OXFORD

Application of deep learning methods in biological networks

Shuting Jin†, Xiangxiang Zeng†, Feng Xia, Wei Huang and Xiangrong Liu

Corresponding author: Xiangrong Liu, Department of Computer Science, Xiamen University, 361005, Xiamen, China.
Tel: 13616048924; E-mail: xrliu@xmu.edu.cn
†These authors contributed equally to this work.

Abstract

The increase in biological data and the formation of various biomolecule interaction databases enable us to obtain diverse biological networks. These biological networks provide a wealth of raw materials for further understanding of biological systems, the discovery of complex diseases and the search for therapeutic drugs. However, the increase in data also increases the difficulty of biological networks analysis. Therefore, algorithms that can handle large, heterogeneous and complex data are needed to better analyze the data of these network structures and mine their useful information. Deep learning is a branch of machine learning that extracts more abstract features from a larger set of training data. Through the establishment of an artificial neural network with a network hierarchy structure, deep learning can extract and screen the input information layer by layer and has representation learning ability. The improved deep learning algorithm can be used to process complex and heterogeneous graph data structures and is increasingly being applied to the mining of network data information. In this paper, we first introduce the used network data deep learning models. After words, we summarize the application of deep learning on biological networks. Finally, we discuss the future development prospects of this field.

Key words: biomolecule; biological information; biological networks; deep learning; deep neural network; graph neural network

Introduction

Since the discovery of the structure of DNA in 1953 (Figure 1), an increasing number of biomolecules and their effects on the characteristics of life have been discovered. In 1958, Crick proposed that RNA was the intermediate carrier of genetic information, and various types and functions of RNA molecules were discovered. After the molecular network regulation of cells was proposed by Jacob and Monod in 1961, an increasing number of biological network regulation patterns were discovered and constructed. Since the 1990s, with the development of various genome sequencing programs and breakthroughs in molecular structure measurement technology and the popularity of the Internet, hundreds of biological databases have sprung up and grown [1]. In the twenty-first century, bioinformatics has become one of the core areas of the natural sciences, and its research involves genomics, proteomics and molecular biology [2]. Biological systems have many networks of different layers and tissue forms, including gene transcriptional regulatory networks, biological metabolism signaling networks and protein interaction networks. In addition, more and more biological networks have been obtained by various experimental research methods according to the correlation between structures,

Shuting Jin is a PhD student in Xiamen University. Her research interest is information mining of biological networks in bioinformatics.
**Xiangxiang Zeng** is a professor in Hunan University. His research interests include bio-computing and bioinformatics.
Feng Xia is a graduate student in Xiamen University. Her research interest is bioinformatics and data mining.
Wei Huang is a graduate student in Xiamen University. His research interest is bio-computing and bioinformatics.
Xiangrong Liu is a professor in Xiamen University. His research interests include bioinformatics and data mining.
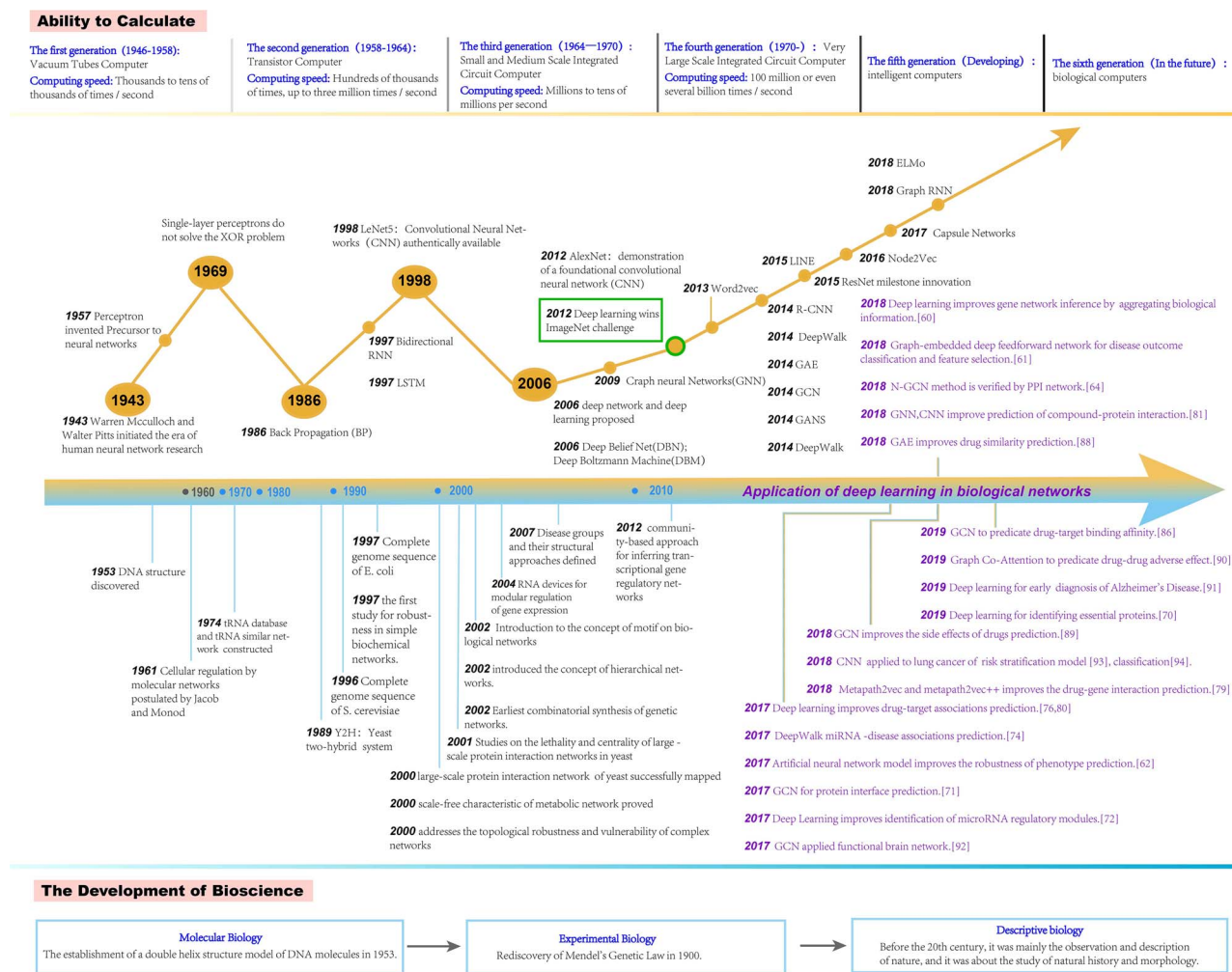
Figure 1. A timeline of milestones in biomolecular networks and deep learning. The development of deep learning and biomolecular networks are illustrated in this figure. The purple font indicates the research on the application of deep learning in biological networks mentioned in the review article. The development of ability of computer is marked at the top and the development of bioscience is marked at the bottom.

relevant features and the interaction between them. Such networks include drug–target association networks and disease–biomolecule association networks. Biological networks research also involves biomolecules and various aspects related to biomolecules. Nodes in biological networks can represent proteins, genes, diseases and some drugs associated with targets. The edges in the network correspond to various biochemical physical or functional interactions between nodes.

Biological networks provide a map of the entire cell and organism, and a large number of collective biological behaviors and co-expressed features can be systematically studied through biological networks. Calculations of biological networks can be used to predict associations between biomolecules, such as interactions between genes, associations between RNAs and diseases and interactions between compound molecules and target proteins. However, due to the complexity and high latitude nature of biological data itself, analyzing these data reasonably and obtaining information that humans can understand and use from the data are major challenges. This issue requires starting a multidisciplinary approach to find more ways to deal with biological big data and to analyze the information contained in the data. Machine learning, a method of analyzing and reasoning which learns from data, trains models and then uses

