Computational Biology and Chemistry 89 (2020) 107406

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac

Research Article

LMI-DForest: A deep forest model towards the prediction of lncRNA-miRNA interactions

Wei Wang [a], Xiaoqing Guan [b], Muhammad Tahir Khan [c], Yi Xiong [d,*], Dong-Qing Wei [d,e,**]

[a] School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China
[b] Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai, China
[c] Institute of Molecular Biology and Biotechnology, The University of Lahore Pakistan, Pakistan
[d] State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China
[e] Peng Cheng Laboratory, Shenzhen, Guangdong, China

ARTICLE INFO

Keywords:
Deep learning
DeepForest
lncRNAs
miRNAs
lncRNA-miRNA interaction

ABSTRACT

The interactions between miRNAs and long non-coding RNAs (lncRNAs) are subject to intensive recent studies due to its critical role in gene regulations. Computational prediction of lncRNA-miRNA interactions has become a popular alternative strategy to the experimental methods for identification of underlying interactions. It is desirable to develop the machine learning-based models for prediction of lncRNA-miRNA based on the experimentally validated interactions between lncRNAs and miRNAs. The accuracy and robustness of existing models based on machine learning techniques are subject to further improvement.

Considering that the attributes of lncRNA and miRNA contribute key importance in the interaction between these two RNAs, a deep learning model, named LMI-DForest, is proposed here by combining the deep forest and autoencoder strategies. Systematic comparison on the experiment validated datasets for lncRNA-miRNA interaction datasets demonstrates that the proposed method consistently shows superior performance over the other machine learning models in the lncRNA-miRNA interaction prediction.

1. Introduction

Long non-coding RNAs (lncRNAs) is one kind of non-coding RNAs (ncRNAs), whose lengths are more than 200 nucleotides (Hung and Chang, 2010; Zhao et al., 2018; Ji et al., 2019). Previous studies (Fatica and Bozzoni, 2014; Turner et al., 2014) have shown that lncRNAs are involved in the regulation of gene expression in different levels, such as transcriptional, posttranscriptional, and epigenetic regulation, and many biological processes, such as chromatin remodeling, gene imprinting, immune response, etc. On the other side, as the endogenous small and non-coding RNA molecules, microRNAs (miRNAs) post-transcriptionally can regulate gene expression (Berezikov et al., 2006; Xie et al., 2019; Zhang et al., 2019a; Huang et al., 2020; Yang et al., 2020a). With accumulated wet experiments, it has been widely approved that lncRNAs and miRNAs play key roles in cell proliferation and cell differentiation, and the interactions between lncRNA and miRNA can lead to some diseases (Kallen et al., 2013; Zhang et al., 2014; Tang et al., 2018; Kuang et al., 2019; Wang et al., 2019a). In some cases, lncRNAs can act as decoys or sponges to regulate the behavior of miRNAs, and miRNAs can trigger lncRNAs decay. Therefore, identification of the interactions between lncRNAs and miRNAs is essential to understand their functions in diseases (Jalali et al., 2013; Veneziano et al., 2019).

LncRNA-miRNA interactions between these two RNAs form a complex regulation network for controlling gene expression on transcriptional, post-transcriptional, and post-translational levels. There are a few methods in the prediction of lncRNA-miRNA interactions. With data collected from some biological experiments, most of these proposed methods are based on the network statistical. Such as the key dysregulation of ncRNA expression in pathogenesis is inferred with the constructed miRNA-mediated network of coding and non-coding RNA interactions. LncRNA-miRNA-RNA interaction network with sensitivity correlation for each triplet is obtained from the breast cancer data (Conte et al., 2017). Based on the knonwn miRNA-lncRNA interaction

* Corresponding author.
** Corresponding author at: Peng Cheng Laboratory, Shenzhen, Guangdong, China.
E-mail addresses: xiongyi@sjtu.edu.cn (Y. Xiong), dqwei@sjtu.edu.cn (D.-Q. Wei).

https://doi.org/10.1016/j.compbiolchem.2020.107406
Received 16 July 2020; Received in revised form 12 October 2020; Accepted 15 October 2020
Available online 20 October 2020

