Advances in the Development of Representation Learning and Its Innovations against COVID-19

Peng Li [1,2], **Mosharaf Md Parvej** [1,2], **Chenghao Zhang** [1,2], **Shufang Guo** [1,2] **and Jing Zhang** [1,2,*]

[1] School of Information Science and Engineering, University of Jinan, Jinan 250024, China; pengli@stu.ujn.edu.cn (P.L.); parvej.mosharaf@outlook.com (M.M.P.); dpl_zch@stu.ujn.edu.cn (C.Z.); 202221100443@stu.ujn.edu.cn (S.G.)

[2] Shandong Provincial Key Laboratory of Network-Based Intelligent Computing, University of Jinan, Jinan 250024, China

[*] Correspondence: ise_zhangjing@ujn.edu.cn

Abstract: In bioinformatics research, traditional machine-learning methods have demonstrated efficacy in addressing Euclidean data. However, real-world data often encompass non-Euclidean forms, such as graph data, which contain intricate structural patterns or high-order relationships that elude conventional machine-learning approaches. Representation learning seeks to derive valuable data representations from enhancing predictive or analytic tasks, capturing vital patterns and structures. This method has proven particularly beneficial in bioinformatics and biomedicine, as it effectively handles high-dimensional and sparse data, detects complex biological patterns, and optimizes predictive performance. In recent years, graph representation learning has become a popular research topic. It involves the embedding of graphs into a low-dimensional space while preserving the structural and attribute information of the graph, enabling better feature extraction for downstream tasks. This study extensively reviews representation learning advancements, particularly in the research of representation methods since the emergence of COVID-19. We begin with an analysis and classification of neural-network-based language model representation learning techniques as well as graph representation learning methods. Subsequently, we explore their methodological innovations in the context of COVID-19, with a focus on the domains of drugs, public health, and healthcare. Furthermore, we discuss the challenges and opportunities associated with graph representation learning. This comprehensive review presents invaluable insights for researchers as it documents the development of COVID-19 and offers experiential lessons to preempt future infectious diseases. Moreover, this study provides guidance regarding future bioinformatics and biomedicine research methodologies.

Keywords: representation learning; graph embedding; graph neural network; deep learning; COVID-19; healthcare

1. Introduction

Following the emergence of COVID-19 at the end of 2019, the virus went through an 11-month phase of relative evolutionary stagnation [1]. Since the end of 2020, COVID-19 has experienced approximately two mutations per month, causing the emergence of new variants of concern (VOCs). Various organizations worldwide have sequenced the COVID-19 genome, finding many mutant variants of the virus. The phylogenetic tree depicted in Figure 1 [2] demonstrates the evolutionary relationships of COVID-19 in time and space. Due to low vaccination rates in many countries, the classical herd immunization goal of eradicating or eliminating COVID-19 has not been achieved [3]. While COVID-19 is unlikely to cause death at present, elderly people with cancer or underlying metabolic illnesses are at a higher risk of serious illness and death [4–6]. The mechanisms underlying this disease still remain unknown. The virus is widely distributed in humans and other mammals [7], as confirmed by a study utilizing machine learning unsupervised clustering methods [8]. Figure 2 demonstrates the structure of the SARS-CoV-2 [9]. A total of four significant

structural proteins are encoded by the SARS-CoV-2 genome: spike (S) protein, envelope (E) protein, membrane (M) glycoprotein, and nucleocapsid (N) protein, in addition to 16 nonstructural proteins and five to eight accessory proteins [10]. According to recent research, the S protein has undergone notable mutations in its amino acid sequence since 2020, which have resulted in changes to its distinct functions [11]. The D614G mutation (amino acid change from aspartate to glycine at position 614) has been found to increase viral infectivity [12], while other mutations, such as K417N, G446S, E484A, Q493R, and N440K, lead to immune escape [13]. In addition, N501Y enhances the binding of the S protein to ACE2, N679K increases viral transmission, and P681H enhances the binding affinity of the S protein [14]. Natural selection can drive the emergence of advantageous mutations in COVID-19 that confer higher fitness benefits, such as pathogenicity, infectivity, transmissibility, ACE2 binding affinity, and antigenicity [15]. Another variant of concern in the ongoing evolution of COVID-19 is XBB, a mutant strain of BA.2 whose prevalence is increasing globally [16]. The availability of publicly accessible and diverse COVID-19 genome sequence data is essential for analyzing the virus's evolutionary trends, which, in turn, can inform drug discovery, vaccine development, and treatment guidance.

Representation learning, a subfield of machine learning, has gained significant attention in various domains owing to its ability to extract meaningful high-level representations from raw data. In the fields of bioinformatics and biomedicine, representation learning techniques have been widely employed to address complex problems and uncover hidden patterns in biological and biomedical data. In representation learning, neural-network-based methods capture intricate relationships and hierarchical structures in data, performing nonlinear transformations on high-dimensional data to solve classification and regression problems [17]. Deep learning models are a prevalent neural network architecture for representation learning, and have been successfully applied to tasks such as gene expression analysis [18], drug discovery [19], and disease diagnosis [20]. Significant advancements in representation learning have been observed in recent years, with its application in bioinformatics and biomedicine growing rapidly. Graph representation learning techniques, in particular, have advanced rapidly and excel at handling unstructured data [21]. These techniques have enabled breakthroughs in areas such as drug design [22], protein interaction [23], and disease prediction [24]. Graphs are powerful structures for modeling complex relationships between entities. In bioinformatics and biomedical research, biological networks such as gene regulatory networks and disease networks provide valuable insights into the functionality of biological systems and the mechanisms of diseases [25]. Methods based on graph representation offer effective means of utilizing this network information for representation learning. Among them, graph embedding is a method of mapping nodes, links, and subgraphs in a graph to low-dimensional representations that capture the nodes' topological structure and intrinsic properties [26]. When solving downstream graph tasks like node classification, link prediction, and visualization, these low-dimensional representations can be used as inputs [27]. However, graph neural networks (GNNs) update the node representations using message aggregation and propagation, integrating contextual information of the nodes. This is typically end-to-end and can directly learn low-dimensional representations of nodes and links from the raw graph data without relying on handcrafted features [28,29]. GNNs have emerged as a promising approach for representation learning on graph-structured data. GNNs can capture the structural and topological properties of biological networks, enabling them to learn representations of both nodes and the entire graph. By propagating information along graph edges and utilizing graph convolutional operations, GNNs can extract hidden features and patterns from biological networks [30]. GNNs have a broad range of applications in bioinformatics and biomedicine, including drug discovery [22], prediction of drug–target interactions [31], prioritization of disease-associated genes [32], and prediction of molecular properties [33]. By integrating network information and leveraging the power of graph representation learning, GNNs have demonstrated significant performance in these tasks, contributing to the discovery of new biological insights and potential therapeutic targets.

According to the World Health Organization, COVID-19 is no longer a public health emergency of global significance [34]. However, experts maintain that it remains an ongoing health threat. The COVID-19 pandemic has provided valuable insights and lessons for humanity. In order to effectively mitigate the impact of potential future outbreaks, it is crucial to make significant advancements in technological capabilities, bolster global vaccine research and production capacities, and optimize the speed and efficiency of our response to emerging viruses. Despite the urgent need for these developments, the complexity and diversity of COVID-19 data make it challenging to handle. Furthermore, the rapid progress of representation learning, particularly graph representation learning, requires a comprehensive and systematic summary of its application in COVID-19.

In this work, a comprehensive review of the advancements in representation learning methods is provided. Firstly, the related methods of neural-network-based language model representation learning are introduced. This is followed by a detailed summary and discussion of graph representation methods, particularly graph neural network-related methods. The representative applications of representation learning in COVID-19 are also introduced. Furthermore, the challenges and opportunities of graph representation methods in the fields of bioinformatics and biomedicine are discussed. This work comprehensively reviews the research on representation learning methods since the emergence of COVID-19 and presents future applications in bioinformatics and biomedicine. It is worth noting that the methods based on neural network models are specifically summarized in this article. An overview of relevant techniques for graph representation learning is the main goal of this article. Therefore, we only introduce a few traditional models in the section on "Neural-network-based language model representation learning". GNNs unquestionably belong under the category of deep learning techniques, and irrespective of certain concepts that overlap with other parts, we still include it individually for discussion.
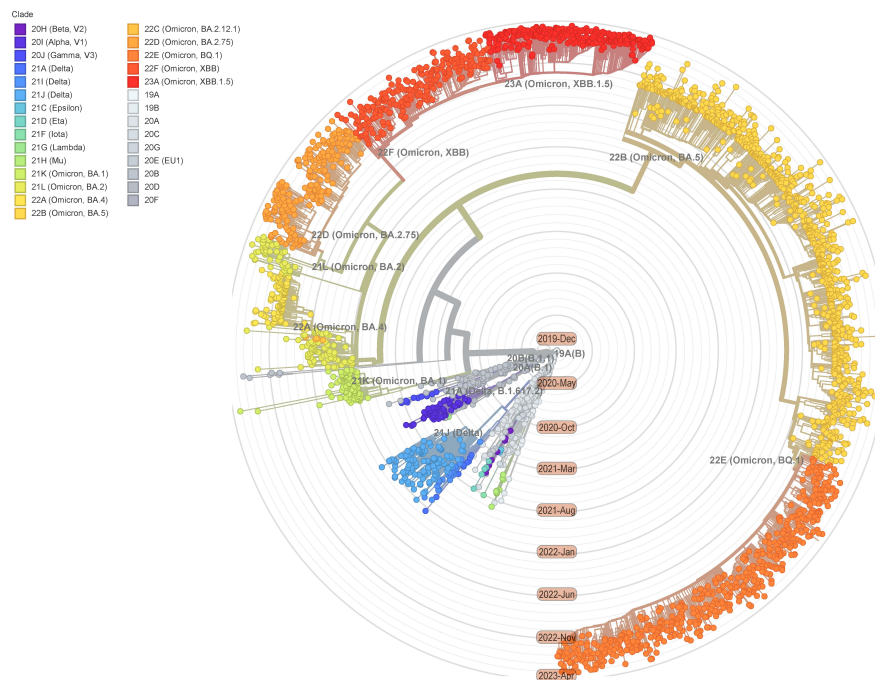

Figure 1. Modified on a phylogenetic tree built on Nextstarin (27 March 2023) [2], consisting of 2775 SARS-CoV-2 genomes from the GISAID database evolving in chronological order, with bright colors representing the most noteworthy variants of the moment, such as red dots.
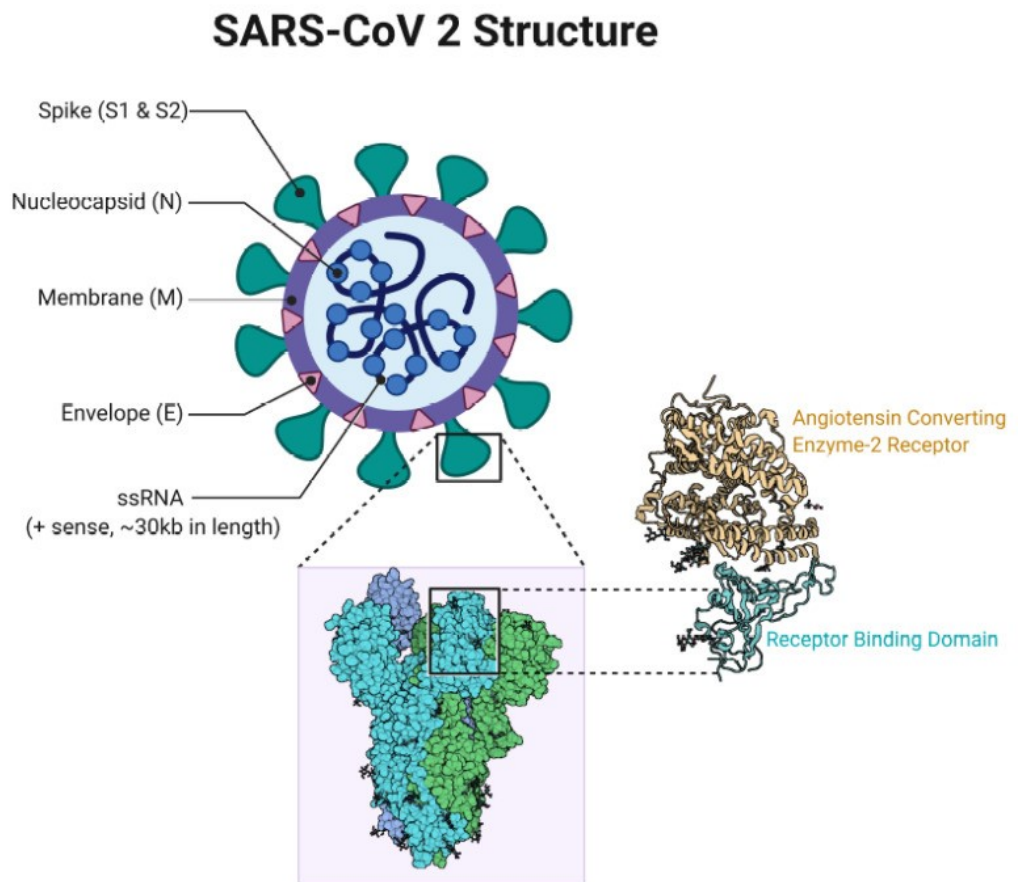
Figure 2. SARS-CoV-2 Structure [9].