the models to make predictions, has made good progress in recent years. Machine learning is becoming a crucial part of modern biology research because it can construct data analysis techniques of predictable models from multidimensional and complex data-sets. Machine learning research on biological networks data includes [3–15]. Deep learning is a subfield of machine learning, and its concept comes from the research of artificial neural networks [16]. The framework of deep learning is the inner law and representation level of learning sample data. Through multilayer processing, the initial low-level feature representation is gradually transformed into a high-level feature representation, and a simple model can be used to complete complex classification and other learning tasks. Deep learning systems can understand and learn complex representations directly from raw data, making them useful in many disciplines [17]. At present, the application of deep learning in biological data is mostly based on the sequence of biomolecules, the text mining of medical records and the calculation of disease images, among others. The applications of deep learning on biological networks remain few. Therefore, understanding and comparing existing deep learning methods can help us propose additional deep learning models to mine biological networks information.
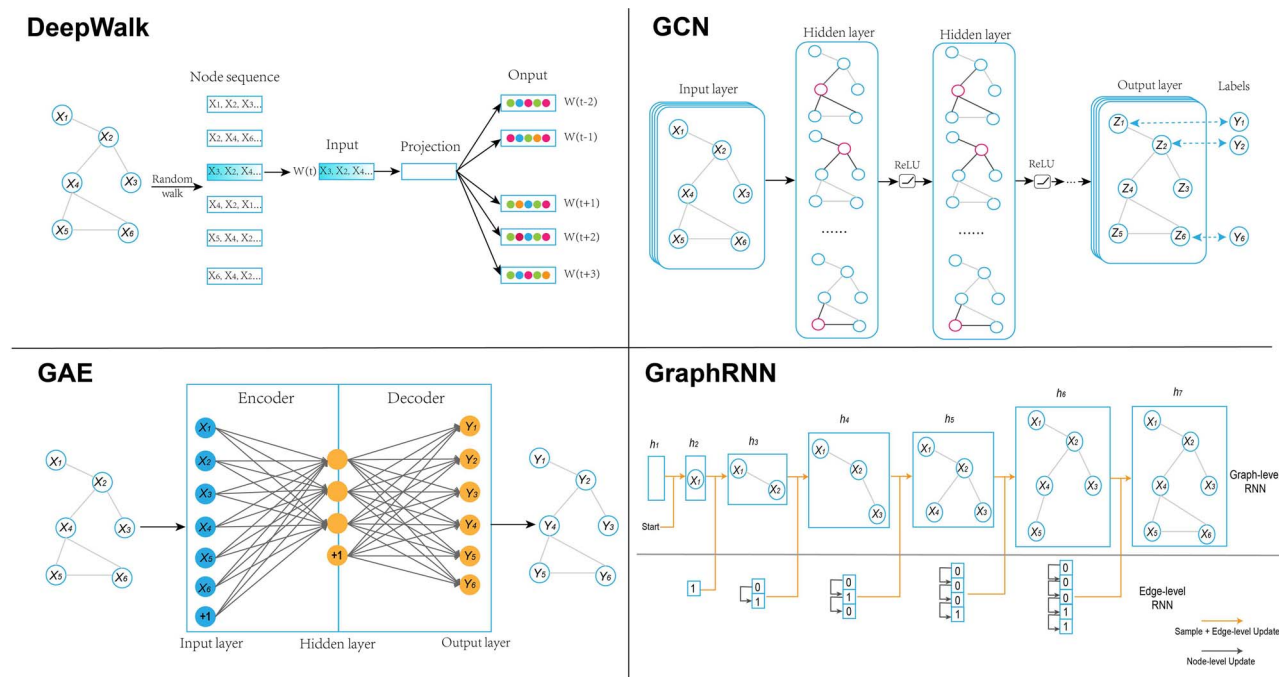
Figure 2. Key models of deep learning techniques for network data. The figure contains the four classic deep learning models for network data introduced in this article. Including DeepWalk, Graph AutoEncoder (GAE), Graph Convolution Network (GCN) and Graph Recurrent Neural Network (Graph RNN).

The main purpose of this review is to introduce the existing applications of deep learning on biological networks and the prospects for their possible applications, providing a reference for researchers interested in this field. The remainder of this paper is summarized as follows: in the section entitled 'Basics of deep learning models on graphs', we will introduce several deep learning models that are applied in network data. In the section 'Application of deep learning on biological networks', we will introduce some of the main works about deep learning on biological networks research. This review is summarized and discussed in the last section.

Basics of deep learning models on graphs

A key aspect of deep learning is that deep neural network realizes the transformation of data at each layer through training and iteratively adjusting its internal parameters to minimize prediction errors. It can understand and learn complex representations from raw data, and it has shown obvious effects in speech recognition, image processing, natural language processing and other fields. Abstracting real-world network as a graph data structure is the modeling of a group of objects (nodes) and their relations (edges). The early model of using deep learning techniques for network data computation is DeepWalk. Based on the DeepWalk, algorithms such as node2vec and Metapath2vec have also been developed. Recently, a large amount of research on the direction of the data structure of the graph has greatly promoted the analysis technique of the graph data structure [18]. Many studies have proposed the model for calculating the weighted average of node neighbor information based on the neural network processing method. These models of processing the graph data structure, using the neural network are collectively referred to as the Graph Neural Networks (GNNs). The concept of GNN was first proposed in 2009, Scarselli et al. [19] extended the existing neural network for processing graph structure data.

Later, a number of GNN methods have been developed for graph data structures, including Graph Auto-encoders (GAEs) [20–24], Graph Convolutional Neural Networks (GCNs) [25–30] and Graph Recurrent Neural Networks (Graph RNNs) [31–33], etc. A brief introduction to the DeepWalk [34] and several typical GNN models (Figure 2) is as follows. The advantages, disadvantages, type, application scenarios and extended models of each model are summarized in the Table 1.

DeepWalk

Based on the Woed2vec node vectorization model, DeepWalk is the first network embedding algorithm using deep learning technology [34]. It was used to learn a latent space representation of social interactions. The main idea is to use the method of random walk on the network data structure to obtain the sequence of the node. This sequence is equivalent to the sentence in the language model, and each node in the sequence is equivalent to the word in the sentence. Then Skip-gram model is used to learn the representation of each node, that is, get the feature vector of the node. Using the feature vector of the nodes, we can classify the nodes or predict the possible associations of the nodes. The DeepWalk algorithm makes full use of the information of the random walk sequence only depending on the local information in the network structure. Therefore, it can be applied to distributed and online systems, avoiding the problem of high computation time and space consumption by using the adjacency matrix. Network Embedding algorithms similar to DeepWalk are Node2vec [35], Walklets [36], Maxmargin DeepWalk (MMDW) [37], metapath2ve/metapath2vec++ [38], etc.

Graph AutoEncoders (GAEs)

The most primitive AutoEncoder network is a three-layer feedforward neural network structure, which is composed of input

Table 1. Some main distinctions of deep learning models on graphs

Model	Advantages	Disadvantages	Type	Application scenarios	Extended models
DeepWalk	Performing well when the data volume is relatively sparse; ability to implement parallel operations; strong scalability.	Lack of generalization capabilities; not applicable to dynamic graphs where the nodes in the graph are constantly changing; computationally expensive and nonoptimal for large graphs.	Unsupervised	Undirected graph; sparse network.	Node2vec [35], Walklets [36], Maxmargin DeepWalk (MMDW) [37], metapath2ve/ metapath2vec++ [38], etc.
GAEs	Fast calculation; ability to discover underlying attributes in the data; easier to expand into deep structures.	The effect of embedding may not be good enough. Many autoencoder-based models need to store the neighbor information of all nodes, which is less efficient for big graphs.	Unsupervised	Graph node representation suitable for learning unsupervised information.	SAE [40], SDEE [21], DNGR [22], GC-MC [23], DRNE [24], etc.
GCNs	Weight sharing, parameter sharing; local connectivity; complexity is greatly reduced.	Poor scalability; limited to shallow layers.	Semi-supervised	The application scenarios are pervasive. In addition to undirected graphs, the improved GCN can also be used for directed graphs, which can handle such as physical systems, molecular structure, knowledge graphs, etc.	(ChebNet) [26], (AGCN) [27],(LGCN) [28], (DCNN) [29], DGCN [30], etc.
Graph RNNs	Good robustness and versatility; capable of representing graphs with different sizes and different sequence lengths; capturing complex dependencies between nodes.	RNN's computing power will be constrained by memory, bandwidth, etc.; low efficiency.	Semi-supervised/ Unsupervised	Suitable for learning the node representation of dynamic graphs; Can be used for graph generation problems; can be combined with other models to obtain space and time graph information.	Graph RNN [32], (DGNN) [33], (RGCNN) [44], (sRGCNN) [45], etc.

layer, hidden layer and output layer. It is an unsupervised neural network model that uses a backpropagation algorithm to compress the input into a latent spatial representation and then reconstruct the output through this representation. The core function of this model is to be able to learn the deep representation of the input data, its input and output are X. AutoEncoder and its variants are widely used in unsupervised learning [39] and are also very useful for node representation in graph data structures that learn unsupervised information. The first to use AutoEncoder in the graphs is Sparse AutoEncoder, SAE [40]. The basic idea of SAE is to use the adjacency matrix or its variant as the original feature of the nodes, so that AutoEncoder is used as the dimension reduction method to learn the low-dimensional node representation. The number of nodes in the input and output of the SAE model is the same, which can be defined as the reconstruction of the input data with the neural network and the extraction of the essence. Experiments have shown that SAE is superior to nondeep learning baseline models [31]. In addition,

SDEE [21], DNGR [22], GC-MC [23], DRNE [24] and other models for processing graph information based on AutoEncoder were proposed.

Graph convolution networks (GCNs)

GCNs is a general architecture that most graph neural network models have, applying convolution operations to graph data structures. The GCNs framework is mainly divided into two categories: Spectral Method and Spatial Method (Nonspectral Method) [41]. The Spectral Method uses the spectral decomposition method to apply the Laplacian matrix decomposition of the graph to collect information from the nodes. The Spatial Method directly uses the topology of the network graph to collect information based on the neighbor information of the graph. The goal of these models is to learn a mapping of signals or features on the graph data structure. Its input is a characterization of

each node, and is trained through the model to produce a node-level output. Graph-level output can be modeled by introducing some pooling operations [41]. Compared with the basic GNN, GCN only has some subtle changes in the aggregate function. In addition, GCNs also includes models like Chebyshev Spectral CNN (ChebNet) [26], Adaptive Graph Convolution Network (AGCN) [27], Large-scale Graph Convolution Networks (LGCN) [28], Diffusion Convolution Neural Networks (DCNN) [29], DGCN [30], etc.

Graph recurrent neural networks (Graph RNNs)

RNN is a special neural network structure, which is proposed according to the idea 'human cognition is based on past experience and memory'. RNN is called a recurrent neural network because the current output of a sequence is also related to the previous output in the training model of RNN [42]. It not only considers the input of the previous moment, but also gives the network a 'memory' function for the previous content. The specific form of expression is that the network memorizes the previous information and applies it to the calculation of the current output. It also includes the output of the hidden layer at the previous moment. The latest research gradually applies RNN to the model of the graph neural network. Zhang et al. [31] called the architecture of applying RNN at the graph level as Graph RNNs, which solved the problem of nonuniqueness, high-dimensionality and complex nonlocal dependence between edges of a given network graph and approximated any distribution of the network graph with the minimum structure assumption. Graph RNN models the graph recursively as a sequence of new nodes and edges to capture the complex joint probability of all nodes and edges in the graph. In addition, You et al. [32] applied Graph RNN to the graph generation problem in 2018. In the same year, Ma et al. [33] proposed the Dynamic Graph Neural Network (DGNN), that is, learn the node representation in dynamic graphs by uses time-aware LSTM [43]. RNN can also be combined with other architectures such as Recurrent Graph CNN (RGCNN) [44] and separable Recurrent Graph CNN (sRGCNN) [45], etc.

Application of deep learning in biological networks

As a hot research topic in recent years, deep learning is suitable for high-dimensional and complex data, so it can be used for biological information mining. For instance, Alipanahi et al. [46] proposed DeepBind, using CNN to predict sequence specificities of DNA- and RNA-binding proteins based on *in vitro* and *in vivo* data. The study shows that sequence specificities can be ascertained from experimental data with deep learning techniques, which offer a scalable, flexible and unified computational approach for pattern discovery. Cuperus et al. [47] trained a convolutional neural network (CNN) on the random library and showed that it performs well at predicting the protein expression of both a held-outset of the random 5' UTRs as well as native *S. cerevisiae* 5' UTRs, and the model additionally was used to computationally evolve highly active 5' UTRs. Before this, the ability to predict protein expression from DNA sequence alone remains poor, so the application of this model will contribute to the development of synthetic biology. Wallach et al. [48] proposed AtomNet, the first CNN designed to predict the bioactivity of small molecules for drug discovery applications. The work has played a great role in the discovery of new medicines and our further understanding of biology. Park et al. [49] proposed a novel method, deepMiRGene, based on RNN learning the palindromic

secondary structure of precursor miRNA. Conventional methods for precursor miRNA identification exploit hand crafted feature sets are obtained by laborious feature engineering,the most important contribution of the approach is that it does not require any painful manual feature engineering. In addition, there are many deep learning applications, which have greatly promoted the development of the biological field. Specific introduction can refer to [50–55], here we mainly introduce the application of deep learning in the biological network data structure.

With the development of research, many deep learning algorithms applied to network data structures have been proposed. Biological networks contain a lot of information among organisms. The exploration of biological networks is important for understanding the internal correlation of biomolecules, drug discovery, disease treatment and the mechanism of action of microorganisms. The deep learning model of these applications on the network data can represent the network results in a multilayered manner, capture the topological features of the known biological networks and combine other heterogeneous information. Through the summary analysis of the existing research, it can provide ideas for mining the information contained in the biological networks by using the deep learning technology in the future. In this section, we will introduce the application of deep learning in biological networks through Genomics data research, Proteomics data research, Transcriptomics data research, Drug discovery, Disease biology and Microbiome data research (Figure 3).

Genomics data research

Genes support the basic structure and performance of biological life and are the basic genetic unit that controls biological traits. Studies have shown that genes work through network integration [56]. Computational methods to discover the function of genes, the interaction between genes and the correlation between genotypes and phenotypes can help better understand the response of drugs, the genetic basis of diseases, etc.

Gene interaction data

The study of gene interactions can provide a way to determine the functional association between genes and their corresponding products and provide a deeper understanding of the potentially important biological phenomena that produces some phenotypes [57–59]. Kishan et al. [60] proposed a deep learning framework, GNE (Gene Network Embedding), to learn embedded representations to unify known gene interaction network and expression data for gene interactions prediction. This method not only finds similar genes, but also infers the characteristics of unknown genes. It also has good performance when the gene network structure is missing. Because the interaction and expression of genes are closely related to disease, Kong et al. [61] proposed a new deep feedforward network classifier embedded feature map information, using gene expression and network data to obtain disease classification. In order to fuse the information of the known feature map into the Deep Neural Network (DNN), a Graph-Embedded Deep Feedforward Network (GEDFN) is proposed. The key idea of this method is to integrate external relational of features into the DNN architecture so that a fixed information sparse connection can be achieved. This method can solve the problem that the genetic network information has thousands of features and only hundreds of training samples, and the scale-free structure of the gene network is not friendly to the establishment of the convolutional neural network.
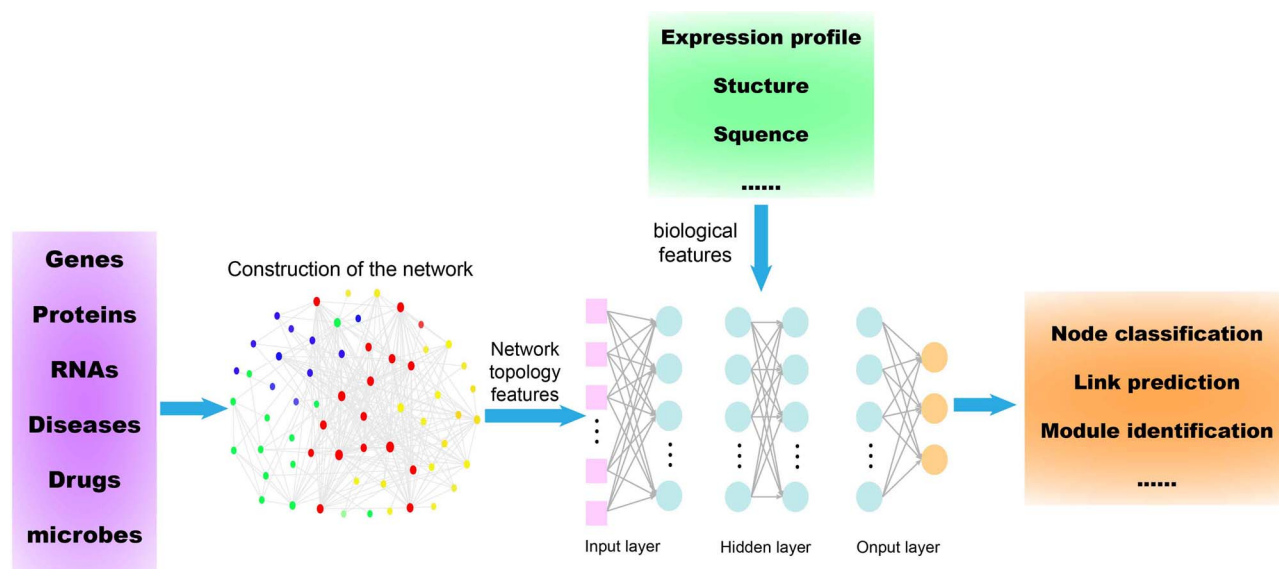
Figure 3. Application framework of deep learning in biological networks. The figure is the basic flowchart of the application of deep learning on biological networks.

Genotype-phenotype association

In addition to the study of gene interactions, research on the association between genes and phenotypes is also the focus of genomic research. Kang et al. [62] developed a regularized Artificial Neural Network (ANN) that encodes the interdependence between genes and their regulatory factors into the classifier structure specifically for predicting phenotypes from gene expression data (e.g. response to therapy). The starting point of this method is a kind of network of gene-regulated interactions. The nodes of the constructed ANN input layer correspond to the genes, the nodes of the hidden layer correspond to the regulators in the network, the output layer consists of a node for binary classification and each node in the hidden layer is connected to the output node. In this method, the induced sparsity in the connections is used to effectively avoid the shortcomings that the conventional ANNs algorithm that requires a large amount of training sample data and over-fitting and can well determine the positive regulatory mechanism and its potential relationship with the phenotype. This kind of transformation of network data nodes into neural network nodes provides a good research idea for dealing with the correlation predictions of regulatory relationships in the biological network.

Proteomics data research

Protein is the material basis of life and is closely related to various life activities. By exploring the mode of action, functional mechanisms, regulation and control of protein population, we can gain a more comprehensive understanding of the disease process, the physiological and pathological processes of cells and the regulatory network and reveal the basic rules of life activities. It can also provide theoretical basis for clinical diagnosis, drug screening, new drug development and personalized medicine.

Protein-protein interactions (PPIs) and protein function

Many of the key functions and life processes in biology are maintained to some extent by different types of PPIs. The popularity of high-throughput experimental methods of protein has

produced a large number of large-scale molecular and functional interaction networks. The connectivity of these networks provides a rich source of information for inferring functional annotation of proteins. PPI datasets as the one of the benchmark datasets and many deep learning models use the PPI network to verify the performance. For example, Grover et al. [35] proposed the Node2vec method; the PPI network was used to perform multilabel classification and link prediction. Hamilton et al. [63] proposed GraphSAGE, a general inductive framework that leverages node feature information to efficiently generate node embedding for unseen nodes. The framework can solve the issue of heterogeneous, large and changing networks. In the approach, the model generalizes completely unseen graphs to use PPI dataset for classification of protein function. Sami et al. [64] proposed the Multiscale Graph Convolution for Semi-supervised Node Classification (N-GCN) was also used in the PPI network data. This method generalized GraphSAGE and proposed the N-SAGE, and resilience to adversarial input perturbations. Chen et al. [65] developed a novel algorithm based on control variables to optimize GCN. The GCN is enhanced by a preprocessing step that uses a method similar to the dropout operation. The verification of the use of data sets also confirmed that the prediction accuracy for classification of protein function has also been greatly improved. Liu et al. [66] proposed an extensible learning method, GeniePath, which has an adaptive path layer with adaptive breadth and depth functions to guide the receptive paths for learning adaptive receptive fields of neural networks defined on permutation invariant graph data. Experiments on large-scale benchmark data prove that this method is superior to the most advanced methods at present, and the classification of PPI network has also achieved good results. The success of GeniePath also shows that it is important to calculate the appropriate data for different nodes to choose the appropriate acceptance path.

In addition to these general deep learning models, there are many studies that improve existing deep learning algorithms to better adapt them to functional prediction of proteins. Kulmanov et al. [67] presented a novel method for predicting protein function, which uses deep learning to learn feature representation from protein sequences and PPI network. This method
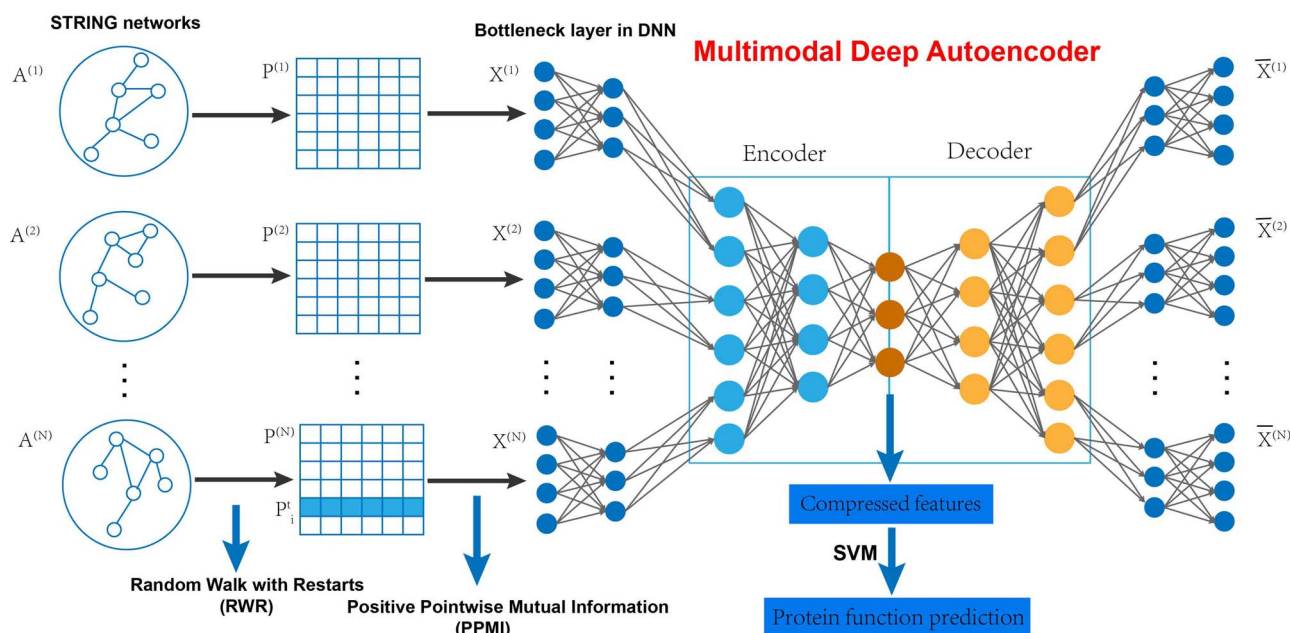
Figure 4. The case study of AutoEncoder model in biological network. The figure shows the deepNF method based on Multimodal Deep Autoencoders (MDA). In the first step, the STRING networks are converted into vectors by the RWR (where P_i^t is a row vector of protein i, whose k-th entry indicates the probability of reaching the k-th protein after t steps) and constructed PPMI matrix. In the second step, the networks are combined via MDA. Finally, low-dimensional features are extracted from the middle layer of the MDA and are used to predict protein function.

is to output information in the Gene Ontology (GO) structure and uses the dependency relationship between GO classes as background information to construct a deep learning model. In order to combine different networks, Gliorijevic et al. [68] proposed deepNF based on Multimodal Deep Autoencoders (MDA), which can extract advanced features of proteins from multiple heterogeneous interactive networks. Firstly, the framework used a Random Walk with Restart (RWR) to convert each network into vectors and then constructed a Positive Pointwise Mutual Information (PPMI) matrix; secondly, it combined PPMI matrices of networks by MDA and extracted the low-dimensional features from the middle layer of the MDA; finally, it predicted protein function by training an SVM classifier on the low-dimensional features (Figure 4). This deep learning-dependent technique has good performance for protein functional prediction and can accurately capture relevant proteins features from complex non-linear networks. Trivodaliev et al. [69] proposed DeePin, using PPI network to predict protein function. This method adapts node2vec to PPI network and preserves the local and global topological information of the PPI network to the maximum extent. All three models predict the function of proteins, and the main difference is that Kulmanov et al. proposed a model that does not rely on manually crafted features but is entirely data-driven, Gliorijevic et al. proposed a model that integrates multiple networks, which is the first method that uses a deep multimodal technique to integrate diverse biological networks, Trivodaliev et al. proposed model performance that is optimal, the Area Under the Curve values are high (>0.79) even though the experimental setup is very simple.

Essential protein

Identify the essential proteins can understand the minimum requirements for cell survival and evolution. Zeng et al. [70] proposed an end-to-end model of deep learning framework to identify the essential proteins. This framework consists of two parts: feature extraction and classification of biological information. Three types of biological data, PPI network, gene expression and subcellular localization are utilized. This framework uses the node2vec technology to process the PPI network. In the study, the author also studied the role of each biological information by deleting a certain kind of data, indicating that the three types of biological information play different roles in identifying essential proteins, but they are of different importance. The most important of these is the embedded data of the PPI network, while the other two data can be used as auxiliary data to improve the recognition performance of the basic protein. This method also illustrates the importance of biomolecule network information and combines different bio-information data with each other to better explore the intrinsic mechanism of action of biomolecules.

Protein interface

Proteins are chains of amino acid residues folded into a three-dimensional structure, giving them biochemical functions. Prediction of the interface between proteins in protein-related research has always been a challenging problem. In order to solve this problem, Fout et al. [71] proposed an image convolution-based method to predict the possible association between protein interfaces. In this study, they represent the protein as a network graph structure where each amino acid residue is a node whose features represent the properties of the residue, and the spatial relationships between residues (distances, angles) are represented as features of the edges. The so-called protein interface prediction problem is simply described as if one residue is derived from a ligand protein and the other residue is derived from a receptor protein, and the task of prediction is to classify node pairs from the representation of two independent network maps of these proteins. The authors classify the amino acid residue nodes in the protein network structure by a graph convolutional network, which makes the
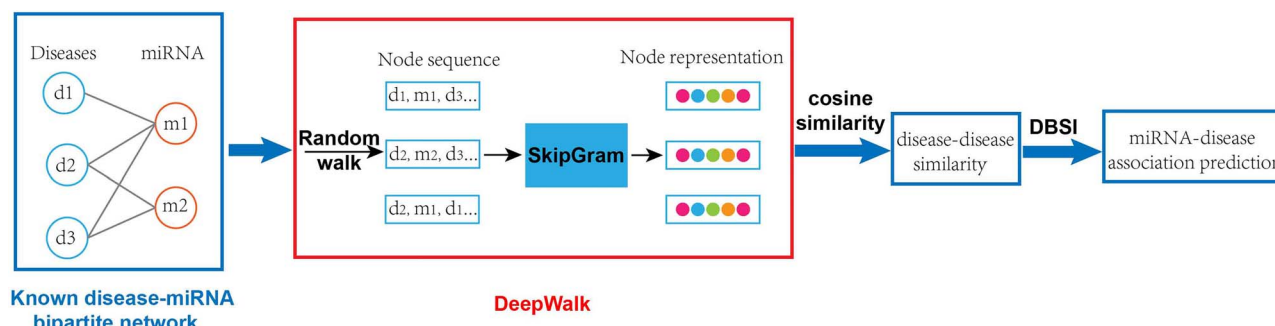
Figure 5. The case study of DeepWalk model in the biological network. The figure shows the miRNA-disease association prediction based on DeepWalk. Firstly, the method uses DeepWalk to determine the similarity of each disease-disease pair based on a known disease-miRNA bipartite network. Then, the drug-based similarity inference (DBSI) is used to calculate the likelihood of association of each candidate disease-miRNA pair.

prediction accuracy reach 0.89. Compared with the method without convolution, the accuracy of protein interface prediction is improved.

Transcription data research

A transcriptome is the sum of all RNAs that a living cell can transcribe. From the RNA level of gene expression study, the studying cell phenotype and function is an important method. Unlike the genome, the definition of the transcriptome contains time and space constraints. The transcriptome profile can provide information on what gene are expressed under what conditions, infer the function of the corresponding unknown gene and reveal the mechanism of action of specific regulatory genes. With this molecular signature based on gene expression profiles, not only can the phenotypic assignment of cells be distinguished, but also biomarkers for various diseases can be found.

MiRNAs belong to the RNA family and are known to inhibit the expression of their target genes. Many studies have proved that abnormal expression of miRNA plays a key role in various complex diseases. Jindal et al. [72] proposed a new method based on Markov graph clustering and deep learning to identify miRNA regulation modules (MRMs). This method utilizes miRNA-target binding information, gene expression data, miRNA expression profiles and PPI data. In order to better process and utilize four different types of data, the algorithm uses the shift and stitch approach [73] to increase the resolution in deep belief networks. Li et al. [74] proposed a similarity-based approach to predict the possible association of miRNA-disease by using the topology of the miRNA-disease network. In the study, the DeepWalk was used to determine the similarity of each disease-disease pairs based on the known miRNA-disease binary network constructed from known database. Then, by using these similarities, new disease-miRNA pairs are predicted and evaluated (Figure 5). The authors used the 22 complex diseases data set by 5-fold cross-validation to obtain the area under the ROC curve scores ranging from 0.805 to 0.937.

Li et al. [75] proposed a novel computational model, called heterogeneous graph convolutional network for miRNA-disease associations (HGCNMDA). The model is based on human PPIs and integrates four biological networks: miRNA–disease, miRNA–gene, disease–gene and PPI network. The result of average predicting precision of this study is 0.966. Besides, there are many other RNAs in the transcriptome, and more and more studies have proved that many RNA dysfunctions will lead to a series of major diseases. Better use of deep technology to discover these RNAs and to mine their functional modules will certainly become a hot research area.