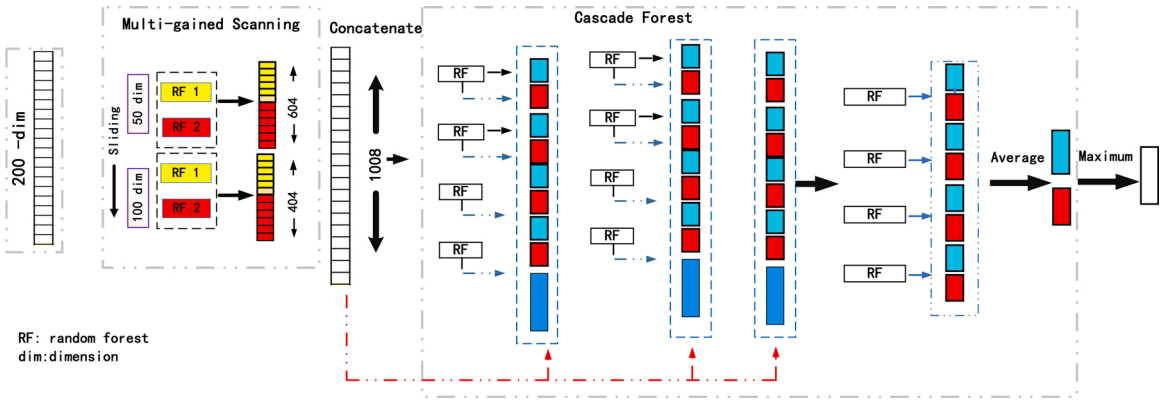

Fig. 1. The whole process of DeepForest. Suppose the sliding windows are 50-dim and 100-dim. In each level of the cascade consists, there are two random forests (RF). In the output of each forest, a two-dimensional class vector, which is then concatenated for re-representation of the original input.

network, group preference Bayesian collaborative filtering model (Huang et al., 2018b) is applied with picked top ranking list for an individual miRNA or lncRNA. INLMI (Hu et al., 2018) is proposed to predict the lncRNA-miRNA interactions according to the integrating the expression similarity network and the sequence similarity network. The prediction of interactions between lncRNA and miRNA is based on logistic matrix factorization (Liu et al., 2020) with neighborhood regularized, and the prediction of lncRNA-miRNA with the expression profile (Huang et al., 2018a). Moreover, there are other available prediction methods (Huang et al., 2019; Ismalia et al., 2019; Zhang et al., 2019b; Zhou et al., 2019; Fan et al., 2020; Hu et al., 2020; Kang et al., 2020; Wang et al., 2020a; Wong et al., 2020; Yang et al., 2020b; Zhang et al., 2020b). However, most of the interactions between lncRNAs and miRNAs are not known until now, a desirable model should be capable of predicting their interactions.

Recently, deep neural networks (DNN) is a hot topic that has achieved great success in various application areas, such as natural language processing, visual recognition, and bioinformatics (Deng et al., 2020; Li et al., 2020a, d; Zhang et al., 2020a). With the high requirements of the amount of training data and hyper-parameter tuning skills, the application of DNN in the task of classification will be limited. As an alternative approach to DNN, deep forest (DeepForest) (Zhou and Feng, 2017) is proposed with two ensemble components: multi-gained scanning, which scan local context from high dimensionality to learn representations of input data according to different random forests, and cascade forest, in which neurons of deep neural networks have been replaced with many different random forests. In DeepForest, compared to DNN, fewer hyper-parameters and less parameter tuning skills in are required the training process.

In this work, a new model called LMI-DForest is proposed for inferring new lncRNA-miRNA interactions on a large scale by combining DeepForest with the autoencoder model. LMI-DForest is based on the known lncRNA-miRNA interactions along with the expression levels of lncRNA and miRNA. With the experimental dataset of lncRNA and miRNA interactions, the model is valiadated by the 2-fold, 5-fold, and 10-fold cross-validation, and the corresponding average AUCs are 0.9933, 0.9940, and 0.9940. Compared to other state-of-the-art methods, LMI-DForest has reached higher prediction performance. In the case study based on lncRNAs or miRNAs, which do not have any known interactions, LMI-DForest helps to predict some novel interactions which do not exist in the dataset.

The rest of this paper is organized as follows. In the section 2, we describe matrial and methods, LMI-DForest, and how the DeepForest is considered in the autoencoder. In section 3 experimental process and results have been given. Finally, In the summaries and discussion of the end, performance of LMI-DForest is summarized and some further applications are also proposed here.

2. Materials and methods

2.1. Datasets

Data containing experimentally confirmed lncRNA-miRNA interactions is very limited, and most of these interactions are inferred from the expression profiles to lncRNA or miRNA. To guranteed true data for our prediction, experimentally confirmed lncRNA-miRNA interactions are from the lncRNASNP2 database (version v2.0) (Miao et al., 2018), which is an updated version to lncRNASNP, and provides comprehensive information to lncRNAs, including lncRNA expression profiling, expanded lncRNA-associated diseases, and noncoding variants in lncRNAs (available at http://bioinfo.life.hust.edu.cn/lncRNASNP). In lncRNASNP2, the database is linked via the IDs of lncRNAs and integrated from a different source of the public database. A total of 18 595 lncRNA-miRNA interactions (Li et al., 2014) between 3521 lncRNAs and 276 miRNAs are verified with high reliance. Considering the superiority of expression file (Huang et al., 2018b), features to lnRNAs and miRNAs are based on the expression file.