2. Representation Learning

Representation learning has facilitated many discoveries in the fields of bioinformatics and biomedicine, ranging from disease interactions to drug discovery. Traditional feature engineering methods require manually preset features to be input into machine learning models, and crafting optimal predictive features for complex networks is a daunting task. In contrast, representation learning learns network features in an automated manner. For instance, network embedding is a common representation learning method that maps nodes or edges in the network to a low-dimensional vector space, capturing the topological structure and node relationships of the network. This method does not require predetermined features, but generates features by learning the structural information of the network.

In addition, deep learning has provided new possibilities for representation learning. The emergence of graph representation learning has laid the foundation for processing and understanding graphical data. It generates useful features by automatically learning hidden patterns in the structure of graphs, thereby being able to directly handle complex graphical data. GNNs have become a leading method for graph representation learning. They can automatically learn complex patterns in networks and generate high-quality feature representations. When dealing with biological data, such as gene expression networks and protein interaction networks, GNNs can capture complex relationships, automatically learn useful features, and eliminate the need for laborious manual feature engineering.

In the aforementioned methods, we can utilize some basic concepts of graph theory. For instance, biological networks can be viewed as a graph $G = (V, E)$, where V is the set of nodes representing genes or proteins, and E is the set of edges representing interactions between genes or proteins. These networks are typically heterogeneous as they contain multiple types of nodes and edges, each type representing different information. For

example, nodes can represent genes, edges can represent interactions between genes, and attributes of nodes and edges can represent gene expression levels or interaction strengths. They can handle graphs with multiple types of nodes and edges, enabling them to capture rich information in the network. Furthermore, GNNs can also handle dynamic graphs, where nodes and edges can be added, deleted, or modified over time. This is particularly useful for handling biological networks that change over time, such as time-varying gene expression networks.

GNNs work by applying neural networks on the nodes of the graph to learn the vector representation of the nodes. These representations can capture the characteristics of the nodes (through the feature vector x_i of the nodes), as well as the position of the nodes in the graph (through the adjacency matrix A). In this way, GNNs can learn higher-order neighborhood information of the nodes, which is achieved by considering both the direct neighbors (first-order neighborhood) and more distant neighbors (higher-order neighborhoods) of the nodes. This can be represented by the higher-order adjacency matrix $M = \hat{E} + \hat{E}^2 + \cdots + \hat{E}^t$, where $\hat{E}$ is the weight matrix of the edges, and t is the maximum order of the neighborhood considered. Furthermore, GNNs can also utilize meta-paths to capture complex relationships between different types of nodes and edges in heterogeneous graphs. A meta-path is a tool that can describe complex relationships between nodes, revealing relationships between different types of nodes in biological systems, thereby providing a deeper understanding of the complexity of biological systems. A meta-path can be represented as $P = v_1^{(1)} \xrightarrow{e_1} v_2^{(1)} \xrightarrow{e_2} v_3^{(2)} \xrightarrow{e_3} ... \xrightarrow{e_{l-1}} v_l^{(k-1)} \xrightarrow{e_l} v_{l+1}^{(k)}$, where $v_i^{(t)} \in V_t$ indicates that the ith node belongs to the tth type of node, and $e_i \in E$ represents the ith type of edge. When learning node representations, GNNs consider not only the topological structure of the graph but also the semantic similarity of the nodes. This is achieved by measuring the similarity or correlation of the node vector representations in a low-dimensional vector space. This semantic similarity can be measured based on the labels of the nodes, the edges between the nodes, or other characteristics.

In precision medicine, the application of GNNs has also undergone significant progress. For instance, the study in [35] was the first to use GNNs to explore drug response prediction based on gene constraint multi-omics integration. Another study [36], utilized deep feature fusion graph convolutional neural network for cerebral arteriovenous malformations. These studies demonstrate that GNNs have tremendous potential in handling complex biological networks, predicting new gene functions, discovering potential drug targets, and understanding disease mechanisms. Although deep learning models are often viewed as "black boxes", the design of GNNs allows us to understand how the model works. For example, we can understand how the model comprehends the characteristics and positions of nodes by examining the vector representations of nodes. This is particularly important for precision medicine research, as it can help us understand the complexity of biological systems and provide powerful tools for the prevention and treatment of diseases.

3. Overview of Representation Learning Methods

In this section, we provide a brief overview of representation learning methods. Representation learning aims to uncover the underlying representations of raw data and project high-dimensional data onto a lower-dimensional space. This space captures the fundamental characteristics of the raw data while retaining the important information that better supports downstream tasks [37,38]. Representation learning can aid researchers in extracting meaningful information and features to attain a more comprehensive representation of biological data. On this basis, we provided a summary of neural-network-based language model representation learning methods and graph representation learning methods and extended other techniques. All the methods reviewed in this section are summarized in Figure 3.
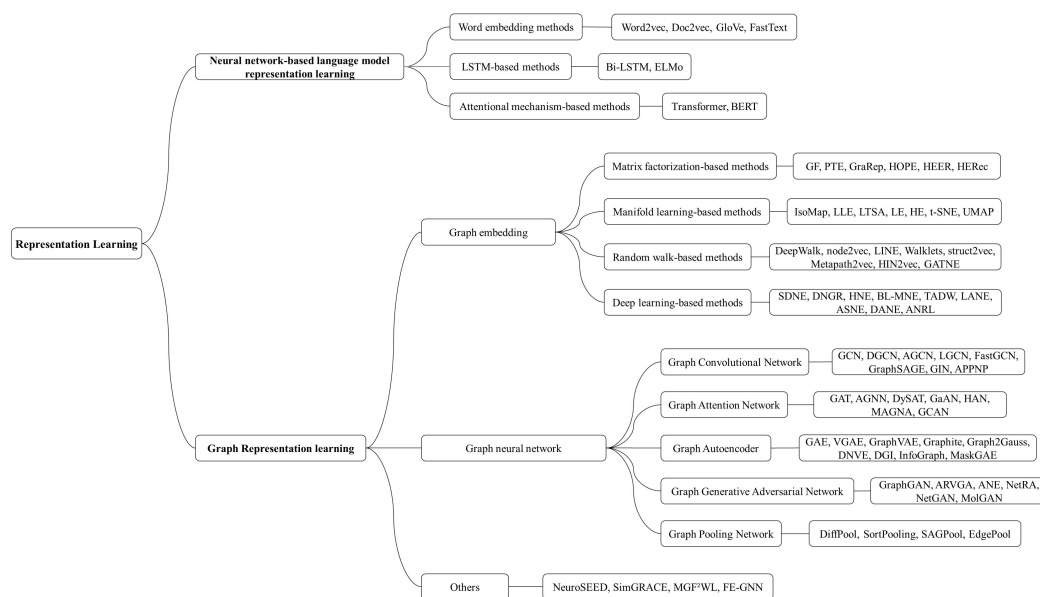
Figure 3. Comprehensive summary of the classification and representation methods of representation learning.

3.1. Neural-Network-Based Language Model Representation Learning

Neural-network-based language model representation learning methods use distributed word vector representations learned through neural networks, especially context-dependent word vectors, which can effectively encode the semantic and syntactic information of words, providing useful language understanding capabilities for downstream natural language processing (NLP) tasks. In this section, we categorize all methods into three major categories: word embedding methods, long short-term memory (LSTM)-based methods and attention mechanism-based methods.

Word embedding methods. A neural network-based method called word2vec [39] is used for natural language processing tasks such as sentiment analysis, language modeling, and text classification. Word2vec uses a neural network to learn a high-dimensional vector representation of each word in the corpus, with similar vectors used to represent words with similar meanings. The method uses a neural network architecture that consists of two learning models: a continuous bag-of-words (CBOW) model that predicts target words using contextual words and a skip-gram model that predicts target words using target words. As a word2vec extension, doc2vec [40] can take sentences of various lengths as training samples. Distributed memory (DM) and distributed bag-of-words are the two models that doc2vec uses to learn paragraph and document embeddings. The DM model predicts words by using vectors associated with passages as well as contextual words. The DBOW model ignores contextual words and predicts words in paragraphs directly using vectors associated with paragraphs. Global vectors for word representation (GloVe) [41] is a word representation tool based on global word frequency statistics that aims to capture semantic relationships between words through global word-word co-occurrence statistics information. The algorithm first creates a word–word co-occurrence matrix from the corpus, which is used to record the number of times each word appears in the context of other words. The matrix is then decomposed to obtain a low-dimensional representation of the words while retaining the co-occurrence statistics information. GloVe is similar to word2vec in that they both learn word embeddings from co-occurrence statistics, but it employs a different approach to capture word meaning. FastText [42] is used to learn vector representations of words and phrases. Unlike word2vec, which operates at the word level, FastText operates at the character n-gram level. By providing embeddings, FastText can represent words outside the vocabulary as the sum of their character n-grams. This enables

FastText to perform better when dealing with uncommon or unknown words. Furthermore, FastText can infer the meaning of unseen words from their subwords.

The ability of word2vec to capture semantic features represents a significant advancement in the field of representation learning, although it is unable to represent the semantic polysemy of words. In biological sequences, the position of a word in a sentence is equivalent to a specific amino acid or a string containing k bases divided by k-mer. For instance, the same amino acid can be encoded by various codons in a DNA sequence, and the same amino acid residue can have different interpretations depending on the context. Therefore, polysemy in biological sequences is significant. Here, we have introduced two methods to deal with contextual associations: one method makes use of LSTM [43], and the other method does so by using the attentional mechanism.

LSTM-based methods. Bidirectional long and short-term memory (Bi-LSTM) [44] is made up of two independent LSTM modules, one of which handles forward sequences and the other reverse sequences. The parameters of the two LSTM modules are independent of one another, and they only share the word embeddings of the input sequences. Bi-LSTM can capture long-term dependencies between elements in a sequence and is frequently used to model contextual information. In Bi-LSTM, the hidden state of the subsequent time step is also a factor that affects the output of each time step in addition to the current time step and the hidden state of the previous time step. This makes it possible to take into account how the forward and reverse information in the sequence depend on one another and better capture contextual information within the sequence. Embeddings from language models (ELMo) [45] is a pre-trained language model for creating word vectors that add a dynamic weight of the pre-trained language model to the output of Bi-LSTM. When using its word vector for a word, the entire text should be input, and the word vector is dynamically generated based on the entirety of contextual information, resulting in different word vector values for the same word in various contexts. To obtain richer feature information, ELMo uses not only the contextual information in Bi-LSTM but also the contextual information in the pre-trained language model.

Attention-mechanism-based methods. The attention mechanism quantifies the degree of dependency between words and obtains a weighted representation by calculating the weights of each input position and weighting the inputs to sum [46,47]. The forward-propagation-based model has a strong representation learning capability, which can naturally introduce attention weights into the model, dynamically calculate the attention weights, and apply the weights to various parts of the input to obtain a more accurate representation. Instead of relying on traditional context-aware architectures when processing sequence data, Transformer [48] achieves a significant improvement in model performance by implementing the attention mechanism and positional encoding [49] via key-value memory neural networks [50,51]. Bidirectional Encoder Representations from Transformers (BERT) [52] is a model with multiple stacks of transformers. The transformer model is further enhanced by the introduction of two new mechanisms, masked language modeling (MLM) and next sentence prediction (NSP). In the pre-training of BERT, [CLS] and [SEP] tokens are added to the start and end of each input sentence, respectively, and then long sequences are spliced together. In the MLM mechanism, some words are masked at random, and the model is asked to predict the words that have been masked to train the model's capacity to model contextual information. In the NSP mechanism, BERT takes two sentences as input and asks the model to determine whether two sentences are adjacent to each other to train the model's ability to model sentence relationships. BERT is more effective at capturing distal word associations than conventional recursive models such as RNN and LSTM [46,53]. Due to the implicit layer state of each word in the recursive model being determined solely by the implicit layer state of neighboring words, the contribution of other words distant from the word diminishes or disappears. By contrast, the attentional mechanism is effective for such problems. For different tasks, the architecture of the BERT pre-trained model can be solved directly without complex retraining or modification of the model architecture. This significantly simplifies the process of migration learning, allowing

the model to be applied to new tasks more quickly. From a bioinformatics point of view, the advantages of BERT are clear, as distal interactions are significant for DNA sequence classification, gene expression prediction, and other tasks.

3.2. Graph Representation Learning

Graph representation learning is a technique that aims to learn low-dimensional vector representations of nodes and edges in graph data. Its overarching objective is to encode the structural and semantic information of graph data in a compact form that can be used for various downstream tasks. In this section, we introduce graph embedding methods and graph neural network methods. In the end, some other methods are introduced and expanded upon.

3.2.1. Graph Embedding

The objective of graph embedding is to find a low-dimensional vector representation of a high-dimensional graph while preserving the connectivity and interactions between nodes in the graph, which typically retains some key information about the nodes in the original graph. We categorize these graph embedding methods into four major types: matrix-factorization-based methods, manifold-learning-based methods, random-walk-based methods and deep-learning-based methods.

Matrix-factorization-based methods. A common graph embedding method is matrix factorization, which learns the low-dimensional representation of nodes by dissecting the adjacency matrix or laplacian matrix of the graph. The graph factorization (GF) [54] learns the low-dimensional representation of nodes in a graph by decomposing the graph's adjacency matrix or Laplacian matrix. While maintaining the advantages of unsupervised text embedding, Predictive Text Embedding (PTE) [55] learns features from labeled data while using both labeled and unlabeled data for representation learning. Matrix factorization is used by both GraRep [56] and High-Order Proximity preserved Embedding (HOPE) [57] to extract the higher-order neighborhood information from the graph. However, GraRep constructs the higher-order proximity matrix by concatenating k-hop transition probability matrices, while HOPE constructs the higher-order proximity matrix by measuring the pairwise similarity between nodes in subgraphs of varying sizes and orders. By generating edge representations and learning heterogeneous metrics, Heterogeneous Information Network Embedding via Edge Representations (HEER) [58] decomposes the network into low-dimensional embeddings and captures both structural and semantic information. A joint embedding matrix can be constructed by HERec [59] using meta-path-guided sampling to capture semantic and structural similarities between nodes of different types, which can then be factorized into low-dimensional embeddings.

Manifold-learning-based methods. In high-dimensional spaces, the sparsity and noise of graph data can pose challenges for model training and lead to overfitting. Manifold learning methods address this issue by mapping high-dimensional graph data to a lower-dimensional manifold space, which preserves the local properties and structural information of the original data while reducing its dimensionality. The main idea of isometric mapping (IsoMap) [60] is to preserve the shortest path between data points to maintain the manifold structure. Both locally linear embedding (LLE) [61] and local tangent space alignment (LTSA) [62] are local information manifold learning methods. A local linear reconstruction is carried out in the neighborhood of each data point by LLE to maintain the flow structure. LTSA preserves the manifold structure by aligning the data in each point's local tangent space. Laplacian eigenmaps (LE) [63] and Hessian eigenmaps (HE) [64] are both manifold learning methods based on spectral analysis. Specifically, LE utilizes the Laplacian matrix to perform local weight assignment on the data's neighborhood for preserving the manifold structure, while HE models the curvature and geometry of the data by leveraging the Hessian matrix. t-Distributed stochastic neighbor embedding (t-SNE) [65] uses a probabilistic approach to model the similarity between points in high-dimensional space and maps them to low-dimensional space while preserving pairwise similarities. Both

uniform manifold approximation and projection (UMAP) [66] and t-SNE use graph layout algorithms to arrange data in low-dimensional space, which makes them very similar. The distinction is that UMAP builds a low-dimensional representation using an optimization technique based on the nearest neighbor graph.

Random-walk-based methods. Inspired by the word vector model in the field of NLP, the researchers consider that the nodes on the graph can be analogized to words and the node sequences to sentences, and then the node sequences can be generated by random walk sampling on the graph to learn the representation vectors of the nodes using the Skip-gram model. DeepWalk [67] is the origin of the random walk approach, employing a simple, unbiased random walk followed by skip-gram representation learning of the walk sequences. Similarly, node2vec [68] is a biased random walk that introduces two parameters to regulate the search strategy of random walks, aiming to achieve a better balance between capturing graph structural features and node similarity features and improving the quality of node representations. By maximizing the likelihood functions of first-order proximity and second-order proximity, large-scale information network embedding (LINE) [69] learns the embedding of nodes. Additionally, LINE employs negative sampling strategies to boost training effectiveness. Walklets [70], which concentrates on higher-order structural information, splits joint random walk sequences of various lengths to obtain structural data on nodes at various scales. The similarity of the structural roles of nodes in the network is taken into account by struct2vec [71], not just the probability of contribution between nodes. Traditional node embedding techniques typically employ a co-occurrence probability-based methodology, which makes it difficult to adequately capture the semantic and structural information between nodes in heterogeneous information networks. By limiting the direction of random walk, meta-paths direct the generation of node sequences related to particular meta-paths, thereby capturing semantic and structural information between nodes. Metapath2vec [72] uses skip-gram to learn node embeddings after defining meta-paths to describe the semantic relationships between nodes and create heterogeneous neighborhoods of nodes. Using the defined meta-paths, HIN2vec [73] also creates diverse node neighborhoods. The crucial distinction is that it reconstructs the neighborhood using a self-encoder model and learns the embedding vector of the nodes by minimizing the reconstruction error. GATNE [74] combines the ideas of graph attention network (GAN) and neighborhood aggregation embedding (NAE). It employs the GAT model for node representation learning in heterogeneous graphs and the NAE method for node neighborhood aggregation. Setting the proper number of walks and walk lengths is crucial when learning node embedding in various heterogeneous networks.

Deep-learning-based methods. Traditional machine learning techniques rely heavily on manually created feature representations, which have a priori knowledge restrictions and limit the expressiveness of the models. Deep learning techniques can enhance data representation by learning multi-level feature representations and reducing the need for manual features. To encode and decode the nodes and obtain the details of the graph structure, SDNE [75] employs a deep self-encoder architecture. First- and second-order proximity are simultaneously optimized, and overfitting is prevented by sparse regularization. The advantage of SDNE is that it learns the relationships between the nodes while retaining the graph's structural information, improving the graph's representation. DNGR [76] directly obtains graph structural information using a random walk model and learns node representations using stacked denoising autoencoders. The nonlinear features of the graph can be learned using these two techniques, but they do not perform well on non-Euclidean graphs. To encode the text, HNE [77] employs recursive neural networks, which can model the hierarchical structure of the text. Both image and text encoders are optimized using a joint training approach to learn better matching. A deep-aligned autoencoder-based embedding method for heterogeneous graphs was introduced by BL-MNE [78] to embed different types of nodes into the same low-dimensional space. TADW [79] learns low-dimensional embedding representations of network nodes using DeepWalk and latent Dirichlet allocation (LDA) methods, and introduces the at-

tention mechanism to weigh the nodes' textual information. Three different attention mechanisms—node attention, attribute attention, and neighbor attention—are introduced by the LANE [80] method, which adaptively adjusts the weights on different embedding layers to better capture the similarities and differences between nodes and improve the quality of embedding. ASNE [81] creates supernodes by combining the structural and attribute information of the nodes, employs an attention-based framework to regulate the weights between the different supernodes in the embedding space, and employs adaptive sparse methods to learn the representation of the nodes. DANE [82] can simultaneously capture highly nonlinear node attributes and topological relationships while preserving their proximity. The expressiveness and generalization performance of the node representation are improved by ANRL [83], which uses adaptive neighborhood regularization to dynamically adjust the weights of each node's neighbors using the relationships between neighboring nodes.

3.2.2. Graph Neural Network-Based Methods

Currently, GNNs are prominent in the field of graph representation learning, particularly for non-Euclidean data. Their ability to model node and edge interactions sets them apart from traditional neural networks, making them highly effective for tasks involving graph-structured data analysis and prediction. Drawing upon prior research [84,85], we present a taxonomy of graph neural networks in this section, organized into five distinct categories based on their architectural design and unique approaches to processing structured graph data: GCNs, GATs, GAEs, GGANs, and GPNs.

GCNs. GCNs define a convolution operation on a graph that aggregates the features of a node's neighboring nodes and the node itself to generate a new node feature. Graph convolution methods can be classified into spectral methods and spatial methods. In graph data, the relationships between nodes are irregular. GCN [86] introduces learnable convolutional parameters to extract and learn features from graph data. By stacking multiple graph convolutional layers, GCN obtains more abundant and complex node feature representations. DGCN [87] employs distinct graph convolution operations on the primal and dual graphs to obtain node representations on both graphs. To address the issue of traditional graph neural networks requiring the definition of static adjacency matrices in advance, AGCN [88] introduces adaptive adjacency matrices and gate mechanisms, avoiding the computational burden brought by multiple convolutional operations. LGCN [89] utilizes a learnable graph convolutional layer (LGCL) and a subgraph training technique for handling large-scale graph data, thereby circumventing the computational limitations of traditional graph convolutional operations that necessitate computing the entire graph. In addition, FastGCN [90] was proposed to address the issue of high computational and memory overheads in traditional graph convolutional methods by introducing importance sampling and mini-batch training. The GraphSAGE [91] approach is considered a significant milestone in the field of graph neural networks. It uses a sampler to randomly sample subgraphs from the graph data that contain the target node, enabling inductive representation learning and prediction on unseen nodes. GIN [92] incorporates a global sorted pooling operation in its aggregation process, endowing the GIN method with a certain degree of graph isomorphism invariance, thereby avoiding the mapping of isomorphic graphs to different vector representations. APPNP [93] uses the personalized PageRank algorithm to achieve personalized propagation and prediction of nodes and utilizes a larger receptive field for node classification tasks.

GATs. To address the issue of information aggregation in traditional graph neural networks, GATs introduced attention mechanisms [48] that enable different weighting of neighboring nodes around each node, thus better capturing the relationships between nodes. The attention mechanism in GAT [94] adaptively calculates the weights between nodes and further incorporates multi-head attention mechanisms to better capture the structural information of graphs and effectively handle large-scale graph data. AGNN [95] utilizes an attention mechanism to dynamically learn the relationships between each node

and its neighboring nodes, thereby enhancing the model's performance. DySAT [96] computes node representations through joint self-attention along the two dimensions of the structural neighborhood and temporal dynamics, more effectively capturing the relationships and features in dynamic graph data. GaAN [97] introduced an adaptive receptive field mechanism to dynamically adjust the neighborhood size of each node, enabling the receptive field size to adaptively fit the density and distribution of its surrounding nodes, thus enhancing the model's flexibility and adaptability to different graph structures. HAN [98] utilizes distinct attention mechanisms to model different types of nodes and edges, learning complex relationships between nodes and the structural information of heterogeneous graphs by adaptively computing weights and representations of nodes. MAGNA [99] adopts a multi-hop attention mechanism that utilizes diffusive priors on attention values to consider all paths between non-connected node pairs, enabling the dynamic capture of complex relationships between nodes. High-order attention mechanisms and adversarial regularization constraints are employed by GCAN [100] to fully utilize both low-order and high-order information of nodes for representation learning.

GAEs. The encoder–decoder architecture has also been widely applied in graph generation and reconstruction tasks. Graph auto-encoder (GAE) [101] initially employs this architecture by utilizing an encoder to compress the raw graph data into low-dimensional vectors and a decoder to decode these low-dimensional vectors back into the original graph. The structures of both the encoder and decoder are composed of multiple layers of GCN. Variational graph auto-encoder (VGAE) [101] is an extension of GAE that introduces a variational inference method on top of GAE, enabling VGAE to learn more robust and interpretable low-dimensional representations. GraphVAE [102] adopts a similar structure to VAE, where the encoder transforms the graph information into a vector that approximates a normal distribution, and the decoder then transforms this vector back into a new graph. Graphite [103] utilizes graph neural networks to parameterize a variational autoencoder and employs a novel iterative graph refinement strategy for decoding. Graph2Gauss [104] builds upon VGAE by modeling node representations using Gaussian distributions and improving the distance metric to better capture complex category structures. In addition, DNVE [105] introduces deep generative models into graph autoencoders, embedding the network through deep encoders and decoders. Contrastive learning has also been applied to graph embedding methods. For example, DGI [106] uses a mutual information maximization method to learn node representations, which can learn high-order relationships between nodes without reconstructing the input graph. Another approach, InfoGraph [107], learns graph-level representations by maximizing the mutual information between the graph-level representation and representations of substructures at different scales within the graph. There is also a novel approach called MaskGAE [108], which combines the Masked Graph Model (MGM) with the graph autoencoder to achieve self-supervised graph representation learning and capture node features and structural information in the graph.

GGANs. GGANs are a type of generative model based on adversarial training that extends conventional generative adversarial networks (GANs) [109] and can be used for generating and learning embedded representations of graph data. The generator network can generate realistic graph data, while the discriminator network can evaluate the quality of the generated graph data. Through adversarial training between the two networks, GGANs can continuously optimize their ability to generate and learn graph data. An earlier method, GraphGAN [110], uses GAN to learn graph embedding representations and employs a loss function based on pairwise similarity to preserve the continuity of the embedding space. ARVGA [111] uses adversarial regularization to regularize node embeddings in the embedding space, thereby improving the continuity and robustness of embedding representations. By introducing an additional regularization term in adversarial training, ANE [112] enhances existing graph embedding methods by treating the prior distribution as real data and embedding vectors as generated samples. NetRA [113] is a network representation learning method based on an encoder–decoder framework and adversarial training. It encodes and decodes random walks on nodes, utilizes adversarial

training to regularize embeddings, and introduces a prior distribution to improve the model's generalization performance. In addition, there is an implicit generative model called NetGAN [114], which is used to simulate real-world networks. It uses GAN to train a generator and utilizes random walk loss in the graph to learn node-level representations. MolGAN [115] is an implicit generative model used for generating small molecule graphs. It combines GAN and policy gradient reinforcement learning methods to improve the quality and diversity of generated molecular graphs.

GPNs. GPNs use two operations, graph pooling and graph convolution, to process input graph data, with the main goal of reducing the graph data size and extracting global features. A general differentiable pooling operation called DiffPool [116] has been proposed, which can optimize node representations and pooling matrices through backpropagation to better capture the hierarchical structure of graph data. Zhang et al. [117] proposed a novel SortPooling operation that sorts each node's feature vector along with the node's degree and then uses the sorted result as the pooling result to better capture the global features of graph data. A method different from other graph pooling methods is SAGPool [118], which dynamically selects a subset of nodes using self-attention scores without the need to specify a fixed-size node subset. This approach enables more accurate learned graph representations that can adapt to different graph structures. Diehl et al. [119] proposed a pooling method called EdgePool, which relies on the edge contraction concept to learn localized and sparse hard pooling transformations. This method calculates importance scores for each edge to select representative edges and compress them together to form new nodes.

In addition, research on embedding methods for non-Euclidean spaces, contrastive learning for graph neural networks, and feature processing is worth noting. Corso et al. [120] proposed a neural distance embedding model, NeuroSEED, which embeds biological sequences into hyperbolic space for improved capture of hierarchical structure. A simple graph contrastive learning framework called SimGRACE [121] was proposed, which does not require data augmentation and enhances the robustness of graph contrastive learning using an adversarial scheme. Tang et al. [122] proposed a novel unsupervised feature selection method, MGF²WL, which combines multiple image fusion and feature weight learning to address poor-quality similarity graphs and designed an effective feature reconstruction model. Sun et al. [123] proposed a feature-expanded graph neural network model, FE-GNN, that utilizes feature subspace flattening and structural principal components to expand the feature space and improve model performance.

We also summarize the representation learning method implementations reviewed in this paper in Table 1, most of which are official implementations.

Table 1. Summary of Representation Learning Methods (accessed on 18 June 2023).

Category	Method Name	Code Link
Neural network-based language model representation learning		
	Word2vec [39]	https://code.google.com/archive/p/word2vec/
	Doc2vec [40]	https://nbviewer.org/github/danielfrg/word2vec/blob/main/examples/doc2vec.ipynb
	GloVe [41]	https://nlp.stanford.edu/projects/glove/
	FastText [42]	https://github.com/facebookresearch/fastText
	Bi-LSTM [44]	-
	ELMo [45]	https://allenai.org/allennlp/software/elmo
	Transformer [48]	https://github.com/tensorflow/tensorflow
	BERT [52]	https://github.com/google-research/bert

Table 1. *Cont.*

Category	Method Name	Code Link
Graph representation learning		
Graph embedding	GF [54]	-
	PTE [55]	https://github.com/mnqu/PTE
	GraRep [56]	https://github.com/ShelsonCao/GraRep
	HOPE [57]	http://git.thumedia.org/embedding/HOPE
	HEER [58]	https://github.com/GentleZhu/HEER
	HERec [59]	https://github.com/librahu/HERec
	IsoMap [60]	https://github.com/scikit-learn/scikit-learn/blob/main/sklearn/manifold/_isomap.py
	LLE [61]	-
	LTSA [62]	-
	LE [63]	-
	HE [64]	-
	t-SNE [65]	https://lvdmaaten.github.io/tsne/
	UMAP [66]	https://github.com/lmcinnes/umap
	DeepWalk [67]	https://github.com/phanein/deepwalk
	node2vec [68]	https://github.com/aditya-grover/node2vec
	LINE [69]	https://github.com/tangjianpku/LINE
	Walklets [70]	https://github.com/benedekrozemberczki/walklets
	struct2vec [71]	https://github.com/leoribeiro/struc2vec
	Metapath2vec [72]	https://ericdongyx.github.io/metapath2vec/m2v.html
	HIN2vec [73]	https://github.com/csiesheep/hin2vec
	GATNE [74]	https://github.com/THUDM/GATNE
	SDNE [75]	https://github.com/suanrong/SDNE
	DNGR [76]	https://github.com/ShelsonCao/DNGR
	HNE [77]	-
	BL-MNE [78]	-
	TADW [79]	https://github.com/thunlp/tadw
	LANE [80]	https://github.com/xhuang31/LANE
	ASNE [81]	https://github.com/lizi-git/ASNE
	DANE [82]	https://github.com/gaoghc/DANE
	ANRL [83]	https://github.com/cszhangzhen/ANRL
Graph neural network	GCN [86]	https://github.com/tkipf/gcn
	DGCN [87]	https://github.com/ZhuangCY/DGCN
	AGCN [88]	https://github.com/yimutianyang/AGCN
	LGCN [89]	https://github.com/divelab/lgcn
	FastGCN [90]	https://github.com/matenure/FastGCN
	GraphSAGE [91]	https://github.com/williamleif/GraphSAGE
	GIN [92]	https://github.com/weihua916/powerful-gnns
	APPNP [93]	https://github.com/gasteigerjo/ppnp
	GAT [94]	https://github.com/PetarV-/GAT
	AGNN [95]	-
	DySAT [96]	https://github.com/aravindsankar28/DySAT
	GaAN [97]	https://github.com/jennyzhang0215/GaAN
	HAN [98]	https://github.com/Jhy1993/HAN
	MAGNA [99]	https://github.com/xjtuwgt/GNN-MAGNA
	GCAN [100]	-
	GAE [101]	https://github.com/tkipf/gae
	VGAE [101]	https://github.com/tkipf/gae
	GraphVAE [102]	https://github.com/snap-stanford/GraphRNN/tree/master/baselines/graphvae
	Graphite [103]	https://github.com/ermongroup/graphite
	Graph2Gauss [104]	https://github.com/abojchevski/graph2gauss
	DNVE [105]	-
	DGI [106]	https://github.com/PetarV-/DGI

Table 1. *Cont.*

Category	Method Name	Code Link
	InfoGraph [107]	https://github.com/fanyun-sun/InfoGraph
	MaskGAE [108]	https://github.com/EdisonLeeeee/MaskGAE
	GraphGAN [110]	https://github.com/hwwang55/GraphGAN
	ARVGA [111]	-
	ANE [112]	-
	NetRA [113]	https://github.com/chengw07/NetRA
	NetGAN [114]	https://github.com/danielzuegner/netgan
	MolGAN [115]	https://github.com/nicola-decao/MolGAN
	DiffPool [116]	https://github.com/RexYing/diffpool
	SortPooling [117]	https://github.com/muhanzhang/DGCNN
	SAGPool [118]	https://github.com/inyeoplee77/SAGPool
	EdgePool [119]	-
Others		
	NeuroSEED [120]	https://github.com/gcorso/NeuroSEED
	SimGRACE [121]	https://github.com/mpanpan/SimGRACE
	MGF²WL [122]	-
	FE-GNN [123]	https://github.com/sajqavril/Feature-Extension-Graph-Neural-Networks

4. Representation Learning Methods for COVID-19

From drug omics to clinical outcomes, representation learning has been widely applied for the representation and modeling of multimodal biological and medical systems. This section summarizes a series of representative methods in the field of representation learning for COVID-19-related domains such as pharmaceuticals, public health, and healthcare, as shown in Figure 4.
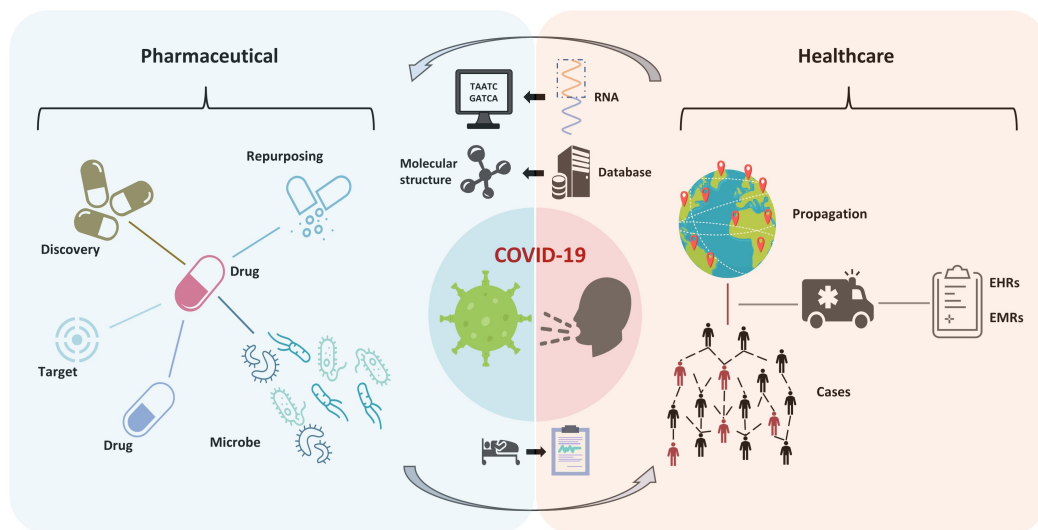

Figure 4. An overview of graphs used in COVID-19 applications. From pharmaceuticals to healthcare, graphs exist in all aspects and influence each other. Drug discovery, repurposing, targets, microbes, and their interactions constitute the pharmaceutical graph. Propagation, cases, electronic health records, and electronic medical records constitute the healthcare knowledge graph. Pharmaceuticals can gain feedback on the effectiveness of drugs through clinical trials, and healthcare can guide drug development by analyzing the obtained virus RNA or drug molecular structure. These multi-modal graphs are interconnected and need to be understood as a whole.

4.1. Pharmaceutical

Traditional drug development is not only characterized by lengthy cycles and exorbitant costs but also by low success rates. Artificial intelligence methods offer advantages such as large-scale data processing, end-to-end learning, handling complex molecular structures, and adaptive learning, which have accelerated drug discovery and drug repurposing. On the other hand, graph representation methods are better suited to handle high-dimensional, non-linear data structures and intuitively display the structural features of drug molecules, making it easier to understand and interpret prediction results.

4.1.1. Drug Discovery

Ton et al. [124] utilized a deep learning platform called Deep Docking to screen over 1.3 billion compounds for inhibitors targeting the SARS-CoV-2 main protease, providing candidate drugs for drug development. Saravanan et al. [125] employed deep learning to predictively screen known compounds, enabling rapid and accurate identification of potential antiviral drugs.

Zhou et al. [126] proposed a nonlinear end-to-end method for predicting interactions between compounds and COVID-19 based on GCN and attention mechanisms in heterogeneous graph representation learning, accelerating drug development. To predict molecular properties and identify new drugs, Wang et al. [127] developed a machine learning tool called AdvProp. AdvProp consists of four modules, including two graph-based and two sequence-based methods. Li et al. [128] integrated chemical, biological, and physical information to propose a multi-physics molecular graph representation and characterization method that can more accurately describe the interactions between drug molecules and proteins, making it better suited for COVID-19 drug discovery. Pi et al. [129] integrated node attention mechanisms and edge attention mechanisms into GCN for predicting the activity of microbial drugs and successfully predicted two drugs related to SARS-CoV-2.

4.1.2. Drug Repurposing

Ge et al. [130] developed a data-driven drug repurposing machine learning and statistical analysis framework to discover potential drug candidates against SARS-CoV-2. Mall et al. [131] proposed a data-driven drug repurposing strategy that successfully screened potential therapeutic drugs against COVID-19. Hooshmand et al. [132] used a multimodal restricted Boltzmann machine approach to connect multimodal information for drug repurposing to treat COVID-19. Anwaar et al. proposed a drug repurposing approach based on deep learning and molecular docking simulations.

Aghdam et al. [133] repurposed existing drugs to treat COVID-19 by utilizing biological network graphs and key proteins to propose multiple informative features for finding potential candidate drugs for COVID-19 treatment. Hsieh et al. [134] constructed a drug repurposing method based on a SARS-CoV-2 knowledge graph and used deep graph neural networks to determine the priority of repurposable drugs. Pham et al. [135] proposed a drug screening method utilizing graph neural networks and multi-head attention mechanisms for high-throughput mechanism-based phenotypic screening and applied it to COVID-19 drug repurposing. By integrating multiple lines of evidence to infer potential COVID-19 treatments, Hsieh et al. [136] used a graph neural network-based model to predict the potential therapeutic effects of existing compounds. Doshi et al. [137] proposed an end-to-end model based on graph neural networks for drug repurposing. The cold-start problem of unverified virus-drug associations (VDAs) for novel viruses is also an urgent issue in need of a solution. Su et al. [138] proposed a drug repositioning model that integrates sequence features and network topology features, providing a viable computational approach for the rapid discovery of potential antiviral drugs against novel viruses. Additionally, they introduced a deep learning model based on constrained multi-view non-negative matrix factorization and graph convolutional networks [139], offering a new perspective on drug repurposing for SARS-CoV-2.

4.1.3. Drug–Target Interaction Prediction

Beck et al. [140] proposed a deep-learning-based drug–target interaction model that can predict the interactions between drugs and specific protein targets for the discovery of commercially available antiviral drugs that can treat SARS-CoV-2. Saha et al. [141] employed a machine-learning-based approach to predict the interactions between the COVID-19 virus and human proteins, identifying key proteins and metabolic pathways relevant to COVID-19 as well as potential drug targets and drug combinations.

To predict the affinity between COVID-19 ion channel targets and drugs, Wang et al. [142] proposed a hybrid graph network model that uses an attention mechanism and GCN to connect drugs and targets. Zhang et al. [143] incorporated multi-layer graph information into transformer networks to explore the molecular structure of drug compounds in drug-target interaction prediction. To alleviate the sparsity of bipartite graphs and obtain better node representations, Li et al. [144] proposed a novel prediction model based on a variational graph autoencoder combined with a dual Wasserstein generative adversarial network gradient penalty strategy according to the prior knowledge graph. This enhances the representation of edge information for better drug–target interaction.

4.1.4. Drug–Drug Interaction Prediction

Tang et al. [145] introduced domain-invariant learning to address the issue of domain shift between different datasets. They improved the generalization performance of drug–drug interaction prediction methods by minimizing the distributional discrepancy between the source and target domains. Sefidgarhoseini et al. [146] effectively improved the accuracy and robustness of drug-drug interaction extraction by combining multiple pre-trained Transformer models, entity tagging strategies, and a voting-based ensemble strategy.

Ren et al. [147] proposed a drug–drug interaction prediction method based on a biomedical knowledge graph, which utilizes a combination of local and global features to enhance predictive performance. Chen et al. [148] designed a multi-scale feature fusion approach and a bi-level cross strategy to optimize the feature fusion method, thereby enhancing the performance of drug–drug interaction prediction. This method also fully utilizes the features extracted from drug molecular structure graphs and biomedical knowledge graphs and can be applied to different types of DDI prediction tasks. Pan et al. [149] integrated the self-attention mechanism, cross-attention mechanism, and graph attention network to construct a multi-source feature fusion network, which enables the model to better capture drug–drug interaction information and further predict drug-drug interaction-related events. Li et al. [150] proposed a novel multi-view substructure learning method that fuses multiple views, such as chemical structures, pharmacophores, and targets, to learn key substructure information for drug-drug interaction. An improved and more precise model for predicting drug–drug interactions was proposed by Ma et al. [151] using a dual graph neural network model that takes into account both molecular-level and substructure-level data.

4.1.5. Bio-Drug Interaction Prediction

Dey et al. [152] constructed a dataset with rich sequence features and utilized machine learning techniques to predict virus–host interactions between SARS-CoV-2 and human proteins. To predict virus–host interactions at the protein and organism levels using machine learning techniques, Du et al. [153] constructed a complex network with rich biological information.

Yang et al. [154] proposed a graph convolutional network-based method to uncover potential associations between human microbiota and drugs. The method employed a multi-kernel fusion strategy to integrate biological information from different sources, thereby enhancing prediction capability. Das et al. [155] proposed an innovative geometric deep learning model that effectively predicts and displays potential drug–virus interactions against SARS-CoV-2 by utilizing message passing neural networks (MPNN) and graph-structured features.

4.2. Public Health and Healthcare

Representation learning has also been applied to case prediction, propagation prediction, and analysis of multi-modal medical system data, such as electronic health records (EHRs) and electronic medical records (EMRs). The first two are important for predicting epidemic trends and guiding public health decisions, while EHRs and EMRs are crucial for improving clinical decision-making and optimizing medical resources.

4.2.1. Case Prediction

Shahid et al. [156] evaluated multiple predictive models and demonstrated the superior performance of the Bi-LSTM model in COVID-19 case forecasting. Abbasimehr et al. [157] employed a combination of multi-head attention, LSTM, and CNN, along with the Bayesian optimization algorithm, to forecast the number of confirmed COVID-19 cases soon. Sinha et al. [158] compared the performance of multiple deep learning models for analyzing and forecasting confirmed cases of COVID-19 and investigated the impact of data preprocessing and model tuning on predictive performance.

Gao et al. [159] proposed a spatiotemporal attention network based on real-world evidence for predicting the number of cases in a fixed number of days in the future. They also designed a dynamic loss term to enhance long-term forecasting. To effectively capture potential information about virus transmission and capture the linearity and nonlinearity present in time series, Ntemi et al. [160] proposed a hybrid model composed of a GCN, an LSTM, and an autoregressive filter to more accurately predict the number of cases. Li et al. [161] proposed a prediction model that combines the lioness optimization algorithm with the graph convolutional network, which can capture spatiotemporal information from feature data to achieve accurate predictions of COVID-19 case numbers. Skianis et al. [162] developed a multi-scale graph model utilizing demographic data, medical facilities, socio-economic indicators, and other related information to improve the prediction accuracy of COVID-19 positive cases and hospitalizations. The model is capable of automatically learning the interrelationships between different features, thereby enhancing its predictive capability.

4.2.2. Propagation Prediction

Malki et al. [163] utilized multiple machine learning models, including random forest, support vector machine, and artificial neural network, to predict COVID-19 transmission trends. Liu et al. [164] employed the Deep Neural Network (DNN) to predict the impact of social distancing on the transmission of COVID-19. They developed the Improved Particle Swarm Optimization (IPSO) algorithm, which introduced a generalized opposition-based learning strategy and an adaptive strategy to optimize the hyperparameters of the DNN. Ayris et al. [165] proposed a deep sequential prediction model and a machine-learning-based nonparametric regression model to predict the transmission of COVID-19.

La et al. [166] proposed an epidemiological model based on graph neural networks that effectively predicted and analyzed the transmission of the COVID-19 epidemic using dynamic graph-structured data. Hy et al. [167] proposed a method that captures multiscale patterns in spatiotemporal data and combines multiscale neighborhood aggregation, temporal convolutions, and multi-task learning to efficiently predict the transmission of COVID-19. To detect spatiotemporal patterns in the transmission of COVID-19, Geng et al. [168] introduced the spectral graph wavelet transform (SGWT) to deeply mine the characteristics of infection data and process COVID-19 data on dynamic graphs. Shan et al. [169] introduced graph neural networks and topological data analysis to propose a Graph Topology Learning method that effectively solves the prediction problem of virus transmission on spatial and temporal scales.

4.2.3. Analysis of EHRs and EMRs

The examination of EHRs and EMRs is integral to the healthcare industry and scientific community, as it enables a comprehensive comprehension of patients' medical conditions and healthcare service utilization. In turn, this facilitates the optimization of medical resources, an enhancement in healthcare service quality, a support system for clinical decision-making, and a catalyst for medical research and innovation. To investigate the clinical characteristics and prognostic factors of COVID-19 patients, Izquierdo et al. [170] employed natural language processing techniques to extract clinical information from patients' EHRs and utilized machine learning methods for modeling. Landi et al. [171] used deep learning techniques for feature extraction from EHRs, enabling effective stratification and classification of patient data, thereby providing robust support for precision medicine and treatment plan development. Wagner et al. [172] proposed an effective method for analyzing clinical records in large-scale EHR systems using deep neural networks, enabling early disease diagnosis based on EHR data. Wanyan et al. [173] developed a deep learning model based on contrastive loss to predict critical events such as severe illness, mortality, and intubation that demonstrates promising performance on imbalanced EHR data. They further proposed a unique EHR data heterogeneous feature design training algorithm and combined it with contrastive positive sampling to predict COVID-19 patient mortality [174]. To better perform prognostic analysis on COVID-19 patients, Ma et al. [175] proposed a distillation transfer learning framework that extracts knowledge from publicly available online EMR data through knowledge distillation.

Wanyan et al. [176] proposed a relationship learning framework based on heterogeneous graph models to predict the mortality rate of COVID-19 patients in intensive care units over different time windows. To address the challenges posed by data heterogeneity and semantic inconsistencies when integrating and analyzing electronic medical record data across institutions, Zhou et al. [177] proposed an integration method based on a multi-view incomplete knowledge graph, which takes into account the relationships between multiple data sources and the connecting information within each view. Gao et al. [178] proposed a new machine learning framework that uses electronic health records to predict hospitalization status and severity of illness in children and fuses clinical domain knowledge via graph neural networks, outperforming data-driven methods in terms of predictability and interpretability.

In Table 2, we summarize the public datasets from the previous application section.

Table 2. Existing Public COVID-19 Datasets (accessed on 8 September 2023).

Dataset	Dataset Type	Available Online
COVID-19db	Drug and target data	http://www.biomedical-web.com/covid19db/home
DrugBank	Drug data	https://go.drugbank.com/
Ensembl	SARS-CoV-2 genomic data	https://COVID-19.ensembl.org/index.html
ESC	SARS-CoV-2 immune escape variants	http://clingen.igib.res.in/esc
GISAID	Genetic sequence; Clinical and epidemiological data	https://gisaid.org/
NCBI	COVID-19 virus sequence	https://www.ncbi.nlm.nih.gov/datasets/taxonomy/2697049/
Our World In Data	COVID-19 cases	https://ourworldindata.org/covid-cases
PDB	Protein Data	https://www.rcsb.org/
RCoV19	COVID-19 information integration	https://ngdc.cncb.ac.cn/ncov/?lang=en
SCoV2-MD	Molecular dynamics of SARS-CoV-2 proteins	https://submission.gpcrmd.org/covid19/
SCovid	Single cell transcriptomics	http://bio-annotation.cn/scovid
T-cell COVID-19 Atlas	CD8 and CD4 T-cell epitopes	https://t-cov.hse.ru
VarEPS	SARS-CoV-2 variations evaluation	https://nmdc.cn/ncovn
World Health Organization	COVID-19 situation reports	https://www.who.int/emergencies/diseases/novel-coronavirus-2019

5. Challenges and Prospects

The graph representation method has demonstrated excellent outcomes in various bioinformatics tasks, indicating huge potential for processing biological data. It can convert biological data into low-dimensional embeddings before using them for downstream tasks. However, there are challenges associated with the process of broadly employing graph representation methods on real-world data. Because graph data are non-Euclidean, its structure is highly irregular in general. In this regard, there exist both opportunities and uncertainties in future developments.

5.1. Data Quality

As researchers delve deeper into the field of GNNs, the quality of graph data emerges as a critical factor that significantly influences the performance and reliability of our models. However, ensuring high-quality graph data is a challenging task due to the potential presence of redundant, erroneous, or missing features and connections, which can adversely impact the performance of GNNs [179]. In bioinformatics and biomedical research, graph data often represent intricate biological entities and their interactions, such as gene regulatory networks, protein–protein interaction networks, and patient similarity networks. The quality of the graph data is of paramount importance, as it directly impacts the accuracy and reliability of downstream analyses and predictions. The variability in the quality of gene sequence data produced by high-throughput sequencing can lead to noise in the sequencing data, causing GNNs to learn incorrect patterns. The sequencing of data for COVID-19 is facing such issues. Several studies have employed training strategies such as decoupled training, joint training, and two-level optimization to address the challenges of graph data augmentation. Therefore, developing reliable methods for graph data representation presents a challenging problem. Furthermore, it is crucial to establish benchmarks and metrics for assessing the quality of biological graph data, which can guide the development and evaluation of methods aimed at improving data quality. For instance, the expert-annotated consultation dialogue dataset established by [180] may provide some assistance in the treatment of COVID-19.

5.2. Hyperparameters and Labels

When employing graph representation models on large datasets, numerous hyperparameters need to be adjusted, and data must be accurately labeled prior to training. The quality of labels and uncertainties in biological data can influence the outcomes of training graph embedding models [181]. In most cases, the quality of labels generated directly by the model is inferior to those manually annotated. In many practical applications, manual label creation remains an indispensable method, albeit time-consuming and requiring some prior knowledge of model hyperparameter settings. Particularly for COVID-19 genomic data, as COVID-19 continues to mutate and produce new variants, it undergoes insertions, deletions, and base substitutions in the genomic sequence, making it challenging to discern patterns and manually create labels. An intriguing topic of research centers around identifying approaches that can enable models in unsupervised learning to generate labels by automatically extracting features of nodes and edges in graph data.

5.3. Interpretability and Extensibility

Although the graph neural network approach has developed quickly, it does not come without some limitations. Graph neural networks are still a "black box" with limited interpretability, but many real-world tasks require higher interpretability and scalability of the model. Interpretability is essential in bioinformatics, where there is a desire to better understand the mechanisms underlying biological processes, to provide accurate guidance for downstream analysis, and to then apply them in bioinformatics [182,183]. It has also proven difficult to develop a highly scalable and generic framework for different biological sequence data in order to handle larger and more complex data, adapt to various application scenarios, and better meet practical needs.

6. Conclusions

Representation learning, especially graph representation learning, provides the ability to handle complex data, integrate multi-modal data, and process large-scale datasets in bioinformatics and biomedical research. It has been successfully applied to identifying gene mutations for complex traits, assisting in disease diagnosis and treatment, and developing safe and effective drugs. In this work, we conducted a comprehensive survey of the applications of representation learning in bioinformatics and biomedicine. We summarized neural-network-based language model representation learning and graph representation learning, and discussed the relevant graph neural network methods in detail. Additionally, we introduced applications in pharmaceuticals, public health, and healthcare for COVID-19 based on these two approaches. Finally, we summarized the challenges faced by those in the popular research field of graph representation learning. It is anticipated that this work could promote the research of graph representation learning in the fields of bioinformatics and biomedicine, provide guidance for researchers in reviewing the entire development process of COVID-19, and prepare preventive measures for potential highly contagious epidemics in the future.

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

References

1. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; COVID-19 Genomics UK (COG-UK) Consortium; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424.
2. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **2018**, *34*, 4121–4123.
3. Morens, D.M.; Folkers, G.K.; Fauci, A.S. The concept of classical herd immunity may not apply to COVID-19. *J. Infect. Dis.* **2022**, *226*, 195–198.
4. Fu, L.; Wang, B.; Yuan, T.; Chen, X.; Ao, Y.; Fitzpatrick, T.; Li, P.; Zhou, Y.; Lin, Y.-f.; Duan, Q.; et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in china: A systematic review and meta-analysis. *J. Infect.* **2020**, *80*, 656–665.
5. Williamson, E.J.; Walker, A.J.; Bhaskaran, K.; Bacon, S.; Bates, C.; Morton, C.E.; Curtis, H.J.; Mehrkar, A.; Evans, D.; Inglesby, P.; et al. Opensafely: Factors associated with COVID-19 death in 17 million patients. *Nature* **2020**, *584*, 430.
6. Guo, Y.-R.; Cao, Q.-D.; Hong, Z.-S.; Tan, Y.-Y.; Chen, S.-D.; Jin, H.-J.; Tan, K.-S.; Wang, D.-Y.; Yan, Y. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—An update on the status. *Mil. Med. Res.* **2020**, *7*, 11.
7. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *Lancet* **2020**, *395*, 497–506.
8. Nguyen, T.T.; Abdelrazek, M.; Nguyen, D.T.; Aryal, S.; Nguyen, D.T.; Reddy, S.; Nguyen, Q.V.H.; Khatami, A.; Hsu, E.B.; Yang, S. Origin of novel coronavirus (COVID-19): A computational biology study using artificial intelligence. *bioRxiv* **2020**. [CrossRef]
9. Cascella, M.; Rajnik, M.; Aleem, A.; Dulebohn, S.C.; Napoli, R.D. Features, Evaluation, and Treatment of Coronavirus (COVID-19). *Statpearls [Internet]*, 2022. Available online: https://www.ncbi.nlm.nih.gov/books/NBK554776/ (accessed on 10 July 2023).
10. Jiang, S.; Hillyer, C.; Du, L. Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Trends Immunol.* **2020**, *41*, 355–359.
11. Shrestha, L.B.; Foster, C.; Rawlinson, W.; Tedla, N.; Bull, R.A. Evolution of the SARS-CoV-2 omicron variants ba. 1 to ba. 5: Implications for immune escape and transmission. *Rev. Med. Virol.* **2022**, *32*, e2381.

12. Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking changes in SARS-CoV-2 spike: Evidence that d614g increases infectivity of the COVID-19 virus. *Cell* **2020**, *182*, 812–827. [PubMed]

13. Cao, Y.; Wang, J.; Jian, F.; Xiao, T.; Song, W.; Yisimayi, A.; Huang, W.; Li, Q.; Wang, P.; An, R.; et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **2022**, *602*, 657–663. [CrossRef] [PubMed]

14. Bhattacharya, M.; Sharma, A.R.; Dhama, K.; Agoramoorthy, G.; Chakraborty, C. Omicron variant (b. 1.1. 529) of SARS-CoV-2: Understanding mutations in the genome, s-glycoprotein, and antibody-binding regions. *GeroScience* **2022**, *44*, 619–637. [CrossRef] [PubMed]

15. Mannar, D.; Saville, J.W.; Zhu, X.; Srivastava, S.S.; Berezuk, A.M.; Tuttle, K.S.; Marquez, A.C.; Sekirov, I.; Subramaniam, S. SARS-CoV-2 omicron variant: Antibody evasion and cryo-em structure of spike protein–ace2 complex. *Science* **2022**, *375*, 760–764.

16. Parums, D.V. The xbb. 1.5 ('kraken') subvariant of omicron SARS-CoV-2 and its rapid global spread. *Med. Sci. Monit.* **2023**, *29*, e939580-1. [CrossRef]

17. Basheer, I.A.; Hajmeer, M. Artificial neural networks: Fundamentals, computing, design, and application. *J. Microbiol. Methods* **2000**, *43*, 3–31. [CrossRef]

18. Chen, Y.; Li, Y.; Narayan, R.; Subramanian, A.; Xie, X. Gene expression inference with deep learning. *Bioinformatics* **2016**, *32*, 1832–1839. [CrossRef]

19. Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **2018**, *23*, 1241–1250.

20. Bakator, M.; Radosav, D. Deep learning and medical diagnosis: A review of literature. *Multimodal Technol. Interact.* **2018**, *2*, 47. [CrossRef]

21. Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review of methods and applications. *AI Open* **2020**, *1*, 57–81.

22. Xiong, J.; Xiong, Z.; Chen, K.; Jiang, H.; Zheng, M. Graph neural networks for automated de novo drug design. *Drug Discov. Today* **2021**, *26*, 1382–1393. [CrossRef]

23. Yang, F.; Fan, K.; Song, D.; Lin, H. Graph-based prediction of protein–protein interactions with attributed signed graph embedding. *BMC Bioinform.* **2020**, *21*, 323. [CrossRef]

24. Zhang, X.-M.; Liang, L.; Liu, L.; Tang, M.-J. Graph neural networks and their current applications in bioinformatics. *Front. Genet.* **2021**, *12*, 690049. [CrossRef] [PubMed]

25. Mercatelli, D.; Scalambra, L.; Triboli, L.; Ray, F.; Giorgi, F.M. Gene regulatory network inference resources: A practical overview. *Biochim. Biophys. Acta (BBA)-Gene Regul. Mech.* **2020**, *1863*, 194430.

26. Cai, H.; Zheng, V.W.; Chang, K.C.-C. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 1616–1637. [CrossRef]

27. Xu, M. Understanding graph embedding methods and their applications. *SIAM Rev.* **2021**, *63*, 825–853. [CrossRef]

28. Kotary, J.; Fioretto, F.; Hentenryck, P.V.; Wilder, B. End-to-end constrained optimization learning: A survey. *arXiv* **2021**, arXiv:2103.16378.

29. Wang, X.; Bo, D.; Shi, C.; Fan, S.; Ye, Y.; Philip, S.Y. A survey on heterogeneous graph embedding: Methods, techniques, applications and sources. *IEEE Trans. Big Data* **2022**, *9*, 415–436. [CrossRef]

30. Muzio, G.; O'Bray, L.; Borgwardt, K. Biological network analysis with deep learning. *Briefings Bioinform.* **2021**, *22*, 1515–1530. [CrossRef]

31. Zhang, Z.; Chen, L.; Zhong, F.; Wang, D.; Jiang, J.; Zhang, S.; Jiang, H.; Zheng, M.; Li, X. Graph neural network approaches for drug-target interactions. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102327. [CrossRef]

32. Ata, S.K.; Wu, M.; Fang, Y.; Ou-Yang, L.; Kwoh, C.K.; Li, X. Recent advances in network-based methods for disease gene prediction. *Briefings Bioinform.* **2021**, *22*, bbaa303. [CrossRef]

33. Wieder, O.; Kohlbacher, S.; Kuenemann, M.; Garon, A.; Ducrot, P.; Seidel, T.; Langer, T. A compact review of molecular property prediction with graph neural networks. *Drug Discov. Today Technol.* **2020**, *37*, 1–12. [CrossRef] [PubMed]

34. World. Statement on the Fifteenth Meeting of the IHR (2005) Emergency Committee on the COVID-19 Pandemic. May 2023. Available online: https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(COVID-19)-pandemic (accessed on 10 July 2023).

35. Feng, R.; Xie, Y.; Lai, M.; Chen, D.Z.; Cao, J.; Wu, J. Agmi: Attention-guided multi-omics integration for drug response prediction with graph neural networks. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1295–1298.

36. Zhu, Y.; Qian, P.; Zhao, Z.; Zeng, Z. Deep feature fusion via graph convolutional network for intracranial artery labeling. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, 11–15 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 467–470.

37. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]

38. Hamilton, W.L.; Ying, R.; Leskovec, J. Representation learning on graphs: Methods and applications. *arXiv* **2017**, arXiv:1709.05584.

39. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*; Burges, C.J., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2013; Volume 26.

40. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the 31st International Conference on Machine Learning, PMLR, Bejing, China, 22–24 June 2014; Xing, E.P., Jebara, T., Eds.; Volume 32 of Proceedings of Machine Learning Research, pp. 1188–1196.

41. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.

42. Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 135–146. [CrossRef]

43. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

44. Huang, Z.; Xu, W.; Yu, K. Bidirectional lstm-crf models for sequence tagging. *arXiv* **2015**, arXiv:1508.01991.

45. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep contextualized word representations. *arXiv* **2018**, arXiv:1802.05365.

46. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.

47. Kim, Y.; Denton, C.; Hoang, L.; Rush, A.M. Structured attention networks. *arXiv* **2017**, arXiv:1702.00887.

48. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł .; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.

49. Gehring, J.; Auli, M.; Grangier, D.; Yarats, D.; Dauphin, Y.N. Convolutional sequence to sequence learning. In Proceedings of the 34th International Conference on Machine Learning, PMLR, International Convention Centre, Sydney, Australia, 6–11 August 2017; Precup, D., Teh, Y.W., Eds.; Volume 70 of Proceedings of Machine Learning Research, pp. 1243–1252.

50. Sukhbaatar, S.; Szlam, A.; Weston, J.; Fergus, R. End-to-end memory networks. In *Advances in Neural Information Processing Systems*; Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28.

51. Miller, A.; Fisch, A.; Dodge, J.; Karimi, A.-H.; Bordes, A.; Weston, J. Key-value memory networks for directly reading documents. *arXiv* **2016**, arXiv:1606.03126.

52. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.

53. Luong, M.-T.; Pham, H.; Manning, C.D. Effective approaches to attention-based neural machine translation. *arXiv* **2015**, arXiv:1508.04025.

54. Ahmed, A.; Shervashidze, N.; Narayanamurthy, S.; Josifovski, V.; Smola, A.J. Distributed large-scale natural graph factorization. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 37–48.

55. Tang, J.; Qu, M.; Mei, Q. Pte: Predictive text embedding through large-scale heterogeneous text networks. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 1165–1174.

56. Cao, S.; Lu, W.; Xu, Q. Grarep: Learning graph representations with global structural information. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 19–23 October 2015; pp. 891–900.

57. Ou, M.; Cui, P.; Pei, J.; Zhang, Z.; Zhu, W. Asymmetric transitivity preserving graph embedding. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1105–1114.

58. Shi, Y.; Zhu, Q.; Guo, F.; Zhang, C.; Han, J. Easing embedding learning by comprehensive transcription of heterogeneous information networks. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2190–2199.

59. Shi, C.; Hu, B.; Zhao, W.X.; Philip, S.Y. Heterogeneous information network embedding for recommendation. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 357–370. [CrossRef]

60. Tenenbaum, J.B.; de Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [CrossRef]

61. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef]

62. Zhang, Z.; Zha, H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Sci. Comput.* **2004**, *26*, 313–338. [CrossRef]

63. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [CrossRef]

64. Donoho, D.L.; Grimes, C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5591–5596. [CrossRef]

65. der Maaten, L.V.; Hinton, G. Visualizing data using t-sne. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

66. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.

67. Perozzi, B.; Al-Rfou, R.; Skiena, S. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 701–710.

68. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.

69. Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; Mei, Q. Line: Large-scale information network embedding. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 1067–1077.

70. Perozzi, B.; Kulkarni, V.; Chen, H.; Skiena, S. Do not walk, skip! online learning of multi-scale network embeddings. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, Sydney, Australia, 31 July–3 August 2017; pp. 258–265.

71. Ribeiro, L.F.R.; Saverese, P.H.P.; Figueiredo, D.R. struc2vec: Learning node representations from structural identity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 385–394.

72. Dong, Y.; Chawla, N.V.; Swami, A. metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 135–144.

73. Fu, T.; Lee, W.-C.; Lei, Z. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 1797–1806.

74. Cen, Y.; Zou, X.; Zhang, J.; Yang, H.; Zhou, J.; Tang, J. Representation learning for attributed multiplex heterogeneous network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 1358–1368.

75. Wang, D.; Cui, P.; Zhu, W. Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1225–1234.

76. Cao, S.; Lu, W.; Xu, Q. Deep neural networks for learning graph representations. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.

77. Chang, S.; Han, W.; Tang, J.; Qi, G.-J.; Aggarwal, C.C.; Huang, T.S. Heterogeneous network embedding via deep architectures. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015; pp. 119–128.

78. Zhang, J.; Xia, C.; Zhang, C.; Cui, L.; Fu, Y.; Philip, S.Y. Bl-mne: Emerging heterogeneous social network embedding through broad learning with aligned autoencoder. In 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 605–614.

79. Yang, C.; Liu, Z.; Zhao, D.; Sun, M.; Chang, E.Y. Network representation learning with rich text information. *IJCAI* **2015**, *2015*, 2111–2117.

80. Huang, X.; Li, J.; Hu, X. Label informed attributed network embedding. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017; pp. 731–739.

81. Liao, L.; He, X.; Zhang, H.; Chua, T.-S. Attributed social network embedding. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 2257–2270. [CrossRef]

82. Gao, H.; Huang, H. Deep attributed network embedding. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI), Stockholm, Sweden, 13–19 July 2018.

83. Zhang, Z.; Yang, H.; Bu, J.; Zhou, S.; Yu, P.; Zhang, J.; Ester, M.; Wang, C. Anrl: Attributed network representation learning via deep neural networks. *IJCAI* **2018**, *18*, 3155–3161.

84. Iuchi, H.; Matsutani, T.; Yamada, K.; Iwano, N.; Sumi, S.; Hosoda, S.; Zhao, S.; Fukunaga, T.; Hamada, M. Representation learning applications in biological sequence analysis. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3198–3208. [CrossRef]

85. Yi, H.-C.; You, Z.-H.; Huang, D.-S.; Kwoh, C.K. Graph representation learning in bioinformatics: Trends, methods and applications. *Briefings Bioinform.* **2022**, *23*, bbab340. [CrossRef] [PubMed]

86. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.

87. Zhuang, C.; Ma, Q. Dual graph convolutional networks for graph-based semi-supervised classification. In Proceedings of the 2018 World Wide Web Conference, Lyon, Francel, 23–27 April 2018; pp. 499–508.

88. Li, R.; Wang, S.; Zhu, F.; Huang, J. Adaptive graph convolutional neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.

89. Gao, H.; Wang, Z.; Ji, S. Large-scale learnable graph convolutional networks. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 1416–1424.

90. Chen, J.; Ma, T.; Xiao, C. Fastgcn: fast learning with graph convolutional networks via importance sampling. *arXiv* **2018**, arXiv:1801.10247.

91. Hamilton, W.; Ying, Z.; Leskovec, J. Inductive representation learning on large graphs. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.

92. Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How powerful are graph neural networks? *arXiv* **2018**, arXiv:1810.00826.

93. Gasteiger, J.; Bojchevski, A.; Günnemann, S. Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv* **2018**, arXiv:1810.05997.

94.	Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *Stat* **2017**, *1050*, 10–21.

95.	Thekumparampil, K.K.; Wang, C.; Oh, S.; Li, L.-J. Attention-based graph neural network for semi-supervised learning. *arXiv* **2018**, arXiv:1803.03735.

96.	Sankar, A.; Wu, Y.; Gou, L.; Zhang, W.; Yang, H. Dynamic graph representation learning via self-attention networks. *arXiv* **2018**, arXiv:1812.09430.

97.	Zhang, J.; Shi, X.; Xie, J.; Ma, H.; King, I.; Yeung, D.-Y. Gaan: Gated attention networks for learning on large and spatiotemporal graphs. *arXiv* **2018**, arXiv:1803.07294.

98.	Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; Yu, P.S. Heterogeneous graph attention network. In Proceedings of the The World Wide Web Conference, WWW '19, San Francisco, CA, USA, 13–17 May 2019; pp. 2022–2032.

99.	Wang, G.; Ying, R.; Huang, J.; Leskovec, J. Multi-hop attention graph neural network. *arXiv* **2020**, arXiv:2009.14332.

100.	Xu, H.; Zhang, S.; Jiang, B.; Tang, J. Graph context-attention network via low and high order aggregation. *Neurocomputing* **2023**, *536*, 152–163. [CrossRef]

101.	Kipf, T.N.; Welling, M. Variational graph auto-encoders. *arXiv* **2016**, arXiv:1611.07308.

102.	Simonovsky, M.; Komodakis, N. Graphvae: Towards generation of small graphs using variational autoencoders. In Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial N. Networks, Rhodes, Greece, 4–7 October 2018; Proceedings, Part I 27; Springer: Berlin/Heidelberg, Germany, 2018; pp. 412–422.

103.	Grover, A.; Zweig, A.; Ermon, S. Graphite: Iterative generative modeling of graphs. In Proceedings of the 36th International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; Chaudhuri, K., Salakhutdinov, R., Eds; Volume 97 of Proceedings of Machine Learning Research; pp. 2434–2444.

104.	Bojchevski, A.; Günnemann, S. Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking. *arXiv* **2017**, arXiv:1707.03815.

105.	Zhu, D.; Cui, P.; Wang, D.; Zhu, W. Deep variational network embedding in wasserstein space. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2827–2836.

106.	Velickovic, P.; Fedus, W.; Hamilton, W.L.; Liò, P.; Bengio, Y.; Hjelm, R.D. Deep graph infomax. *ICLR (Poster)* **2019**, *2*, 4.

107.	Sun, F.-Y.; Hoffmann, J.; Verma, V.; Tang, J. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. *arXiv* **2019**, arXiv:1908.01000.

108.	Li, J.; Wu, R.; Sun, W.; Chen, L.; Tian, S.; Zhu, L.; Meng, C.; Zheng, Z.; Wang, W. Maskgae: Masked graph modeling meets graph autoencoders. *arXiv* **2022**, arXiv:2205.10053.

109.	Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]

110.	Wang, H.; Wang, J.; Wang, J.; Zhao, M.; Zhang, W.; Zhang, F.; Xie, X.; Guo, M. Graphgan: Graph representation learning with generative adversarial nets. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.

111.	Pan, S.; Hu, R.; Fung, S.; Long, G.; Jiang, J.; Zhang, C. Learning graph embedding with adversarial training methods. *IEEE Trans. Cybern.* **2019**, *50*, 2475–2487. [CrossRef] [PubMed]

112.	Dai, Q.; Li, Q.; Tang, J.; Wang, D. Adversarial network embedding. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.

113.	Yu, W.; Zheng, C.; Cheng, W.; Aggarwal, C.C.; Song, D.; Zong, B.; Chen, H.; Wang, W. Learning deep network representations with adversarially regularized autoencoders. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2663–2671.

114.	Bojchevski, A.; Shchur, O.; Zügner, D.; Günnemann, S. NetGAN: Generating graphs via random walks. In Proceedings of the 35th International Conference on Machine Learning, PMLR, Stockholmsmässan, Stockholm, Sweden, 10–15 July 2018; Dy, J., Krause, A., Eds; Volume 80 of Proceedings of Machine Learning Research, pp. 610–619.

115.	Cao, N.D.; Kipf, T. Molgan: An implicit generative model for small molecular graphs. *arXiv* **2018**, arXiv:1805.11973.

116.	Ying, Z.; You, J.; Morris, C.; Ren, X.; Hamilton, W.; Leskovec, J. Hierarchical graph representation learning with differentiable pooling. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 1–11.

117.	Zhang, M.; Cui, Z.; Neumann, M.; Chen, Y. An end-to-end deep learning architecture for graph classification. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.

118.	Lee, J.; Lee, I.; Kang, J. Self-attention graph pooling. In Proceedings of the International Conference on Machine Learning; PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 3734–3743.

119.	Diehl, F.; Brunner, T.; Le, M.T.; Knoll, A. Towards graph pooling by edge contraction. In *ICML 2019 Workshop on Learning and Reasoning with Graph-Structured Data*, Long Beach Convention Center, Long Beach, CA, USA, 2019.

120.	Corso, G.; Ying, Z.; Pándy, M.; Veličković, P.; Leskovec, J.; Liò, P. Neural distance embeddings for biological sequences. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 18539–18551.

121.	Xia, J.; Wu, L.; Chen, J.; Hu, B.; Li, S.Z. Simgrace: A simple framework for graph contrastive learning without data augmentation. In Proceedings of the ACM Web Conference 2022, Lyon, France, 25–29 April 2022; pp. 1070–1079.

122.	Tang, C.; Zheng, X.; Zhang, W.; Liu, X.; Zhu, X.; Zhu, E. Unsupervised feature selection via multiple graph fusion and feature weight learning. *Sci. China Inf. Sci.* **2023**, *66*, 1–17. [CrossRef]

123.	Sun, J.; Zhang, L.; Chen, G.; Zhang, K.; U, P.X.; Yang, Y. Feature expansion for graph neural networks. *arXiv* **2023**, arXiv:2305.06142.

124. Ton, A.-T.; Gentile, F.; Hsing, M.; Ban, F.; Cherkasov, A. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol. Inform.* **2020**, *39*, 2000028. [CrossRef]

125. Saravanan, K.M.; Zhang, H.; Hossain, M.T.; Reza, M.S.; Wei, Y. Deep learning-based drug screening for COVID-19 and case studies. In *In Silico Modeling of Drugs against Coronaviruses: Computational Tools and Protocols*; Humana, New York, NY, USA, 2021; pp. 631–660.

126. Zhou, D.; Peng, S.; Wei, D.-Q.; Zhong, W.; Dou, Y.; Xie, X. Lunar: Drug screening for novel coronavirus based on representation learning graph convolutional network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 1290–1298. [CrossRef]

127. Wang, Z.; Liu, M.; Luo, Y.; Xu, Z.; Xie, Y.; Wang, L.; Cai, L.; Qi, Q.; Yuan, Z.; Yang, T.; et al. Advanced graph and sequence neural networks for molecular property prediction and drug discovery. *Bioinformatics* **2022**, *38*, 2579–2586. [CrossRef] [PubMed]

128. Li, X.-S.; Liu, X.; Lu, L.; Hua, X.-S.; Chi, Y.; Xia, K. Multiphysical graph neural network (mp-gnn) for COVID-19 drug design. *Briefings Bioinform.* **2022**, *23*, bbac231. [CrossRef] [PubMed]

129. Pi, J.; Jiao, P.; Zhang, Y.; Li, J. Mdgnn: Microbial drug prediction based on heterogeneous multi-attention graph neural network. *Front. Microbiol.* **2022**, *13*, 819046. [CrossRef] [PubMed]

130. Ge, Y.; Tian, T.; Huang, S.; Wan, F.; Li, J.; Li, S.; Yang, H.; Hong, L.; Wu, N.; Yuan, E.; et al. A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *bioRxiv* **2020**. [CrossRef]

131. Mall, R.; Elbasir, A.; Meer, H.A.; Chawla, S.; Ullah, E. Data-driven drug repurposing for COVID-19. *ChemRxiv* **2020**. [CrossRef]

132. Hooshmand, S.A.; Ghobadi, M.Z.; Hooshmand, S.E.; Jamalkandi, S.A.; Alavi, S.M.; Masoudi-Nejad, A. A multimodal deep learning-based drug repurposing approach for treatment of COVID-19. *Mol. Divers.* **2021**, *25*, 1717–1730. [CrossRef]

133. Aghdam, R.; Habibi, M.; Taheri, G. Using informative features in machine learning based method for COVID-19 drug repurposing. *J. Cheminform.* **2021**, *13*, 70. [CrossRef]

134. Hsieh, K.; Wang, Y.; Chen, L.; Zhao, Z.; Savitz, S.; Jiang, X.; Tang, J.; Kim, Y. Drug repurposing for COVID-19 using graph neural network with genetic, mechanistic, and epidemiological validation. *Res. Sq.* 2020, *preprint*.

135. Pham, T.-H.; Qiu, Y.; Zeng, J.; Xie, L.; Zhang, P. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nat. Mach. Intell.* **2021**, *3*, 247–257. [CrossRef]

136. Hsieh, K.; Wang, Y.; Chen, L.; Zhao, Z.; Savitz, S.; Jiang, X.; Tang, J.; Kim, Y. Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence. *Sci. Rep.* **2021**, *11*, 23179. [CrossRef] [PubMed]

137. Doshi, S.; Chepuri, S.P. A computational approach to drug repurposing using graph neural networks. *Comput. Biol. Med.* **2022**, *150*, 105992. [CrossRef] [PubMed]

138. Su, X.; You, Z.; Wang, L.; Hu, L.; Wong, L.; Ji, B.; Zhao, B. Sane: A sequence combined attentive network embedding model for COVID-19 drug repositioning. *Appl. Soft Comput.* **2021**, *111*, 107831. [CrossRef]

139. Su, X.; Hu, L.; You, Z.; Hu, P.; Wang, L.; Zhao, B. A deep learning method for repurposing antiviral drugs against new viruses via multi-view nonnegative matrix factorization and its application to SARS-CoV-2. *Briefings Bioinform.* **2022**, *23*, bbab526. [CrossRef]

140. Beck, B.R.; Shin, B.; Choi, Y.; Park, S.; Kang, K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 784–790. [CrossRef] [PubMed]

141. Saha, S.; Chatterjee, P.; Halder, A.K.; Nasipuri, M.; Basu, S.; Plewczynski, D. Ml-dtd: Machine learning-based drug target discovery for the potential treatment of COVID-19. *Vaccines* **2022**, *10*, 1643. [CrossRef] [PubMed]

142. Wang, X.; Li, Q.; Liu, Y.; Du, Z.; Jin, R. Drug repositioning of COVID-19 based on mixed graph network and ion channel. *Math. Biosci. Eng.* **2022**, *19*, 3269–3284. [CrossRef]

143. Zhang, P.; Wei, Z.; Che, C.; Jin, B. Deepmgt-dti: Transformer network incorporating multilayer graph information for drug–target interaction prediction. *Comput. Biol. Med.* **2022**, *142*, 105214. [CrossRef]

144. Li, G.; Sun, W.; Xu, J.; Hu, L.; Zhang, W.; Zhang, P. Ga-ens: A novel drug–target interactions prediction method by incorporating prior knowledge graph into dual wasserstein generative adversarial network with gradient penalty. *Appl. Soft Comput.* **2023**, *139*, 110151. [CrossRef]

145. Tang, Z.; Chen, G.; Yang, H.; Zhong, W.; Chen, C.Y. Dsil-ddi: A domain-invariant substructure interaction learning for generalizable drug–drug interaction prediction. In *IEEE Transactions on Neural Networks and Learning Systems*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–9.

146. Sefidgarhoseini, S.; Safari, L.; Mohammady, Z. Drug-Drug Interaction Extraction Using Transformer-Based Ensemble Model. *Res. Sq.* 2023, *preprint*. [CrossRef]

147. Ren, Z.-H.; You, Z.-H.; Yu, C.-Q.; Li, L.-P.; Guan, Y.-J.; Guo, L.-X.; Pan, J. A biomedical knowledge graph-based method for drug–drug interactions prediction through combining local and global features with deep neural networks. *Briefings Bioinform.* **2022**, *23*, bbac363. [CrossRef] [PubMed]

148. Chen, Y.; Ma, T.; Yang, X.; Wang, J.; Song, B.; Zeng, X. Muffin: Multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics* **2021**, *37*, 2651–2658. [CrossRef]

149. Pan, D.; Quan, L.; Jin, Z.; Chen, T.; Wang, X.; Xie, J.; Wu, T.; Lyu, Q. Multisource attention-mechanism-based encoder–decoder model for predicting drug–drug interaction events. *J. Chem. Inf. Model.* **2022**, *62*, 6258–6270. [CrossRef]

150. Li, Z.; Zhu, S.; Shao, B.; Liu, T.-Y.; Zeng, X.; Wang, T. Multi-view substructure learning for drug-drug interaction prediction. *arXiv* **2022**, arXiv:2203.14513.

151. Ma, M.; Lei, X. A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions. *PLoS Comput. Biol.* **2023**, *19*, e1010812. [CrossRef] [PubMed]

152. Dey, L.; Chakraborty, S.; Mukhopadhyay, A. Machine learning techniques for sequence-based prediction of viral–host interactions between SARS-CoV-2 and human proteins. *Biomed. J.* **2020**, *43*, 438–450. [CrossRef] [PubMed]

153. Du, H.; Chen, F.; Liu, H.; Hong, P. Network-based virus–host interaction prediction with application to SARS-CoV-2. *Patterns* **2021**, *2*, 100242. [CrossRef] [PubMed]

154. Yang, H.; Ding, Y.; Tang, J.; Guo, F. Inferring human microbe–drug associations via multiple kernel fusion on graph neural network. *Knowl.-Based Syst.* **2022**, *238*, 107888. [CrossRef]

155. Das, B.; Kutsal, M.; Das, R. A geometric deep learning model for display and prediction of potential drug-virus interactions against SARS-CoV-2. *Chemom. Intell. Lab. Syst.* **2022**, *229*, 104640. [CrossRef]

156. Shahid, F.; Zameer, A.; Muneeb, M. Predictions for COVID-19 with deep learning models of lstm, gru and bi-lstm. *Chaos Solitons Fractals* **2020**, *140*, 110212. [CrossRef]

157. Abbasimehr, H.; Paki, R. Prediction of COVID-19 confirmed cases combining deep learning methods and bayesian optimization. *Chaos Solitons Fractals* **2021**, *142*, 110511. [CrossRef] [PubMed]

158. Sinha, T.; Chowdhury, T.; Shaw, R.N.; Ghosh, A. Analysis and prediction of COVID-19 confirmed cases using deep learning models: A comparative study. In *Advanced Computing and Intelligent Technologies: Proceedings of the ICACIT 2021, New Delhi, India, 20–21 March 2021*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 207–218.

159. Gao, J.; Sharma, R.; Qian, C.; Glass, L.M.; Spaeder, J.; Romberg, J.; Sun, J.; Xiao, C. Stan: Spatio-temporal attention network for pandemic prediction using real-world evidence. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 733–743. [CrossRef] [PubMed]

160. Ntemi, M.; Sarridis, I.; Kotropoulos, C. An autoregressive graph convolutional long short-term memory hybrid neural network for accurate prediction of COVID-19 cases. *IEEE Trans. Comput. Soc. Syst.* **2022**, *10*, 724–735. [CrossRef]

161. Li, D.; Ren, X.; Su, Y. Predicting COVID-19 using lioness optimization algorithm and graph convolution network. *Soft Comput.* **2023**, *27*, 5437–5501. [CrossRef]

162. Skianis, K.; Nikolentzos, G.; Gallix, B.; Thiebaut, R.; Exarchakis, G. Predicting COVID-19 positivity and hospitalization with multi-scale graph neural networks. *Sci. Rep.* **2023**, *13*, 5235. [CrossRef]

163. Malki, Z.; Atlam, E.-S.; Ewis, A.; Dagnew, G.; Ghoneim, O.A.; Mohamed, A.A.; Abdel-Daim, M.M.; Gad, I. The COVID-19 pandemic: Prediction study based on machine learning models. *Environ. Sci. Pollut. Res.* **2021**, *28*, 40496–40506. [CrossRef]

164. Liu, D.; Ding, W.; Dong, Z.S.; Pedrycz, W. Optimizing deep neural networks to predict the effect of social distancing on COVID-19 spread. *Comput. Ind. Eng.* **2022**, *166*, 107970. [CrossRef]

165. Ayris, D.; Imtiaz, M.; Horbury, K.; Williams, B.; Blackney, M.; See, C.S.H.; Shah, S.A.A. Novel deep learning approach to model and predict the spread of COVID-19. *Intell. Syst. Appl.* **2022**, *14*, 200068. [CrossRef]

166. Gatta, V.L.; Moscato, V.; Postiglione, M.; Sperli, G. An epidemiological neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak. *IEEE Trans. Big Data* **2020**, *7*, 45–55. [CrossRef]

167. Hy, T.S.; Nguyen, V.B.; Tran-Thanh, L.; Kondor, R. Temporal multiresolution graph neural networks for epidemic prediction. In *Workshop on Healthcare AI and COVID-19*; PMLR: Baltimore, MD, USA, 22 July 2022; pp. 21–32.

168. Geng, R.; Gao, Y.; Zhang, H.; Zu, J. Analysis of the spatio-temporal dynamics of COVID-19 in massachusetts via spectral graph wavelet theory. *IEEE Trans. Signal Inf. Process. Over Netw.* **2022**, *8*, 670–683. [CrossRef]

169. Shan, B.; Yuan, X.; Ni, W.; Wang, X.; Liu, R.P. Novel graph topology learning for spatio-temporal analysis of COVID-19 spread. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 2693–2704. [CrossRef]

170. Izquierdo, J.L.; Ancochea, J.; Savana COVID-19 Research Group; Soriano, J.B. Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: Retrospective study using machine learning and natural language processing. *J. Med. Internet Res.* **2020**, *22*, e21801. [CrossRef]

171. Landi, I.; Glicksberg, B.S.; Lee, H.-C.; Cherng, S.; Landi, G.; Danieletto, M.; Dudley, J.T.; Furlanello, C.; Miotto, R. Deep representation learning of electronic health records to unlock patient stratification at scale. *NPJ Digit. Med.* **2020**, *3*, 96. [CrossRef] [PubMed]

172. Wagner, T.; Shweta, F.N.U.; Murugadoss, K.; Awasthi, S.; Venkatakrishnan, A.J.; Bade, S.; Puranik, A.; Kang, M.; Pickering, B.W.; O'Horo, J.C.; et al. Augmented curation of clinical notes from a massive ehr system reveals symptoms of impending COVID-19 diagnosis. *eLife* **2020**, *9*, e58227. [CrossRef]

173. Wanyan, T.; Honarvar, H.; Jaladanki, S.K.; Zang, C.; Naik, N.; Somani, S.; Freitas, J.K.D.; Paranjpe, I.; Vaid, A.; Zhang, J.; et al. Contrastive learning improves critical event prediction in COVID-19 patients. *Patterns* **2021**, *2*, 100389. [CrossRef] [PubMed]

174. Wanyan, T.; Lin, M.; Klang, E.; Menon, K.M.; Gulamali, F.F.; Azad, A.; Zhang, Y.; Ding, Y.; Wang, Z.; Wang, F.; et al. Supervised pretraining through contrastive categorical positive samplings to improve COVID-19 mortality prediction. In Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Northbrook, IL, USA, 7–10 August 2022; pp. 1–9.

175. Ma, L.; Ma, X.; Gao, J.; Jiao, X.; Yu, Z.; Zhang, C.; Ruan, W.; Wang, Y.; Tang, W.; Wang, J. Distilling knowledge from publicly available online emr data to emerging epidemic for prognosis. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 3558–3568.

176. Wanyan, T.; Vaid, A.; Freitas, J.K.D.; Somani, S.; Miotto, R.; Nadkarni, G.N.; Azad, A.; Ding, Y.; Glicksberg, B.S. Relational learning improves prediction of mortality in COVID-19 in the intensive care unit. *IEEE Trans. Big Data* **2020**, *7*, 38–44. [CrossRef] [PubMed]

177. Zhou, D.; Gan, Z.; Shi, X.; Patwari, A.; Rush, E.; Bonzel, C.-L.; Panickan, V.A.; Hong, C.; Ho, Y.-L.; Cai, T.; et al. Multiview incomplete knowledge graph integration with application to cross-institutional ehr data harmonization. *J. Biomed. Inform.* **2022**, *133*, 104147. [CrossRef]

178. Gao, J.; Yang, C.; Heintz, J.; Barrows, S.; Albers, E.; Stapel, M.; Warfield, S.; Cross, A.; Sun, J. Medml: Fusing medical knowledge and machine learning models for early pediatric COVID-19 hospitalization and severity prediction. *Iscience* **2022**, *25*, 104970. [CrossRef]

179. Ding, K.; Xu, Z.; Tong, H.; Liu, H. Data augmentation for deep graph learning: A survey. *ACM SIGKDD Explor. Newsl.* **2022**, *24*, 61–77. [CrossRef]

180. Wu, Z.; Balloccu, S.; Kumar, V.; Helaoui, R.; Recupero, D.R.; Riboni, D. Creation, analysis and evaluation of annomi, a dataset of expert-annotated counselling dialogues. *Future Internet* **2023**, *15*, 110. [CrossRef]

181. Ching, T.; Himmelstein, D.S.; Beaulieu-Jones, B.K.; Kalinin, A.A.; Do, B.T.; Way, G.P.; Ferrero, E.; Agapow, P.-M.; Zietz, M.; Hoffman, M.M.; et al. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **2018**, *15*, 20170387. [CrossRef] [PubMed]

182. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. *Briefings Bioinform.* **2018**, *19*, 1236–1246. [CrossRef] [PubMed]

183. Zampieri, G.; Vijayakumar, S.; Yaneske, E.; Angione, C. Machine and deep learning meet genome-scale metabolic modeling. *PLoS Comput. Biol.* **2019**, *15*, e1007084. [CrossRef] [PubMed]