Drug discovery

Drugs are generally designed by identifying a biomolecule target that the drug can function, such as a protein whose activity can be modified by a compound to achieve a beneficial therapeutic effect. Some complex or concomitant diseases usually require treatment with multiple drugs, and the combined use of these drugs may also increase the risk of side effects. Therefore, the identification of drugs-targets, compounds-proteins and drug-drug interactions is very important for drug research and design.

Drug-target

Pharmaceutical compounds may act on targets to potentially treat a variety of diseases. Zong et al. [76] proposed a similarity-based drug-target prediction method. The method uses the DeepWalk method to calculate in the known network of drugs and targets and obtains the feature vectors of the nodes and calculates the similarity correlation between the nodes and then adopts drug-based similarity inference (DBSI) and target-based similarity inference (TBSI) to discover the drug–target associations [77, 78]. Zhu et al. [79] introduced the learning methods into biological heterogeneous network and utilized the representation learning models metapath2vec and metapath2vec++ to predict drug–gene associations through heterogeneous networks constructed by genes, drugs and adverse drug reactions. However, these methods of calculation using an associated network constructed from known drug–target associations can only predict the association between drugs and targets already present in the drug-target associated network and cannot be used to discover new drugs and targets. And these studies did not take advantage of the molecular structural features of the compounds. The development of deep learning has made it possible to characterize and model compounds. For a molecular map of a compound, the atoms are labeled as nodes and the bonds are labeled as edges. The current neural networks can manipulate molecular maps directly to learn new chemical characterizations. In order to improve the representational and predictive power for molecular systems, Gomes et al. [80] proposed the Atomic Convolutional Neural Network (ACNN) to predict protein-ligand binding affinity, which is a general 3-dimensional spatial convolution operation for learning atomic-level chemical interactions directly from atomic
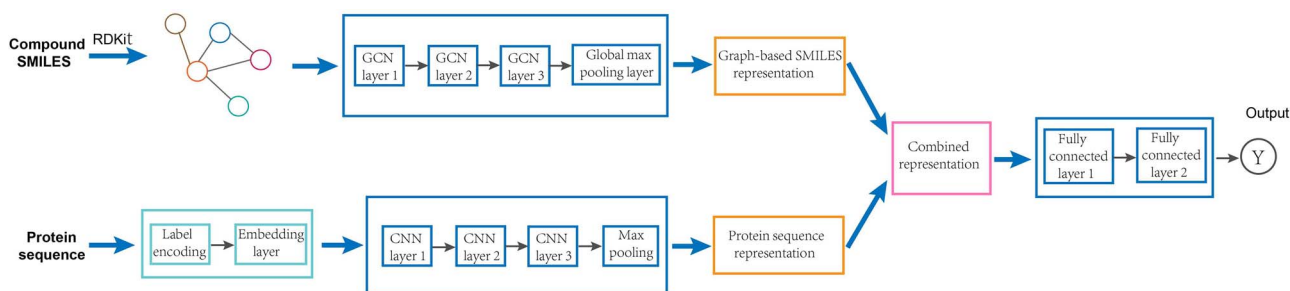
Figure 6. The case study of GCN model in biological network. The figure shows the GraphDTA method based on GCN to predict drug-target binding affinity. Compound SMILES is first converted to a molecular graph by RDKit software, and then GCN is used to learn molecular graph representation. Protein sequence is first categorically encoded and embedded, and then several CNN layers are applied to learn sequence representation. Finally, the two representation vectors are concatenated and undergo various fully connected layers, ended by a regression layer to estimate the output as the drug-target affinity value.

coordinates and demonstrating its application to structure-based bioactivity prediction. Tsubaki et al. [81] developed a new CPI prediction approach based on end-to-end learning by combining a graph neural network (GNN) [82, 83] and a convolutional neural network (CNN) [84] to predict compound-protein interactions. This work also uses the neural attention mechanism [85] to solve the difficult problem of deep model analysis. This mechanism allows considering which subsequences in the protein are more important to the drug when predicting drug interactions. This research provides new insights into end-to-end representation learning and builds a common machine learning model in bioinformatics. The verification of the data set proves that this method is better than the existing end-to-end method, and it also has good performance on the unbalanced data set. Thin et al. [86] proposed GraphDTA to capture the structural information of drugs, possibly enhancing the predictive power of the drug–target binding affinity (DTA). The DTA prediction problem is converted into a regression task where the input is a pair of protein and drug representation, and the output is a continuous value reflecting the affinity binding of the protein to the drug. In this method, a corresponding molecular map is constructed for each input compound string of the drug compound, and the nodes in the figure adopt the atom feature design in DeepChem [87]. The processing of the graphical input to the compound is to represent the input through multiple GCN layers and Graph pooling (Figure 6). These methods preserve the image form of the compound and capture the information about the bonds between the atoms. Studies have also shown that this representation of a molecule as a graph can significantly improve the predictive performance of the drug–target association.

Drug-drug

Association studies between drugs can be used to support clinical medicine, such as inferring some of the side effects, indications and interactions of new drugs. Ma et al. [88] treated each type of drug feature as a view and learned the integrated drug similarity using a multiview Graph AutoEncoder (GAE). In this work, they modeled each drug as a node in the drug association network. For each drug node, the multiview node and edge embedding were obtained by extended graph convolutional networks (GraphCNN) to predict the similarity between drugs. Zitnik et al. proposed Decagon [89], a new multirelational edge prediction model that is a convolutional graph neural network operating in a multi-relational environment to predict drug-drug interactions and their types by constructing PPIs, drug-protein interactions, and drug-drug interactions that side-type

labeled with effect types. Decagon is a general-purpose graph convolutional neural network designed to run on a large multi-modal graph where nodes can be connected through a number of different relation types or applied to link predictions in other domains and problems. Decagon further enhances predictive power by including additional information related to protein targets of interest. Although protein-related ancillary information is very beneficial to the use of algorithms, the cost of obtaining such ancillary information is also high. Deac et al. [90] proposed another graph neural network architecture that relies solely on the molecular structure information of drugs and uses a synergistic attention mechanism to discover possible side effects between drugs. Moreover, this method can set the result to binary or multilabel output, which means that the binary output can be used to determine whether the two drugs will produce a certain side effect, and the multilabel output can also be used to predict the presence or absence of all side effects considered. This approach opens up a new and promising direction for the use of machine learning to analyze multidrug side effects, which is applied to any type of interactive discovery task between structured inputs (such as molecules) in principle, especially when a large number of tagged examples are available.

Disease biology

Making full use of the information contained in biological networks can help us better understand complex diseases. Traditional methods for exploring disease-related biomolecules rely heavily on the identification and characterization of specific aspects of the disease and are unable to fully understand the pathogenesis of the disease and the associated biomolecule associations. The use of biological network data structures for analysis can reveal a more comprehensive pathogenesis and disease-related biomolecules, and can also explore the relationship between diseases through molecular network data.

In the case of Alzheimer's Disease (AD), deep learning has great advantages in diagnosing brain diseases and providing clinical decision support. Ju et al. [91] provided support for clinical diagnosis of AD based on deep learning and brain networks. They represent areas in the brain network as nodes, while edges represent functional links between different areas. The deep AutoEncoder network model is used to extract the recognition features from the brain network. Compared with the traditional methods, this method greatly improves the discriminating ability of MCIs and NCs, and provides decision support for the clinical diagnosis of neurodegenerative diseases, especially AD. Nowadays, many studies model functional connections between

neural pathways or brain regions as graphics, and the design of appropriate graphical-like functions is critical to uncovering patterns of disruption associated with certain brain diseases. Ktena et al. [92] proposed a novel method for learning similarity measure between irregular graphs with known node correspondences relations. This method utilizes the capability of convolutional neural networks and the concept in spectral theory to compute irregular graphs. The authors applied the proposed model for functional brain connectivity graphs from the Autism Brain Imaging Data Exchange (ABIDE) database to distinguish different categories of subjects, proving that this model is feasible and applicable to any problem involving the comparison of the graphs.

Deep learning methods are also used to classify lung cancer using biological networks. Choi et al. [93] developed a new system-level risk stratification model of lung cancer based on gene co-expression networks, which replaces the previous models based on individual prognosis genes. The construction of this model reflects the co-expression pattern of survival-related network module genes through deep CNN. The representative genes of these networks are used to establish a risk stratification model based on deep learning. This study provides a new perspective for integrating gene co-expression network into gene expression characteristics and clinical application of deep learning in genomic data science to predict prognosis. Teppei et al. [94] proposed a convolutional network lung cancer classification method that combines proteomic and transcriptome data and uses convolutional network frameworks to process network information. It is hopeful that gene expression and protein network data can be used to classify lung cancer and achieve better results. These studies will provide a good theoretical support for clinically more targeted treatment of different types of diseases.

Microbiome data research

Microbial populations exist in dynamic interconnected ecosystems. Studying how microbial species coexist and interact in host-related or natural environments is critical to advancing the basic science of microbiology and understanding human health and disease. Related studies have used large microbiome data to infer common interspecies interactions and species-phenotype relationships. However, the rational use of microbiological data related data, the discovery of the principle of association in the microbial world and understanding its impact on environmental diseases are still major challenges.

Deep learning as a powerful form of machine learning has also been used in the field of microbiology. Fredericksen et al. [95] used three-dimensional visualization and deep learning models to reveal that the complex parasitic fungal network in the ant can control the ant's behavior. Reiman et al. [96] proposed a framework PopPhy-CNN that uses the structure of CNNs to predict disease from the abundance distribution of microorganisms. Nguyen et al. [97] proposed that the Met2Img approach also uses the CNN framework, relying on embedding classification abundance as color pixels into images, and uses CNN to predict disease based on the images created. Asgari et al. [98] proposed the MicroPheno framework, a method for predicting the microbial community sample environment or host phenotype based on the k-mer distribution in 16S rRNA data. Lo et al. [99] proposed that MetaNN uses a neural network framework and a new data augmentation technique to mitigate the effects of data overfitting and predict host and position point phenotypes. Ditzler et al. [100] applied DNN and recurrent neural networks (RNN) to host and environmental phenotypic

prediction. Le et al. [101] proposed a model to predict metabolite abundances using only microbe abundances, and that is the first application of neural encoder-decoder for the interpretable of multiomics biological data. However, none of these studies utilized network topology information between microorganisms or between microorganisms and their hosts. Many biologically relevant studies have demonstrated the importance of network topology information for association mining.

Khan et al. [102] extended previous work on phenotypic prediction of microbial communities and proposed three different machine learning models: random forests, DNN, and a novel GCN architecture. The experimental results demonstrate that the use of GCN has better performance for multiclass disease classification of microbial community genomes. The success of the GCN model also means that the geometry involved in microbial phylogeny contains information meaningful for disease classification. Microorganisms as a community in which group growth plays a role, better use of the topology information associated with the data will be of great significance to research in this field. This is also an area worth exploring research work on microorganisms in the future.

Summary of deep learning applied to biological network mentioned

We summarize the application in Table 2.

Challenge and opportunities

Data processing

The development of high-throughput technology makes the acquisition of biological data easier and makes it possible to obtain thousands of data from biomolecules and intermolecular interactions. However, a large amount of data is generated quickly and low cost, which also increases the error rate. Moreover, many biological datasets also have imbalances in sample categories. The quality of biological data requires professional researchers to obtain more accurate data. We should also find ways to overcome data redundancy, imbalance and incomplete calculations to improve the accuracy of prediction.

Heterogeneous information

Biological networks may contain a variety of biomolecules. In addition to biological networks, it is also necessary to combine different biological information to improve the accuracy of calculation, such as binging gene expression profiles, molecular sequence of proteins, molecular structure of drugs, CT images of diseases, etc. It is a key and difficult point for future development to propose effective calculation methods to process these different types of data information and combine them to perform calculations. At present, combining the vector representation of different information learned by using different algorithms is a common method, but we should consider different information processing methods and the difference in the importance of information in the calculation and propose more effective computational models that combine heterogeneous information.

Limitation of deep learning models

Many deep learning algorithms require a large number of training sets. Although the data available in biological systems have grown tremendously, the order of magnitude is still small for many deep learning frameworks, and the advantages of deep

Table 2. Summary of deep learning applied to biological networks mentioned

Application	Classification	Research aim	Data type	DL methods	Result
Genomics Data research	Gene Interaction Data	Gene interaction prediction [60]	Gene interaction network data, gene expression data	GNE (Skip-gram model)	Yeast: AUROC 0.825; AUPR 0.821 E.coil: AUROC 0.940; AUPR: 0.939
		Disease outcome classification [61]	Gene expression matrix, Gene network	Graph-Embedded Deep Feedforward Networks	AUC 0.938
	Genotype-Phenotype Association	Robust phenotype prediction [62]	PPIs; protein-gene interactions, Gene expression data from two clinical phenotypes	ANN	(Balanced Accuracy): cross-validation: ≥0.8 independent test sets: ≥0.6 and ≤0.75
Proteomics research	PPIs and Protein Function	Proposed node2vec [35]	BlogCatalog, PPIs, Wikipedia	node2vec	PPI: Macro-F1 0.1791; AUC_Average 0.7543
		Purposed model marries GCN and random walk for classification objective [64]	Pubmed; Citeseer, Cora, PPIs	N-GCN N-SAGE	PPI (Micro-F1): N-GCN: 0.468 N-SAGE: 0.650
		Developed a novel algorithm to optimize GCN [65]	Citeseer, Cora, PubMed, NELL, Reddit, PPIs	GCN	PPI: Micro-F1 around 0.976
		Proposed GeniePath [66]	Pubmed, BlogCatalog[1], BlogCatalog[2], Alipay dataset, PPIs	GCN	PPI: Micro-F1 0.979
		Protein functions prediction [67]	Protein sequence, PPIs	CNN	DeepGO_selected terms(F max): Biological Processes (BP) 0.40; Molecular Functions (MF) 0.50; Cellular Components (CC) 0.64;
		Protein functions prediction [68]	Human and yeast STRING networks	Multimodal Deep Autoencoder	Human: MF [101–300]: m-AUPR around 0.54; M-AUPR aruond 0.49; Micro-averaged F1 around 0.42; Accuracy around 0.16;
		Protein functions prediction [69]	PPIs	node2vec	The Area Under the Curve values >0.79
	Essential protein	Identifying essential proteins [70]	PPI network, Essential proteins, Gene expression dataset, Subcellular localization dataset	node2vec,RNNs	Accuracy 0.850; Precision 0.680; Recall 0.505; F-measure 0.579; AUC 0.83
	Protein interface	Protein Interface Prediction [71]	Protein as a network graph structure	GCN	AUCs around 0.89
Transcription research	unclassified	Identification microRNA Regulatory Modules [72]	MiRNA-target binding information, Gene expression data, MiRNA expression profiles, PPIs	DBN	able to distinguish 34 enriched modules from the datasets
		MicroRNA-Disease Association prediction [74]	Disease-miRNA association	DeepWalk	5-fold ROC ranging from 0.805 to 0.937
		MicroRNA-Disease Association prediction [75]	MiRNA–disease, MiRNA–gene, Disease–gene, PPI network	GNN	AUC 0.9626; Average precision 0.9660
Drug discovery	Drug-Target	Drug-target association prediction [76]	Drugs, Targets, Drug–target associations, Disease-gene associations	DeepWalk	10-fold AUC ROC score 98.96%; Monte Carlo cross-validation AUC ROC score with LTN 99.25%
		Drug-gene interaction prediction [79]	Similarity between drugs, Similarity between genes, Drug–gene interaction, Drug–ADR interaction	Metapath2vec, Metapath2vec++	(AUROC): Metapath2vec 0.8093; Metapath2vec++ 0.8367;

(Continued)

Table 2. Continued.

Application	Classification	Research aim	Data type	DL methods	Result
		Protein-ligand binding affinity prediction [80]	3D crystal structures, protein-ligand complexes	ACNN	PDBBind core test sets (MUE [kcal/mol]): Random 0.774; Stratiified 0.997; Scaffold 0.993; Temporal 0.974;
		Compound-protein Interaction prediction [81]	Compound molecular map, Protein sequence	GNN, CNN	C.elegans dataset_Negative ratio 1: AUC 0.978; Precision 0.938; Recall 0.929
		Drug–target binding affinity prediction [86]	Compound molecular graph, Protein sequence	GCN, CNN, GAT	CI measures: GAT and GIN 0.892 and 0.893
Drug-Drug	Drug Similarity Integration [88]	DDI, Label Side Effect, Off-Label Side Effect, Chemical Structure		GAE, GraphCNN	Dataset 1_Multiple Views_Test Split (25%): (ROC-AUC): AttSemiGAE around 0.798; AttTransGAE around 0.785; (PR-AUC): AttSemiGAE around 0.655; AttTransGAE around 0.687;
		Modeling polypharmacy side effects [89]	The human PPI network, Proteins and drugs, Drug-drug, Drug-protein, Drug-side effect	Graph convolutional encoder	AUROC 0.872; AUPRC 0.832; AP@50 0.803
		Drug-drug Adverse Effect Prediction [90]	The datasets used by [89]	Graph Co-Attention	(10-fold cross-validation): MHCADDI 0.882; MHCADDI-ML 0.819
Disease biology	unclassified	Early Diagnosis of Alzheimer's Disease [91]	Alzheimer's disease neuroimaging (ADNI) association data base	DNN, Autoencoder	(AUC): R-fMRI time series data 0.6191; Correlation coefficient data 0.9164
		Purposed distance metric learning using GCN: Application to Functional Brain Networks [92]	The Autism Brain Imaging Data Exchange (ABIDE) database	GCN	(AUC): All site 0.58; site 6 0.70; site 9 0.71; site 14 0.66; site 16 0.57; site 18 0.98;
		Risk Stratification Model for Lung Cancer [93]	Gene Expression Data	CNN	DL-based model_test set 1: index 0.709 ± 0.042, p = 0.004; NetScore_test set 2: C-index = 0.651 ± 0.042, p = 0.003;
		Lung cancer classification [94]	PPIs, Gene expression, BEC specimens, Gene expression profiles for lung cancer	Spectral clustering-CNN	Accuracy: 0.832; Sensitivity: 0.890; Specificity: 0.773; Precision:0.797
Microbiome data research	unclassified	Multiclass disease classification from microbial whole-community metagenomes [102]	Utilizing 5643 aggregated, annotated whole-community metagenomes from 19 different diseases	GCN	Average test-set accuracy 75%; Average AUC 92.1%; Average AUPR 50%;

learning training cannot be exploited. In addition to selecting more orders of magnitude and more accurate information, we also need to improve the deep learning models to solve some small and sparse problems in the biological data set. Deep learning methods also have difficulty in explaining the output of their models from a biological point of view. But many calculations for biological networks not only want to get the final result but also hope to find potential biological mechanisms and functional networks system that form this result. So how to turn these methods into a biological interpretation and analysis is also an important challenge.

In future research, we can combine deep learning methods with other methods to solve existing problems. For example, the Synthetic Minority Oversampling Technique (SMOTE) [103, 104] algorithm is used to deal with the imbalance of data. Generative Adversarial Networks (GAN) [105–107] is used to generate more data similar to the training set to meet deep learning that requires a large amount of training data. Text mining algorithms can be used to obtain relevant information of nodes in biological networks to improve computing performance [108, 109]. Combining knowledge graphs [110–112] with deep learning models can improve the interpretability of model calculations.

Conclusion

With the deepening of omics research, the development of high-throughput screening research methods and technologies has led to the rapid development of biomolecule network research.

The formation of various biomolecule interaction libraries provides an opportunity to use biological network research as well as the challenge, because the increase of data also increases the difficulty of extracting useful information from these huge biological data. Therefore, finding more suitable algorithm technology has become the focus of research in this field. Deep learning has been developed to handle the functions of speech recognition, image recognition and other functions of millions of inputs. It has also been used in many biological fields for drug discovery, mining biomolecule markers and so on.

This paper mainly summarizes the application of deep learning in biological network data. Previous review papers have explored the availability of deep learning for biological networks information mining [113, 114], but due to the limitations of actual research, there are no detailed introduction examples of research. Here, through the introduction of recent research methods, people can better understand and discover the available fields and practicality of deep learning on biological network data. From these studies, it can be seen that in addition to the typical deep learning method using random walk or matrix vector processing network structure, there many other studies map nodes in the biological network into the input layer of the neural network according to the special properties of the biometric network. Nodes and hidden layer nodes, through the weights obtained by neural network training, obtain the possibility of the existence of regulatory associations between biomolecules. Some studies combine the topology properties of network with the feature information of node. The deep learning model that processes network structure is targeted to learn the topological information of biological network, and other methods are used to process molecular characteristic information, which is combined to improve computing performance. The idea is also the trend of future research and development, which integrates multiple heterogeneous information, such as network structure, molecular properties and so on. Existing researches illustrate the role of deep learning in biological network analysis. The use of deep learning methods to process biological networks can improve the predictive power of some applications in different biological fields. Through the review of the application of deep learning in biological networks, we also summarize the challenges faced by current deep learning applications on biological networks, and look forward to the future development direction.

Key Points

- The use of biological network data can capture the association properties between biomolecules. Combining biological network data with other biological information can improve predictive performance.
- The introduction of graph deep learning brings a new direction of network data structure processing, and can handle large, multidimensional, complex biological data.
- Combining other algorithms combined with deep learning models can overcome the impact of some data quality problems and improve the applicability of deep learning algorithms.

Funding

References

1. Serena C, Lorenzo G, Taratufolo MC, *et al*. Development of a multiple loci variable number of tandem repeats analysis (MLVA) to unravel the intra-Pathovar structure of pseudomonas syringae pv. Actinidiae populations worldwide. *PLoS One* 2015;**10**:2018–25.
2. Kanehisa M, Bork P. Bioinformatics in the post-sequence era. *Nat Genet* 2003;**33**(3):305–10.
3. Plaimas K, Eils R, Konig R. Identifying essential genes in bacterial metabolic networks with machine learning methods. *BMC Syst Biol* 2010;**4**:56.
4. Hor CY, Yang CB, Yang ZJ, *et al*. Prediction of protein essentiality by the support vector machine with statistical tests. *Evol Bioinformatics Online* 2013;**9**:387–416.
5. Nandi S, Subramanian A, Sarkar RR. An integrative machine learning strategy for improved prediction of essential genes in Escherichia coli metabolism using flux-coupled features. *Mol BioSyst* 2017;**13**:1584.
6. Luo P, Tian LP, Ruan J, *et al*. Identifying disease genes from PPI networks weighted by gene expression under different conditions. In: *IEEE International Conference on Bioinformatics & Biomedicine*. 2017.
7. Zhang Z, Song J, Tang J, *et al*. Detecting complexes from edge-weighted PPI networks via genes expression analysis. *BMC Syst Biol* 2018;**12**:40.
8. Wang YH, Zeng JY. Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics* 2013;**29**:126–34.
9. Schrynemackers M, Küffner R, Geurts P. On protocols and measures for the validation of supervised methods for the inference of biological networks. *Front Genet* 2013;**4**: e169–74.
10. Xu S, Barnes RO, Li C, *et al*. BMRF-net: a software tool for identification of protein interaction subnetworks by a bagging Markov random field-based method. *Bioinformatics* 2015;**31**:2412–4.
11. Mangan NM, Brunton SL, Proctor JL, *et al*. Inferring biological networks by sparse identification of nonlinear dynamics, IEEE transactions on molecular. *Biol Multi-Scale Commun* 2016;**2**:52–63.
12. Fogelberg C, Palade V. GreenSim: a network simulator for comprehensively validating and evaluating new machine learning techniques for network structural inference. 2010.
13. Jeng B, Chen JX, Liang TP. Applying data mining to learn system dynamics in a biological model. *Expert Systems with Applications* 2006;**30**(1):50–58.
14. Cho H, Berger B, Peng J. Diffusion component analysis: Unraveling functional topology in biological networks. *Comput Therm Sci* 2016;**9029**:62–4.
15. Yates PD, Mukhopadhyay ND. An inferential framework for biological network hypothesis tests. *BMC Bioinf* 2013;**14**:94.
16. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;**61**:85–117.
17. Najafabadi MM, Villanustre F, Khoshgoftaar TM, *et al*. Deep learning applications and challenges in big data analytics. *J Big Data* 2015;**2**:1.

18. Zhou J, Cui G, Zhang Z, *et al*. Graph neural networks: a review of methods and applications.

19. Scarselli F, Gori M, Tsoi AC, *et al*. The graph neural network model. *IEEE Trans Neural Netw* 2009;**20**:61–80.

20. Kipf TN, Welling M. Variational graph auto-encoders. 2016.

21. 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. ACM* 2016:1225–34.

22. CaoWL S, and Xu Q. Deep neural networks for learning graph representations. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* 2016:1145–52.

23. R v d B, Thomas NK, Welling M. Graph convolutional matrix completion. arXiv preprint arXiv:1706.02263, 2017.

24. Tu K. P C, Wang X., Yu P. S, Zhu W. Deep recursive network embedding with regular equivalence. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM* 2018:2357–66.

25. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2016.

26. Defferrard M, Xavier B, Vandergheynst P. *Convolutional neural networks on graphs with fast localized spectral filtering in NIPS*, 2016, 3844–52.

27. Li R, Sheng W, Zhu F, *et al*. Adaptive graph convolutional neural networks. 2018.

28. 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*. 2018, p. 1416–24.

29. Atwood J, Towsley D. Diffusion-convolutional neural networks. *Comput Therm Sci* 2015.

30. Chenyi Zh, Qiang Ma. Dual graph convolutional networks for graphbased semi-supervised classification. In: *Proceedings of the 2018 World Wide Web Conference* 2018:pp. 499–508.

31. Zhang Z, Cui P, Zhu W. *Deep Learning on Graphs: A Survey*. 2018.

32. You J, Ying R, Ren X, *et al*. *GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models* 2018.

33. Yao Ma, Ziyi Guo, Zhaochun Ren, Yihong Eric Zhao, Jiliang Tang, Dawei Yin. Dynamic graph neural networks. arXiv preprint 2018;arXiv:1810.10627.

34. Perozzi B, Al-Rfou R, Skiena S. DeepWalk:online learning of social representations. In: *Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*. 2014.

35. Grover A, Leskovec J. node2vec: scalable feature learning for networks. *KDD* 2016;**2016**:855–64.

36. Perozzi B, Kulkarni V, Chen H, *et al*. Don't Walk, Skip! Online learning of multi-scale network embeddings. In: 2016.

37. Tu C, Zhang W, Liu Z, *et al*. Max-margin deepwalk: discriminative learning of network representation. In: *International Joint Conference on Artificial Intelligence*. 2016.

38. Dong Y, Chawla NV, Swami A, *et al*. metapath2vec: Scalable representation learning for heterogeneous networks. In: *Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*. 2017.

39. Vincent P, Larochelle H, Bengio Y, and Manzagol PA. Extracting and composing robust features with denoising autoencoders. In: *International Conference on Machine Learning* 2008:1096–103.

40. Tian F., Gao B, Cui Q, Chen E, and Liu TY. Learning deep representations for graph clustering. In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* 2014.

41. Wu Z, Pan S, Chen F, *et al*. *A Comprehensive Survey on Graph Neural Networks*. 2019.

42. Graves A, Mohamed AR, Hinton GJA, *et al*. Speech recognition with deep recurrent neural networks. Acoustics, Speech, and Signal Processing, 1988. ICASSP-88. 1988 International Conference on, 2013;**38**.

43. Baytas IM, Xiao C, Zhang X, *et al*. Patient Subtyping via Time-Aware LSTM Networks. In: *Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*. 2017.

44. Te G, Hu W, Zheng A, *et al*. Rgcnn: Regularized graph cnn for point cloud segmentation. In: *Proceedings of the 26th ACM international conference on Multimedia*. 2018, p. 746–54.

45. Monti F, Bronstein MM, Bresson X. *Geometric Matrix Completion with Recurrent Multi-Graph Neural Networks*, 2017.

46. Alipanahi B, Delong A, Weirauch MT, *et al*. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol* 2015;**33**:831–8.

47. Cuperus JT, Groves B, Kuchina A, *et al*. Deep learning of the regulatory grammar of yeast 5′ untranslated regions from 500,000 random sequences. *Genome Research* 2017;**27**(12).

48. Wallach I, Dzamba M. Heifets AJapa. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery [J]. arXiv preprint arXiv:1510.02855, 2015.

49. Park S, Min S, Choi H, *et al*. deepMiRGene: deep neural network based precursor microrna prediction. arXiv preprint arXiv:1605.00017, 2016.

50. Yue T, Wang H. Deep Learning for Genomics: A Concise Overview. arXiv preprint arXiv:1802.00810, 2018.

51. Mamoshina P, Vieira A, Putin E, *et al*. Applications of deep learning in biomedicine. *ACS Mol Pharmaceut*. 5b00982.