In the lncRNAs database, NONCODE (Fang et al., 2018) is an integrated knowledge database of non-coding RNAs (ncRNAs), including the ncRNA sequences and related information (e.g. function, cellular role, cellular location, chromosomal information, etc.). Features of expression file to lncRNAs are present with expression levels to 22 different tissues/cell lines. Features to miRNA are from microrna.org (http://www.microrna.org/microrna/home.do) (Betel et al., 2008), which is a comprehensive database of microRNA expression profiles and target prediction. Expression files to each miRNA are for 172 various tissues and cell lines in the human body, which are derived from a comprehensive sequencing project of a large set of human tissues and cell lines of normal and disease origins.

The dataset is constructed with entire lncRNA and miRNA interactions. In the constructed dataset, positive data is based on the known interactions between lncRNAs and miRNAs, and the negative samples contain all unknown lncRNA and miRNA interactions.

2.2. DeepForest

There are two components in the DeepForest (Zhou and Feng, 2017): multi-gained scanning and cascade forest. In the multi-gained scanning, which tries to extract local features by scanning raw features to generate a series of local low-dimensional feature vectors. With such low-dimensional vectors, the class distribution of the input features will be learned with a series of forests. Suppose there are 200 features, with the sliding window of 50, 151 feature vectors will be generated when sliding the feature by the window for one feature. With features from these positive/negative instances, the random forest will be trained for each 100-dimensional vector and generate two-dimensional class

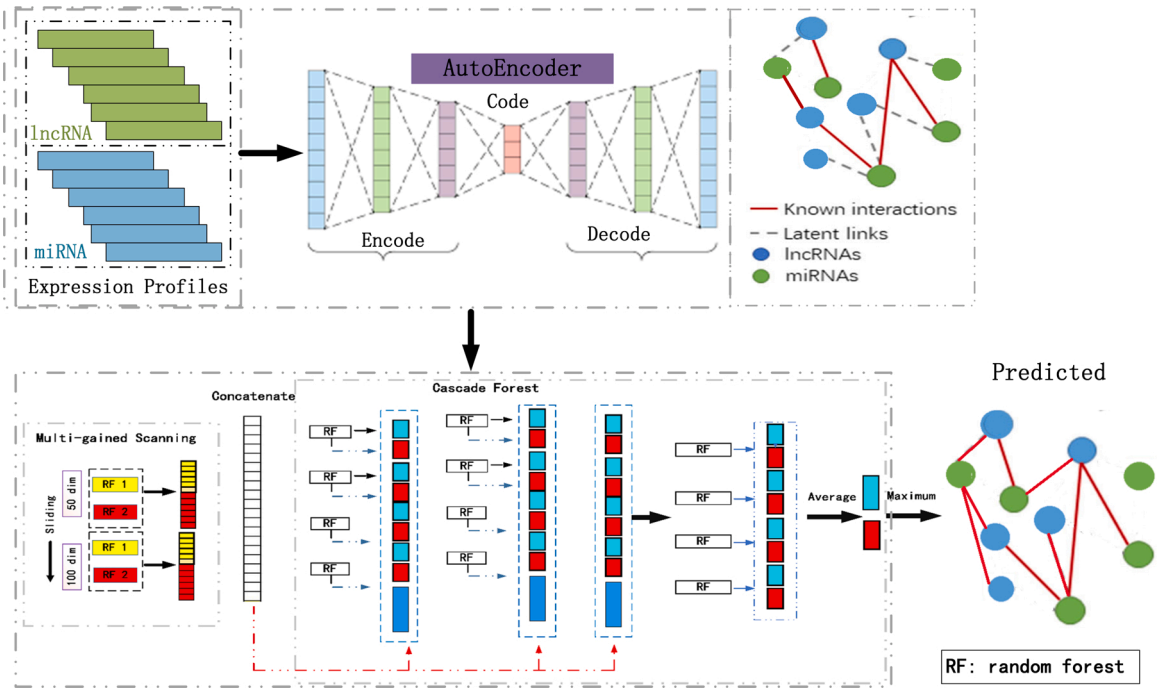

Fig. 2. Flow chart of LMI-DForest.

vectors for each forest/classifier.

As an alternative to deep neural networks, DeepForest is very powerful in hyper-level representations at low expense, which has various applications in bioinformatics field recently (Chu et al., 2019; Su et al., 2019; Wang et al., 2020b; Yu et al., 2020; Zeng et al., 2020; