *Proc. IEEE/CVF CVPR*, 2023, pp. 11 536–11 546.
[124] B. Kim, Y. Oh, and J. C. Ye, "Diffusion adversarial representation learning for self-supervised vessel segmentation," *Proc. ICLR*, 2023.
[125] X. Guo, J. W. Gichoya, S. Purkayastha *et al.*, "CVAD: An anomaly detector for medical images based on cascade VAE," in *Proc. MILLanD*. Springer, 2022, pp. 187–196.
[126] T. Nakao, S. Hanaoka, Y. Nomura *et al.*, "Unsupervised deep anomaly detection in chest radiographs," *J. Digit. Imaging*, vol. 34, pp. 418–427, 2021.
[127] J. Wolleb, F. Bieder, R. Sandkühler *et al.*, "Diffusion models for medical anomaly detection," in *Proc. MICCAI*. Springer, 2022, pp. 35–45.
[128] J. Irvin, P. Rajpurkar, M. Ko *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI*, vol. 33, no. 01, 2019, pp. 590–597.
[129] A. Johnson, L. Bulgarelli, T. Pollard *et al.*, "MIMIC-IV," *PhysioNet*, 2021.
[130] W. Hu and S. Y. Wang, "Predicting glaucoma progression requiring surgery using clinical free-text notes and transfer learning with transformers," *Transl. Vis. Sci. Techn.*, vol. 11, no. 3, pp. 37–37, 2022.
[131] J. Devlin, M.-W. Chang, K. Lee *et al.*, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
[132] J. Lee, W. Yoon, S. Kim *et al.*, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.
[133] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
[134] V. Sanh, L. Debut, J. Chaumond *et al.*, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.
[135] Z. Wang, S. Stavrakis, and B. Yao, "Hierarchical deep learning with generative adversarial network for automatic cardiac diagnosis from ECG signals," *Comput. Biol. Med.*, vol. 155, p. 106641, 2023.
[136] A. Staffini, T. Svensson, U.-i. Chung *et al.*, "A disentangled VAE-BiLSTM model for heart rate anomaly detection," *Bioengineering*, vol. 10, no. 6, p. 683, 2023.
[137] S. Park, K. H. Lee, B. Ko *et al.*, "Unsupervised anomaly detection with generative adversarial networks in mammography," *Sci. Rep.*, vol. 13, no. 1, p. 2925, 2023.
[138] C. Han, L. Rundo, K. Murao *et al.*, "GAN-based multiple adjacent brain MRI slice reconstruction for unsupervised alzheimer's disease diagnosis," in *Proc. CIBB*. Springer, 2019, pp. 44–54.
[139] T. N. Jarada, J. G. Rokne, and R. Alhajj, "SNF–CVAE: computational method to predict drug–disease interactions using similarity network fusion and collective variational autoencoder," *Knowl. Based Syst.*, vol. 212, p. 106585, 2021.
[140] F. Xu, S. Liu, Y. Xiang *et al.*, "Prediction of the short-term therapeutic effect of anti-VEGF therapy for diabetic macular edema using a generative adversarial network with OCT images," *J. Clin. Med.*, vol. 11, no. 10, p. 2878, 2022.
[141] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu *et al.*, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proc. IEEE CVPR*, 2018, pp. 8798–8807.
[142] C. Mennella, U. Maniscalco, G. De Pietro *et al.*, "Generating a novel synthetic dataset for rehabilitation exercises using pose-guided conditioned diffusion models: A quantitative and qualitative evaluation," *Comput. Biol. Med.*, vol. 167, p. 107665, 2023.
[143] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proc. IEEE/CVF ICCV*, 2023, pp. 3836–3847.
[144] I. Boukhennoufa, D. Jarchi, X. Zhai *et al.*, "A novel model to generate heterogeneous and realistic time-series data for post-stroke rehabilitation assessment," *IEEE Trans. Neural Syst. Rehabilitation Eng.*, vol. 31, pp. 2676–2687, 2023.
[145] X. Du, Y. Liu, Z. Lu *et al.*, "A low-latency communication design for brain simulations," *IEEE Netw.*, vol. 36, no. 2, pp. 8–15, 2022.
[146] W. Lu *et al.*, "The human digital twin brain in the resting state and in action," *arXiv preprint arXiv:2211.15963*, 2022.