52. Mahmud M, Kaiser MS, Hussain A, *et al*. *Applications of Deep Learning and Reinforcement Learning to Biological Data*. 2018.

53. Sarah W. Deep learning for biology [J]. *Nature* **554** (7693):555–7.

54. Jing Y, Bian Y, Hu Z, *et al*. Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era. *Aaps Journal* 2018;**20**:58.

55. Li P, Peng M, Bo L, *et al*. The advances and challenges of deep learning application in biological big data processing. **13**:352–9.

56. Forsberg SK, Bloom JS, Sadhu MJ, *et al*. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat Genet* 2017;**49**:497–503.

57. Boucher B, Jenna S. Genetic interaction networks: better understand to better predict. *Front Genet* 2013;**4**:290.

58. Mani R, St Onge RP, Hartman JL, *et al*. Defining genetic interaction. *Proc Natl Acad Sci U S A* 2008;**105**: 3461–6.

59. Lage K. Protein-protein interactions and genetic diseases: the interactome. *Biochim Biophys Acta Mol Basis Dis* 1842;**2014**:1971–80.

60. Kishan KC, Li R, Cui F, *et al*. GNE: a deep learning framework for gene network inference by aggregating biological information. *BMC Syst Biol* 2019;**13**.

61. Kong Y, Yu T. A graph-embedded deep feedforward network for disease outcome classification and feature selection using gene expression data. *Bioinformatics* 2018;**34**(21):3727–37.

62. Kang T, Ding W, Zhang L, *et al*. A biological network-based regularized artificial neural network model for robust phenotype prediction from gene expression data. *BMC Bioinf* 2017;**18**:565.

63. Hamilton WL, Ying R, Leskovec J. *Inductive Representation Learning on Large Graphs* 2017.

64. Abu-El-Haija S, Kapoor A, Perozzi B, *et al*. *N-GCN: Multi-scale Graph Convolution for Semi-supervised Node Classification*. 2018.

65. Chen J, Zhu J, Song L. Stochastic Training of Graph Convolutional Networks with Variance Reduction.

66. Liu Z, Chen C, Li L, *et al*. GeniePath: Graph Neural Networks with Adaptive Receptive Paths. 2018.

67. Kulmanov M, Khan MA, Hoehndorf R. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics* 2017;**34**:660–8.

68. Gligorijevic V, Barot M, Bonneau R. deepNF: deep network fusion for protein function prediction. *Bioinformatics* 2017;**34**.

69. JM TK, Kalajdziski S. Deep Learning the Protein Function in Protein Interaction Networks. In: *International Conference on Telecommunications, 2018, Springer, Cham*, 2018, 2185–97.

70. Zeng M, Li M, Fei Z, *et al*. A deep learning framework for identifying essential proteins by integrating multiple types of biological information. *IEEE/ACM Trans Comput Biol Bioinf* 2019;1–1.

71. Fout A, Byrd J, Shariat B, *et al*. Protein Interface prediction using graph convolutional networks. In: *Advances in Neural Information Processing Systems 30* (Nips 2017) 2017, 30.

72. Jindal V. A deep learning framework for identification of microRNA regulatory modules: student research abstract. In: *Symposium on Applied Computing*. 2017.

73. Sermanet P, Eigen D, Zhang X, *et al*. Overfeat: integrated recognition, localization and detection using convolutional networks. arXiv preprint. 2013.

74. Li GH, Luo JW, Xiao Q, *et al*. Predicting MicroRNA-disease associations using network topological similarity based on DeepWalk. *IEEE Access* 2017;**5**:24032–9.

75. Li C, Liu H, Hu Q, *et al*. A novel computational model for predicting microRNA–disease associations based on heterogeneous graph convolutional networks. *Cell* 2019; **8**:977.

76. Zong N, Kim H, Ngo V, *et al*. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics* 2017;**33**.

77. Yamanishi Y, Araki M, Gutteridge A, *et al*. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;**24**: I232–40.

78. Chen X, Liu MX, Yan GY. Drug-target interaction prediction by random walk on the heterogeneous network. *Mol BioSyst* 2012;**8**:1970–8.

79. Zhu S, Bing J, Min X, *et al*. Prediction of drug–gene interaction by using Metapath2vec. *Front Genet* 2018;**9**:248.

80. Gomes J, Ramsundar B, Feinberg EN, *et al*. *Atomic Convolutional Networks for Predicting Protein-Ligand Binding Affinity*, 2017.

81. Tsubaki M, Tomii K, Sese J. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* 2019;**35**:309–18.

82. Scarselli F, Gori M, Tsoi AC, *et al*. The graph neural network model. *IEEE Trans Neural Netw* 2009;**20**:61–80.

83. Kearnes S, McCloskey K, Berndl M, *et al*. Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des* 2016;**30**:595–608.

84. Kim Y. Convolutional neural networks for sentence classififification. arXiv preprint 2014;arXiv:1408.5882.

85. Dzmitry Bahdanau KC, Bengio Y. *Neural Machine Translation by Jointly Learning to Align and Translate*. San Diego, CA: ICLR, 2015.

86. Thin Nguyen HL, Venkatesh S. GraphDTA: prediction of drug-target binding affinity using graph convolutional networks. *bioRxiv* 2019.

87. Bharath Ramsundar PE, Walters P, Pande V. Deep Learning for the Life Sciences. In: *Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More*. Sebastopol, CA: O'Reilly Media, 2019.

88. Ma T, Cao X, Zhou J, *et al*. *Drug Similarity Integration Through Attentive Multi-view Graph Auto-Encoders*, 2018.

89. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;**34**:i457–66.

90. Deac A, Huang YH, Veličković P, *et al*. *Drug-Drug Adverse Effect Prediction with Graph Co-Attention*, 2019.

91. Ju R, Hu C, Zhou P, *et al*. *Early Diagnosis of Alzheimer's Disease Based on Resting-State Brain Networks and Deep Learning, IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019.

92. Ktena SI, Parisot S, Ferrante E, *et al*. Distance Metric Learning Using Graph Convolutional Networks: Application to Functional Brain Networks. In: *International Conference on Medical Image Computing & Computer-assisted Intervention*. 2017.

93. Choi H, Na KJ. A risk stratification model for lung cancer based on gene Coexpression network and deep learning. *Biomed Res Int* 2018;**2018**:2914280.

94. Matsubara T, Ochiai T, Hayashida M, *et al*. Convolutional Neural Network Approach to Lung Cancer Classification Integrating Protein Interaction Network and Gene Expression Profiles. In: *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering* (BIBE). 2018.

95. Fredericksen MA, Zhang YZ, Hazen ML, *et al*. Three-dimensional visualization and a deep-learning model reveal complex fungal parasite networks in behaviorally manipulated ants. *Proc Natl Acad Sci U S A* 2017;**114**: 12590–5.

96. AAM DR, Dai Y. Popphy-cnn: a phylogenetic tree embedded architecture for convolution neural networks for metagenomic data. *bioRxiv* 2018;257931.

97. EP THN, Chevaleyre Y, Sokolovska N, *et al*. Disease classification in metagenomics with 2d embeddings and deep learning. arXiv preprint;arXiv:1806.09046.

98. Asgari E, Garakani K, Mchardy AC, *et al*. MicroPheno: predicting environments and host phenotypes from 16S rRNA gene sequencing using a k-mer based representation of shallow sub-samples. *Bioinformatics* 2018;**34**:i32–42.

99. Lo C, Marculescu R. MetaNN: accurate classification of host phenotypes from metagenomic data using neural networks. *Bmc Bioinformatics* 2019;**20**:314.

100. Ditzler G, Polikar R, Rosen G. Multi-layer and recursive neural networks for metagenomic classification. *IEEE Trans Nanobiosci* 2015;**14**:608.

101. Le V, Quinn TP, Tran T, *et al*. Deep in the bowel: highly interpretable neural encoder-decoder networks predict gut metabolites from gut microbiome [J]. *bioRxiv* 2019: 686394.

102. Khan SKL. Multiclass disease classification from microbial whole-community Metagenomes using graph convolutional neural networks. *BioRxiv* 2019.

103. Chawla NV, Bowyer KW, Hall LO, *et al*. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;**16**:321–57.

104. More A. Survey of resampling techniques for improving classification performance in unbalanced datasets. arXiv preprint arXiv:1608.06048, 2016.

105. Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial networks. *Advances in Neural Information Processing Systems* 2014;**3**:2672–80.

106. Creswell A, White T, Dumoulin V, *et al*. Generative adversarial networks: an overview. *IEEE Signal Processing Magazine* 2017;**35**:53–65.

107. Yu L, Zhang W, Wang J, *et al*. Seqgan: Sequence generative adversarial nets with policy gradient. In: *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.

108. Jia G, Li Y, Zhang H, *et al*. Estimating heritability and genetic correlations from large health datasets in the absence of genetic data. *Nat Commun* 2019;**10**:5508.

109. Lee J, Yoon W, Kim S, *et al*. BioBERT: pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020;**36**(4):1234–40.

110. Nickel M, Murphy K, Tresp V, *et al*. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE* 2016;**104**:11–33.

111. Xie R, Liu Z, Jia J, *et al*. Representation learning of knowledge graphs with entity descriptions. In: *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.

112. Mohamed SK, Nováček V, AJB N. Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics* 2020;**36**:603–10.

113. Mamoshina P, Vieira A, Putin E, *et al*. Applications of deep learning in biomedicine. *Mol Pharm* 2016;**13**: 1445–54.

114. Camacho DM, Collins KM, Powers RK, *et al*. Next-generation machine learning for biological networks. *Cell* 2018;**173**:1581–92.