*Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi*
*Sayı 9, S. 69-91, 10, 2025*

***Araştırma Makalesi***

IJANSER
https://as-proceeding.com/index.php/ijanser
ISSN:2980-0811

*International Journal of Advanced Natural Sciences and Engineering Researches*
*Volume 9, pp. 69-91, 10, 2025*

***Research Article***

# Machine Learning in Biosciences: A Review of Applications

Ahmet TOPRAK[1*], Esma ERYILMAZ DOGAN[2]

[1]*Department of Electricity and Energy, Selcuk University, Konya, Turkey.*

*https://orcid.org/0000-0003-3337-4917*

[2]*Department of Biomedical Engineering, Faculty of Technology, Selcuk University, Konya, Turkey.*

*https://orcid.org/0000-0001-6809-7513*

[*]*(atoprak@selcuk.edu.tr) Email of the corresponding author*

*(Received: 25 September 2025, Accepted: 01 October 2025)*

(5th International Conference on Frontiers in Academic Research ICFAR 2025, September 25-26, 2025)

**ATIF/REFERENCE:** Toprak, A. & Eryılmaz Dogan, E. (2025). Machine Learning in Biosciences: A Review of Applications, *International Journal of Advanced Natural Sciences and Engineering Researches*, 9(10), 69-91.

***Abstract*** **–** Studies have shown that only about 2% of the genome encodes proteins, while the remaining 98% consists of non-coding RNAs (ncRNAs). Based on length, ncRNAs are classified as small (<200 nt) or long (>200 nt) and play key roles in biological processes. Experimentally verified associations between ncRNAs (miRNAs, lncRNAs, circRNAs) and diseases remain limited, since laboratory studies are costly and time-consuming. Thus, computational approaches have become essential for predicting disease-related ncRNAs.

Similarly, drug-target interactions are vital for drug discovery, as drugs act by binding to and inhibiting target molecules. Yet, experimental identification of these interactions is expensive, driving the development of computational prediction methods.

Microbes also influence human health, with microbiomes playing essential physiological roles. Identifying disease-related microbes is crucial, but experimental approaches are limited by cost and time. Hence, computational methods are widely employed.

Overall, computational strategies can be grouped into score functions, network-based algorithms, multi-source biological integration, and machine learning. This review highlights machine learning approaches for predicting ncRNA-disease associations, drug-target interactions, and disease-related microbes. It also summarizes key databases and successful methodologies, serving as a guide for future research in this field.

*Keywords – miRNA, lncRNA, circRNA, miRNA-disease Associations, lncRNA-disease Associations, circRNA-disease Associations, Drug, Drug-Target Interaction, Drug Repurposing, Drug Design, Microbe-Disease Association, Machine Learning*

## I. INTRODUCTION

Machine learning, can emulate human intelligence, is a computational algorithm that uses input data to perform a desired task. Machine learning generally has four subgroups as unsupervised learning, semi-supervised learning, supervised learning, and reinforcement learning. Machine learning is widely used in

many fields today. It is widely used in prediction of miRNA-disease relationship [1-4], lncRNA-disease relationship [5], circRNA-disease relationship [6], microbe-disease relationships [7], and drug-target interaction [8] especially in Bioinformatics.

With the completion of the human genome, it was revealed that about <2% of the genome consists of protein-coding RNAs, and the remaining approximately 98% consists of non-protein-coding RNA [9, 10]. As the complexity of organisms increases, the proportion of non-protein-coding RNA also increases. Until now, this large part of DNA that did not encode was thought to be "junk DNA", meaning garbage with no use and no function. Studies have revealed that these non-coding parts actually take on great tasks and functions in cell and living life.

Until recently, the vast majority of the genetic information was thought to be processed by proteins, but it is now known that the majority of the mammalian genome is transcribed into ncRNAs [11]. The assumption that creates the simple formula of the transfer of biological information from DNA to protein is called "Central dogma" [12] and the steps of the process are as follows, as can be seen in Figure 1. It is the DNA replication, the copying of DNA information into RNA (transcription), and the synthesis of proteins (translation) using the information in RNA.

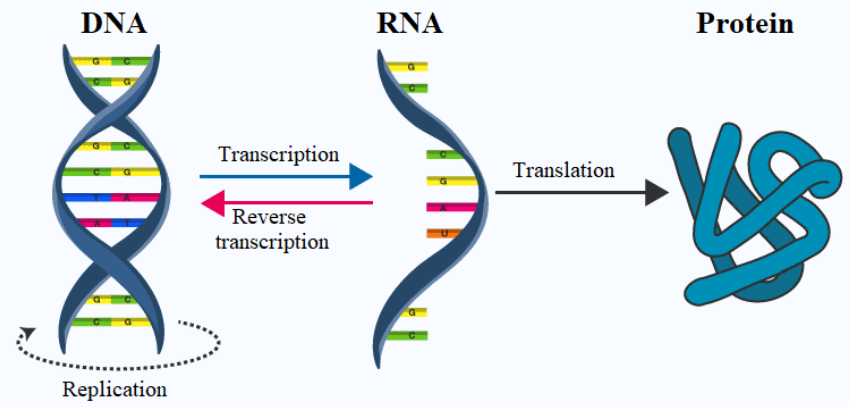


Figure 1. Schematic representation of DNA, RNA, Protein, and genetic transfer

non-coding RNA (ncRNA) refers to RNA molecules that are transcribed from DNA but do not code for proteins. However, just because they do not encode protein sequences does not mean that they do not have important functions. In fact, there are many different types of ncRNAs, each with its own unique function in the cell. Recent studies have revealed that a large proportion of the human genome is transcribed into ncRNAs, suggesting that these molecules may play a much more important role in cellular function than previously thought [11]. ncRNAs can be divided into two main groups based on their length: (i) Small ncRNAs: These are ncRNAs that are less than 200 nucleotides in length. (ii) Long ncRNAs: These are ncRNAs that are 200 nucleotides or longer in length [13].

***Micro RNA:*** In 1993, in the Victor Ambros laboratory, Lee and his team, discovered the first microRNA, which was transcribed by the lin-4 gene in the nematode worm Caenorhabditis elegans. This was a groundbreaking discovery because at the time, it was thought that all functional RNAs were coding proteins. This was the first example of a non-coding RNA regulating gene expression, and it opened up a whole new field of research in molecular biology. The term "microRNA" was coined in 2001 to describe small non-coding RNAs that are now known to play critical roles in gene regulation in many different organisms [14, 15]. This gene first discovered, by regulating the expression levels of lin-14 and lin-28, acts a crucial part in the development of nematode larvae [16]. microRNA (miRNA), a subclass of single-stranded small non-coding RNA molecule, is about 18-24 nucleotides in length, and regulates post-transcriptional gene expression by controlling the translation of mRNA into proteins [17-19]. It is estimated that miRNAs regulate the translation of more than 60% of protein-coding genes. miRNAs play a major role in the regulation of many processes such as cell proliferation, development, differentiation, death, apoptosis, metabolism, aging, signal transduction, and viral infection [17]. While some miRNAs regulate only certain individual targets, others can act as master regulators of a process. Thus, significant miRNAs regulate the expression levels of hundreds of genes simultaneously [20]. Although miRNAs

function within cells, they are also abundant in the blood. Thus, they can move to all cells, indicating that they can mediate intercellular communication [21]. Therefore, miRNAs especially the relationships between miRNAs and human diseases have received much attention from researchers.

***Long non-coding RNA:*** The first lncRNAs emerged with the discovery of lncRNAs involved in epigenetic regulation, such as H19 and X-inactive specific transcript (Xist) in early 1990s [22-24]. lncRNAs, which are considered as a class of non-protein-coding transcripts, are over 200 nucleotides in length [25]. The majority of the mammalian non-coding transcriptome consists of lncRNAs. Although lncRNAs do not code protein, they play major roles in in epigenetic regulation, apoptosis, proliferation, and cell differentiation [26]. Dysregulation of lncRNAs, just like miRNAs, can cause many human diseases such as breast cancer, lung cancer, prostate cancer, colon cancer, bladder cancer, ovarian cancer, leukemia, diabetes, and Alzheimer's [27-29]. Studies have shown that cancer cell antigen presentation and intrinsic tumor suppression are down-regulated by oncogenic lncRNAs [30]. The most expressed lncRNA H19 plays a crucial role in tumor initiation, progression, and relapse in many types of cancers such as thyroid cancer, liver cancer, and so on [31]. For example, H19 controls cell cycle progression by regulating RB-E2F signaling in colorectal cancer [32] and contributes to cell proliferation by regulating p53 activity in bladder cancer [33]. Moreover, lncRNA Xist has been shown to bind PRC2 to initiate X chromosome inactivation [34] and has proven to be associated with human glioblastoma stem cells [35]. Additionally, it was also observed that HOTAIR expression level was higher in primary breast tumors and metastases [36]. For this reason, the associations between lncRNAs and diseases have attracted the attention of researchers and they have focused their studies on this subject.

***Circular RNA:*** More than 40 years ago, the first circular RNA (circRNA) molecules named Viroids, were discovered [37]. With the development of bioinformatics, new circRNAs have been discovered in mammals, insects, plants, and eukaryotes. circRNAs are a class of non-coding single-stranded RNA molecules. The expression level of circRNA is generally low, but it has been experimentally confirmed that some circRNAs are highly expressed in specific cells or tissues [38, 39]. Experimental results show that circRNAs play significant role in many biological processes and in the emergence of human complex diseases such as cancer [40]. For example, circRNA CDR1as is differentially expressed in colorectal cancer [41], hepatocellular carcinoma [42], and neurological disorders [43], and also it can regulate miRNAs in tumor cells [44]. Likewise, CircRNA ciRS-7 efficiently regulates the activity of miRNA miR-7 [45]. Therefore, identifying potential circRNA-disease association is important both for discovering therapeutic targets and for understanding the complex mechanism of disease. The primary goal of biological research is to understand the mechanisms that cause complex human diseases. For this reason, research on disease into genes have been extended to ncRNAs [46]. Increasing arguments show that ncRNAs interact with many targets. Studies indicates that ncRNAs are associated with various complex human such as neurological diseases, diabetes, cardiovascular diseases (CVD), cancers, Alzheimer's and immune deficiency syndrome [47, 48]. Especially, it has been observed that ncRNAs can act as tumor suppressors or oncogenes especially in colon cancer, lung cancer, breast cancer, ovarian cancer, and prostate cancer. Therefore, ncRNAs can be used as biomarkers to predict and analyze many different types of cancer. Uncovering ncRNA and disease associations is useful for diagnosing, treating, and preventing diseases, as well as understanding the molecular mechanism of diseases. It also contributes to personalized drug treatment [49, 50].

***Drug:*** Drugs are the general term for substances given to humans or animals for the treatment, prevention, or diagnosis of a disease. In other words, drugs are chemical compounds that provide desired therapeutic effects by interacting with certain targets in humans or animals. Drugs can either directly target disease-associated genes or target disease-causing proteins [51]. Developing drugs to treat a particular disease is a very costly and time-consuming process [52]. The process of selecting an appropriate target for the drug being development and identifying the compound to bind to the target requires a lot of time [53]. But more importantly, drugs must go through clinical trials before they are released to the market. However, many drugs do not pass the clinical trial stage. The success rate of passing the trial stages of drugs developed according to the data of pharmaceutical companies is only 19% [54]. The main drug development stages are given in Figure 2.


Figure 2. Drug Development

Drug-target interaction (DTI) means that a drug binds to a target location, causing a change in the behavior or function of the target. When a drug is absorbed or injected, the chemical composition of the drug binds to the target molecule and can reacts with the target, preventing the target of functioning. Inhibition of target functions can regulate metabolism or kill pathogens that cause diseases. Identifying drug-target interactions facilitates of drug repositioning [55] [56] [57], drug discovery [58], drug resistance [59], and drug side effect prediction [60] [61]. Interactions between drugs and target proteins can be inferred by wet lab experiments using various techniques [62]. However, biological experiments to discover a new drug are both very expensive, time-consuming, and challenging work. For example, the identification of each new molecular entity requires approximately $1,8 billion [52] and it takes almost ten years for approval of a new drug developed [63]. Therefore, in-silico prediction of drug-target interactions is desirable given the many factors mentioned above [64]. Possible interactions between drugs and targets can be predict effectively and quickly with computational techniques than experimental techniques. DTI prediction methods are basically divided into three categories namely Ligand-based approaches, Insertion-based approaches, and Chemogenomic approaches. Ligands and 3D structures available for some target proteins are still missing. Therefore, chemogenomic approaches are widely used to predict DTI due to the limited applicability of ligand-based and docking-based approaches. Chemogenomic approach also divided in two subcategories as feature-based methods and similarity-based methods. In feature-based methods, drugs and targets are represented by feature vectors. In similarity-based methods, the inputs are the drug similarity matrix, the target similarity matrix, and the DTI matrix that shows which drug and target pairs interact [65, 66]. For this reason, similarity-based methods have been discussed in our study, and powerful and reliable computational techniques are needed to predict interactions between drug and target.

***Drug repurposing:*** also called drug repositioning. Drug development requires a very expensive and time-consuming experimental process. However, about 90% of the drugs developed are rejected by the Food and Drug Administration (FDA). Drug repurposing refers to the process of discovering new therapeutic uses for existing drugs that have already been approved for the treatment of other diseases. The goal of drug repurposing is to identify new indications for drugs that have already been shown to be safe and effective, which can potentially lead to faster and less expensive development of new treatments. For example, Sildenafil, trade name "Viagra", is an important example of drug repurposing. Sildenafil, developed in the 1980s for the treatment of chest pain, has been found to be ineffective in clinical trials. However, some side effects such as prolonged erections have been observed in patients given this drug. That's why researchers repurposed the drug Sildenafil to treat erectile dysfunction. As a result, computational methods can be used to predict drug repositioning [67].

***Drug Design:*** Drug design, also known as rational drug design or computer-aided drug design, is the process of creating new pharmaceuticals based on the molecular structure of diseases and their targets. The goal of drug design is to develop drugs that are safe, effective, and have fewer side effects. The process of drug design involves several stages, including target identification, lead discovery, lead optimization, preclinical testing, and clinical trials. In target identification, researchers identify the specific molecular target of a disease [68]. Lead discovery involves identifying potential compounds that can interact with the target. Lead optimization involves testing and modifying the selected compounds to improve their effectiveness and reduce their toxicity. Preclinical testing is performed on animals to assess the safety and efficacy of the drug candidate, while clinical trials are conducted on humans to determine its safety, efficacy, and optimal dosing. Drug design utilizes various computational and experimental techniques, such as molecular modeling, virtual screening, and high-throughput screening, to identify and optimize drug candidates. Advances in computer technology, computational biology, and structural biology have made drug design a more efficient and effective process. The drug discovery process is a complex and challenging endeavor that requires significant time, resources, and expertise. The low

success rates and high costs associated with the traditional drug discovery process have driven the development of computer-aided drug molecular design techniques. These techniques leverage computational methods, artificial intelligence, and machine learning to predict the potential efficacy and safety of new drugs, helping to reduce the time and costs associated with the traditional drug discovery process. In recent years, the use of computer-aided drug molecular design has shown promising results, with many drug discovery projects utilizing these methods to streamline the development process and improve the chances of success. The use of these techniques can help identify promising drug candidates early in the discovery process, reducing the number of failed experiments and ultimately leading to faster and more efficient drug development. In addition, computer-aided drug molecular design techniques also offer a valuable tool for understanding the underlying mechanisms of diseases and can help to identify new targets for drug development. This can lead to the development of innovative treatments for previously untreatable diseases and can help to advance our understanding of the biological processes underlying these conditions. Overall, the integration of computer-aided drug molecular design techniques into the drug discovery process represents a major step forward in the industry, offering the potential to revolutionize the way that new drugs are developed and discovered [69].

***Protein-Protein interactions:*** One of the most common molecules, proteins play important roles in many biological processes such as any cell's function and regulation. Proteins can associate with DNA and RNA in the genome to initiate transcription and the production of proteins, and monomeric chains of protein can lead to functional complexes that are stable. Understanding the interactions between proteins can help us to understand the function of each protein, and also aids in the understanding of cellular pathways, this information is crucial in developing effective treatments for human diseases, additionally, it also helps in the design of new drugs [70]. As a result, the process of predicting PPIs is fundamental to research and has gained increased attention in recent years. Laboratory experiments that seek to find PPIs are typically time-consuming and expensive. To address this issue, multiple computational models have been proposed that allow the systematic identification of protein pairs that interact. With the increasing rate of artificial intelligence, the potential for machine learning to predict protein-protein interactions has increased significantly [71].

***Microbes:*** Microbes are indeed small organisms, including bacteria, archaea, fungi, viruses, and other microorganisms, that could exist as cell groups or single cells. They are found in nearly every environment on Earth, including soil, water, air, and within living organisms. Studies have shown that some microbes can be parasitic and cause infections or diseases in different tissues, such as urogenital system, skin, and lung. However, it is important to note that not all microbes are harmful, and many play important beneficial roles in the human body [72-74].

Microbes, which were parasites in the human body for millions of years, evolved over time and established close and complex relationships with the immune system [75]. Studies have revealed that there is a very close relationship between diseases and microbes. For example, a major rising in the *Enterobacteriaceae's* expression level was observed in colorectal carcinoma patients [76]. Also, *Gordonibacter pamelae* and *Bifidobacterium catenulatum* are less abundant in colorectal carcinoma patients than in healthy people [77]. In addition, studies have shown that *Veillonella* and *Streptococcus* levels increase in liver cirrhosis patients, while *Eubacterium* and *Alistipes* are dominant in healthy people. For this reason, we can say that *Veillonella* and *Streptococcus* have an effect on the progression of liver cirrhosis [78].

Therefore, it is necessary to reveal the associations among microbes and human diseases to understand disease pathogenesis. However, determining the relationships between diseases and microbes through experimental studies is a very expensive and time-consuming process, just like identifying the ncRNA-disease relationship and drug-target interactions. Therefore, calculation techniques are being developed for prediction of disease related microbes. In this study, several successful machine learning-based computational methods are explained.

## II. DATABASES

With technology' and bioinformatics' development, various databases have been developed for miRNA, lncRNA, and drug-target interaction storage. Some of these databases are described below.

**miRbase** contains all miRNA species discovered so far and provides nomenclature for newly discovered miRNAs. The current version of miRbase contains miRNA sequences: 48.860 mature miRNAs and 38.589 hairpin precursors [79].

**HMDD** includes 53.530 miRNA-disease relationships with 2.360 diseases and 1.817 miRNAs information from 37.090 articles. The human disease and miRNA association data in this database have been experimentally proven [80, 81].

**miR2Disease** contains detailed information of associations among miRNAs and diseases, including miRNA-disease relationships description, disease name, miRNA ID, expression pattern, detection method for miRNA expression, and experimentally validated miRNA target genes [82].

**deepBase** contains information on the expression levels and functions of ncRNAs. This database provides a comprehensively expression information of lncRNAs and small RNAs by combining thousands of data approximately 50 cancer tissues and 80 normal tissues [83].

**miRCancer** is a human cancer database containing miRNA expression level information such as up or down regulation of miRNA. Currently miRCancer documents 9.080 relationships between 57.984 microRNAs and 196 human cancers from more than 7.288 publications [84].

**LncRNADisease** database contains 13.191 experimentally proven associations among 6.066 lncRNAs and 484 diseases, also 12.249 experimentally proven associations among 10.732 circRNAs and 262 diseases [85].

**Lnc2Cancer** provides comprehensively experimentally validated lncRNA-cancer relationships and circRNA-cancer relationships. This database contains 10.303 relationships information between 216 cancers, 743 circRNAs, and 2.659 lncRNAs from more than 1.500 published papers [86].

**DrugBank** provides comprehensive cheminformatics and bioinformatics information by combining comprehensive drug targets with detailed drug data. This frequently updated database contains over 7.800 drug entries, >15.000 drug-target interactions, 340 FDA-approved biotech drugs and 2.200 FDA-approved small molecule drugs [87].

**KEGG** is a database containing knowledge about diverse chemical compounds, drugs, and diseases, as well as genomic, chemical, and systemic functional information obtained from experimental studies [88].

**STITCH** contains information about predicted and known interactions between proteins and chemicals. This database contains, from 2.031 organisms, 1,6 billion interactions among 9,6 million proteins, and 500.000 chemicals [89].

**TDR targets** is a chemogenomic database containing information on neglected tropical diseases and an important resource for discovering drug. TDR Targets chemogenomic includes genomic information from pathogens and detailed knowledge about bioactive components [90].

**HMDAD**, the Human Microbe-Disease Association Database, is a database that collects microbe-disease relationships from studies of microbiota. This database includes 483 empirically proven associations between 292 microbes and 39 diseases [91].

**MorCVD** is a database containing host-pathogen protein-protein interactions data i.e. a total of 19 cardiovascular diseases including Endocarditis, Pericarditis, and Myocarditis. This database contains a total of 23.377 host-pathogen protein-protein interactions along [92].

**PHI** (The pathogen-host interactions database) is a database that includes biological and molecular information on genes verified to affect the outcome of interactions between pathogens and hosts. There is information of 296 pathogens, 9.973 genes, 249 hosts, and 22.408 interactions in the latest version of PHI database [93].

**STRING** database regularly collects both protein-protein physical interactions and protein-protein functional interactions [94].

## III. COMPUTATIONAL MODELS

Because experimental studies are time-consuming and expensive, reliable computational tools are needed to predict ncRNA-disease associations, microbe-disease associations, drug-target interactions. There are four subgroups of computational methods used to determine relationships to date: Computational methods used to date to determine these relationships have four subgroups: score function models, complex network algorithm model, multiple biological information models, and machine learning models [95]. We comprehensively examined machine learning algorithms for forecasting ncRNA related disease and also drug-target interactions.

### *Regularized Least Squares*

This method is a prediction model for disease related miRNAs based on the regularized least squares algorithm (RLSMDA), shown in Figure 3, which can be used for prediction without using any negative samples [96]. However, this model is highly dependent on parameters. This computational method is a semi-supervised learning model developed for inferring miRNA-disease associations. In the RLSMDA model, relationships between miRNAs and diseases are given by a combined classifier in the miRNA space and disease space. The Regularized Least Squares (RLS) method was carried out to form two optimum classifiers in the RLSMDA method. For miRNA space and disease space, optimal classification functions can be obtained by solving the following optimization problem:

$$F_M^* = SM * (SM + \eta_M * I_M)^{-1} * A^T \quad (1)$$

$$F_D^* = SD * (SD + \eta_D * I_D)^{-1} * A \quad (2)$$

where, similarity networks of miRNA functional and disease semantic were represented with $SM$ and $SD$, and miRNA-disease relationship network was represented with an adjacency matrix $A$, $I_M$ is the identity matrix with the same size as the $SM$, and the $I_D$ is the identity matrix with the same dimensions as the $SD$. $\eta_M$ and $\eta_D$ parameters are for trade-off. Lastly, optimal classifier results from space of miRNA and disease are integrated to get new relationships between miRNAs and diseases.

$$F^* = w * F_M^{*T} + (1 - w) * F_D^* \quad (3)$$

$F(i, j)$ is the predicted scores that indicates the possibility of relationship among miRNA $i$ and disease $j$.

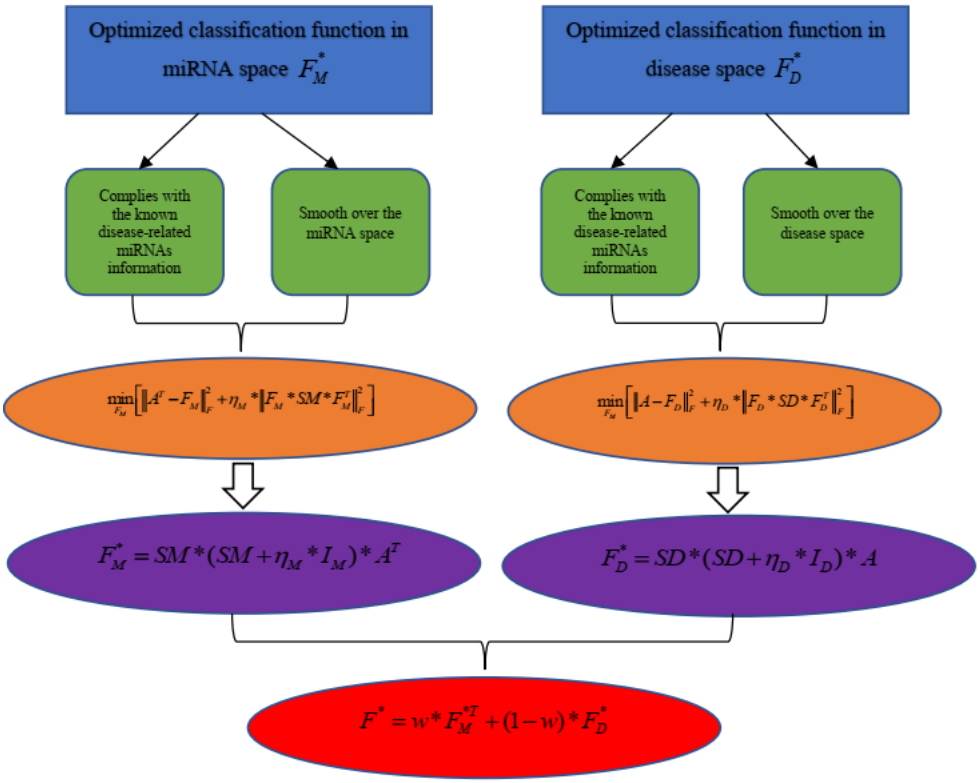


Figure 3. RLSMDA

RLSMDA method has some limitations. First of these limitations, it is not clear how to decide parameter values in RLSMDA. Secondly, reconstruction of miRNA functional similarity and disease semantic similarity will increase the predictive talent. The advantage of RLSMDA is that it can be applied to diseases that do not have any known related miRNAs.

***Random Forest***

A Random Forest computation model for miRNA-disease association (RFMDA) prediction developed based on machine learning is shown in Figure 4 [97]. The training set used in this study was obtained from the HMDD database. miRNA functional similarity, disease semantic similarity, and Gaussian interaction profile kernel similarity were integrated to create feature vectors to represent miRNA-disease samples. The features of miRNA functional similarity and integrated semantic similarity are represented by $SD$ and $SM$. An 878-dimensional vector represented by $F$ is created using $SD$ and $SM$ with the following equation.

$$F = \left(SM(m(i)), SD(d(u))\right) \tag{4}$$

Then, using the following equation, each feature's final score can be calculated.

$$Sco(i) = \frac{FF_p(i)}{FF_n(i)} i = 1, 2, \ldots, 878. \tag{5}$$

Where $F_p$ demonstrate the feature in the positive sample set and $F_n$ represent negative sample set.

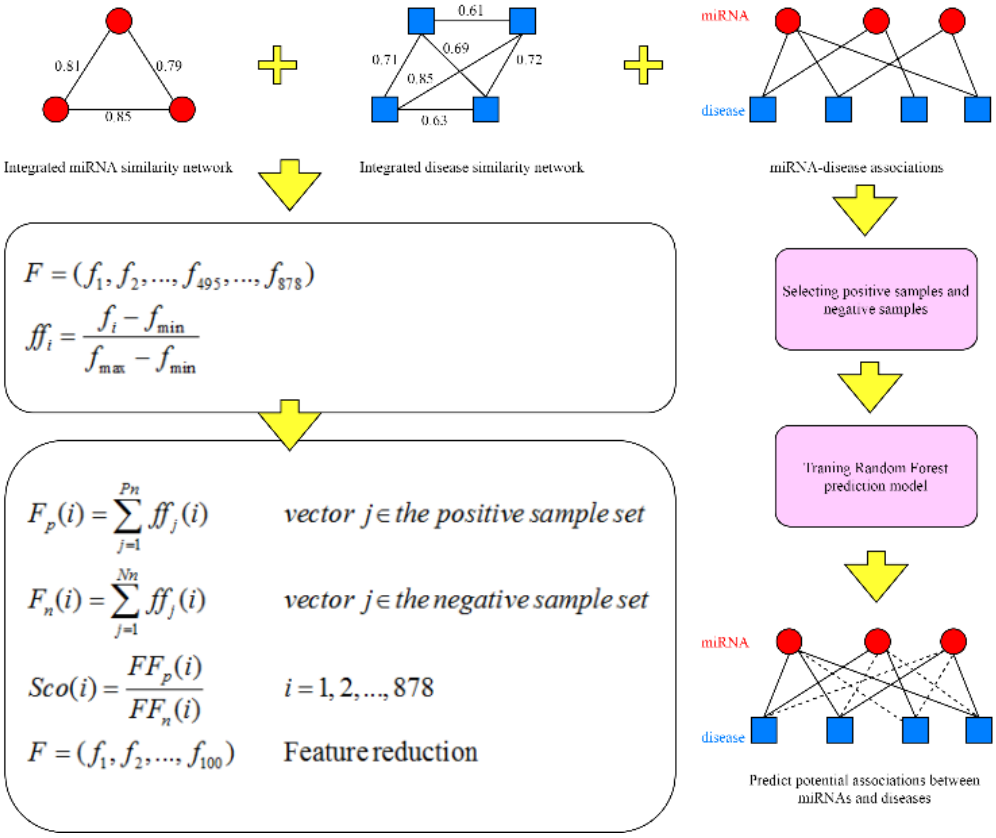


Figure 4. FRMDA

The RFMDA method also has some limitations. RFMDA requires both positive and negative training examples. As known, obtaining reliable negative samples, is very difficult or even impossible. Negative samples of unknown associations between miRNA and disease were selected by a random selection method. This random selection method can affect the prediction result.

### *K-Nearest Neighbors (KNN)*

As seen in Figure 5, a calculational technique based on K-nearest neighbor (KNN), namely RKNNMDA, is proposed [98]. In this technique, for re-ranking, the prediction scores calculate with using a support vector mechanism. RKNNMDA is a computational method developed to forecast new disease-associated miRNAs, by combining known disease-miRNA relationships, similarities of disease semantic, miRNA functional, and Gaussian interaction profile kernel.

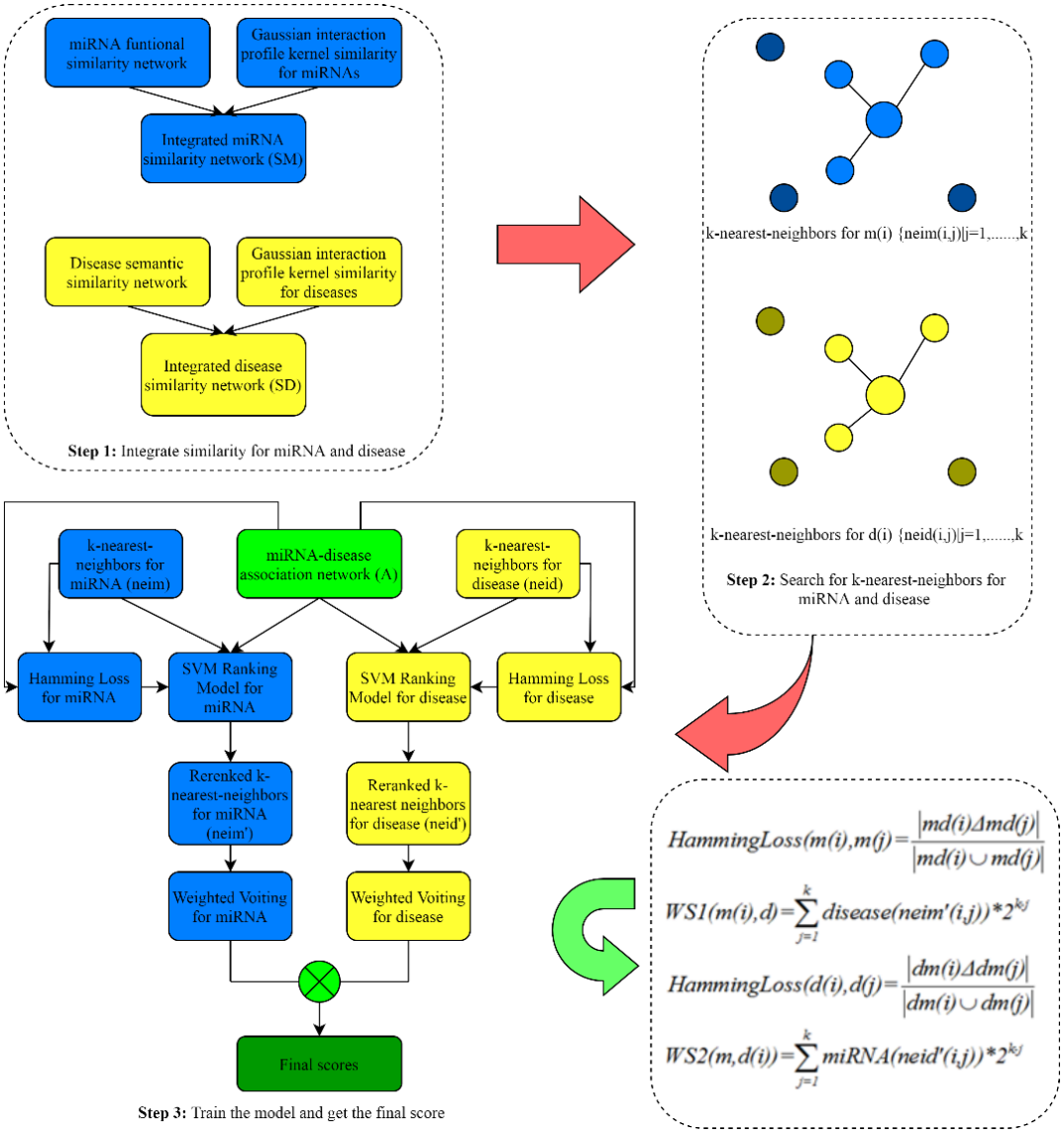


Figure 5. RKNNMDA

Each miRNA $m(i)$'s k-nearest-neighbors $neim(i)$ are determined based on the KNN algorithm. These determined k-nearest-neighbors are reranked with the SVM Ranking model. The weight value is represented by $WS$, and the $WS1$ value between $m(i)$ and disease $d$ is calculated by the following equation:

$$WS1(m(i),d) = \sum_{j=1}^{k} disease\big(neim'(i,j)\big) * 2^{k-j} \quad (6)$$

where $neim'(i,j)$ represents $j$th neighbor miRNA of miRNA $m(i)$, and $disease\big(neim'(i,j)\big)$ indicates feature score of disease $d$ with regard to miRNA $m(i)$ and its neighbor $m(j)$.
Similar manner, $WS2$ between $d(i)$ and miRNA $m$ is calculated by the following equation:

$$WS2(m,d(i)) = \sum_{j=1}^{k} miRNA\big(neid'(i,j)\big) * 2^{k-j} \quad (7)$$

where $neid'(i,j)$ represents $j$th neighbor disease of disease $d(i)$, and $miRNA\big(neid'(i,j)\big)$ indicates feature score of miRNA $m$ with regard to disease $d(i)$ and its neighbor $d(j)$. To determined potential relationships between miRNAs and disease, $WS1$ and $WS2$ have been integrated.

### *Graph Regression*

A Graph Regression method developed using singular value decomposition and partial least squares regression for miRNA-disease association prediction (GRMDA) is shown in Figure 6 [99]. In this study, a graph regression between several similarities data, and known disease-miRNA relationships were used to forecast new disease-miRNA relationships. Since graph regression in disease similarity, miRNA similarity, and disease-miRNA relationship space is performed simultaneously, the following equation is obtained.

$$\{A_r^*, A_d^*, F_r^*, F_d^*, B_r^*, B_d^*\} = \arg min \left\{ \begin{matrix} \|A - A_r A_d^T\|^2 + \|S_m - F_r F_r^T\|^2 + \|S_d - F_d F_d^T\|^2 \\ + \|A_r - F_r B_r\|^2 + \|A_d - F_d B_d\|^2 \end{matrix} \right\} \quad (8)$$

where $F_r, F_d, B_r$, and $B_d$ represents the features of miRNA, feature of diseases, association between $A$ and $F_r$, association between $A$ and $F_d$, respectively.
In the following equation to create $A_r, A_d, F_r$, and $F_d$, Singular Value Decomposition was applied for low-rank decompositions.

$$M \overset{SVD}{\Longrightarrow} U\Sigma V^T = \left(U\sqrt{\Sigma}\right)\left(V\sqrt{\Sigma}\right)^T = LR^T \quad (9)$$

Disease-related miRNA candidates can be calculated in the following equation.

$$C = F_r B_r B_d^T F_d^T \quad (10)$$

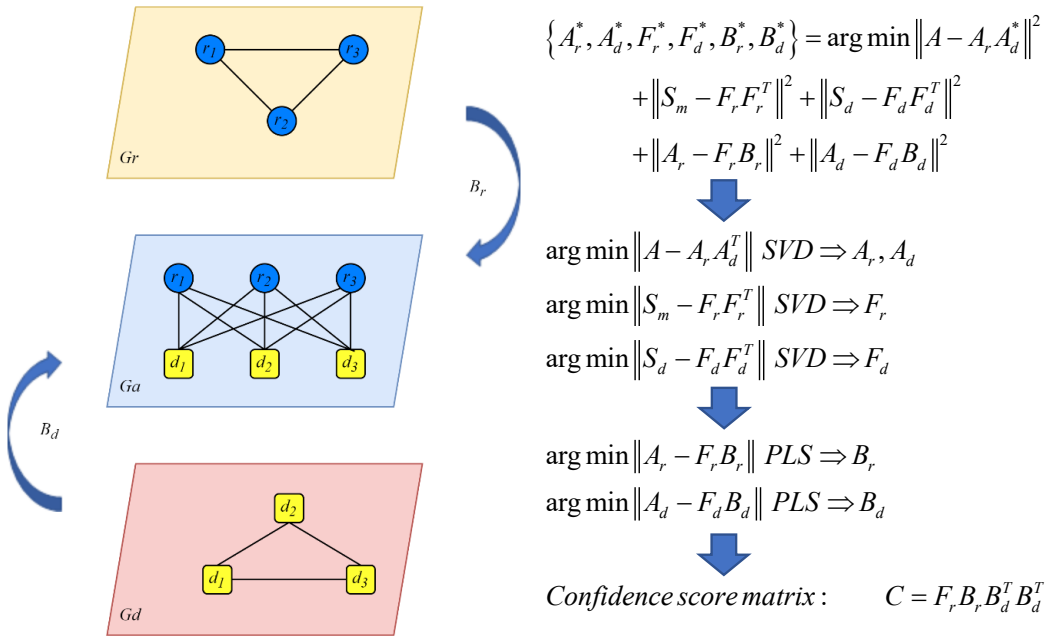


Figure 6. GRMDA

As in the previous methods, the GRMDA method has some weaknesses. Firstly, existing studies utility from known miRNA-disease relationships data. However, the collection of data has not yet reached the final result. This means that our estimation is always in the absence of data. That is, there will always be a shortcoming in our predictions. Secondly, in Singular Value Decomposition and Partial Least Squares Regression there are difficulties in parameter selection due to the size of the matrices.

### *Support Vector Machine*

In this approach, a computational method of ILDMSF was proposed to forecast possible lncRNA-disease relationships, as shown in Figure 7 [5]. ILDMSF combined multiple lncRNA-lncRNA similarity and disease-disease similarity with a network fusion method. Support Vector Machine (SVM) is

employed to predict relationships between lncRNAs and diseases, and also the bagging method is used to deal with imbalance data.
The SVM function can be defined in the following eq.

$$\begin{cases} max\ \lambda & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j K(x_i, x_j) + \sum_{i=1}^{N}\lambda_i \\ s.t. & \sum_{i=1}^{N}\lambda_i y_i = 0,\ \ 0 \le \lambda_i \le W,\ \ i = 1,2,\ldots\ldots,N \end{cases} \tag{11}$$

where $N, K, W,$ and $\lambda$ parameters represent the number of samples, kernel function, soft margin, and Lagrange multiplier, respectively. $x_i$ is the feature vector for $i$th sample. The label corresponding to $x_i$, is $y_j$, and usually 0 or 1.

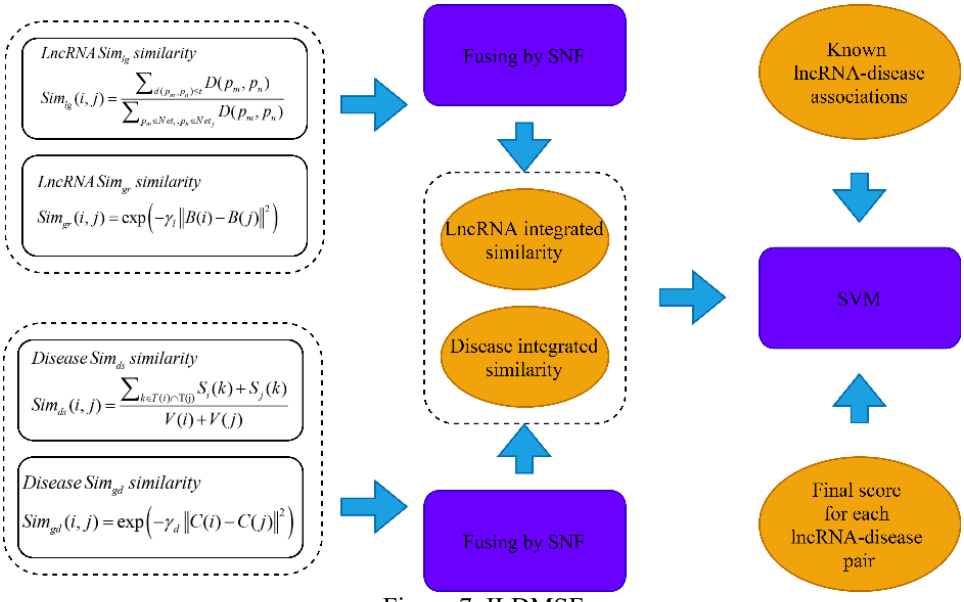


Figure 7. ILDMSF

**Kernel Fusion and Deep Auto-Encoder**

This method is a prediction model for prediction of circRNA-disease associations based on the Kernel Fusion and Deep Autoencoder (KFDAE) [100]. Primarily, each circRNA-diseases pair's feature vectors are obtained. These resulting sets are used as the input data of the DAE. The DAE's structure is depicted in Figure 8. $Y$ is denoted of the encoder output, can be calculated follows:

$$Y = \Phi(wX + b) \tag{12}$$

$$\Phi(x) = \frac{1}{1+exp(-x)} \tag{13}$$

where, $w$, $X$, and $b$ represent weight, bias, and input, respectively. Then, feature vectors obtained from DAE are used as the input of the multilayer perceptron to train the model.

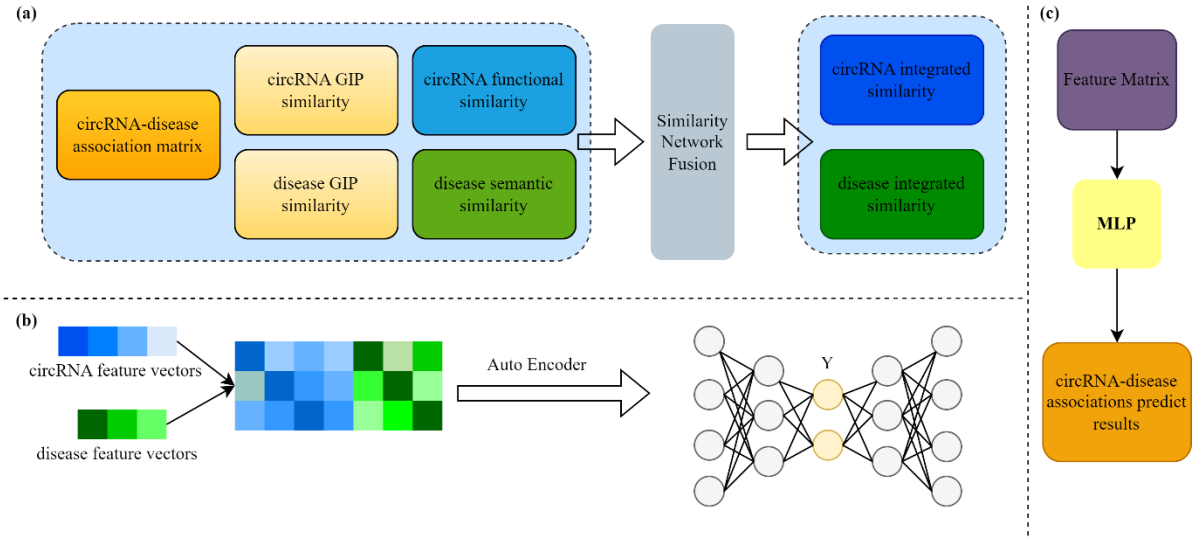


Figure 8. KFDAE

## *Kernel Regression Method*

Yamanishi et al. suggested a supervised learning technique in this study, shown in Figure 9, that maps targets in genomic space and drugs in chemical space into pharmacological space. This technique forecast possible drug-target interactions with integrating known DTI network, target proteins' sequence information, and chemical structure [101].

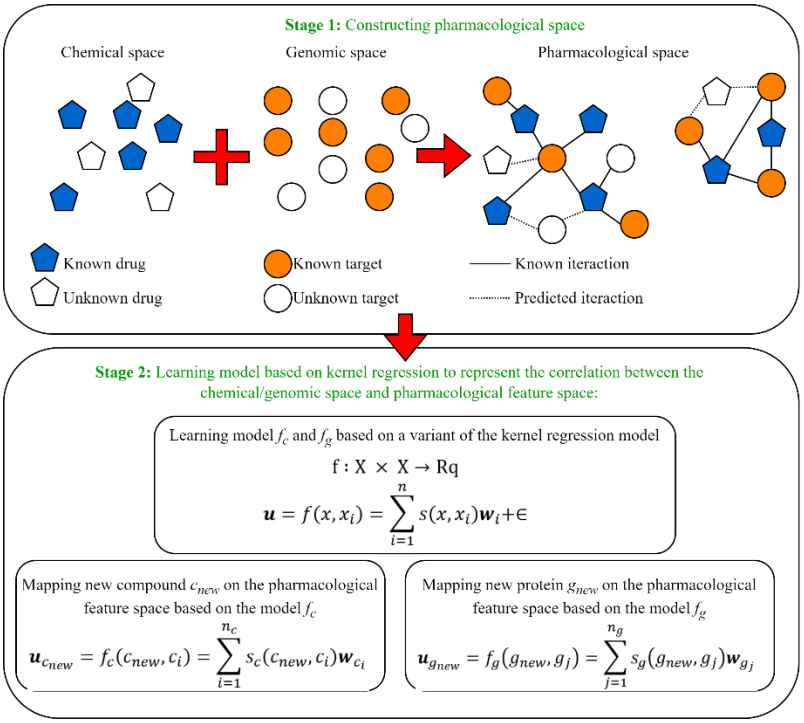


Figure 9. Kernel Regression Method

$$u_{c_{new}} = \sum_{i=1}^{n_c} S_c\,(c_{new}, c_i) w_{c_i} \tag{14}$$

$$u_{g_{new}} = \sum_{j=1}^{n_g} S_g\,(g_{new}, g_i) w_{g_j} \tag{15}$$

where, $S_c(.,.)$ is the chemical structure similarity score, $S_g(.,.)$ is the sequence similarity score, and $w_{c_i}$ and $w_{g_i}$ are the weight vector. In this pharmacological space, feature-based similarity scores were calculated, and potential compound-protein interactions were predicted.

### *DeepDTIs*

A deep-learning-based method named DeepDTIs was developed by Wen et al. [102]. Without separating the targets into different classes, DeepDTIs aimed to infer new interactions between approved drugs and targets. The flowchart of DeepDTIs is shown in Figure 10. DeepDTIs method predicts new interactions in three sections. In first section, most common features are calculated to identify drugs and targets. Second section consists of the second, third and fourth layers called the hidden layer. Third section is the output layer, where a classification is made with known label drug-target interactions. Based on the Deep Belief Network, the joint distribution between $l$ hidden layers and training sample vector $x$ is modeled as follows.

$$P(x, h^1, h^2, \dots h^l) = \left(\prod_{k=0}^{l-2} P(h^k | h^{k+1})\right) P(h^{l-1}, h^l) \quad (16)$$

where, $x = h^0, P(h^{k-1}|h^k)$, is a k-level hidden-visible conditional probability distribution and $P(h^{l-1}, h^l)$ is the visible-hidden joint distribution in the top-level Restricted Boltzmann Machine.

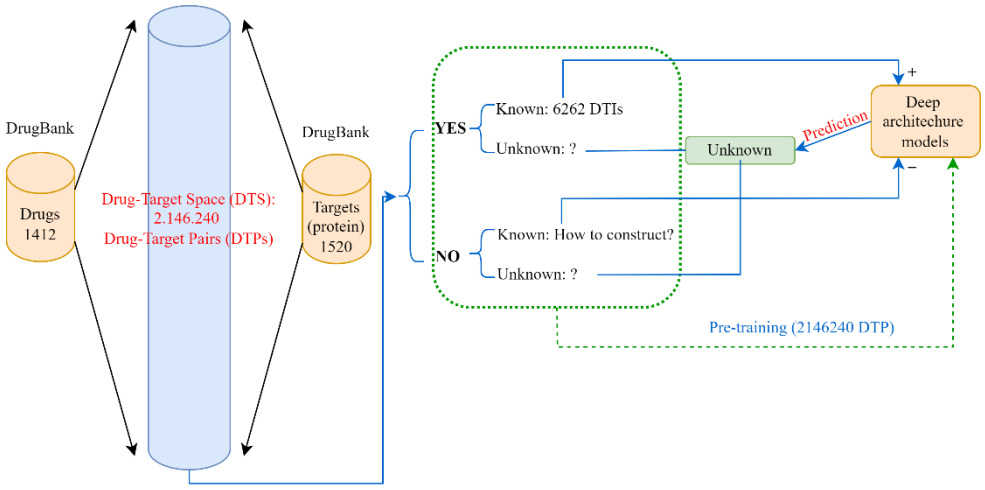


Figure 10. DeepDTIs

### *Lakizadeh's proposed model*

Lakizadeh et al. proposed a method called DRSE [67] for drug repurposing in order to integrate various heterogeneous data. This model for predicting drug-related diseases takes into account the side-effect characteristics of drugs and is relies on the combining the more data. The proposed method includes random walk with restart (RWR), diffusion component analysis (DCA), and matrix factorization (MF). RWR technique combine drug and disease features, DCA technique extract features, and MF technique estimate final prediction.

Jaccard similarity was used to calculate similarity matrices.

$$J(A, B) = \frac{Q_{11}}{Q_{01} + Q_{10} + Q_{11}} \quad (17)$$

Where, $Q_{01}, Q_{10}$, and $Q_{11}$ are the number of features, which is 0 in $A$ and 1 in $B$, which is equal to 1 in $A$ and 0 in $B$, and which is equal to 1 for both vectors $A$ and $B$.

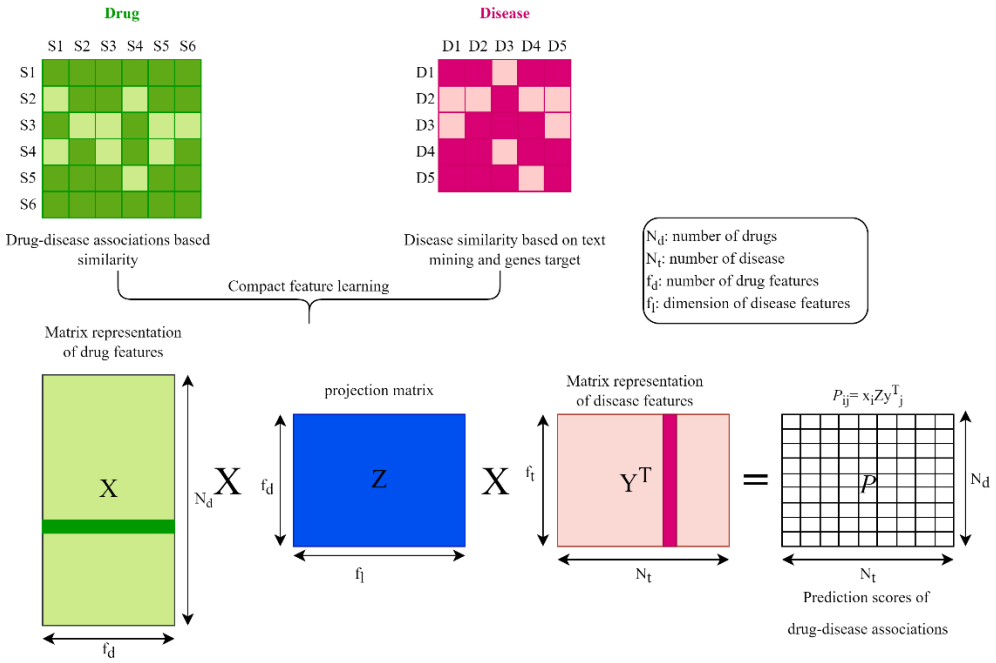


Figure 11. DRSE

$P \in R^{N_d \times N_t}$ shows matrix of relations between drugs and diseases, $X \in R^{N_d \times f_d}$ is drugs feature matrix, $Y \in R^{N_t \times f_t}$ diseases feature matrix, and $Z \in R^{f_d \times f_t}$ is projection matrix. Then, the possibility of a relationship between drug $i$ and disease $j$ is computed with the following equation.

$$score(i,j) = x_i Z y_j^T \quad (18)$$

### *MultiPPIs*

Zou et al. proposed the DeepWalk method, a graph representation learning method, to extract multi-source relationship information of proteins with other biomolecules and named it MultiPPIs [103]. The flowchart of MultiPPIs is shown in Figure 12.

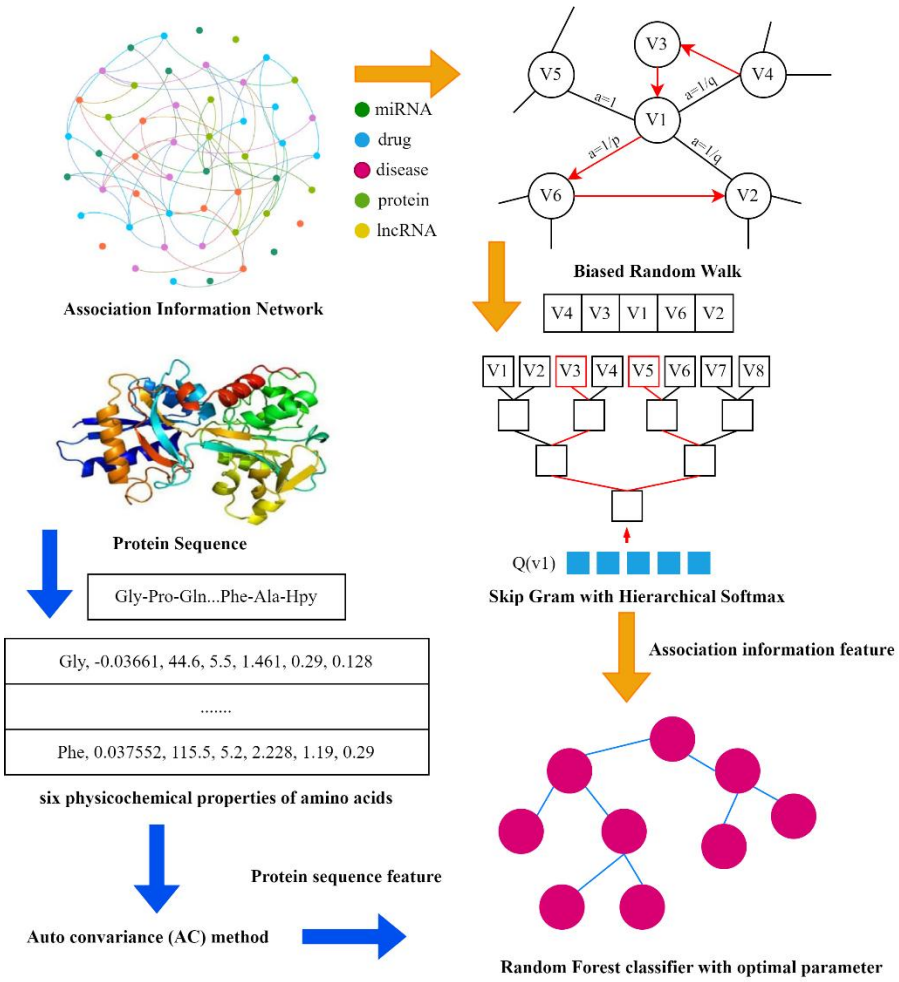


Figure 12. MultiPPIs

Firstly, the protein sequence features of amino acids were obtained by the autocovariance method (AC) with the following equations.

$$P'_{ij} = \frac{P_{ij} - \overline{P_j}}{S_j}, (i = 1,2,\dots,6; j = 1,2,\dots,20) \tag{19}$$

$$\overline{P_j} = \frac{\sum_{i=1}^{20} P_{ij}}{20} \tag{20}$$

$$S_j = \sqrt{\frac{\sum_{i=1}^{20} (P_{ij} - \overline{P_j})^2}{20}} \tag{21}$$

$$AC = \frac{1}{N-d} \sum_{j=1}^{N-d} \left(X_{i,j} - \frac{1}{n}\sum_{i=1}^{n} X_{i,j}\right)\left(X_{i+d,j} - \frac{1}{n}\sum_{i=1}^{n} X_{i,j}\right) \tag{22}$$

Here the length of the protein sequence is represented by $N$, and the $j$th descriptor value of the $i$th amino acid is represented by $X_{i,j}$. Secondly, the relationships between miRNAs, lncRNAs, drugs, proteins, and diseases were integrated to obtain a multi-source relationship network. Then, Random Forest classifier was used to predict protein-protein interactions.

### *Adaptive Boosting*

An Adaptive Boosting model named ABHMDA [104], was developed for uncovering new microbe-disease relationships with calculating the probability of association between microbes and diseases using a powerful classifier. ABHMDA's flowchart is shown in Figure 13.

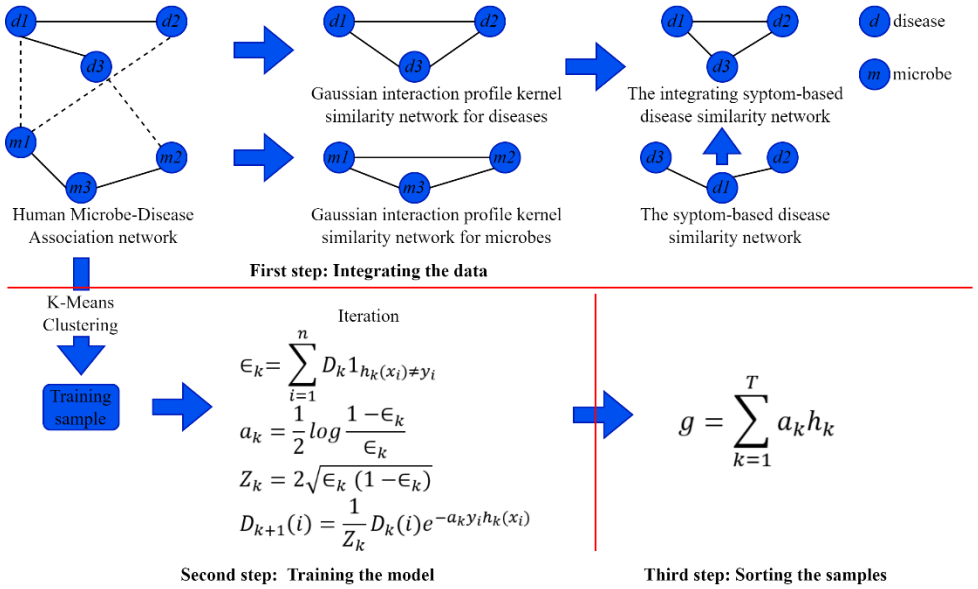


Figure 13. ABHMDA

Firstly, similarities of Gaussian interaction kernel were computed for both microbe space and disease space rely on the known microbe-disease relationship network. Then, Gaussian interaction profile kernel similarity matrix and disease symptom similarity matrix integrated. Also, the feature vector is determined for each microbe-disease relationship. By applying k-means clustering according to their feature vectors, candidate samples are divided into 23 clusters.

Second, the training samples are classified using decision trees and the weights of the classifiers are calculated. The error function $\epsilon_i$, the variate $Z_i$, and the ratio of the weak classifier in the strong classifier is computed by following equation:

$$\epsilon_i = \sum_{j=1}^{n} D_i 1_{h(i)j \neq Y_j} \quad (23)$$

$$\alpha_i = \frac{log\frac{1-\epsilon_i}{\epsilon_i}}{2} \quad (24)$$

$$Z_i = 2[\epsilon_i\ (1-\epsilon_i)]^2 \quad (25)$$

The sample's weight is updated by equation 20:

$$D_{i+1}(j) = \frac{1}{Z_i} D_i(j) e^{-\alpha_i Y_j h(i)_j} \quad (26)$$

Finally, the probability of new microbe-disease association could be calculated as follows.

$$p = \sum_{i=1}^{n_c} \alpha_i H(i) \quad (27)$$

### *Laplacian Regularized Least Squares*

Laplacian regularized least squares classifier, a new semi-supervised computational model called LRLSHMDA [105], was used to forecast the possible microbe-disease interactions. LRLSHMDA's flowchart is shown in Figure 14.

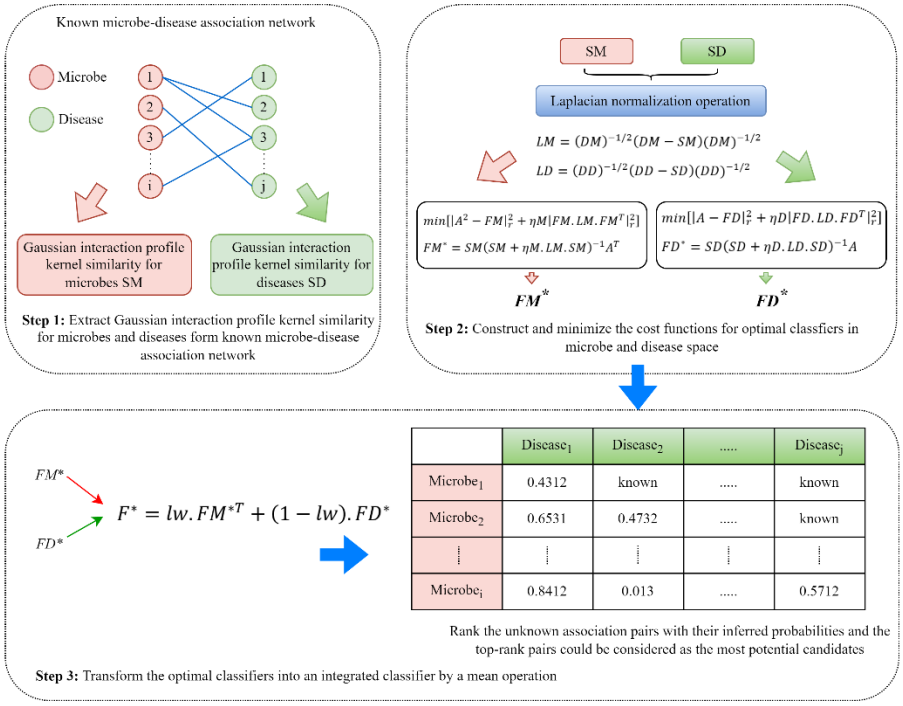


Figure 14. LRLSHMDA

Firstly, Gaussian interaction kernel similarities were calculated for both microbes and diseases based on the known microbe-disease relationship network. Then to normalize these matrices, Laplace operation used, shown as following:

$$LM = (DM)^{-1/2}(DM - KM)(DM)^{-1/2} \quad (28)$$

$$LD = (DD)^{-1/2}(DD - SD)(DD)^{-1/2} \quad (29)$$

Then, the cost functions could be calculated as follows:

$$min_{FM}[\|A^T - FM\|_F^2 + \eta M\|FM.LM.FM^T\|_F^2] \quad (30)$$

$$min_{FD}[\|A - FD\|_F^2 + \eta D\|FD.LD.FD^T\|_F^2] \quad (31)$$

The above two formulas are converted into optimal classification functions.

$$FM^* = SM(SM + \eta M.LM.SM)^{-1}A^T \quad (32)$$

$$FD^* = SM(SD + \eta D.LD.SD)^{-1}A \quad (33)$$

$F^*$ was the new microbe-disease association probability matrix.

$$F^* = lw.FM^{*T} + (1 - lw).FD^* \quad (34)$$

## IV. CONCLUSION

Diseases are the leading causes that negatively affect human life or result in death. It has been determined that the main causes of diseases are non-coding RNAs (i.e., miRNA, lncRNAs, and circRNA) and microbes. miRNAs play a major role in the regulation of many processes such as cell proliferation, development, differentiation, death, apoptosis, metabolism, aging, signal transduction, and viral infection.

While some miRNAs regulate only certain individual targets, others can act as master regulators of a process. Thus, significant miRNAs regulate the expression levels of hundreds of genes simultaneously. lncRNAs do not code protein, but they play diverse roles in gene expression regulation, including transcriptional regulation, post-transcriptional regulation, and epigenetic regulation. Dysregulation of lncRNAs, just like miRNAs, can cause many human diseases such as breast cancer, lung cancer, prostate cancer, colon cancer, bladder cancer, ovarian cancer, leukemia, diabetes, and Alzheimer's. circRNAs play significant role in many biological processes and in the emergence of human complex diseases such as cancer. Microbes have been found to be parasitic in diverse human body texture, such as the urogenital tract, skin, and lungs. For this reason, recent research has focused on determining disease-related non-coding RNAs and disease-related microbes. The discovery of drugs used for the prevention and treatment of detected diseases and the repositioning of drugs are also very important. Drugs may directly target disease-related genes or disease-causing proteins. Drug repurposing is the identification of new indications for approved drugs beyond the initial indications.

The identification of disease-related ncRNAs, drug-target interactions, and drug repurposing is crucial for disease diagnosis, treatment, prevention, and personalized medicine. However, since the experimental processes used to determine these relationships are very expensive and time-consuming, the tendency to computational methods has increased and machine learning-based algorithms have gained popularity. This article summarizes some of the recently widely used based on machine learning models for prediction of ncRNA-disease associations, microbe-disease associations, drug-target interactions, and protein-protein interactions. In addition, it also includes databases of miRNA-disease associations, lncRNA-disease associations, circRNA-disease associations, drug-target interactions, and protein-protein interactions. To further improve the prediction performance of the computational methods mentioned, it is necessary to make full use of different types of heterogeneous data sources and integrate new association networks.

## REFERENCES


[1] A. Toprak and E. Eryilmaz Dogan, "Prediction of Potential MicroRNA-Disease Association Using Kernelized Bayesian Matrix Factorization," *Interdiscip Sci,* vol. 13, no. 4, pp. 595-602, Dec 2021, doi: https://doi.org/10.1007/s12539-021-00469-w.

[2] A. Toprak and E. Eryilmaz, "Prediction of miRNA-disease associations based on Weighted k-Nearest known neighbors and network consistency projection," *J Bioinform Comput Biol,* vol. 19, no. 1, p. 2050041, Feb 2021, Art no. 2050041, doi: https://doi.org/10.1142/S0219720020500419.

[3] A. Toprak, "Identification of disease-related miRNAs based on weighted k-nearest known neighbours and inductive matrix completion," *International Journal of Data Mining and Bioinformatics,* vol. 27, no. 4, pp. 231-251, 2023, doi: https://doi.org/10.1504/ijdmb.2023.134297.

[4] A. Toprak, "Predicting human miRNA disease association with minimize matrix nuclear norm," *Scientific Reports,* vol. 14, no. 1, p. 30815, 2024, doi: https://doi.org/10.1038/s41598-024-81213-4.

[5] Q. Chen *et al.*, "ILDMSF: Inferring Associations Between Long Non-Coding RNA and Disease Based on Multi-Similarity Fusion," *IEEE/ACM Trans Comput Biol Bioinform,* vol. 18, no. 3, pp. 1106-1112, May-Jun 2021, doi: https://doi.org/10.1109/TCBB.2019.2936476.

[6] A. Toprak, "circRNA-disease association prediction with an improved unbalanced Bi-Random walk," *Journal of Radiation Research and Applied Sciences,* vol. 17, no. 2, p. 100858, 2024, doi: https://doi.org/10.1016/j.jrras.2024.100858.

[7] J. Qu, Y. Zhao, and J. Yin, "Identification and Analysis of Human Microbe-Disease Associations by Matrix Decomposition and Label Propagation," *Front Microbiol,* vol. 10, p. 291, 2019, doi: https://doi.org/10.3389/fmicb.2019.00291.

[8] M. A. Thafar *et al.*, "DTiGEMS+: drug-target interaction prediction using graph embedding, graph mining, and similarity-based techniques," *J Cheminform,* vol. 12, no. 1, p. 44, Jun 29 2020, doi: https://doi.org/10.1186/s13321-020-00447-2.

[9] M. C. Frith, M. Pheasant, and J. S. Mattick, "The amazing complexity of the human transcriptome," *Eur J Hum Genet,* vol. 13, no. 8, pp. 894-7, Aug 2005, doi: https://doi.org/10.1038/sj.ejhg.5201459.

[10] J. E. Wilusz, H. Sunwoo, and D. L. Spector, "Long noncoding RNAs: functional surprises from the RNA world," *Genes Dev,* vol. 23, no. 13, pp. 1494-504, Jul 1 2009, doi: https://doi.org/10.1101/gad.1800909.

[11] J. S. Mattick and I. V. Makunin, "Non-coding RNA," *Hum Mol Genet,* vol. 15 Spec No 1, no. suppl_1, pp. R17-29, Apr 15 2006, doi: https://doi.org/10.1093/hmg/ddl046.

[12] F. Crick, "Central dogma of molecular biology," *Nature,* vol. 227, no. 5258, pp. 561-3, Aug 8 1970, doi: https://doi.org/10.1038/227561a0.
[13] D. Veneziano, G. Nigita, and A. Ferro, "Computational Approaches for the Analysis of ncRNA through Deep Sequencing Techniques," *Front Bioeng Biotechnol,* vol. 3, p. 77, 2015, doi: https://doi.org/10.3389/fbioe.2015.00077.
[14] F. Saydam, İ. Değirmenci, and H. V. Güneş, "MikroRNA'lar ve kanser," *Dicle Tıp Dergisi,* vol. 38, no. 1, 2011.
[15] R. C. Lee, R. L. Feinbaum, and V. Ambros, "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14," *Cell,* vol. 75, no. 5, pp. 843-54, Dec 3 1993, doi: https://doi.org/10.1016/0092-8674(93)90529-y.
[16] B. J. Reinhart *et al.*, "The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans," *Nature,* vol. 403, no. 6772, pp. 901-6, Feb 24 2000, doi: https://doi.org/10.1038/35002607.
[17] L. Zhang, X. Chen, and J. Yin, "Prediction of Potential miRNA-Disease Associations Through a Novel Unsupervised Deep Learning Framework with Variational Autoencoder," *Cells,* vol. 8, no. 9, Sep 6 2019, doi: https://doi.org/10.3390/cells8091040.
[18] A. Li, Y. Deng, Y. Tan, and M. Chen, "A novel miRNA-disease association prediction model using dual random walk with restart and space projection federated method," *PLoS One,* vol. 16, no. 6, p. e0252971, 2021, doi: https://doi.org/10.1371/journal.pone.0252971.
[19] S. Chandra, D. Vimal, D. Sharma, V. Rai, S. C. Gupta, and D. K. Chowdhuri, "Role of miRNAs in development and disease: Lessons learnt from small organisms," *Life Sciences,* vol. 185, pp. 8-14, Sep 15 2017, doi: https://doi.org/10.1016/j.lfs.2017.07.017.
[20] M. Esteller, "Non-coding RNAs in human disease," *Nat Rev Genet,* vol. 12, no. 12, pp. 861-74, Nov 18 2011, doi: https://doi.org/10.1038/nrg3074.
[21] X. Chen, N. N. Guan, Y. Z. Sun, J. Q. Li, and J. Qu, "MicroRNA-small molecule association identification: from experimental results to computational models," *Brief Bioinform,* Oct 16 2018, doi: https://doi.org/10.1093/bib/bby098.
[22] C. I. Brannan, E. C. Dees, R. S. Ingram, and S. M. Tilghman, "The product of the H19 gene may function as an RNA," *Mol Cell Biol,* vol. 10, no. 1, pp. 28-36, Jan 1990, doi: https://doi.org/10.1128/mcb.10.1.28-36.1990.
[23] N. Brockdorff *et al.*, "The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus," *Cell,* vol. 71, no. 3, pp. 515-26, Oct 30 1992, doi: https://doi.org/10.1016/0092-8674(92)90519-i.
[24] G. Borsani *et al.*, "Characterization of a murine gene expressed from the inactive X chromosome," *Nature,* vol. 351, no. 6324, pp. 325-9, May 23 1991, doi: https://doi.org/10.1038/351325a0.
[25] A. Garitano-Trojaola, X. Agirre, F. Prosper, and P. Fortes, "Long non-coding RNAs in haematological malignancies," *Int J Mol Sci,* vol. 14, no. 8, pp. 15386-422, Jul 24 2013, doi: https://doi.org/10.3390/ijms140815386.
[26] M. Guttman *et al.*, "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals," *Nature,* vol. 458, no. 7235, pp. 223-7, Mar 12 2009, doi: https://doi.org/10.1038/nature07672.
[27] J. Yu, P. Ping, L. Wang, L. Kuang, X. Li, and Z. Wu, "A Novel Probability Model for LncRNA-Disease Association Prediction Based on the Naive Bayesian Classifier," *Genes (Basel),* vol. 9, no. 7, Jul 8 2018, doi: https://doi.org/10.3390/genes9070345.
[28] C. Lu *et al.*, "Prediction of lncRNA-disease associations based on inductive matrix completion," *Bioinformatics,* vol. 34, no. 19, pp. 3357-3364, Oct 1 2018, doi: https://doi.org/10.1093/bioinformatics/bty327.
[29] X. Chen, Y.-A. Huang, X.-S. Wang, Z.-H. You, and K. C. Chan, "FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model," *Oncotarget,* vol. 7, no. 29, p. 45948, 2016, doi: https://doi.org/10.18632/oncotarget.10008.
[30] Q. Hu *et al.*, "Oncogenic lncRNA downregulates cancer cell antigen presentation and intrinsic tumor suppression," *Nat Immunol,* vol. 20, no. 7, pp. 835-851, Jul 2019, doi: https://doi.org/10.1038/s41590-019-0400-7.
[31] E. Raveh, I. J. Matouk, M. Gilon, and A. Hochberg, "The H19 Long non-coding RNA in cancer initiation, progression and metastasis - a proposed unifying theory," *Mol Cancer,* vol. 14, no. 1, p. 184, Nov 4 2015, doi: https://doi.org/10.1186/s12943-015-0458-2.
[32] P. R. Chowdhury, S. Salvamani, B. Gunasekaran, H. B. Peng, and V. Ulaganathan, "H19: An Oncogenic Long Non-coding RNA in Colorectal Cancer," *Yale J Biol Med,* vol. 96, no. 4, pp. 495-509, Dec 2023, doi: https://doi.org/10.59249/TDBJ7410.
[33] C. Liu, Z. Chen, J. Fang, A. Xu, W. Zhang, and Z. Wang, "H19-derived miR-675 contributes to bladder cancer cell proliferation by regulating p53 activation," *Tumor Biology,* vol. 37, no. 1, pp. 263-270, 2016, doi: https://doi.org/10.1007/s13277-015-3779-2.
[34] G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, and N. Brockdorff, "Requirement for Xist in X chromosome inactivation," *Nature,* vol. 379, no. 6561, pp. 131-7, Jan 11 1996, doi: https://doi.org/10.1038/379131a0.
[35] Y. Yao *et al.*, "Knockdown of long non-coding RNA XIST exerts tumor-suppressive functions in human glioblastoma stem cells by up-regulating miR-152," *Cancer Lett,* vol. 359, no. 1, pp. 75-86, Apr 1 2015, doi: https://doi.org/10.1016/j.canlet.2014.12.051.
[36] R. A. Gupta *et al.*, "Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis," *Nature,* vol. 464, no. 7291, pp. 1071-6, Apr 15 2010, doi: https://doi.org/10.1038/nature08975.
[37] H. L. Sanger, G. Klotz, D. Riesner, H. J. Gross, and A. K. Kleinschmidt, "Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures," *Proc Natl Acad Sci U S A,* vol. 73, no. 11, pp. 3852-6, Nov 1976, doi: https://doi.org/10.1073/pnas.73.11.3852.

[38] C. C. Wang, C. D. Han, Q. Zhao, and X. Chen, "Circular RNAs and complex diseases: from experimental results to computational models," *Brief Bioinform,* vol. 22, no. 6, p. bbab286, Nov 5 2021, doi: https://doi.org/10.1093/bib/bbab286.
[39] L. S. Kristensen, M. S. Andersen, L. V. W. Stagsted, K. K. Ebbesen, T. B. Hansen, and J. Kjems, "The biogenesis, biology and characterization of circular RNAs," *Nat Rev Genet,* vol. 20, no. 11, pp. 675-691, Nov 2019, doi: https://doi.org/10.1038/s41576-019-0158-7.
[40] S. Ghosal, S. Das, R. Sen, P. Basak, and J. Chakrabarti, "Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits," *Front Genet,* vol. 4, p. 283, 2013, doi: https://doi.org/10.3389/fgene.2013.00283.
[41] W. Weng *et al.*, "Circular RNA ciRS-7—a promising prognostic biomarker and a potential therapeutic target in colorectal cancer," *Clinical Cancer Research,* vol. 23, no. 14, pp. 3918-3928, 2017, doi: https://doi.org/10.1158/1078-0432.CCR-16-2541.
[42] L. Yu, X. Gong, L. Sun, Q. Zhou, B. Lu, and L. Zhu, "The circular RNA Cdr1as act as an oncogene in hepatocellular carcinoma through targeting miR-7 expression," *PloS one,* vol. 11, no. 7, p. e0158347, 2016, doi: https://doi.org/10.1371/journal.pone.0158347.
[43] G. Floris, L. Zhang, P. Follesa, and T. Sun, "Regulatory Role of Circular RNAs and Neurological Disorders," *Mol Neurobiol,* vol. 54, no. 7, pp. 5156-5165, Sep 2017, doi: https://doi.org/10.1007/s12035-016-0055-4.
[44] L. Peng, X. Q. Yuan, and G. C. Li, "The emerging landscape of circular RNA ciRS-7 in cancer (Review)," *Oncol Rep,* vol. 33, no. 6, pp. 2669-74, Jun 2015, doi: https://doi.org/10.3892/or.2015.3904.
[45] J. Li, Y. Zheng, G. Sun, and S. Xiong, "Restoration of miR-7 expression suppresses the growth of Lewis lung cancer cells by modulating epidermal growth factor receptor signaling," *Oncol Rep,* vol. 32, no. 6, pp. 2511-6, Dec 2014, doi: https://doi.org/10.3892/or.2014.3519.
[46] L. He and G. J. Hannon, "MicroRNAs: small RNAs with a big role in gene regulation," *Nat Rev Genet,* vol. 5, no. 7, pp. 522-31, Jul 2004, doi: https://doi.org/10.1038/nrg1379.
[47] A. Esquela-Kerscher and F. J. Slack, "Oncomirs - microRNAs with a role in cancer," *Nat Rev Cancer,* vol. 6, no. 4, pp. 259-69, Apr 2006, doi: https://doi.org/10.1038/nrc1840.
[48] G. A. Calin and C. M. Croce, "MicroRNA signatures in human cancers," *Nat Rev Cancer,* vol. 6, no. 11, pp. 857-66, Nov 2006, doi: https://doi.org/10.1038/nrc1997.
[49] J. N. Goh *et al.*, "microRNAs in breast cancer: regulatory roles governing the hallmarks of cancer," *Biol Rev Camb Philos Soc,* vol. 91, no. 2, pp. 409-28, May 2016, doi: https://doi.org/10.1111/brv.12176.
[50] C. Xu *et al.*, "Prioritizing candidate disease miRNAs by integrating phenotype associations of multiple diseases with matched miRNA and mRNA expression profiles," *Mol Biosyst,* vol. 10, no. 11, pp. 2800-9, Nov 2014, doi: https://doi.org/10.1039/c4mb00353e.
[51] M. A. Yildirim, K. I. Goh, M. E. Cusick, A. L. Barabasi, and M. Vidal, "Drug-target network," *Nat Biotechnol,* vol. 25, no. 10, pp. 1119-26, Oct 2007, doi: https://doi.org/10.1038/nbt1338.
[52] S. M. Paul *et al.*, "How to improve R&D productivity: the pharmaceutical industry's grand challenge," *Nat Rev Drug Discov,* vol. 9, no. 3, pp. 203-14, Mar 2010, doi: https://doi.org/10.1038/nrd3078.
[53] L. Yao, J. A. Evans, and A. Rzhetsky, "Novel opportunities for computational biology and sociology in drug discovery," *Trends Biotechnol,* vol. 28, no. 4, pp. 161-70, Apr 2010, doi: https://doi.org/10.1016/j.tibtech.2010.01.004.
[54] J. A. DiMasi, L. Feldman, A. Seckler, and A. Wilson, "Trends in risks associated with new drug development: success rates for investigational drugs," *Clin Pharmacol Ther,* vol. 87, no. 3, pp. 272-7, Mar 2010, doi: https://doi.org/10.1038/clpt.2009.295.
[55] J. T. Dudley, T. Deshpande, and A. J. Butte, "Exploiting drug-disease relationships for computational drug repositioning," *Brief Bioinform,* vol. 12, no. 4, pp. 303-11, Jul 2011, doi: https://doi.org/10.1093/bib/bbr013.
[56] S. J. Swamidass, "Mining small-molecule screens to repurpose drugs," *Brief Bioinform,* vol. 12, no. 4, pp. 327-35, Jul 2011, doi: https://doi.org/10.1093/bib/bbr028.
[57] F. Moriaud *et al.*, "Identify drug repurposing candidates by mining the protein data bank," *Brief Bioinform,* vol. 12, no. 4, pp. 336-40, Jul 2011, doi: https://doi.org/10.1093/bib/bbr017.
[58] A. L. Hopkins, "Drug discovery: Predicting promiscuity," *Nature,* vol. 462, no. 7270, pp. 167-8, Nov 12 2009, doi: https://doi.org/10.1038/462167a.
[59] A. Masoudi-Nejad, Z. Mousavian, and J. H. Bozorgmehr, "Drug-target and disease networks: polypharmacology in the post-genomic era," *In Silico Pharmacol,* vol. 1, no. 1, p. 17, 2013, doi: https://doi.org/10.1186/2193-9616-1-17.
[60] E. Lounkine *et al.*, "Large-scale prediction and testing of drug activity on side-effect targets," *Nature,* vol. 486, no. 7403, pp. 361-7, Jun 10 2012, doi: https://doi.org/10.1038/nature11159.
[61] E. Pauwels, V. Stoven, and Y. Yamanishi, "Predicting drug side-effect profiles: a chemical fragment-based approach," *BMC Bioinformatics,* vol. 12, no. 1, p. 169, May 18 2011, doi: https://doi.org/10.1186/1471-2105-12-169.
[62] T. Takenaka, "Classical vs reverse pharmacology in drug discovery," *BJU Int,* vol. 88 Suppl 2, pp. 7-10; discussion 49-50, Sep 2001, doi: https://doi.org/10.1111/j.1464-410x.2001.00112.x.
[63] M. Dickson and J. P. Gagnon, "Key factors in the rising cost of new drug discovery and development," *Nat Rev Drug Discov,* vol. 3, no. 5, pp. 417-29, May 2004, doi: https://doi.org/10.1038/nrd1382.
[64] A. Ezzat, M. Wu, X. L. Li, and C. K. Kwoh, "Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey," *Brief Bioinform,* vol. 20, no. 4, pp. 1337-1357, Jul 19 2019, doi: https://doi.org/10.1093/bib/bby002.

[65] A. Ezzat, M. Wu, X. L. Li, and C. K. Kwoh, "Drug-target interaction prediction using ensemble learning and dimensionality reduction," *Methods,* vol. 129, pp. 81-88, Oct 1 2017, doi: https://doi.org/10.1016/j.ymeth.2017.05.016.
[66] A. Ezzat, M. Wu, X. L. Li, and C. K. Kwoh, "Drug-target interaction prediction via class imbalance-aware ensemble learning," *BMC Bioinformatics,* vol. 17, no. Suppl 19, p. 509, Dec 22 2016, doi: https://doi.org/10.1186/s12859-016-1377-y.
[67] A. Lakizadeh and S. M. H. Mir-Ashrafi, "Drug repurposing improvement using a novel data integration framework based on the drug side effect," *Informatics in Medicine Unlocked,* vol. 23, p. 100523, 2021, doi: https://doi.org/10.1016/j.imu.2021.100523.
[68] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nat Rev Drug Discov,* vol. 3, no. 8, pp. 673-83, Aug 2004, doi: https://doi.org/10.1038/nrd1468.
[69] L. Wang, J. Ding, L. Pan, D. Cao, H. Jiang, and X. Ding, "Artificial intelligence facilitates drug design in the big data era," *Chemometrics and Intelligent Laboratory Systems,* vol. 194, 2019, doi: https://doi.org/10.1016/j.chemolab.2019.103850.
[70] Y. Murakami, L. P. Tripathi, P. Prathipati, and K. Mizuguchi, "Network analysis and in silico prediction of protein-protein interactions with applications in drug discovery," *Curr Opin Struct Biol,* vol. 44, pp. 134-142, Jun 2017, doi: https://doi.org/10.1016/j.sbi.2017.02.005.
[71] R. Casadio, P. L. Martelli, and C. Savojardo, "Machine learning solutions for predicting protein–protein interactions," *WIREs Computational Molecular Science,* vol. 12, no. 6, 2022, doi: https://doi.org/10.1002/wcms.1618.
[72] F. Sommer and F. Backhed, "The gut microbiota--masters of host development and physiology," *Nat Rev Microbiol,* vol. 11, no. 4, pp. 227-38, Apr 2013, doi: https://doi.org/10.1038/nrmicro2974.
[73] E. Holmes, A. Wijeyesekera, S. D. Taylor-Robinson, and J. K. Nicholson, "The promise of metabolic phenotyping in gastroenterology and hepatology," *Nat Rev Gastroenterol Hepatol,* vol. 12, no. 8, pp. 458-71, Aug 2015, doi: https://doi.org/10.1038/nrgastro.2015.114.
[74] M. Ventura *et al.*, "Genome-scale analyses of health-promoting bacteria: probiogenomics," *Nat Rev Microbiol,* vol. 7, no. 1, pp. 61-71, Jan 2009, doi: https://doi.org/10.1038/nrmicro2047.
[75] J. C. Clemente, L. K. Ursell, L. W. Parfrey, and R. Knight, "The impact of the gut microbiota on human health: an integrative view," *Cell,* vol. 148, no. 6, pp. 1258-70, Mar 16 2012, doi: https://doi.org/10.1016/j.cell.2012.01.035.
[76] J. C. Arthur *et al.*, "Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer," *Nat Commun,* vol. 5, no. 1, p. 4724, Sep 3 2014, doi: https://doi.org/10.1038/ncomms5724.
[77] A. M. Thomas *et al.*, "Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation," *Nat Med,* vol. 25, no. 4, pp. 667-678, Apr 2019, doi: https://doi.org/10.1038/s41591-019-0405-7.
[78] N. Qin *et al.*, "Alterations of the human gut microbiome in liver cirrhosis," *Nature,* vol. 513, no. 7516, pp. 59-64, Sep 4 2014, doi: https://doi.org/10.1038/nature13568.
[79] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones, "miRBase: from microRNA sequences to function," *Nucleic Acids Res,* vol. 47, no. D1, pp. D155-D162, Jan 8 2019, doi: https://doi.org/10.1093/nar/gky1141.
[80] C. Cui, B. Zhong, R. Fan, and Q. Cui, "HMDD v4.0: a database for experimentally supported human microRNA-disease associations," *Nucleic Acids Res,* vol. 52, no. D1, pp. D1327-D1332, Jan 5 2024, doi: https://doi.org/10.1093/nar/gkad717.
[81] N. K. Singh, "microRNAs Databases: Developmental Methodologies, Structural and Functional Annotations," *Interdiscip Sci,* vol. 9, no. 3, pp. 357-377, Sep 2017, doi: https://doi.org/10.1007/s12539-016-0166-7.
[82] Q. Jiang *et al.*, "miR2Disease: a manually curated database for microRNA deregulation in human disease," *Nucleic Acids Research,* vol. 37, no. Database issue, pp. D98-104, Jan 2009, doi: https://doi.org/10.1093/nar/gkn714.
[83] F. Xie *et al.*, "deepBase v3.0: expression atlas and interactive analysis of ncRNAs from thousands of deep-sequencing data," *Nucleic Acids Research,* vol. 49, no. D1, pp. D877-D883, 2021, doi: https://doi.org/10.1093/nar/gkaa1039.
[84] B. Xie, Q. Ding, H. Han, and D. Wu, "miRCancer: a microRNA-cancer association database constructed by text mining on literature," *Bioinformatics,* vol. 29, no. 5, pp. 638-44, Mar 1 2013, doi: https://doi.org/10.1093/bioinformatics/btt014.
[85] X. Lin *et al.*, "LncRNADisease v3.0: an updated database of long non-coding RNA-associated diseases," *Nucleic Acids Res,* vol. 52, no. D1, pp. D1365-D1369, Jan 5 2024, doi: https://doi.org/10.1093/nar/gkad828.
[86] Y. Gao *et al.*, "Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data," *Nucleic acids research,* vol. 49, no. D1, pp. D1251-D1258, 2021, doi: https://doi.org/10.1093/nar/gkaa1006.
[87] D. S. Wishart *et al.*, "DrugBank 5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Res,* vol. 46, no. D1, pp. D1074-D1082, Jan 4 2018, doi: https://doi.org/10.1093/nar/gkx1037.
[88] M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, and M. Ishiguro-Watanabe, "KEGG for taxonomy-based analysis of pathways and genomes," *Nucleic Acids Res,* vol. 51, no. D1, pp. D587-D592, Jan 6 2023, doi: https://doi.org/10.1093/nar/gkac963.
[89] D. Szklarczyk, A. Santos, C. von Mering, L. J. Jensen, P. Bork, and M. Kuhn, "STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data," *Nucleic Acids Res,* vol. 44, no. D1, pp. D380-4, Jan 4 2016, doi: https://doi.org/10.1093/nar/gkv1277.
[90] L. Uran Landaburu *et al.*, "TDR Targets 6: driving drug discovery for human pathogens through intensive chemogenomic data integration," *Nucleic Acids Res,* vol. 48, no. D1, pp. D992-D1005, Jan 8 2020, doi: https://doi.org/10.1093/nar/gkz999.

[91] W. Ma *et al.*, "An analysis of human microbe-disease associations," *Brief Bioinform,* vol. 18, no. 1, pp. 85-97, Jan 2017, doi: https://doi.org/10.1093/bib/bbw005.
[92] N. Singh, V. Bhatia, S. Singh, and S. Bhatnagar, "MorCVD: A Unified Database for Host-Pathogen Protein-Protein Interactions of Cardiovascular Diseases Related to Microbes," *Sci Rep,* vol. 9, no. 1, p. 4039, Mar 11 2019, doi: https://doi.org/10.1038/s41598-019-40704-5.
[93] M. Urban *et al.*, "PHI-base in 2022: a multi-species phenotype database for Pathogen–Host Interactions," *Nucleic Acids Research,* vol. 50, no. D1, pp. D837-D847, 2022, doi: https://doi.org/10.1093/nar/gkab1037.
[94] D. Szklarczyk *et al.*, "The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest," *Nucleic Acids Res,* vol. 51, no. D1, pp. D638-D646, Jan 6 2023, doi: http://doi.org/10.1093/nar/gkac1000.
[95] X. Chen, D. Xie, Q. Zhao, and Z.-H. You, "MicroRNAs and complex diseases: from experimental results to computational models," *Briefings in Bioinformatics,* vol. 20, no. 2, pp. 515-539, 2019, doi: https://doi.org/10.1093/bib/bbx130.
[96] X. Chen and G. Y. Yan, "Semi-supervised learning for potential human microRNA-disease associations inference," *Sci Rep,* vol. 4, p. 5501, Jun 30 2014, doi: https://doi.org/10.1038/srep05501.
[97] X. Chen, C. C. Wang, J. Yin, and Z. H. You, "Novel Human miRNA-Disease Association Inference Based on Random Forest," *Mol Ther Nucleic Acids,* vol. 13, pp. 568-579, Dec 7 2018, doi: https://doi.org/10.1016/j.omtn.2018.10.005.
[98] X. Chen, Q. F. Wu, and G. Y. Yan, "RKNNMDA: Ranking-based KNN for MiRNA-Disease Association prediction," *RNA Biology,* vol. 14, no. 7, pp. 952-962, Jul 3 2017, doi: https://doi.org/10.1080/15476286.2017.1312226.
[99] X. Chen, J. R. Yang, N. N. Guan, and J. Q. Li, "GRMDA: Graph Regression for MiRNA-Disease Association Prediction," *Front Physiol,* vol. 9, p. 92, 2018, doi: https://doi.org/10.3389/fphys.2018.00092.
[100] W. Y. Kang, Y. L. Gao, Y. Wang, F. Li, and J. X. Liu, "KFDAE: CircRNA-Disease Associations Prediction Based on Kernel Fusion and Deep Auto-Encoder," *IEEE J Biomed Health Inform,* vol. 28, no. 5, pp. 3178-3185, May 2024, doi: https://doi.org/10.1109/JBHI.2024.3369650.
[101] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics,* vol. 24, no. 13, pp. i232-40, Jul 1 2008, doi: https://doi.org/10.1093/bioinformatics/btn162.
[102] M. Wen *et al.*, "Deep-Learning-Based Drug-Target Interaction Prediction," *J Proteome Res,* vol. 16, no. 4, pp. 1401-1409, Apr 7 2017, doi: https://doi.org/10.1021/acs.jproteome.6b00618.
[103] H. T. Zou, B. Y. Ji, and X. L. Xie, "A multi-source molecular network representation model for protein-protein interactions prediction," *Sci Rep,* vol. 14, no. 1, p. 6184, Mar 14 2024, doi: https://doi.org/10.1038/s41598-024-56286-w.
[104] L. H. Peng, J. Yin, L. Zhou, M. X. Liu, and Y. Zhao, "Human Microbe-Disease Association Prediction Based on Adaptive Boosting," *Front Microbiol,* vol. 9, p. 2440, 2018, doi: https://doi.org/10.3389/fmicb.2018.02440.
[105] F. Wang *et al.*, "LRLSHMDA: laplacian regularized least squares for human microbe–disease association prediction," *Scientific reports,* vol. 7, no. 1, p. 7601, 2017, doi: https://doi.org/10.1038/s41598-017-08127-2.