future internet

MDPI

Review

Artificial Intelligence to Reshape the Healthcare Ecosystem

Gianluca Reali [1,2,*,†] and Mauro Femminella [1,2,†]

1 Department of Engineering, University of Perugia, Via G. Duranti 93, 06125 Perugia, Italy; mauro.femminella@unipg.it
2 Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), 43124 Parma, Italy
* Correspondence: gianluca.reali@unipg.it
† The authors contributed equally to this work.

Abstract: This paper intends to provide the reader with an overview of the main processes that are introducing artificial intelligence (AI) into healthcare services. The first part is organized according to an evolutionary perspective. We first describe the role that digital technologies have had in shaping the current healthcare methodologies and the relevant foundations for new evolutionary scenarios. Subsequently, the various evolutionary paths are illustrated with reference to AI techniques and their research activities, specifying their degree of readiness for actual clinical use. The organization of this paper is based on the interplay three pillars, namely, algorithms, enabling technologies and regulations, and healthcare methodologies. Through this organization we introduce the reader to the main evolutionary aspects of the healthcare ecosystem, to associate clinical needs with appropriate methodologies. We also explore the different aspects related to the Internet of the future that are not typically presented in papers that focus on AI, but that are equally crucial to determine the success of current research and development activities in healthcare.

Keywords: artificial intelligence; ML algorithms; healthcare methodologies; information technologies

check for updates

Citation: Reali, G.; Femminella, M. Artificial Intelligence to Reshape the Healthcare Ecosystem. *Future Internet* **2024**, *16*, 343. https://doi.org/10.3390/fi16090343

Academic Editor: Petros Patias

Received: 26 August 2024
Revised: 9 September 2024
Accepted: 19 September 2024
Published: 20 September 2024

1. Introduction

Artificial intelligence (AI) refers to many computationally intensive technologies that are profoundly changing many aspects of daily life. Market research on AI-driven healthcare growth is in agreement that the forecast is double-digit and increasing over the next decade [1–3].

The goal of this paper is to provide readers with a comprehensive overview of how AI is pervading healthcare services through the use of the available techniques and technologies. The paper is intended for technical personnel who intend to deepen their skills in AI techniques in the various application directions in the healthcare sector. However, some insights into the possible use of these techniques in an operational setting may also be appreciated by healthcare professionals.

The potential of this rapidly evolving scenario is changing the expectations of both patients and healthcare professionals. The former are inclined to use increasingly advanced devices, such as smartphones, wearables, cameras, home appliances, and even drones. The latter are increasingly eager to use advanced tools both to speed up the time for providing services and to improve medical performance in terms of reliability and efficiency.

While some features may suggest that this process is a natural evolution of a progressive penetration of digital technologies into medical practices, some disruptive features, both in medical services and in the related underlying technologies, are absolute novelties capable of revolutionizing healthcare.

In fact, since the introduction of the first technologies in clinical instrumentation, such as ultrasound and electrocardiographs, healthcare professionals have progressively used new diagnostic and interventional techniques to improve their services. The digitization of information has produced a significant leap forward, so much so that today advanced services such as telemedicine [4,5], remote surgery [6], and genomics [7,8] characterize the current state of the art. Nowadays, the use of electronic health records (EHRs), which allow managing patients' health data, is part of everyday practice. So-called health information exchange (HIE) systems [9] are used to exchange health information between different organizations.

The importance of using these technologies is demonstrated by the creation of new working groups and standards dedicated to health data management, such as HL7 (Health Level Seven [10]) for the exchange, integration, sharing, and retrieval of electronic health information and FHIR (Fast Healthcare Interoperability Resources), which consists of a set of standards developed by HL7 to facilitate the exchange of electronic health data.

Again, during the COVID-19 pandemic, the need for immunization information systems (IISs), which allow for updating and consulting vaccine data in a timely and accurate manner, and vaccine surveillance systems, used to monitor post-vaccination effects, has clearly emerged.

Clearly, as the level of expectations increases, so does the level of technological challenge. Luckily, new medical technologies can also benefit from research findings produced in other contexts. For example, the volume of data generated by some medical applications can be managed due to efficient and scalable technologies for storage, management, and processing [7]. As a further example, other applications that require remote patient monitoring can benefit from the increasing development of the so-called Internet of Things (IoT), with the related standardization activities. Again, Release 13 of the 3rd Generation Partnership Project (3GPP) specifications includes *enhanced machine-type communication* (eMTC), which extends IoT coverage by the use of low-complexity devices and Long-Term Evolution (LTE) base stations. This process has in turn generated significant results in the development of real-time monitoring techniques of vital signs collected by biosensors, the development of new sensors through the introduction of stimuli-sensitive materials, and techniques for assisted drug delivery [11].

The introduction of AI techniques has produced an acceleration in the innovation process. This can be illustrated in two directions. One is the short-term improvement in current practices, obtained by updating existing applications. The other direction leads to the introduction of new research clinical practices that also require the cultural updating of medical personnel [12,13]. Moreover, some disrupting technological novelties, such as digital twins [14], virtual reality or ultrahigh-quality imaging, which will be enabled by forthcoming 6G communications systems [15], have the potential to deeply reshape the services of the future.

The AI techniques that have played a more or less significant role in healthcare research are classified in Figure 1, according to a structure reflecting the presentation of this paper. The figure is divided into four large families of algorithms, even if it frequently happens that they are combined. Specifically, the classification includes the class of algorithms that use artificial neural networks (ANNs) [16–19], those of the evolutionary type [20,21], those alternatives to ANN that are able to provide data classification and prediction solutions [22,23], and those related to Bayesian networks [24,25]. This paper provides a comprehensive overview of AI techniques that, according to the ongoing research and application developments, represent the bulk of efforts in designing future networked systems for healthcare.

In particular, we focus on the techniques based on the use of ANNs. The following areas are considered, with reference to medical applications in which they are able to generate the most appreciated results:

- Classical deep learning, including convolutional neural networks and U-Net architectures;
- Graph neural networks;
- Recurrent neural networks;
- Generative AI;
- Diffusion models;

- Reinforcement learning.

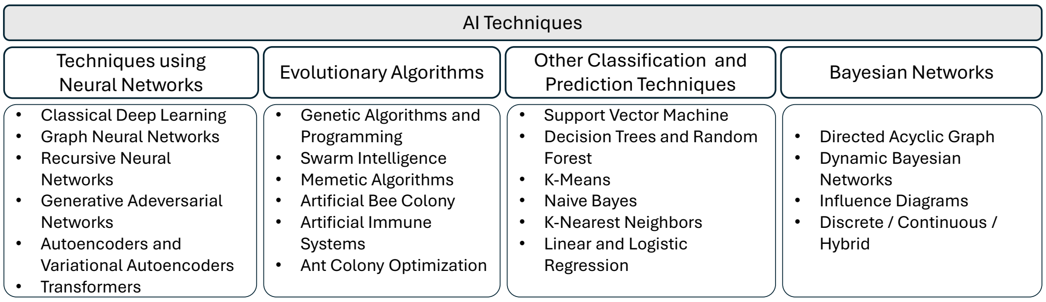

Figure 1. General classification of AI techniques involved in healthcare.

These AI techniques are first presented with a sufficient level of depth to understand their potential. Subsequently, their applicability in the various healthcare sectors is demonstrated by indicating specific applications that refer to basic healthcare activities, such as diagnosis, clinical management, and also administration.

Subsequently, technological areas are also identified that are enabling for the effective introduction of AI in healthcare. The evolutionary lines in these areas are also illustrated. In particular, the following areas are considered: *Human–computer interaction*, *explainability*, *wearable sensors*, *privacy and security*, *network and computing infrastructure*, *bias and equity*, and *regulation and governance*.

The overall architecture of the paper is shown in Figure 2.

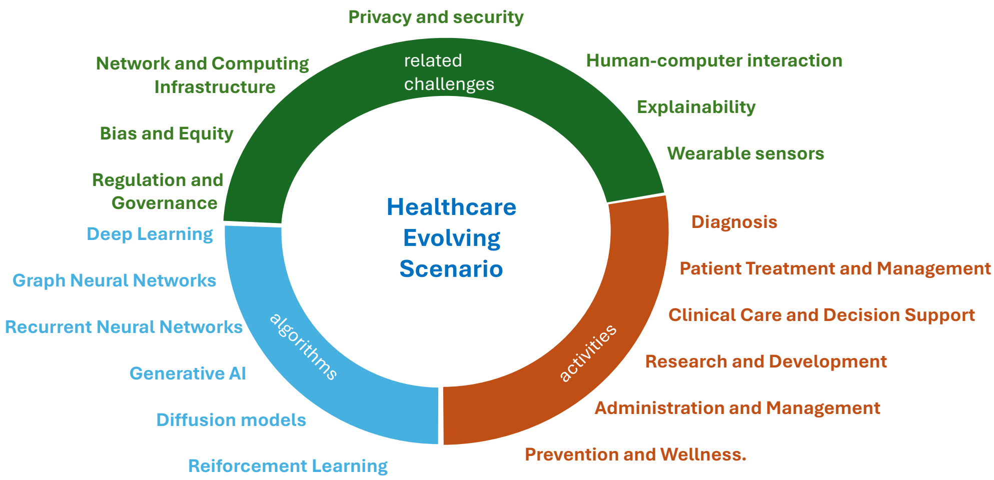

Figure 2. Graphical representation of the organization of the paper.

To the best of our knowledge, this is the first survey paper to follow this approach, systematically embracing all aspects, namely, algorithmic, medical, organizational, and technological, that are necessary for the success of the ongoing evolutionary processes in healthcare.

The paper is organized as follows. In Section 2, we provide some background on AI technologies for healthcare. In Section 3, we describe the baseline areas where the effectiveness of the presented AI technologies are demonstrated through a literature review. Section 4 presents the AI-related research challenges in future medical networked systems. Finally, Section 5 reports our conclusions.

2. AI Techniques for Healthcare

The current state of the use of machine learning (ML) algorithms in healthcare is already quite complex [22]. In general, both supervised and unsupervised learning algorithms are used. The algorithms in the first class use tagged training data to configure the parameters of the algorithms, while unsupervised algorithms discover detailed information from untagged input data. Supervised algorithms, such as *decision trees* [22], are typically used for classification and regression tasks, while unsupervised ML is suitable for discovering patterns in data, such as for clustering tasks.

In the following, we focus on the machine learning techniques that underpin much of the current state and progress of AI in healthcare. Some tools already incorporate AI techniques, such as QIIME2 [26] and Nextflow [27], that are widely used [28–32]. In the following, we focus on the AI techniques that are driving the major innovations in healthcare.

A complete mathematical treatment of each of these and all the variants of each algorithm is beyond the scope of this paper. However, in addition to presenting the rationale for each technique, we also present basic mathematical modeling and provide the interested reader with directions for further study in the references.

The techniques we discuss in this paper are those that refer to artificial neural networks (ANNs). These networks implement a computational mathematical model composed of a group of interconnections between nodes, which are vaguely inspired by the functioning of biological neurons. In order for a neural network to implement the desired mathematical function, it is necessary to train it to configure its parameters. These parameters make a neural network an adaptive system, that is, one that changes its non-linear structure (nodes and interconnections) of statistical data based on information that passes through it during the learning phase. A generic neural network includes three types of processing units, as shown in Figure 3.

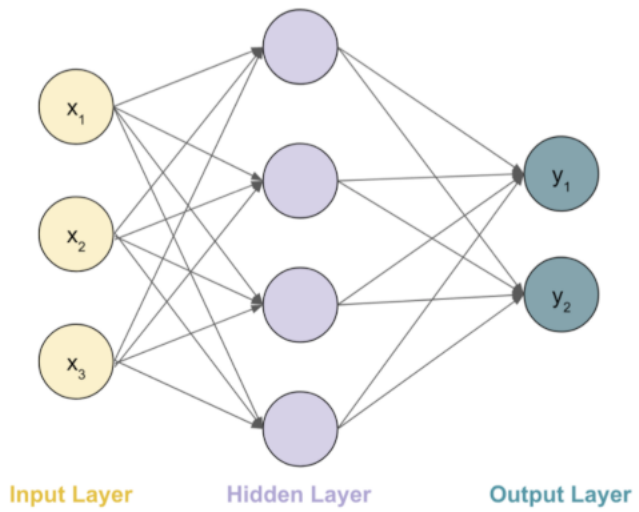

Figure 3. Basic structure of a multi-layer neural network.

- I—Input nodes: These receive and process external data by adapting them to the internal nodes, to which each node I is connected;
- H—Hidden (internal) nodes: Organized in multiple levels, each of these nodes processes the received signals and transmits the result to the subsequent nodes. They actually perform, independently of each other, the data processing process through the use of algorithms;
- O—Output nodes: Final layer of nodes that collect the processing results of the H-layer and adapt them to the next neural network block.

Each connection in a neural network is associated with a numerical trainable weight which controls the amount of information carried by the connection downstream. In addition, each node has a numerical bias associated with it. An activation function is used for taking in the input received by a node and processing it to obtain the desired output. Common non-linear activation functions used in ANNs include [33]:

- **ReLU (rectified linear unit)**:

$$\text{ReLU}(x) = \max(0, x) \tag{1}$$

  It is widely used due to both its simplicity and its effectiveness in mitigating the vanishing gradient problem.
- **Leaky ReLU**:

$$\text{Leaky ReLU}(x) = \max(\alpha x, x) \tag{2}$$

  where α is a small constant that allows a small, non-zero gradient when the unit is inactive. It aims at mitigating the so-called *dying ReLU* problem.
- **Sigmoid**:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{3}$$

  It is an S-shaped curve, that maps input to a range between 0 and 1. It is prone to vanishing gradients.
- **Tanh (hyperbolic tangent)**:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4}$$

  It is an S-shaped curve that maps input to a range between -1 and 1. It is also prone to vanishing gradients.
- **Swish**:

$$\text{Swish}(x) = x \cdot \sigma(\beta x) \tag{5}$$

  An activation function that was shown to outperform ReLU in particular situations, with β as a learnable parameter.

The interested reader may find a technical comparison of activation functions in [33].

Through the weights and activation functions used, a neural network implements a global non-linear mathematical function. The output of the network is therefore the value of the implemented function, regardless of the semantic nature of data. Although the general structure of neural networks, such as the one shown in Figure 3, is extremely flexible, the computational cost of training them when the number of nodes is very large and training data are extremely complex and voluminous is impractical. Therefore, when the data have a known structure, this can be used to adapt the neural network architecture to obtain network training times compatible with the needs of the application.

2.1. Deep Learning: Convolutional Neural Networks and U-Net Architectures

Convolutional neural networks (CNNs) are specialized neural networks that overcome the intrinsic limitations of a classical ANN for image processing. They analyze images through artificial neurons (i.e., nodes) organized in three dimensions, called channels: width, height, and depth [16]. In particular, they can detect and classify input images by extracting their features, such as edges and corners. For example, they have been proposed to detect objects such as circulating tumor cells or for the early diagnosis of different pathology and syndromes, such as tumors and strokes [17–19,34]. The CNN structure consists of multiple layers for feature detection, sketched in Figure 4:

- Convolutional layers: These use filters (or *kernels*) that slide over the input data and multiply by their values to capture local patterns. Each trained filter is expected to focus on a particular region of the input image file. In other words, its role is to offer subsequent layers a local feature, such an edge or a particular shape pattern, to either compose a global feature or an immediate detection.
- Non-linear activation functions in nodes: These functions are applied after each convolutional layer. Non-linearity allows the network to learn and implement complex functions, and is essential for processing images and implement tasks such as recognition and classification.

- Pooling layers: They play a crucial role by performing down-sampling operations along the spatial dimensions of the input images. The benefits of this operation include spatial dimensionality reduction, which reduces the number of parameters in the network and memory requirements to store them, invariance to small translations, rotations, and distortions in the input images, and reduction in the sensitivity of the implemented function to both noise and random variations in input data. Descriptions of some types of pooling functions follow:
  - **Max pooling**: Takes the maximum value within a defined window (e.g., 3×3):

$$y_{i,j} = \max_{m,n}(x_{i+m,j+n}) \tag{6}$$

    where x is the input feature map, y is the output pooled map, and m, n define the window size.
  - **Average pooling**: Computes the average value within a defined window.

$$y_{i,j} = \frac{1}{k \times k} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} x_{i+m,j+n} \tag{7}$$

    where k is the window size.
  - **Global pooling**: Applies pooling over the entire spatial dimensions of the feature map. This way, each feature map is reduced to a single value, often used before the following fully connected layers to flatten the feature maps.
- Fully connected layers (also known as dense layers): They take the aspect of the layers of the general ANN shown in Figure 3. They are used to learn complex, non-linear combinations of the features. For example, in a classification network, the fully connected layers map the features to class scores.
- Dropout layer: During training, dropout is often applied to fully connected layers to prevent overfitting. Dropout randomly sets a fraction of the neurons to zero at each training step, forcing the network to learn redundant representations and improving generalization.

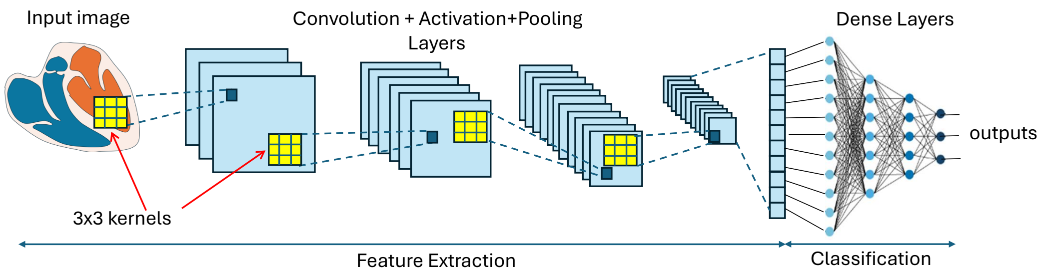

Figure 4. Example of a convolutional neural network for image classification.

For some medical applications, it is not enough to classify images and identify objects in them, it is also necessary to identify the boundaries of the detected objects. This is a process known as segmentation. For example, manually segmenting internal organs from X-ray images typically takes a very long time, and the use of dedicated neural networks is essential to improve the efficiency of medical practices that require it. In the pioneering paper [35], the authors present U-Net, a novel CNN architecture suitable for automatic segmentation of medical images. Since a CNN can learn the set of features of an input image and represent them as a numerical vector, this information is used for segmenting objects within the image through a reconstruction process. This task is quite difficult. Although it is relatively simple to represent an image as a vector, the reverse process is significantly more complex. The idea of a U-Net architecture is to make the functionality mapping learned during the representation of the input image into a vector, and use it to reconstruct the portion of interest of the output image. In this way, the structure of the input image is preserved and distortion introduced by reconstruction is reduced. The U-Net structure includes the typical elements of deep learning architectures, such as convolutional and pooling layers. Figure 5 shows the structure of a sample U-Net, with some parameter values that should be regarded as an example. The basic structure of the U-Net includes a contracting path, a bottom section, and an expanding path. The contracting path consists of some convolutional layers followed by max pooling layers. Each convolutional layer typically makes use of a 3 × 3 kernel, followed by a ReLU activation function. Max pooling layers are used to reduce the spatial dimensions of the feature maps. The bottom of the U-shape typically consists of two 3 × 3 convolutions followed by a ReLU activation function. The expanding path consists of up-sampling (or transposed convolution) layers that increase the spatial dimensions of the feature maps. Each up-sampling step is followed by a concatenation with the corresponding feature map from the contracting path, referred to as skip connections, providing the decoder with high-resolution features from the encoder. Following the concatenation, two 3 × 3 convolutions and a ReLU activation function refine the up-sampled feature maps. The number of feature channels halves with each up-sampling step.

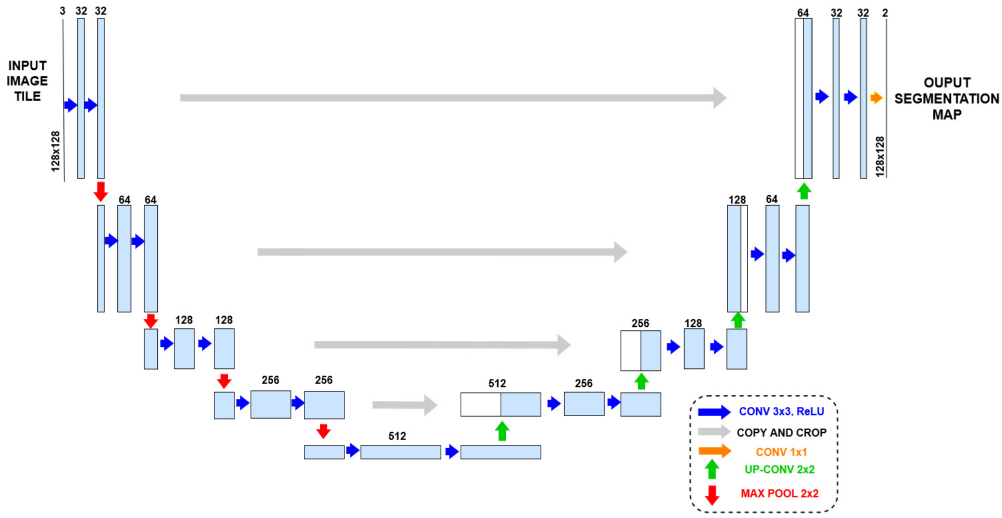

Figure 5. Example of U-Net architecture. The blue rectangles indicate multi-channel feature maps. The number of channels is shown above boxes. The white boxes indicate the copied feature maps for the reconstruction process.

The described basic architecture is the basis of many research proposals. In [36], a deep learning-based segmentation model is proposed. It performs binary segmentation through a P-Net [37] architecture, which draws inspiration from the well-known VGG-16 [38] CNN model. The proposal displays good performance, although it requires a long training phase.

Finally, it is worth mentioning the research activities that make use of 2D segmentation to obtain a three-dimensional representation of organs. In particular, 3D U-Net [39] is an architecture which extends the original U-Net structure [35], with appreciable results.

The reader interested in developing DL applications could make use of different libraries. Among these, we mention TensorFlow [40], PyTorch [41], and CUDA Toolkit [42].

2.2. Graph Neural Networks

In spite of the excellent results generated by CNNs in many fields, when it is necessary to use irregular data, without a predefined or easily identifiable structure, and with intricate logical relationships between different components, graph neural networks (GNNs) are extremely useful [43]. GNNs are a class of neural networks designed to handle graph-structured data. For this reason, their applicability to many healthcare problems is immediate. For example, typical limitations of traditional medical decision systems when heterogeneous medical data are used can be successfully tackled through the use of GNNs. They have paved the way to personalized medical decision algorithms based on high-precision representation models of patient health status, related to living habits, clinical information, and genetic analyses [44–46]. Similarly, brain activity analysis, complex neurodevelopmental disorders, drug interactions, and drug discovery can significantly benefit from GNN modeling [47,48].

Before delving into GNNs' applications, we introduce some basic concepts that illustrate how they work. GNNs extend the concepts of CNNs to graphs, which abstract the context information present in the graph elements for classification and prediction operations. A graph G is represented as $G = (V, E)$, where V is the set of nodes (also known as vertices) and E is the set of edges connecting the nodes. For example, nodes can be logically associated with proteins and vertices could relate to their affinity for establishing a bond. A graph can be illustrated by its adjacency matrix $\mathbf{A}$, which embeds the connections between nodes:

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{if there is an edge between node } i \text{ and node } j, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Each node $v_i \in V$ can have an associated feature vector $\mathbf{x}_i$. The feature matrix for all nodes is denoted as $\mathbf{X}$:

$$\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \ldots; \mathbf{x}_N] \in \mathbb{R}^{N \times F}, \tag{9}$$

where N is the number of nodes and F is the number of features per node. The logical relationship between nodes is of interest not only per se, e.g., for being associated with pathology or chemical interactions, but also because it represents the dynamism with which information is transferred between nodes. This process is captured by *graph convolutional networks* (GCNs) [49]. The GCN performs graph convolutions using the graph adjacency matrix and a node feature matrix. Each layer of the GCN updates node representations by aggregating features received from node neighbors, and then, applying a linear transformation followed by a non-linear activation function. The graph convolution generalizes the convolution operation from CNNs to graph structures. The basic idea of this operation is to aggregate information from node neighbors. A graph convolution operation is defined as

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right), \tag{10}$$

where $\mathbf{H}^{(l)}$ is the node feature matrix at layer l with $\mathbf{H}^{(0)} = \mathbf{X}$; $\tilde{\mathbf{A}}$ is the normalized adjacency matrix with added self-loops, typically defined as $\tilde{\mathbf{A}} = \mathbf{D}^{-1/2}(\mathbf{A} + \mathbf{I})\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is the degree matrix and $\mathbf{I}$ is the identity matrix; $\mathbf{W}^{(l)}$ is a layer-specific trainable weight matrix; and σ is an activation function, such as ReLU.

GNNs often use message passing mechanisms, where information is propagated between nodes. The general framework can be described as follows for a node v_i:

$$\mathbf{m}_i^{(l+1)} = \text{AGGREGATE}^{(l)}\left(\{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\}\right), \tag{11}$$

$$\mathbf{h}_i^{(l+1)} = \text{UPDATE}^{(l)}\left(\mathbf{h}_i^{(l)}, \mathbf{m}_i^{(l+1)}\right), \tag{12}$$

where $\mathbf{h}_i^{(l)}$ is the feature vector of node i at layer l, $\mathcal{N}(i)$ denotes the set of neighbors of node i, $\text{AGGREGATE}^{(l)}$ is a function that aggregates the messages from the neighbors, and $\text{UPDATE}^{(l)}$ is a function that updates the node feature based on its previous feature and the aggregated message.

Instead of aggregating information from all neighbors, *GraphSAGE* [50] samples a fixed-size set of neighbors to maintain computational efficiency. It defines several aggregation functions to combine information from neighbors. Common aggregation functions include *mean aggregator*, *LSTM aggregator* (see Section 2.3.1 for a general LSTM description), and *pooling aggregator*:

$$\mathbf{h}_{\mathcal{N}(v_i)}^{(k)} = \text{mean}\left(\{\mathbf{h}_u^{(k-1)}, \forall u \in \mathcal{N}(v_i)\}\right) \quad (13)$$

$$\mathbf{h}_{\mathcal{N}(v_i)}^{(k)} = \text{LSTM}\left(\{\mathbf{h}_u^{(k-1)}, \forall u \in \mathcal{N}(v_i)\}\right) \quad (14)$$

$$\mathbf{h}_{\mathcal{N}(v_i)}^{(k)} = \max\left(\sigma\left(\mathbf{W}_{\text{pool}}\mathbf{h}_u^{(k-1)} + \mathbf{b}\right), \forall u \in \mathcal{N}(v_i)\right) \quad (15)$$

where the matrix $\mathbf{W}_{\text{pool}}$ and $\mathbf{b}$ are trainable. Concerning the update function, the updated node representation is computed as

$$\mathbf{h}_i^{(k)} = \sigma\left(\mathbf{W}^{(k)}\left[\mathbf{h}_i^{(k-1)} \| \mathbf{h}_{\mathcal{N}(v_i)}^{(k)}\right] + \mathbf{b}^{(k)}\right) \quad (16)$$

where $\|$ denotes concatenation, σ is a non-linear activation function (e.g., ReLU), and $\mathbf{W}^{(k)}$ and $\mathbf{b}^{(k)}$ are trainable parameters for layer k.

Graph pooling layers are used to reduce the number of nodes in the graph, similar to down-sampling in CNNs. A common method is *graph coarsening* [51], which clusters nodes and merges them. We indicate it generically with the following expression:

$$\mathbf{H}^{(l+1)} = \text{POOL}\left(\mathbf{H}^{(l)}, \mathbf{A}^{(l)}\right). \quad (17)$$

For node classification, the output layer produces a label for each node:

$$\mathbf{Z} = \text{softmax}\left(\mathbf{H}^{(L)}\mathbf{W}^{(L)}\right), \quad (18)$$

where L is the final layer, and $\mathbf{Z}$ is the matrix of class probabilities for each node.

For graph classification, the node features from the final layer can be pooled into a graph-level feature vector, followed by a fully connected layer and a softmax activation, as follows:

$$\mathbf{y} = \text{softmax}\left(\text{POOL}\left(\mathbf{H}^{(L)}\right)\mathbf{W}^{(L)}\right). \quad (19)$$

Graph attention networks (GATs) [52] introduce an attention mechanism to graph neural networks, enabling the model to learn the importance (through attention weights) of neighboring nodes when aggregating their features. The attention mechanism is further detailed in Section 2.4.4. This approach allows GATs to weigh the contributions of different neighbors differently, providing a more flexible and powerful solution to capture complex relationships. The *attention mechanism* assigns different weights to different neighbors based on their features and importance. *Self-attention* computes attention scores for each node–neighbor pair using learnable parameters. *Multi-head attention* improves the stability and expressiveness of the model by using multiple attention heads.

The interested reader can find a useful introduction to other similar variants in [45].

Several libraries have gained popularity for implementing and working with GNNs. They provide efficient implementations of different GNN data models and integration with deep learning frameworks, such as TensorFlow and PyTorch. PyTorch Geometric [53] is a library built on PyTorch, providing various methods for implementing GNNs. Its key features include support for various GNN models (e.g., GCN, GAT, GraphSAGE), and utilities for handling large-scale graphs. Deep Graph Library [54] supports multiple backend frameworks, including PyTorch and TensorFlow. It supports various GNN architectures and can handle large-scale graphs. Graph Nets [55] is a library developed by DeepMind, specifically for building graph networks in TensorFlow and Sonnet. It provides a high-level API for constructing GNNs. It integrates with TensorFlow. Spektral [56] is a library built on TensorFlow and Keras. It provides a wide range of GNN models for managing data graphs and is designed to be user-friendly and flexible. StellarGraph [57] is a library focused on graph-structured data. It is built on TensorFlow and Keras, and focuses on node classification, link prediction, and graph classification.

2.3. Recurrent Neural Networks

It often happens that the data to be processed must be considered as a temporal sequence. After all, natural processes evolve according to temporal processes or time series that characterize their lifetime. The models presented above have characteristics that are well suited to modeling snapshots of these processes, and not their evolution. For this reason, recurrent neural networks (RNNs) have been introduced [58].

Given their nature of allowing handling sequential data, they have beem successfully applied in a variety of fields, such as natural language processing (NLP), sentiment analysis of a piece of text, named-entity recognition (NER), speech recognition, text-to-speech and speech-to-text, time-series prediction of stock prices, market trends, and economic indicators, generation of descriptive text for video content, and control of dynamic systems in real time. Clearly, all these features and applications can easily find space in healthcare [59,60]. Some typical examples include aid in medical diagnosis by the analysis of sequential medical data [61–63], monitoring of patient vitals over time to detect anomalies [64], analysis of epidemiological data for tracking infections, and much more [65,66].

The development of recurrent networks has occurred over time, with several proposals that have followed one another. In this paper, we refer to the two most popular architectures, which are long short-term memory (LSTM) [67] and gated recurrent unit (GRU) [68]. The interested reader can find a description of RNN development in [58].

These networks are called ”recurrent” since they perform the same task for every element of an input sequence in a chain of operations, providing an output depending on previous computations. An RNN processes an input sequence $\mathbf{x} = (x_1, x_2, \ldots, x_T)$ by maintaining a hidden state $\mathbf{h}_t$ at each time step t. The hidden state is updated using the previous hidden state and the current input.

The basic RNN operation can be illustrated as follows:

$$\mathbf{h}_t = \sigma_h(\mathbf{W}_{hx}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h) \quad (20)$$

$$\mathbf{y}_t = \sigma_y(\mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y) \quad (21)$$

where $\mathbf{h}_t$ is the hidden state at time step t, $\mathbf{x}_t$ is the input at time step t, $\mathbf{y}_t$ is the output at time step t, $\mathbf{W}_{hx}$, $\mathbf{W}_{hh}$, and $\mathbf{W}_{hy}$ are weight matrices, $\mathbf{b}_h$ and $\mathbf{b}_y$ are bias vectors, and σ_h and σ_y are activation functions, such as tanh or ReLU.

RNNs can have different architectures. A “one-to-many” architecture refers to a situation where the RNN receives a single input and produces a sequence of outputs. A “many-to-one” architecture refers to the situation where the network processes a sequence of inputs and generates a single output. Finally, a “many-to-many” architecture refers to processing a sequence of inputs and the generation of a corresponding sequence of outputs.

The most popular RNN libraries are offered by TensorFlow, PyTorch, and Keras. However, Apache MxNet [69], Microsoft Cognitive Toolkit [70], and Chainer [71] are widely used in the research and development of RNN models, each offering unique features that cater to different requirements and preferences.

2.3.1. Long Short-Term Memory

LSTM [72] is an RNN variant designed to overcome some limitations of traditional RNNs in learning long-term dependencies. For this purpose, LSTM networks include memory cells with gates to control the flow of information. The LSTM cell is governed by the following equations:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \tag{22}$$
$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \tag{23}$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \tag{24}$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) \tag{25}$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \tag{26}$$

where $\mathbf{f}_t$, $\mathbf{i}_t$, and $\mathbf{o}_t$ are the forget, input, and output gates, respectively, $\mathbf{c}_t$ is the cell state at time step t, $\mathbf{W}_*$ and $\mathbf{U}_*$ are weight matrices, $\mathbf{b}_*$ is the bias vectors, σ is the sigmoid activation function, and $\odot$ denotes element-wise multiplication.

2.3.2. Gated Recurrent Unit

The gated recurrent unit is an RNN type that has the same objectives as LSTM but with a simpler architecture, combining forget and input gates into a single update gate. The GRU operation [73] is given by

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z) \tag{27}$$
$$\mathbf{r}_t = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r) \tag{28}$$
$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h (\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h) \tag{29}$$
$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t \tag{30}$$

where $\mathbf{z}_t$ is the update gate, $\mathbf{r}_t$ is the reset gate, $\tilde{\mathbf{h}}_t$ is the candidate hidden state, $\mathbf{h}_t$ is the hidden state at time step t, $\mathbf{W}_*$ and $\mathbf{U}_*$ are weight matrices, and $\mathbf{b}_*$ is the bias vectors. Figure 6 shows the LSTM and GRU architectures.

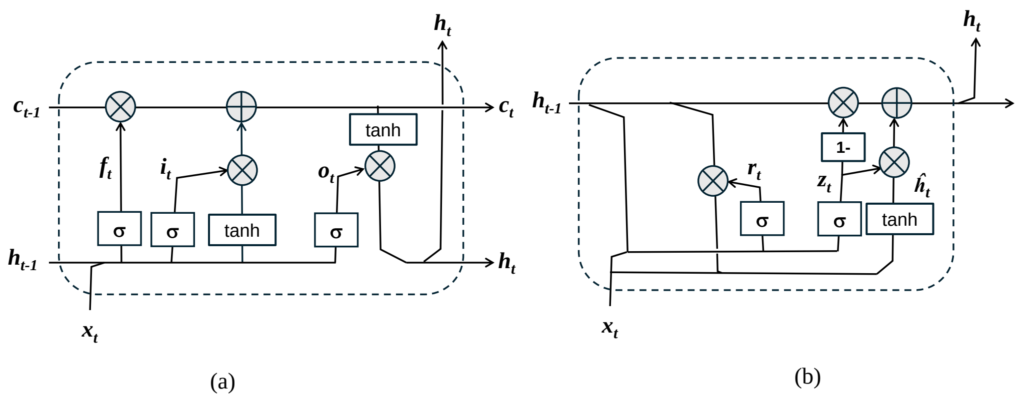

Figure 6. Example of a recurrent neural network: (**a**) LSTM; (**b**) GRU.

2.4. Generative AI

In some situations, it may happen that the available data are not sufficient, or it takes a long time to collect a sufficient volume to train the algorithms, or the data are corrupted by gaps or excessive noise. These issues can frequently occur in the biomedical field. For example, it may happen that the number of images to train a deep learning system for the recognition of a syndrome is not large enough, or the time necessary to acquire such images during clinical activities would requires some years. It may also happen that the data present in the EHR system are incomplete, with some fields not filled in. In these situations, generative AI techniques can provide decisive support [74]. Another situation where generative AI has been successfully applied is in the automatic production of clinical documentation, relieving medical personnel of this burden. Clinical documentation is a very relevant aspect in medical activity, requiring a considerable effort of time and energy. Therefore, the ability to facilitate medical workflow is essential to improve timely treatments for patients. In particular, it is essential to facilitate the transcription of conversations with patients, the process of formulating the diagnosis from symptoms, and to provide a better patient experience. These aspects and possible solutions can be found in the recent papers [75,76] .

In this section, we illustrate the basic concepts of modeling of the following AI algorithms used in this area.

2.4.1. Generative Adversarial Networks

GANs consist of the combined usage of two neural networks, namely, a generator G and a discriminator D. They are trained simultaneously by executing an adversarial competition between them [77].

The generator input is a random noise process z, with distribution $p_z(z)$. Typically, a Gaussian distribution is used. Noise samples are transformed for obtaining synthetic data $G(z)$ that mimic the distribution of the emulated process p_{data}. In turn, the discriminator receives both real data x, sampled from the real data distribution $p_{\text{data}}(x)$, and the synthetic data $G(z)$. It evaluates the probability $D(G(z))$ that represents the likelihood of the input being real data. The objective of the discriminator is to correctly classify the input data as real or fake. The objective of the generator is to fool the discriminator into classifying the synthetic data as real. The training process can be modeled as a minimax game, where D and G are optimized simultaneously. The objective function $V(G,D)$ can be written by using the entropy expression as follows:

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{31}$$

Since D must be trained to maximize the probability of correctly classifying data, we can write

$$\max_D \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{32}$$

Therefore, training can take advantage of the gradient ascent on the discriminator parameters θ_D:

$$\theta_D \leftarrow \theta_D + \nabla_{\theta_D}\left(\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))]\right) \tag{33}$$

Since G must be trained in order to compromise the classification performance of D, that is, to maximize the probability of the discriminator being fooled, we can write

$$\min_G \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{34}$$

or

$$\max_G \mathbb{E}_{z\sim p_z(z)}[\log D(G(z))] \tag{35}$$

Also, in this case, the gradient descent algorithm can be applied to the generator parameters θ_G:

$$\theta_G \leftarrow \theta_G + \nabla_{\theta_G}\mathbb{E}_{z\sim p_z(z)}[\log D(G(z))] \tag{36}$$

By iteratively updating the discriminator and the generator, the generated data tend to be indistinguishable from real data to the discriminator.

As mentioned above, this approach proves advantageous in different situations. For example, GANs can generate synthetic patient data for privacy protection or to augment training datasets in order to improve the robustness and accuracy of predictive models [78–80]. A notable example of GAN technologies is StyleGAN, developed by NVIDIA researchers [81].

2.4.2. Autoencoders and Variational Autoencoders

Autoencoders and variational autoencoders (VAEs) [82–84] are types of artificial neural networks used to learn efficient coding of input data. Efficient coding is essential for compressing data representation for autoencoders and generating new content, detecting anomalies, and removing noise for VAEs. Both autoencoders and VAEs include two essential components, an *encoder* and a *decoder*. The encoder transform the data from a higher- to a lower-dimensional space. The decoder converts the latent space back to higher-dimensional space. In both cases, although there are some important differences, which are illustrated below, the decoder is trained so that most of the information in the dataset space is preserved in the lower-dimensional, also called *latent*, space.

Autoencoders

Let $\mathbf{x} \in \mathbb{R}^d$ be the input data. They are processed by the encoder function $\mathbf{z} = f_\theta(\mathbf{x})$, where f_θ is typically a neural network parameterized by θ. $\mathbf{z} \in \mathbb{R}^m$ is therefore the input data representation in the latent space. The decoder takes $\mathbf{z}$ as input and converts the data by the function $\hat{\mathbf{x}} = g_\phi(\mathbf{z})$, where g_ϕ is typically again a neural network parameterized by ϕ. In this way, the input data are reported back into $\mathbb{R}^d$. The autoencoder is trained to minimize the reconstruction error between $\mathbf{x}$ and $\hat{\mathbf{x}}$. A commonly used loss function is the root mean squared error (MSE):

$$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \|\mathbf{x} - g_\phi(f_\theta(\mathbf{x}))\|^2 \tag{37}$$

The parameters θ and ϕ are optimized during the training phase in order to minimize the loss:

$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\left[\|\mathbf{x} - g_\phi(f_\theta(\mathbf{x}))\|^2\right] \tag{38}$$

Variational Autoencoders

VAEs include a probabilistic component in their operation. Essentially, the encoder outputs parameters of a probability distribution rather than a simple 'estimate, which is the basis for generating new content. This distribution is referred to as $q_\phi(\mathbf{z}|\mathbf{x})$, and approximates the posterior distribution $p_\theta(\mathbf{z}|\mathbf{x})$. The decoder, in turn, evaluates the likelihood of the data given the latent variables. The VAE's goal is to maximize the so-called evidence lower bound (ELBO):

$$\log p_\theta(\mathbf{x}) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x})\|p_\theta(\mathbf{z})) \tag{39}$$

where $\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})]$ is the expected log-likelihood of the data, and $\mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x})\|p_\theta(\mathbf{z}))$ is the Kullback–Leibler divergence between the approximate posterior and the prior. In more detail, assuming a Gaussian input data distribution, the encoder outputs the mean $\mu_\phi(\mathbf{x})$ and the standard deviation $\sigma_\phi(\mathbf{x})$ of the latent variable distribution:

$$q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \mu_\phi(\mathbf{x}), \sigma^2_\phi(\mathbf{x})) \tag{40}$$

In this way, it is not possible to compare two samples: one input and one output. Therefore, the distance must consider the parameters being estimated to allow back-propagation during training. For this purpose, the latent variable $\mathbf{z}$ is sampled as

$$\mathbf{z} = \mu_\phi(\mathbf{x}) + \sigma_\phi(\mathbf{x}) \odot \epsilon \tag{41}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$. The decoder can sample the reconstructed data $\hat{\mathbf{x}}$ from the likelihood:

$$p_\theta(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\hat{\mathbf{x}}; g_\theta(\mathbf{z}), \sigma^2\mathbf{I}) \tag{42}$$

Finally, the VAE loss function combines the reconstruction loss and the Kullback–Leibler (KL) divergence [85]:

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = -\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] + \mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})) \tag{43}$$

The parameters θ and ϕ are optimized by minimizing this loss:

$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})}[\mathcal{L}(\theta, \phi; \mathbf{x})] \tag{44}$$

The schematic architectures of autoencoders and variational autoencoders are shown in Figure 7.

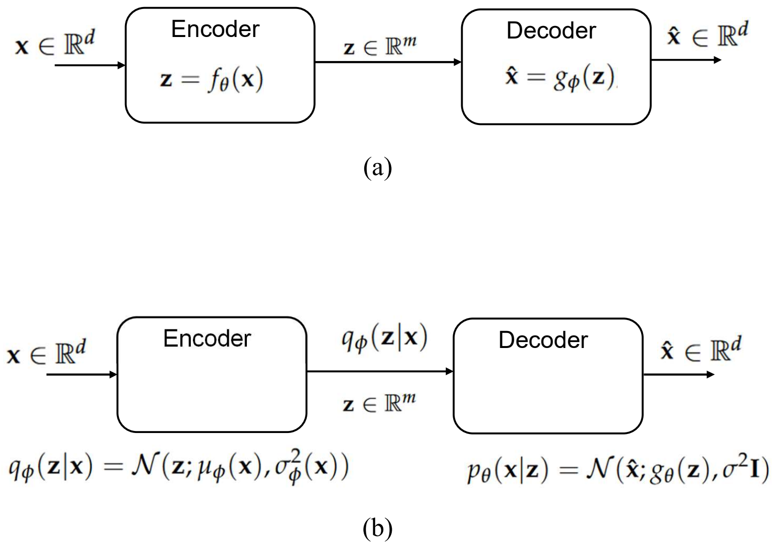

Figure 7. Schemes of autoencoder (**a**) and variational autoencoder (**b**) GRUs.

Autoencoders and VAEs can be used in different healthcare contexts [82,84]. For example, they can generate realistic emulations of disease evolution, aiding in understanding complex conditions and finding suitable treatment. Autoencoders can be used to learn compact representations of medical data, aiding in essential operations such as anomaly detection, identification of outliers, and data compression. For this reason, autoencoders are often regarded as valid alternatives to other dimensionality reduction and feature extraction methods in data analysis, such as *principal component analysis (PCA)*, *linear discriminant analysis (LDA)*, and *t-distributed stochastic neighbor embedding (t-SNE)* [22]. VAEs can also help both in generating and reconstructing medical images, facilitating the training of systems used to implement tasks such as difficult segmentation of organs.

2.4.3. Recurrent Neural Networks (RNNs)

RNNs are illustrated in Section 2.3. In this section, we only add the role that these networks can have as a generative source for useful information in the healthcare field. In particular, their usage has been proposed for predictive modeling [86]. These models are used for generating sequences of medical data, such as patient health records, used to train systems able to predict future health events and outcomes [87,88]. In addition, they are also used for text generation to generate medical reports from raw clinical data [86,89].

2.4.4. Transformers

Natural language processing (NLP) has proven to be one of the most promising areas in influencing the future of healthcare [90]. For this purpose, Transformer-based models, such as the Generative Pre-trained Transformer (GPT), can be used to generate and understand medical texts, assist in medical coding, summarization, and question-answering tasks [1,91,92].

The Transformer model was proposed in [93]. It is the core of many popular models for NLP, including BERT [94], the already mentioned GPT [95], and T5 [96]. The main innovation of the Transformer model is the introduction of the so-called *attention* mechanisms. Essentially, they analyze the relationships between words in a sentence without using other models such as RNNs. The added value of the introduction of Transformers includes, essentially, parallelization, capture of long-range dependencies, and scalability. This means that unlike previous approaches, such as RNNs, that process input data sequentially, Transformers allow for parallel processing for speeding up training times. Moreover, self-attention mechanisms allow for consideration of the entire input data sequence simultaneously, improving the capability to capture contextual relationships between data. Finally, the Transformer architecture scales well with increasing data volume and model size. Figure 8 shows the basic elements of Transformers, which are illustrated in what follows.

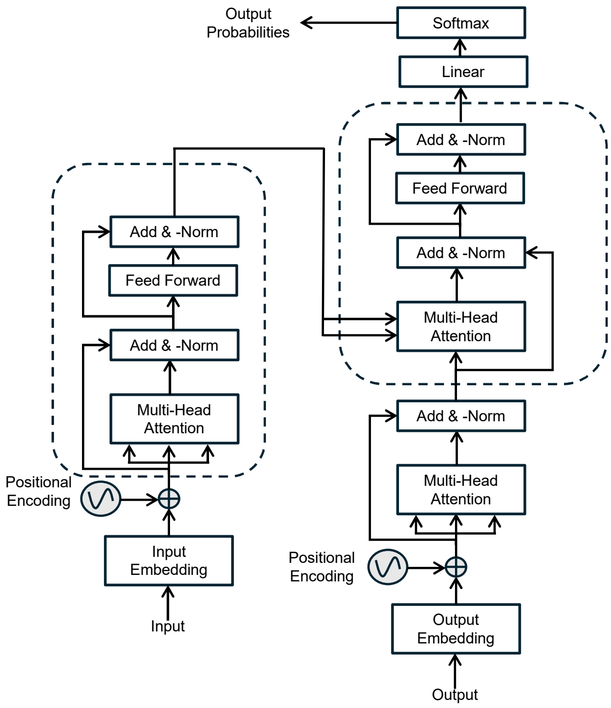

Figure 8. Transformer architecture.

The self-attention mechanism allows the model to weigh the importance of different words in a sentence when encoding a particular word. It is defined by the following elements.

The scaled dot-product attention is defined as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (45)$$

where Q (queries), K (keys), and V (values) are the input matrices, and d_k is the dimension of the keys. In more detail, consider an input sequence of length n. This sequence includes tokens, that could be words, subwords, characters, or even whole sentences, depending on the level of granularity needed for text processing. Each token is encoded by an *embedding* of dimension d_{model}. Data are aggregated as a matrix $X \in \mathbb{R}^{n \times d_{\text{model}}}$, referred to as the input embedding matrix. The Q, K, and V matrices are obtained by linearly transforming X as follows:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V \quad (46)$$

where $W_Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$ is the weight matrix for queries, $W_K \in \mathbb{R}^{d_{\text{model}} \times d_k}$ is the weight matrix for keys, and $W_V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ is the weight matrix for values. Thus, the resulting matrices Q, K, and V have dimensions $Q \in \mathbb{R}^{n \times d_k}$, $K \in \mathbb{R}^{n \times d_k}$, and $V \in \mathbb{R}^{n \times d_v}$. Clearly, $d_k = d_v = d_{\text{model}}/h$, where h is the number of attention heads, which refers to the *multi-head attention* operation of the model, which allows for focusing on different parts of the input sequence in parallel, enhancing its ability to capture different aspects of the relationships between tokens. Essentially, the multi-head attention mechanism extends the self-attention mechanism by applying it multiple times in parallel. Hence, given the input embedding X, the multi-head attention mechanism is modeled as follows:

1. First it is necessary to linearly project the input embedding X into multiple sets of keys, queries, and values by using the learned weight matrices:
$$Q_i = XW_Q^i, \quad K_i = XW_K^i, \quad V_i = XW_V^i, \quad \forall i \in \{1, \ldots, h\} \tag{47}$$
where h is the number of attention heads.
2. Then, the self-attention mechanism is applied to each set of keys, queries, and values, as follows:
$$\text{head}_i = \text{Attention}(Q_i, K_i, V_i) \tag{48}$$
3. After this, the outputs of all attention heads are concatenated:
$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h) \tag{49}$$
4. Finally, the result of the concatenation is linearly projected through the learned weight matrix W^O:
$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{50}$$

Positional encoding is added to the input embedding to include in the model the position of words in the sequence. This encoding is calculated by using sine and cosine functions with different frequencies, as follows:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{51}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{52}$$

where pos is the position and i is the dimension.

These equations ensure that each position pos is mapped to a unique encoding vector, with alternating sine and cosine functions applied to even and odd dimensions, respectively. This approach to positional encoding has some important properties. In particular, each pos value in the sequence has a unique encoding, the difference between encodings reflects the relative positions of tokens, thus helping the model to capture their order. Furthermore, the use of trigonometric functions introduces periodic patterns, which can help to identify repeating structures in the data. The positional encoding is simply added to the input embedding before sending it to the Transformer model. Specifically, given an input embedding $X \in \mathbb{R}^{n \times d_{\text{model}}}$, the positional encoding $PE \in \mathbb{R}^{n \times d_{\text{model}}}$ is added as follows:

$$Z = X + PE \tag{53}$$

where Z is the resulting input to the Transformer.

The further component worthy of attention is the *layer normalization and residual connections*. Each layer in the Transformer architecture is followed by this operation to ensure stable and efficient training:

$$\text{LayerNorm}(x + \text{Sublayer}(x)) \tag{54}$$

Finally, a fully connected feed-forward neural network processes each position in the sequence independently. This network consists of two linear transformations with a ReLU activation between them:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{55}$$

A very important aspect to consider from these pre-trained models is their flexibility to be subsequently adapted, through subsequent training, to generate text in different application contexts. In the healthcare area, notable examples, which demonstrate the usefulness of these models, are BioGPT [91], GPT-4 Medprompt [97], and MediTron-70B [98], trained on large-scale biomedical literature for generating fluent descriptions using biomedical terms.

2.5. Diffusion Models

Diffusion models are probabilistic generative models. They are effective in generating high-quality output data, typically images [99–101]. The basic idea of these models is to generate new content by reversing a diffusion process, that is, loss of information due to noise. Gaussian noise distribution is typically used. Due to this intrinsic capability, diffusion models can be useful in medical imaging applications, such as MRI and CT scans.

Diffusion models consist of two phases. First, in the *forward diffusion process*, a neural network is trained to introduce noise into the dataset, then this process is reversed. The forward diffusion process gradually adds noise to data over a number of time steps until the original data are indistinguishable from the Gaussian cloud distribution. Let x_0 be a sample from the data distribution $q(x_0)$. The result of the forward process is a sequence of latent variables $x_1, x_2, \ldots, x_T$, where T denotes the total number of steps.

Since at each step t noise is added to the sample, it follows that

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}) \tag{56}$$

where β_t is a variance schedule that controls the amount of noise added at each step. The full forward process can be written as

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}) \tag{57}$$

The *reverse diffusion process* aims to recover the original data from the noise by reversing the forward process. This operation is represented as

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{58}$$

where μ_θ and Σ_θ are the parameters for the model to learn in order to tune the reverse process, that can be expressed as

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \tag{59}$$

where $p(x_T)$ is typically a standard Gaussian distribution.

The entire process, sketched in Figure 9, can be modeled as a Markov chain. The model is trained to minimize the difference between the forward and reverse processes. The resulting training objective can be derived as

$$L_{\text{vlb}} = \mathbb{E}_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] \tag{60}$$

that can be expressed as the sum of KL divergences between the forward and reverse distributions at each step.

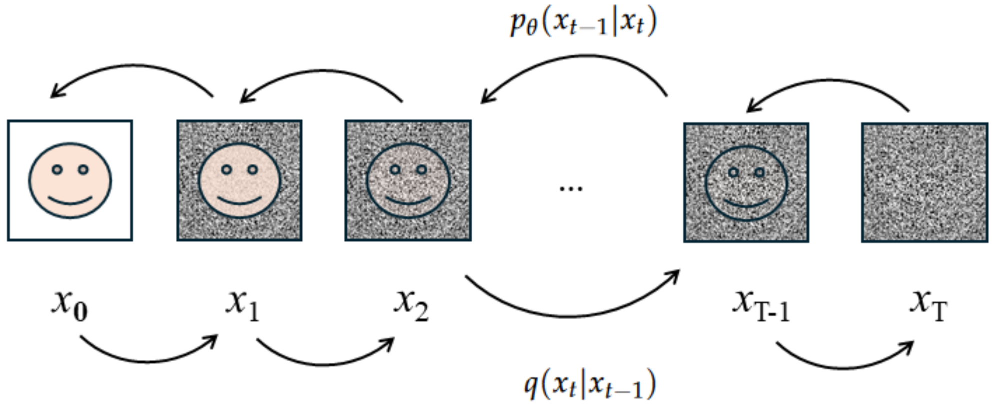

Figure 9. Schematic representation of the forward and reverse diffusion processes.

The complexity of this operation induced the definition of a simplified training objective, often used, which aims to reduce the noise in the corresponding score. This objective can be expressed as

$$L_{\text{simple}} = \mathbb{E}_{x_0,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(x_t,t)\|^2\right] \tag{61}$$

where ϵ is the noise added at step t and ϵ_θ is the model prediction of the noise.

To generate samples from the model, a noise sample $x_T \sim \mathcal{N}(0, \mathbf{I})$ is used and processed by applying the learned reverse process iteratively:

$$x_{t-1} \sim p_\theta(x_{t-1}|x_t) \tag{62}$$

The process ends when a sample x_0 from the learned data distribution is obtained.

As mentioned above, diffusion models can generate high-quality medical images, which can be used for different purposes. For example, it is possible to increase the volume and diversity of medical image datasets for training robust machine learning models. They have proved useful for enhancing the clarity and detail of diagnostic images. As a further example, diffusion models can also generate novel molecular structures with desirable properties, aiding in drug discovery [102].

Concerning the implementation of applications in the context of healthcare based on diffusion models, PyTorch and TensorFlow, are extremely popular. However, it is worth mentioning Diffusers [103] and NVIDIA Clara [104].

A notable research example of the use of diffusion models is illustrated in [105]. The authors present a study on the motion in hydrogels of mucins, using a Bayesian analysis for comparing different diffusion models for various tracer trajectories.

2.6. Reinforcement Learning

When dealing with particularly complex systems, characterized by non-deterministic evolutionary dynamics, reinforcement learning (RL) techniques can be very useful. This can happen in various healthcare sectors. For example, in the definition of personalized medical and care techniques, evolutionary patterns may occur that cannot be framed in general models that can be trained through a dataset collected from different patients. Another difficult situation is clinical and care robotics, in which robots must behave in accordance with the needs of the environment in which they operate [106]. In situations of this type, RL techniques may be crucial.

The general objective of RL techniques is to allow an agent to learn through continuous interaction with the environment where it resides and has to interact [107,108]. Learning consists of discovering suitable strategies by making decisions and receiving related rewards, either positive or negative, according to the results due to the selected actions. Thus, the objective can be synthesized as learning the strategy that maximizes the final reward. Learning *iterations* consist of consecutive elementary interactions between the agent and the environment, called *steps*. Interactions, along with their rewards, are associated with a *state* of the system. In such systems, a state consists of a collection of parameter values that are sufficient to model how the system evolves over time, that is, the system state sequence. Thus, the association between states, actions, and rewards is the basis of training. In summary, at each step, the agent makes decisions, interacts with the environment, characterized by a state, receives the reward, and adapts its behavior based on the collected reward. Steps are repeated over time until the agent's performance is acceptable. RL algorithms include *policy* functions. These functions determine the strategies that map states with the actions taken by the agent. Therefore, the policy must determine the agent's behavior in all states of the environment. Policies can be of different natures. Essentially they can be deterministic or statistical. In the context of healthcare AI systems, statistical policies are typically used. They are indicated as $\pi_\theta(a|s)$, and represent the probability of selecting the action a in the state s. The parameters in the vector θ represent the learned parameters determining the optimal policy.

In recent years, RL research has produced numerous algorithms characterized by different properties in terms of state representation, stability, and convergence speed. The interested reader can find details in [109–112].

Research on the application of RL techniques in different fields in healthcare is still in its early phase, but some contributions are promising. For example, Ref. [113] shows an application of RL techniques for discovering new treatments and personalizing existing ones. Application to dynamic treatment regimes in chronic diseases for automated medical diagnosis is investigated in [114].

Robotics is one of the main fields of application of RL. Therefore, robotic systems used in healthcare are largely influenced by the applications of these techniques. Ref. [106] presents an application of RL to precision robotics and the determination of an optimal strategy driving the growth of cell cultures and the development of therapeutic solutions.

However, recent results also show how research is still in progress. For example, in [115] the authors focus on oncology and identify current challenges and pitfalls to be considered in order to successfully implement RL-based decision support systems for precision oncology.

Concerning the available technologies for implementing the RL model, in addition to TensorFlow and PyTorch it is worth mentioning Gymnasium [116], which allows developers to emulate complex environments and easily interact with them, and Stable-Baselines3 [117], a useful library of RL algorithms.

Table 1 reports a summary of the presented AI methods, specifying the best-suited for healthcare tasks, critical limitations of their usage, essential requirements, input data features, some available popular technologies, and their readiness level.

Table 1. Summary of AI methods in healthcare and relevant features.

AI Method	Best-Suited Healthcare Tasks	Critical Limitations	Essential Requirements	Input Data	Available Technologies	Readiness Level
DL	Segmentation activity of internal organs, early and advanced diagnosis for different pathologies and syndromes	High computational cost for training, high training data volumes	Annotated data, high-performance computing infrastructure, domain expertise for training	Medical images (X-rays, CTs), EHRs, omics data	TensorFlow, PyTorch, Keras, CUDA Toolkit	Medium to High

Table 1. *Cont.*

AI Method	Best-Suited Healthcare Tasks	Critical Limitations	Essential Requirements	Input Data	Available Technologies	Readiness Level
GNN	Drug discovery, protein affinity prediction, modeling complex relationships in EHRs	Scalability issues, difficult interpretability of results, complex data preprocessing	Need for graph-structured data, high-performance computing infrastructure, graph processing tools	Molecular structures, biomedical pathways, EHR graph data	TensorFlow, PyTorch, Deep Graph Library, Spektral, StellarGraph	Low to medium with rapid growth rate
RNN	Time-series analysis of health data (e.g., EHR), continuous patient monitoring, telemedicine, protein affinity prediction	Vanishing gradient, long training times, highly variable health status	Data preprocessing and alignment, high-quality data, significant computational resources	Time-series data, EHR sequences, wearable sensor data	TensorFlow, PyTorch, Keras	Medium
Generative AI	Clinical documentation, conversational agents, synthetic medical data generation	Possible misleading data, high computational requirements, ethical concerns	Reliable validation of results, high-quality training data, ethical oversights	Clinical images, EHRs, omics data	BERT, GPT, StyleGAN, BioGPT, GPT-4 Medprompt, MediTron-70B	Medium to high, with growing usage for synthetic data for training purposes
Diffusion Models	Image reconstruction, denoising of medical images, generation of high-resolution medical images	High computational cost, training complexity, privacy leaks for federated learning	High computational resources, suitable process initialization, high-quality training data	Medical images (MRI, CT scans), noisy or incomplete training images	PyTorch, TensorFlow, Diffusers, NVIDIA Clara	Low, with rapid growth rate
RL	Personalized treatments, robotic surgery, support for clinical decision making	Slow training, complex experimental setup, ethical concerns, safety concerns	Emulated environments, real-world feedback data, ethical compliance and safety assessment	Patient data for state model, wearable sensor data, treatment outcomes	TensorFlow, PyTorch, Gymnasium, Stable-Baselines3	Medium to High

3. AI Penetration in Baseline Healthcare Services

Section 2 includes many bibliographic references that show several advances in the context of healthcare. The aim of this section is to show the versatility of the techniques presented and the potential of their joint use. For this purpose, we consider some baseline activities that characterize the healthcare and highlight the existing contributions to them by all the techniques presented in Section 2.

The considered baseline activities are *diagnosis*, *patient treatment and management*, *clinical care and decision support*, *research and development*, *administration and management*, and *prevention and wellness*. For each baseline, Table 2 reports some selected examples of existing contributions in order to show the flexibility and adaptability of the AI techniques presented.

The selected examples in the identified baselines are often overlapping, and some contributions can be considered in common between them. Furthermore, a good part of the contributions are characterized by a technology readiness level that is still insufficient for clinical practice. However, they are sufficient to demonstrate how AI is penetrating and reshaping related activities.

Table 2. AI penetration in some baseline healthcare services.

Area	Deep Learning	GNN	RNN	Generative AI	Diffusion Models	Reinforcement Learning
Diagnosis	Analysis of radiological images, such as X-rays or CT scans [35,118–122]. Interpretation of laboratory test results, such as blood tests and genetic tests. In [123], the performance of U-Net for segmenting COVID-19 lesions on lung CT-scans is analyzed.	Symptom-based diagnosis: Use of algorithms to diagnose diseases based on symptoms reported by patients [45].	Aid in medical diagnosis by the analysis of sequential medical data [61–63].	Symptom-based diagnosis: Use of algorithms to diagnose diseases based on symptoms reported by patients. Ref. [124] adds the attention mechanism into the original U-Net architecture for improving the ability to segment small items.	Non-invasive prediction of tumor growth rate by using diffusion models is shown in [125].	Ref. [113] shows an application of RL techniques for discovering new treatments and personalizing existing ones.
Patient Treatment and Management	Processing of histopathology of nasal polyps prognostic information by deep learning for patient treatment is shown in [126].	Hierarchical GNN for patient treatment preference prediction, integrating doctors' information and their viewing activities as external knowledge with EMRs to construct the graph [127].	Virtual assistant that monitors patients' vitals over time to detect anomalies [64].	Management of EHR system when incomplete [74].	Diffusion model application in EHR. Solution to perform class-conditional sampling for preserving label information [128].	Application to dynamic treatment regimes in chronic diseases [114].
Clinical Care and Decision Support	A study about the utility of machine learning in pediatrics is presented in [129].	GNN model that learns patient representation using different network configurations and feature modes [130].	Solutions based on the analysis of sequential medical data [61–63].	Analysis of the accuracy of ChatGPT-derived patient counseling responses based on clinical care guidelines in urology [131].	Diffusion models generating high-quality realistic mixed-type tabular EHRs, preserving privacy, used for data augmentation [132].	Survey including recommendation systems to physicians based on clinical guidelines and patient data analysis [133].
Research and Development	Blood vessel segmentation is investigated in [134,135]. In [134,135] a U-Net is used for coronary artery stenosis detection on X-ray coronary angiograms.	Drug discovery: Using AI to identify new drug compounds and predict their efficacy and safety [136].	Analysis of epidemiological data and for tracking infections, and much more [65,66].	Research for improving images in healthcare for challenging situations [137].	Generation of high-quality data for training AI algorithms [138].	Precision robotics application to healthcare [106]. Ref. [115] focuses on precision oncology and identifies current challenges and pitfalls.
Administration and Management	The potentials of deep learning in hospital administration and management, including ethical and legal issues are presented in [139].	Implementation of a knowledge graph for discovering insights in medical subject headings [140].	Management of resources, scheduling activities, and improving the operational efficiency of hospitals. For example, see [141].	Transformative healthcare for automating clinical documentation and processing of patient information [142].	Diffusion models have also been proposed within the life cycle of innovation [143].	Many examples in the great survey in [144].

Table 2. *Cont.*

Area	Deep Learning	GNN	RNN	Generative AI	Diffusion Models	Reinforcement Learning
Prevention and Wellness	A deep learning algorithm for sleep stage scoring, making use of a single EEG channel, is presented in [145].	Use of a patient graph structure with basic information like age, gender, and diagnosis, and the trained GNN models for identifying therapies [146].	Analysis of sequential medical data [61–63].	Empowering workers and anticipating harms in integrating large language models with workplace technologies [147].	Use of 3D avatars in applications for fitness and wellness [148].	RL in patients with mood and anxiety disorders [149].

The scenario that emerges from analysis of the table is characterized by the use of sophisticated AI models suitably complementing human expertise. Beyond enabling advanced services, such as precision and personalized care, the AI techniques and technologies will be able to improve the effectiveness of doctors' work by relieving them of a large part of administrative and repetitive tasks. It is worth considering the increasing presence of generative AI systems. Their expected contribution is quite complex. On the one hand, these systems can contribute to the interaction with patients during medical consultations. On the other, during consultations generative AI may also help doctors define patient treatment through the analysis of vast datasets to recommend personalized treatment plans. AI-assisted diagnosis is the sector that seems most mature and already presents examples of clinical application. Image analysis and early pathology detection already make extensive use of deep learning techniques. However, even in this case, generative AI promises to improve accuracy and speed through the processing of learning datasets.

A further extraordinary contribution from AI is accelerating drug development. Algorithms can help researchers identify disease markers and find new combinations of chemicals to create and rapidly screening new pharmaceutical compounds, predicting drug interactions and even repurposing existing drugs.

Administration and management may appear less impressive than individualized diagnoses and treatments. However, from the point of view of the sustainability of the healthcare system, the contribution of AI is equally valuable and can be pursued by making use of all the techniques analyzed.

4. Other Related Challenges

To bring AI into professional healthcare activities, it is not sufficient to create and implement sophisticated algorithms and have an adequate volume of data. There are aspects of a technical, ethical, and regulatory nature that are equally decisive and which require further research and implementation efforts. Most of them are common to all services based on the future Internet, although with peculiarities related to the specific application area.

We have identified the following areas, which are analyzed below: *Human–computer interaction*, *explainability*, *wearable sensors*, *privacy and security*, *network and computing infrastructure*, *bias and equity*, *regulation and governance*.

These areas are the pillars that allow full connectivity between the areas and the tools for the healthcare practices resulting in Table 2, as shown in Figure 10. In other words, this full connectivity can lead to real progress only if it is fueled by the development of the identified areas. In this way, the developed AI-based tools, along with the relevant regulatory support, can truly generate a new healthcare which significantly exploits the potential and opportunities provided by AI.

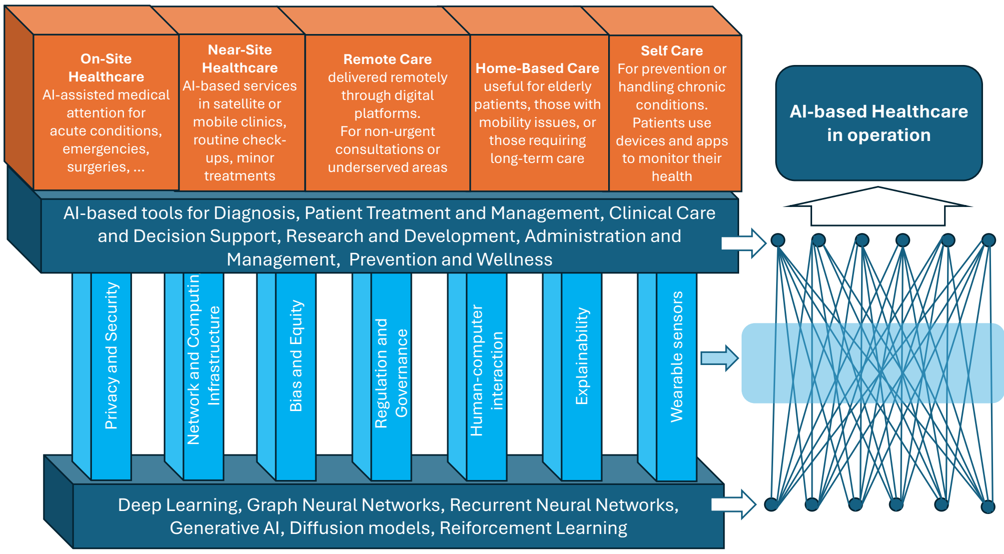

Figure 10. Schematic representation of the forward and reverse diffusion processes.

4.1. Human–Computer Interaction

Development of advanced and intuitive user interfaces is essential to allow doctors and patients to easily interact with AI systems [150]. Augmented reality (AR) and virtual reality (VR) can contribute to improving medical training, surgical planning, and patient care [151]. Human–computer interaction (HCI) plays an enabling role in the healthcare sector, enabling the effective use of digital technologies. However, it must be considered that the different healthcare practices and the types of problems to be addressed require specific solutions. For example, the need to ensure coordination and treatment in mental health and suicide prevention services is addressed in [152,153]. HCI has evolved through several phases, the latest of which focuses on emotional aspects and user experience for Transformer-based generative AI [150]. This is to optimize patient-centered care and supported decision making [154]. Furthermore, HCI is essential to train healthcare workers in the use of digital tools that enable easier access to AI technologies, as well as to ensure adequate cybersecurity to protect patients, healthcare workers, facilities, and companies [155].

4.2. Explainability

Explainable AI can be seen as an enrichment of AI techniques, achieved by introducing analytical processes and methods that allow users to consciously use the results of algorithms by understanding the results themselves. Basically, in many sectors, including healthcare, the uncritical use of algorithmic results is difficult to accept, especially when the consequences of choices have a significant impact on people [156,157]. This subject has already been present for a few years. For example, Ref. [158] provides an examination of the role of explainability in medical artificial intelligence and proposes an ethical evaluation of its adoption in tools that include AI components in clinical practice.

A similar objective is pursued in Ref. [159], where indications are provided to guide developers and researchers for future activities on clinical topics, in particular on applications of medical imaging.

The interested reader can find recent developments on this topic in the two valuable survey papers [160,161], where a deep bibliographical analysis covers all the major publishing houses and draws specific conclusions in relation to various pathologies.

4.3. Wearable Sensors

The transformative potential of AI is also crossing the area of wearable sensing. Through purpose-specific artificial intelligence algorithms and machine learning methodologies, next-generation wearable devices can achieve precise and personalized health monitoring. An overview of recent advances in wearable sensors, including biosensors, is provided in [162]. The authors focused on materials, structural configurations and transduction mechanisms. Ref. [163] details the availability of wearable sensors to monitor essential parameters such as respiratory and heart rates, sweat, and tears. Implantable sensors for cardiovascular treatments, for the collection of nervous signals and neurotransmitters, are also covered.

These technological advances might seem like a normal evolutionary process for enhancing already established medical practices such as telemedicine. Indeed, there is something more, which has the characteristics of a revolution in interaction with the patient, both in person and remotely: *digital twins*. The development of patient digital twins consists in creating a digital model of the patient themselves, making extensive use of AI, in order to be able to interact and receive feedback sufficiently close to what would be obtained by interacting directly with the patient. It is clear that this is a medical practice that allows treatments to be adapted in a way that was not possible before, enhance prevention, and maintain health well-being; it revisits telemedicine at its roots and also has a role in decongesting the healthcare system. To achieve this result, not only is it necessary to be able to create sufficiently adherent models of the patient, but it is necessary to keep them updated through continuous interaction, which requires the most advanced sensors. An overview of the current applications of DTs in healthcare is given in [164,165].

4.4. Privacy and Security

The increasingly massive use of AI in the healthcare sector requires the use and exchange of sensitive data, including patients' personal data, such as medical history, test results, and treatment plans. Therefore, security and privacy protection are exposed to threats that require the adoption of adequate solutions. Such solutions include advanced encryption, through the development of more secure encryption methods to protect sensitive medical data during transmission and storage, and anonymization techniques, which are essential for secure data sharing between different entities [166,167]. In this context, blockchain technologies are valid solutions to ensure the integrity, traceability, and security of medical transactions and records [168,169].

4.5. Network and Computing Infrastructure

In the ever-evolving landscape of AI healthcare, all the components of the Internet of the future are expected to have a significant contribution in terms of advanced data transfer, flexibility, and efficiency. This contribution can be declined based on the needs of health services. For example, the availability of digital twins requires the presence of a widespread, pervasive, and continuously available network. For these applications, the 5G system has already represented a technological turning point due to its performance, flexibility and adaptability [170–172]. As regards image-based diagnostic services, access to the applications available in the cloud must be constant, broadband, and secure. Furthermore, to adequately support remote surgery, in addition to the aforementioned characteristics, it is essential to have an adequate latency in data transfer, on the order of ms. Many other examples could be given, but it seems clear that the presence of AI is also indispensable in the creation and provision of network services. In this case, it is necessary to mention both research in Beyond 5G (B5G) and 6G [15]. It is also necessary to consider the considerable research effort in the intelligent management of edge computing services to reduce latency and improve data privacy by enabling local processing of sensitive data before they are sent to the cloud [173,174].

Concerning computing power, *quantum computing* is considered a game-changer. In fact, due to the increasing computation complexity of medical computing, for handling complex situations computation can still be a significant bottleneck [175]. For this reason, a quantum computing service could make a difference [176]. For example, since DNA information enables personalized medicine through the development of new therapies and drugs, being able to achieve this quickly, at speeds that are orders of magnitude faster, would make the level of detail of models much higher. Even in medical imaging and digital twins, quantum computers have the potential to create efficient rendering systems that can provide doctors with greater fine-grained clarity in real time [176].

4.6. Bias and Equity

This point addresses fundamental ethical aspects related to the design and use of AI. Bias and fairness can be pursued through targeted methods to identify and mitigate unfair and discriminatory outcomes. An interesting survey on fairness and bias in AI, also in healthcare, can be found in [177]. The presentation includes origins of issues, ramifications, and possible mitigation strategies. An experimental study on the influence of a large language model (LLM) generative AI system—ChatGPT—on the accuracy of physician decision making and bias is presented in [178]. It emerges that physicians are prone to change their initial evaluation following AI assistance. It also appears that it provides a significant improvement in clinical decision making without causing any race or gender biases. In [179], the potential impacts caused by socioeconomic inequalities in three key areas, namely, work, education, and health, are analyzed. It emerges that the healthcare sector could benefit greatly from the diagnostic and predictive capabilities of AI, in terms of accessibility of healthcare. However, the risk of increasing existing inequalities for under-resourced and marginalized communities is expected if appropriate countermeasures are not taken. The need to ensure the reliability of artificial intelligence systems before their implementation is underlined in [180]. The paper provides an example that demonstrates how reliable artificial intelligence can be used to eliminate otherwise existing prejudices.

4.7. Regulation and Governance

This section concerns the need to regulate and govern AI processes in order to avoid problems that could compromise their correct and acceptable use. It has been known for some time that AI research often evolves at a faster pace than the rules that govern it and that action is needed to fill a regulatory gap [181]. We believe that the first thing to consider in order to illustrate the current situation and the forthcoming possible scenario is a recent guide published by the World Health Organization (WHO) [182]. This guide has been released to assist Member States in mapping the benefits and challenges associated with the use of large multimodal transport models (LMMs), which can accept one or more types of data inputs and also generate different types of outputs, for health.

A global picture on regulatory issues, that integrates the WHO vision with Standards Development Organizations and National Regulation Authorities, is presented in [183]. A similar perspective is present in [184], where the authors hope for greater harmonization of ethical rules at an international level, currently considered inadequate. This conclusion derives from the analysis of different jurisdictions from different continents, even if a global player such as the European Union was not considered. The same need is also underlined in [185]. In addition to expressing the centrality of the WHO, it presents an analysis of the existing literature using data science techniques, focused on articles relating to the use of artificial intelligence in the healthcare sector, with particular attention to regulations, policies, and guidelines implemented by the EU or by the WHO.

5. Conclusions

This survey paper illustrates the evolutionary processes that are reshaping the healthcare sector. In presenting the evolutionary framework, we have mainly focused on recent contributions to technical research and service innovation. The paper shows both the already-consolidated AI techniques and the promising ones, subject to intense research, for which applicability is expected in the future. Initially, the paper illustrates the AI techniques that are having the greatest success in terms of research and implementation, with basic technical details for the reader interested in improving their knowledge. In particular, the following techniques are illustrated: Deep learning, including convolutional neural networks and U-Net architectures; graph neural networks; recurrent neural networks; generative AI; diffusion models; and reinforcement learning.

The techniques illustrated are then associated with the healthcare practices most suited to them, specifying the limitations and available technologies. Finally, a bibliographical analysis shows the possible contribution of all the techniques presented in different basic healthcare activities.

The first part of the paper is therefore designed to be useful to an IT researcher who intends to understand the potential of the various AI techniques and their mutual contribution. The rest of the paper could also be useful to a physician or healthcare professional who wants to improve their awareness of what can be achieved using the new available AI tools without going into deep technical details.

In this way, we believe we offer the reader a complete picture of the processes that are reconfiguring the healthcare sector, how to possibly contribute to this development, and what to expect in the near future.

Author Contributions: Conceptualization, G.R.; bibliography, G.R.; organization, M.F.; writing—original draft preparation, G.R.; writing—review and editing, M.F.; funding acquisition, M.F. and G.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the Fondazione Perugia under the project "Stroke AI-based Detection in Emergencies—SAIDE", project number: 21876 (2023.0537).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Maslej, N.; Fattorini, L.; Perrault, R.; Parli, V.; Reuel, A.; Brynjolfsson, E.; Etchemendy, J.; Ligett, K.; Lyons, T.; Manyika, J.; et al. *The AI Index 2024 Annual Report*; AI Index Steering Committee, Institute for Human-Centered AI, Stanford University: Stanford, CA, USA, 2024.
2. Artificial Intelligence in Healthcare Market. 2024. Available online: https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-in-healthcare-market-100534 (accessed on 22 August 2024).
3. Artificial Intelligence in Healthcare: Market Size, Growth, and Trends. 2024. Available online: https://binariks.com/blog/artificial-intelligence-ai-healthcare-market/ (accessed on 22 August 2024).
4. Park, H.A. Secure Telemedicine System. In Proceedings of the 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 12–14 December 2018; pp. 732–737. [CrossRef]
5. Picozzi, P.; Nocco, U.; Puleo, G.; Labate, C.; Cimolin, V. Telemedicine and Robotic Surgery: A Narrative Review to Analyze Advantages, Limitations and Future Developments. *Electronics* **2023**, *13*, 124. [CrossRef]
6. Bhattacharya, S.; Rawat, D. Comparative study of remote surgery techniques. In Proceedings of the 2015 IEEE Global Humanitarian Technology Conference (GHTC), Seattle, WA, USA, 8–11 October 2015; pp. 407–413. [CrossRef]
7. Femminella, M.; Reali, G.; Valocchi, D.; Nunzi, E. The ARES Project: Network Architecture for Delivering and Processing Genomics Data. In Proceedings of the 2014 IEEE 3rd Symposium on Network Cloud Computing and Applications (ncca 2014), Rome, Italy, 5–7 February 2014; pp. 23–30. [CrossRef]
8. Reali, G.; Femminella, M.; Nunzi, E.; Valocchi, D. Genomics as a service: A joint computing and networking perspective. *Comput. Netw.* **2018**, *145*, 27–51. [CrossRef]
9. Holmgren, A.J.; Esdar, M.; Hüsers, J.; Coutinho-Almeida, J. Health Information Exchange: Understanding the Policy Landscape and Future of Data Interoperability. *Yearb. Med. Inform.* **2023**, *32*, 184–194. [CrossRef]

10. Health Level Seven. Available online: https://www.hl7.org/ (accessed on 22 August 2024).
11. Yetisen, A.K.; Martinez-Hurtado, J.L.; Ünal, B.; Khademhosseini, A.; Butt, H. Wearables in Medicine. *Adv. Mater.* **2018**, *30*, 1706910. [CrossRef] [PubMed]
12. Nunzi, E.; Iorio, A.M.; Camilloni, B. A 21-winter seasons retrospective study of antibody response after influenza vaccination in elderly (60–85 years old) and very elderly (>85 years old) institutionalized subjects. *Hum. Vaccines Immunother.* **2017**, *13*, 2659–2668. [CrossRef]
13. Renga, G.; D'Onofrio, F.; Pariano, M.; Galarini, R.; Barola, C.; Stincardini, C.; Bellet, M.M.; Ellemunter, H.; Lass-Flörl, C.; Costantini, C.; et al. Bridging of host-microbiota tryptophan partitioning by the serotonin pathway in fungal pneumonia. *Nat. Commun.* **2023**, *14*, 5753. [CrossRef]
14. Cellina, M.; Cè, M.; Alì, M.; Irmici, G.; Ibba, S.; Caloro, E.; Fazzini, D.; Oliva, G.; Papa, S. Digital Twins: The New Frontier for Personalized Medicine? *Appl. Sci.* **2023**, *13*, 7940. [CrossRef]
15. Chataut, R.; Nankya, M.; Akl, R. 6G Networks and the AI Revolution—Exploring Technologies, Applications, and Emerging Challenges. *Sensors* **2024**, *24*, 1888. [CrossRef]
16. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6. [CrossRef]
17. International Atomic Energy Agency. *Diagnostic Radiology Physics—A Handbook for Teachers and Students*; Non-Serial Publications; IAEA: Wien, Austria, 2014. Available online: https://www.iaea.org/publications/8841/diagnostic-radiology-physics (accessed on 18 September 2024).
18. Horowitz, S.L.; Pavlidis, T. Picture Segmentation by a Tree Traversal Algorithm. *J. ACM* **1976**, *23*, 368–388. [CrossRef]
19. Fu, K.; Mui, J. A survey on image segmentation. *Pattern Recognit.* **1981**, *13*, 3–16. [CrossRef]
20. Ghaheri, A.; Shoar, S.; Naderan, M.; Hoseini, S.S. The Applications of Genetic Algorithms in Medicine. *Oman Med. J.* **2015**, *30*, 406–416. [CrossRef] [PubMed]
21. Ling, S.H.; Lam, H.K. Evolutionary Algorithms in Health Technologies. *Algorithms* **2019**, *12*, 202. [CrossRef]
22. An, Q.; Rahman, S.; Zhou, J.; Kang, J.J. A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. *Sensors* **2023**, *23*, 4178. [CrossRef] [PubMed]
23. Langarizadeh, M.; Moghbeli, F. Applying Naive Bayesian Networks to Disease Prediction: A Systematic Review. *Acta Inform. Medica* **2016**, *24*, 364. [CrossRef]
24. McLachlan, S.; Dube, K.; Hitman, G.A.; Fenton, N.E.; Kyrimi, E. Bayesian networks in healthcare: Distribution by medical condition. *Artif. Intell. Med.* **2020**, *107*, 101912. [CrossRef]
25. Kyrimi, E.; Dube, K.; Fenton, N.; Fahmi, A.; Neves, M.R.; Marsh, W.; McLachlan, S. Bayesian networks in healthcare: What is preventing their adoption? *Artif. Intell. Med.* **2021**, *116*, 102079. [CrossRef]
26. QIIME2. Available online: https://qiime2.org/ (accessed on 22 August 2024).
27. Nextflow. Available online: https://www.nextflow.org/ (accessed on 22 August 2024).
28. Tian, Y.; Gou, W.; Ma, Y.; Shuai, M.; Liang, X.; Fu, Y.; Zheng, J.S. The Short-Term Variation of Human Gut Mycobiome in Response to Dietary Intervention of Different Macronutrient Distributions. *Nutrients* **2023**, *15*, 2152. [CrossRef] [PubMed]
29. Costantini, C.; Nunzi, E.; Spolzino, A.; Merli, F.; Facchini, L.; Spadea, A.; Melillo, L.; Codeluppi, K.; Marchesi, F.; Marchesini, G.; et al. A High-Risk Profile for Invasive Fungal Infections Is Associated with Altered Nasal Microbiota and Niche Determinants. *Infect. Immun.* **2022**, *90*, e00048-22. [CrossRef]
30. Nunzi, E.; Mezzasoma, L.; Bellezza, I.; Zelante, T.; Orvietani, P.; Coata, G.; Giardina, I.; Sagini, K.; Manni, G.; Di Michele, A.; et al. Microbiota-Associated HAF-EVs Regulate Monocytes by Triggering or Inhibiting Inflammasome Activation. *Int. J. Mol. Sci.* **2023**, *24*, 2527. [CrossRef]
31. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [CrossRef]
32. Langer, B.E.; Amaral, A.; Baudement, M.O.; Bonath, F.; Charles, M.; Chitneedi, P.K.; Clark, E.L.; Di Tommaso, P.; Djebali, S.; Ewels, P.A.; et al. Empowering bioinformatics communities with Nextflow and nf-core. *bioRxiv* **2024**. [CrossRef]
33. Dubey, S.R.; Singh, S.K.; Chaudhuri, B.B. Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *arXiv* **2021**, arXiv:2109.14545. [CrossRef]
34. Benedetti, P.; Femminella, M.; Reali, G. Mixed-Sized Biomedical Image Segmentation Based on U-Net Architectures. *Appl. Sci.* **2022**, *13*, 329. [CrossRef]
35. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
36. Wang, G.; Li, W.; Zuluaga, M.; Aughwane, R.; Patel, P.; Aertsen, M.; Doel, T.; David, A.; Deprest, J.; Ourselin, S.; et al. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Trans. Med. Imaging* **2018**, *37*, 1562–1573. [CrossRef]
37. Wang, G.; Zuluaga, M.A.; Li, W.; Pratt, R.; Patel, P.A.; Aertsen, M.; Doel, T.; David, A.L.; Deprest, J.A.; Ourselin, S.; et al. DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1559–1572. [CrossRef]

38. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
39. Çiçek, O.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv* **2016**, arXiv:1606.06650. [CrossRef]
40. An End-to-End Platform for Machine Learning. 2024. Available online: https://www.tensorflow.org/?hl=en (accessed on 6 September 2024).
41. Pytorch. 2024. Available online: https://pytorch.org/ (accessed on 6 September 2024).
42. Cuda Toolkit. 2024. Available online: https://developer.nvidia.com/cuda-toolkit (accessed on 6 September 2024).
43. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Trans. Neural Netw.* **2009**, *20*, 61–80. [CrossRef]
44. Yan, Y.; He, S.; Yu, Z.; Yuan, J.; Liu, Z.; Chen, Y. Investigation of Customized Medical Decision Algorithms Utilizing Graph Neural Networks. *arXiv* **2024**, arXiv:2405.17460. [CrossRef]
45. Ahmedt-Aristizabal, D.; Armin, M.A.; Denman, S.; Fookes, C.; Petersson, L. Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future. *Sensors* **2021**, *21*, 4758. [CrossRef]
46. Meng, X.; Zou, T. Clinical applications of graph neural networks in computational histopathology: A review. *Comput. Biol. Med.* **2023**, *164*, 107201. [CrossRef] [PubMed]
47. Saxena, R.R.; Saxena, R. Applying Graph Neural Networks in Pharmacology. *TechRxiv* **2024**. [CrossRef]
48. Paul, S.G.; Saha, A.; Hasan, M.Z.; Noori, S.R.H.; Moustafa, A. A Systematic Review of Graph Neural Network in Healthcare-Based Applications: Recent Advances, Trends, and Future Directions. *IEEE Access* **2024**, *12*, 15145–15170. [CrossRef]
49. Zhang, S.; Tong, H.; Xu, J.; Maciejewski, R. Graph convolutional networks: A comprehensive review. *Comput. Soc. Netw.* **2019**, *6*, 11. [CrossRef]
50. Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. *arXiv* **2017**, arXiv:1706.02216. [CrossRef]
51. Cai, C.; Wang, D.; Wang, Y. Graph Coarsening with Neural Networks. *arXiv* **2021**, arXiv:2102.01350. [CrossRef]
52. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2017**, arXiv:1710.10903. [CrossRef]
53. PyTorch Geometric. Available online: https://pytorch-geometric.readthedocs.io/en/latest/ (accessed on 22 August 2024).
54. Deep Graph Library. Available online: https://www.dgl.ai/ (accessed on 22 August 2024).
55. Graph Nets. Available online: https://github.com/google-deepmind/graph_nets (accessed on 22 August 2024).
56. Spektral. Available online: https://graphneural.network/ (accessed on 22 August 2024).
57. StellarGraph. Available online: https://stellargraph.readthedocs.io/en/stable/ (accessed on 22 August 2024).
58. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv* **2015**, arXiv:1506.00019. [CrossRef]
59. Al-Askar, H.; Radi, N.; MacDermott, A. Recurrent Neural Networks in Medical Data Analysis and Classifications. In *Applied Computing in Medicine and Health*; Elsevier: Amsterdam, The Netherlands, 2016; pp. 147–165. [CrossRef]
60. Al Olaimat, M.; Bozdag, S.; for the Alzheimer's Disease Neuroimaging Initiative. TA-RNN: An attention-based time-aware recurrent neural network architecture for electronic health records. *Bioinformatics* **2024**, *40*, i169–i179. [CrossRef]
61. Choi, E.; Schuetz, A.; Stewart, W.F.; Sun, J. Using recurrent neural network models for early detection of heart failure onset. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 361–370. [CrossRef]
62. Riasi, A.; Delrobaei, M.; Salari, M. A decision support system based on recurrent neural networks to predict medication dosage for patients with Parkinson's disease. *Sci. Rep.* **2024**, *14*, 8424. [CrossRef] [PubMed]
63. Liu, L.J.; Ortiz-Soriano, V.; Neyra, J.A.; Chen, J. KIT-LSTM: Knowledge-guided Time-aware LSTM for Continuous Clinical Risk Prediction. *medRxiv* **2022**. [CrossRef]
64. Julie, G.; Jaisakthi, S.M.; Robinson, Y.H. (Eds.) *Handbook of Deep Learning in Biomedical Engineering and Health Informatics*, 1st ed.; Apple Academic Press Inc.: Palm Bay, FL, USA; CRC Press: Boca Raton, FL, USA, 2022; OCLC: 1237707833.
65. Saha, A.; Samaan, M.; Peng, B.; Ning, X. A Multi-Layered GRU Model for COVID-19 Patient Representation and Phenotyping from Large-Scale EHR Data. In Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Houston, TX, USA, 3–6 September 2023; pp. 1–6. [CrossRef]
66. Kiser, A.C.; Eilbeck, K.; Bucher, B.T. Developing an LSTM Model to Identify Surgical Site Infections using Electronic Healthcare Records. *AMIA Summits Transl. Sci. Proc.* **2023**, *2023*, 330–339.
67. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
68. Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078. [CrossRef]
69. Apache MxNet. Available online: https://mxnet.apache.org/versions/1.9.1/ (accessed on 22 August 2024).
70. Microsoft Cognitive Toolkit. Available online: https://cntk.ai (accessed on 22 August 2024).
71. Chainer. Available online: https://chainer.org/ (accessed on 22 August 2024).
72. Rayan, A.; Holyl Alruwaili, S.; Alaerjan, A.S.; Alanazi, S.; Taloba, A.I.; Shahin, O.R.; Salem, M. Utilizing CNN-LSTM techniques for the enhancement of medical systems. *Alex. Eng. J.* **2023**, *72*, 323–338. [CrossRef]

73. Liu, Y.; Song, Z.; Xu, X.; Rafique, W.; Zhang, X.; Shen, J.; Khosravi, M.R.; Qi, L. Bidirectional GRU networks-based next POI category prediction for healthcare. *Int. J. Intell. Syst.* **2022**, *37*, 4020–4040. [CrossRef]
74. Chen, Y.; Esmaeilzadeh, P. Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. *J. Med. Internet Res.* **2024**, *26*, e53008. [CrossRef]
75. Blease, C.; Torous, J.; McMillan, B.; Hägglund, M.; Mandl, K.D. Generative Language Models and Open Notes: Exploring the Promise and Limitations. *JMIR Med. Educ.* **2024**, *10*, e51183. [CrossRef]
76. Biswas, A.; Talukdar, W. Intelligent Clinical Documentation: Harnessing Generative AI for Patient-Centric Clinical Note Generation. *Int. J. Innov. Sci. Res. Technol. (IJISRT)* **2024**, *9*, 994–1008. [CrossRef]
77. Jabbar, A.; Li, X.; Omar, B. A Survey on Generative Adversarial Networks: Variants, Applications, and Training. *arXiv* **2020**, arXiv:2006.05132. [CrossRef]
78. Gonzalez-Abril, L.; Angulo, C.; Ortega, J.A.; Lopez-Guerra, J.L. Generative Adversarial Networks for Anonymized Healthcare of Lung Cancer Patients. *Electronics* **2021**, *10*, 2220. [CrossRef]
79. Vaccari, I.; Orani, V.; Paglialonga, A.; Cambiaso, E.; Mongelli, M. A Generative Adversarial Network (GAN) Technique for Internet of Medical Things Data. *Sensors* **2021**, *21*, 3726. [CrossRef]
80. Abedi, M.; Hempel, L.; Sadeghi, S.; Kirsten, T. GAN-Based Approaches for Generating Structured Data in the Medical Domain. *Appl. Sci.* **2022**, *12*, 7075. [CrossRef]
81. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. *arXiv* **2018**, arXiv:1812.04948. [CrossRef]
82. Singh, A.; Ogunfunmi, T. An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications. *Entropy* **2021**, *24*, 55. [CrossRef]
83. Zemouri, R.; Levesque, M.; Boucher, E.; Kirouac, M.; Lafleur, F.; Bernier, S.; Merkhouf, A. Recent Research and Applications in Variational Autoencoders for Industrial Prognosis and Health Management: A Survey. In Proceedings of the 2022 Prognostics and Health Management Conference (PHM-2022 London), London, UK, 7–29 May 2022; pp. 193–203. [CrossRef]
84. Chen, S.; Guo, W. Auto-Encoders in Deep Learning—A Review with New Perspectives. *Mathematics* **2023**, *11*, 1777. [CrossRef]
85. Bu, Y.; Zou, S.; Liang, Y.; Veeravalli, V.V. Estimation of KL Divergence: Optimal Minimax Rate. *IEEE Trans. Inf. Theory* **2018**, *64*, 2648–2674. [CrossRef]
86. Zhu, F.; Ye, F.; Fu, Y.; Liu, Q.; Shen, B. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. *Sci. Rep.* **2019**, *9*, 6734. [CrossRef] [PubMed]
87. Morid, M.A.; Sheng, O.R.L.; Dunbar, J. Time Series Prediction Using Deep Learning Methods in Healthcare. *ACM Trans. Manag. Inf. Syst.* **2023**, *14*, 1–29. [CrossRef]
88. Liao, Y.; Liu, H.; Spasić, I. Deep learning approaches to automatic radiology report generation: A systematic review. *Inform. Med. Unlocked* **2023**, *39*, 101273. [CrossRef]
89. Hripcsak, G.; Austin, J.H.M.; Alderson, P.O.; Friedman, C. Use of Natural Language Processing to Translate Clinical Information from a Database of 889,921 Chest Radiographic Reports. *Radiology* **2002**, *224*, 157–163. [CrossRef]
90. Hossain, E.; Rana, R.; Higgins, N.; Soar, J.; Barua, P.D.; Pisani, A.R.; Turner, K. Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review. *Comput. Biol. Med.* **2023**, *155*, 106649. [CrossRef]
91. Luo, R.; Sun, L.; Xia, Y.; Qin, T.; Zhang, S.; Poon, H.; Liu, T.Y. BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining. *arXiv* **2022**, arXiv:2210.10341. [CrossRef]
92. Yenduri, G.; Ramalingam, M.; Selvi, G.C.; Supriya, Y.; Srivastava, G.; Maddikunta, P.K.R.; Raj, G.D.; Jhaveri, R.H.; Prabadevi, B.; Wang, W.; et al. GPT (Generative Pre-Trained Transformer)— A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *IEEE Access* **2024**, *12*, 54608–54649. [CrossRef]
93. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762. [CrossRef]
94. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805. [CrossRef]
95. Wang, Y.; Pan, Y.; Yan, M.; Su, Z.; Luan, T.H. A Survey on ChatGPT: AI–Generated Contents, Challenges, and Solutions. *IEEE Open J. Comput. Soc.* **2023**, *4*, 280–302. [CrossRef]
96. Mastropaolo, A.; Scalabrino, S.; Cooper, N.; Nader Palacio, D.; Poshyvanyk, D.; Oliveto, R.; Bavota, G. Studying the Usage of Text-To-Text Transfer Transformer to Support Code-Related Tasks. In Proceedings of the 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), Madrid, Spain, 22–30 May 2021; pp. 336–347. [CrossRef]
97. Chen, X.; Pun, C.M.; Wang, S. MedPrompt: Cross-Modal Prompting for Multi-Task Medical Image Translation. *arXiv* **2023**, arXiv:2310.02663. [CrossRef]
98. Chen, Z.; Cano, A.H.; Romanou, A.; Bonnet, A.; Matoba, K.; Salvi, F.; Pagliardini, M.; Fan, S.; Köpf, A.; Mohtashami, A.; et al. MEDITRON-70B: Scaling Medical Pretraining for Large Language Models. *arXiv* **2023**, arXiv:2311.16079. [CrossRef]
99. Cao, H.; Tan, C.; Gao, Z.; Xu, Y.; Chen, G.; Heng, P.A.; Li, S.Z. A Survey on Generative Diffusion Models. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 2814–2830. [CrossRef]
100. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 10674–10685. [CrossRef]

101. Croitoru, F.A.; Hondru, V.; Ionescu, R.T.; Shah, M. Diffusion Models in Vision: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10850–10869. [CrossRef]
102. Alakhdar, A.; Poczos, B.; Washburn, N. Diffusion Models in De Novo Drug Design. *arXiv* **2024**, arXiv:2406.08511. [CrossRef]
103. Diffusers. Available online: https://huggingface.co/docs/diffusers/index (accessed on 25 August 2024).
104. NVIDIA Clara. Available online: https://docs.nvidia.com/clara/index.html (accessed on 25 August 2024).
105. Cherstvy, A.G.; Thapa, S.; Wagner, C.E.; Metzler, R. Non-Gaussian, non-ergodic, and non-Fickian diffusion of tracers in mucin hydrogels. *Soft Matter* **2019**, *15*, 2526–2551. [CrossRef]
106. Al-Hamadani, M.; Fadhel, M.; Alzubaidi, L.; Harangi, B. Reinforcement Learning Algorithms and Applications in Healthcare and Robotics: A Comprehensive and Systematic Review. *Sensors* **2024**, *24*, 2461. [CrossRef]
107. Kaelbling, L.P.; Littman, M.L.; Moore, A.W. Reinforcement learning: A survey. *J. Artif. Int. Res.* **1996**, *4*, 237–285. [CrossRef]
108. Shakya, A.K.; Pillai, G.; Chakrabarty, S. Reinforcement learning algorithms: A brief survey. *Expert Syst. Appl.* **2023**, *231*, 120495. [CrossRef]
109. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing Atari with Deep Reinforcement Learning. *arXiv* **2013**, arXiv:1312.5602. [CrossRef]
110. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]
111. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347. [CrossRef]
112. Mnih, V.; Badia, A.P.; Mirza, M.; Graves, A.; Lillicrap, T.; Harley, T.; Silver, D.; Kavukcuoglu, K. Asynchronous Methods for Deep Reinforcement Learning. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; Volume 48, pp. 1928–1937.
113. Coronato, A.; Naeem, M.; De Pietro, G.; Paragliola, G. Reinforcement learning for intelligent healthcare applications: A survey. *Artif. Intell. Med.* **2020**, *109*, 101964. [CrossRef]
114. Yu, C.; Liu, J.; Nemati, S. Reinforcement Learning in Healthcare: A Survey. *arXiv* **2019**, arXiv:1908.08796. [CrossRef]
115. Eckardt, J.N.; Wendt, K.; Bornhäuser, M.; Middeke, J.M. Reinforcement Learning for Precision Oncology. *Cancers* **2021**, *13*, 4624. [CrossRef] [PubMed]
116. OpenAI. Gymnasium—An API Standard for Reinforcement Learning with a Diverse Collection of Reference Environments. 2024. Available online: https://gymnasium.farama.org/index.html (accessed on 6 September 2024).
117. Stable Baselines3. Stable-Baselines3 Docs - Reliable Reinforcement Learning Implementations. 2017. Available online: https://stable-baselines3.readthedocs.io/en/master/ (accessed on 14 July 2020).
118. Li, X.; Qian, W.; Xu, D.; Liu, C. Image Segmentation Based on Improved Unet. *J. Phys. Conf. Ser.* **2021**, *1815*, 012018. [CrossRef]
119. Lu, H.; She, Y.; Tie, J.; Xu, S. Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation. *Front. Neuroinform.* **2022**, *16*, 911679. [CrossRef]
120. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
121. Gao, Y.; Huang, R.; Chen, M.; Wang, Z.; Deng, J.; Chen, Y.; Yang, Y.; Zhang, J.; Tao, C.; Li, H. FocusNet: Imbalanced Large and Small Organ Segmentation with an End-to-End Deep Neural Network for Head and Neck CT Images. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019, Proceedings of the 22nd International Conference, Shenzhen, China, 13–17 October 2019*; Proceedings, Part III; Springer: Berlin/Heidelberg, Germany, 2019; pp. 829–838. [CrossRef]
122. Valindria, V.V.; Lavdas, I.; Cerrolaza, J.J.; Aboagye, E.O.; Rockall, A.G.; Rueckert, D.; Glocker, B. Small Organ Segmentation in Whole-body MRI using a Two-stage FCN and Weighting Schemes. In Proceedings of the MLMI@MICCAI, Granada, Spain, 16 September 2018.
123. Saood, A.; Hatem, I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. *BMC Med. Imaging* **2021**, *21*, 19. [CrossRef]
124. Han, G.; Zhang, M.; Wu, W.; He, M.; Liu, K.; Qin, L.; Liu, X. Improved U-Net based insulator image segmentation method based on attention mechanism. *Energy Rep.* **2021**, *7*, 210–217. [CrossRef]
125. Fujima, N.; Sakashita, T.; Homma, A.; Harada, T.; Shimizu, Y.; Tha, K.K.; Kudo, K.; Shirato, H. Non-invasive prediction of the tumor growth rate using advanced diffusion models in head and neck squamous cell carcinoma patients. *Oncotarget* **2017**, *8*, 33631–33643. [CrossRef] [PubMed]
126. Wang, K.; Ren, Y.; Ma, L.; Fan, Y.; Yang, Z.; Yang, Q.; Shi, J.; Sun, Y. Deep learning-based prediction of treatment prognosis from nasal polyp histology slides. *Int. Forum Allergy Rhinol.* **2023**, *13*, 886–898. [CrossRef] [PubMed]
127. Li, Q.; Chen, L.; Cai, Y.; Wu, D. Hierarchical Graph Neural Network for Patient Treatment Preference Prediction with External Knowledge. In *Advances in Knowledge Discovery and Data Mining*; Kashima, H., Ide, T., Peng, W.C., Eds.; Series Title: Lecture Notes in Computer Science; Springer Nature: Cham, Switzerland, 2023; Volume 13937, pp. 204–215. [CrossRef]
128. He, H.; Zhao, S.; Xi, Y.; Ho, J.C. MedDiff: Generating Electronic Health Records using Accelerated Denoising Diffusion Model. *arXiv* **2023**, arXiv:2302.04355. [CrossRef]

129. Ashton, J.J.; Young, A.; Johnson, M.J.; Beattie, R.M. Using machine learning to impact on long-term clinical care: Principles, challenges, and practicalities. *Pediatr. Res.* **2023**, *93*, 324–333. [CrossRef] [PubMed]
130. Proios, D.; Yazdani, A.; Bornet, A.; Ehrsam, J.; Rekik, I.; Teodoro, D. Leveraging patient similarities via graph neural networks to predict phenotypes from temporal data. In Proceedings of the 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), Thessaloniki, Greece, 9–13 October 2023; pp. 1–10. [CrossRef]
131. Whiles, B.B.; Bird, V.G.; Canales, B.K.; DiBianco, J.M.; Terry, R.S. Caution! AI Bot Has Entered the Patient Chat: ChatGPT Has Limitations in Providing Accurate Urologic Healthcare Advice. *Urology* **2023**, *180*, 278–284. [CrossRef] [PubMed]
132. Ceritli, T.; Ghosheh, G.O.; Chauhan, V.K.; Zhu, T.; Creagh, A.P.; Clifton, D.A. Synthesizing Mixed-type Electronic Health Records using Diffusion Models. *arXiv* **2023**. [CrossRef]
133. Liu, S.; See, K.C.; Ngiam, K.Y.; Celi, L.A.; Sun, X.; Feng, M. Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review. *J. Med. Internet Res.* **2020**, *22*, e18477. [CrossRef]
134. Wu, W.; Zhang, J.; Xie, H.; Zhao, Y.; Zhang, S.; Gu, L. Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint. *Comput. Biol. Med.* **2020**, *118*, 103657. [CrossRef]
135. Jun, T.J.; Kweon, J.; Kim, Y.H.; Kim, D. T-Net: Nested encoder–decoder architecture for the main vessel segmentation in coronary angiography. *Neural Netw.* **2020**, *128*, 216–233. [CrossRef]
136. Moon, S.; Zhung, W.; Yang, S.; Lim, J.; Kim, W.Y. PIGNet: A physics-informed deep learning model toward generalized drug–target interaction predictions. *Chem. Sci.* **2022**, *13*, 3661–3673. [CrossRef]
137. Kazerouni, A.; Aghdam, E.K.; Heidari, M.; Azad, R.; Fayyaz, M.; Hacihaliloglu, I.; Merhof, D. Diffusion models in medical imaging: A comprehensive survey. *Med. Image Anal.* **2023**, *88*, 102846. [CrossRef]
138. Hung, A.L.Y.; Zhao, K.; Zheng, H.; Yan, R.; Raman, S.S.; Terzopoulos, D.; Sung, K. Med-cDiff: Conditional Medical Image Generation with Diffusion Models. *Bioengineering* **2023**, *10*, 1258. [CrossRef]
139. Malik, A.; Solaiman, B. AI in hospital administration and management: Ethical and legal implications. In *Research Handbook on Health, AI and the Law*; Solaiman, B., Cohen, I.G., Eds.; Edward Elgar Publishing: Cheltenham, UK, 2024; pp. 21–40. [CrossRef]
140. Pham, T.; Tao, X.; Zhang, J.; Yong, J.; Zhou, X.; Gururajan, R. MeKG: Building a Medical Knowledge Graph by Data Mining from MEDLINE. In *Brain Informatics*; Liang, P., Goel, V., Shan, C., Eds.; Series Title: Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; Volume 11976, pp. 159–168. [CrossRef]
141. Chen, Z.; Dou, Y.; De Causmaecker, P. Neural networked-assisted method for the nurse rostering problem. *Comput. Ind. Eng.* **2022**, *171*, 108430. [CrossRef]
142. Sai, S.; Gaur, A.; Sai, R.; Chamola, V.; Guizani, M.; Rodrigues, J.J.P.C. Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies, and Limitations. *IEEE Access* **2024**, *12*, 31078–31106. [CrossRef]
143. Wang, C.; Li, X.; Ma, W.; Wang, X. Diffusion models over the life cycle of an innovation: A bottom-up and top-down synthesis approach. *Public Adm. Dev.* **2020**, *40*, 105–118. [CrossRef]
144. Yu, C.; Liu, J.; Nemati, S.; Yin, G. Reinforcement Learning in Healthcare: A Survey. *ACM Comput. Surv.* **2023**, *55*, 1–36. [CrossRef]
145. Kim, D.; Lee, J.; Woo, Y.; Jeong, J.; Kim, C.; Kim, D.K. Deep Learning Application to Clinical Decision Support System in Sleep Stage Classification. *J. Pers. Med.* **2022**, *12*, 136. [CrossRef]
146. Diaz Ochoa, J.G.; Mustafa, F.E. Graph neural network modelling as a potentially effective method for predicting and analyzing procedures based on patients' diagnoses. *Artif. Intell. Med.* **2022**, *131*, 102359. [CrossRef] [PubMed]
147. Das Swain, V.; Saha, K. Teacher, Trainer, Counsel, Spy: How Generative AI can Bridge or Widen the Gaps in Worker-Centric Digital Phenotyping of Wellbeing. In Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work, Newcastle upon Tyne, UK, 25–27 June 2024; pp. 1–13. [CrossRef]
148. Liu, L.; Zhao, K. Report on Methods and Applications for Crafting 3D Humans. *arXiv* **2024**, arXiv:2406.01223. [CrossRef]
149. Pike, A.C.; Robinson, O.J. Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry* **2022**, *79*, 313. [CrossRef]
150. Wang, X.; Abubaker, S.M.; Babalola, G.T.; Tulk Jesso, S. Co-Designing an AI Chatbot to Improve Patient Experience in the Hospital: A human-centered design case study of a collaboration between a hospital, a university, and ChatGPT. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–10. [CrossRef]
151. Qu, J.; Zhang, Y.; Tang, W.; Cheng, W.; Zhang, Y.; Bu, L. Developing a virtual reality healthcare product based on data-driven concepts: A case study. *Adv. Eng. Inform.* **2023**, *57*, 102118. [CrossRef]
152. Balcombe, L.; De Leo, D. Human-Computer Interaction in Digital Mental Health. *Informatics* **2022**, *9*, 14. [CrossRef]
153. Priya, K.V.; Dinesh Peter, J. Enhanced Defensive Model Using CNN against Adversarial Attacks for Medical Education through Human Computer Interaction. *Int. J. Hum.—Comput. Interact.* **2023**, 1–13. [CrossRef]
154. Slegers, K. Human–Computer Interaction. In *The International Encyclopedia of Health Communication*, 1st ed.; Ho, E.Y., Bylund, C.L., Van Weert, J.C.M., Eds.; Wiley: Hoboken, NJ, USA, 2022; pp. 1–7. [CrossRef]
155. Moini, J.; Akinso, O.; Ferdowsi, K.; Moini, M. The role of computers in health care. In *Health Care Today in the United States*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 485–498. [CrossRef]

156. Reddy, S. Explainability and artificial intelligence in medicine. *Lancet Digit. Health* **2022**, *4*, e214–e215. [CrossRef] [PubMed]
157. Beger, J. The crucial role of explainability in healthcare AI. *Eur. J. Radiol.* **2024**, *176*, 111507. [CrossRef]
158. The Precise4Q consortium; Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V.I. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 310. [CrossRef]
159. Chaddad, A.; Peng, J.; Xu, J.; Bouridane, A. Survey of Explainable AI Techniques in Healthcare. *Sensors* **2023**, *23*, 634. [CrossRef] [PubMed]
160. Sadeghi, Z.; Alizadehsani, R.; Cifci, M.A.; Kausar, S.; Rehman, R.; Mahanta, P.; Bora, P.K.; Almasri, A.; Alkhawaldeh, R.S.; Hussain, S.; et al. A review of Explainable Artificial Intelligence in healthcare. *Comput. Electr. Eng.* **2024**, *118*, 109370. [CrossRef]
161. Frasca, M.; La Torre, D.; Pravettoni, G.; Cutica, I. Explainable and interpretable artificial intelligence in medicine: A systematic bibliometric review. *Discov. Artif. Intell.* **2024**, *4*, 15. [CrossRef]
162. Shajari, S.; Kuruvinashetti, K.; Komeili, A.; Sundararaj, U. The Emergence of AI-Based Wearable Sensors for Digital Health Technology: A Review. *Sensors* **2023**, *23*, 9498. [CrossRef]
163. Wang, C.; He, T.; Zhou, H.; Zhang, Z.; Lee, C. Artificial intelligence enhanced sensors - enabling technologies to next-generation healthcare and biomedical platform. *Bioelectron. Med.* **2023**, *9*, 17. [CrossRef]
164. Katsoulakis, E.; Wang, Q.; Wu, H.; Shahriyari, L.; Fletcher, R.; Liu, J.; Achenie, L.; Liu, H.; Jackson, P.; Xiao, Y.; et al. Digital twins for health: A scoping review. *NPJ Digit. Med.* **2024**, *7*, 77. [CrossRef]
165. Johnson, Z.; Saikia, M.J. Digital Twins for Healthcare Using Wearables. *Bioengineering* **2024**, *11*, 606. [CrossRef] [PubMed]
166. Keshta, I. AI-driven IoT for smart health care: Security and privacy issues. *Inform. Med. Unlocked* **2022**, *30*, 100903. [CrossRef]
167. Wang, C.; Zhang, J.; Lassi, N.; Zhang, X. Privacy Protection in Using Artificial Intelligence for Healthcare: Chinese Regulation in Comparative Perspective. *Healthcare* **2022**, *10*, 1878. [CrossRef]
168. Selvanambi, R.; Bhutani, S.; Veauli, K. Security and Privacy for Electronic Healthcare Records Using AI in Blockchain. In *Research Anthology on Convergence of Blockchain, Internet of Things, and Security*; IGI Global: Hershey, PA, USA, 2022; pp. 768–777. [CrossRef]
169. Taherdoost, H. Privacy and Security of Blockchain in Healthcare: Applications, Challenges, and Future Perspectives. *Sci* **2023**, *5*, 41. [CrossRef]
170. Ahad, A.; Ali, Z.; Mateen, A.; Tahir, M.; Hannan, A.; Garcia, N.M.; Pires, I.M. A Comprehensive review on 5G-based Smart Healthcare Network Security: Taxonomy, Issues, Solutions and Future research directions. *Array* **2023**, *18*, 100290. [CrossRef]
171. Abir, S.M.A.A.; Abuibaid, M.; Huang, J.S.; Hong, Y. Harnessing 5G Networks for Health Care: Challenges and Potential Applications. In Proceedings of the 2023 International Conference on Smart Applications, Communications and Networking (SmartNets), Istanbul, Turkiye, 25–27 July 2023; pp. 1–6. [CrossRef]
172. Pradhan, B.; Das, S.; Roy, D.S.; Routray, S.; Benedetto, F.; Jhaveri, R.H. An AI-Assisted Smart Healthcare System Using 5G Communication. *IEEE Access* **2023**, *11*, 108339–108355. [CrossRef]
173. Punugoti, R.; Dutt, V.; Anand, A.; Bhati, N. Exploring the Impact of Edge Intelligence and IoT on Healthcare: A Comprehensive Survey. In Proceedings of the 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 14–16 June 2023; pp. 1108–1114. [CrossRef]
174. Izhar, M.; Naqvi, S.A.A.; Ahmed, A.; Abdullah, S.; Alturki, N.; Jamel, L. Enhancing Healthcare Efficacy Through IoT-Edge Fusion: A Novel Approach for Smart Health Monitoring and Diagnosis. *IEEE Access* **2023**, *11*, 136456–136467. [CrossRef]
175. Alekseeva, D.; Ometov, A.; Lohan, E.S. Towards the Advanced Data Processing for Medical Applications Using Task Offloading Strategy. In Proceedings of the 2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Thessaloniki, Greece, 10–12 October 2022; pp. 51–56. [CrossRef]
176. Ur Rasool, R.; Ahmad, H.F.; Rafique, W.; Qayyum, A.; Qadir, J.; Anwar, Z. Quantum Computing for Healthcare: A Review. *Future Internet* **2023**, *15*, 94. [CrossRef]
177. Shuford, J. Exploring Ethical Dimensions in AI: Navigating Bias and Fairness in the Field. *J. Artif. Intell. Gen. Sci. (JAIGS)* **2024**, *3*, 103–124. ISSN: 3006-4023. [CrossRef]
178. Goh, E.; Bunning, B.; Khoong, E.; Gallo, R.; Milstein, A.; Centola, D.; Chen, J.H. ChatGPT Influence on Medical Decision-Making, Bias, and Equity: A Randomized Study of Clinicians Evaluating Clinical Vignettes. *medRxiv* **2023**. [CrossRef]
179. Capraro, V.; Lentsch, A.; Acemoglu, D.; Akgun, S.; Akhmedova, A.; Bilancini, E.; Bonnefon, J.F.; Brañas-Garza, P.; Butera, L.; Douglas, K.M.; et al. The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *arXiv* **2024**, arXiv:2401.05377. [CrossRef]
180. Alzubaidi, L.; Al-Sabaawi, A.; Bai, J.; Dukhan, A.; Alkenani, A.H.; Al-Asadi, A.; Alwzwazy, H.A.; Manoufali, M.; Fadhel, M.A.; Albahri, A.S.; et al. Towards Risk-Free Trustworthy Artificial Intelligence: Significance and Requirements. *Int. J. Intell. Syst.* **2023**, *2023*, 1–41. [CrossRef]
181. Winter, J.S.; Davidson, E. Governance of artificial intelligence and personal health information. *Digit. Policy Regul. Gov.* **2019**, *21*, 280–290. [CrossRef]
182. World Health Organization. *Ethics and Governance of Artificial INTELLIGENCE for Health*; World Health Organization: Geneva, Switzerland, 2024.
183. Zhou, K.; Gattinger, G. The Evolving Regulatory Paradigm of AI in MedTech: A Review of Perspectives and Where We Are Today. *Ther. Innov. Regul. Sci.* **2024**, *58*, 456–464. [CrossRef] [PubMed]

184. Chakraborty, A.; Karhade, M. Global AI Governance in Healthcare: A Cross-Jurisdictional Regulatory Analysis. *arXiv* **2024**, arXiv:2406.08695. [CrossRef]
185. Bouderhem, R. Shaping the future of AI in healthcare through ethics and governance. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 416. [CrossRef]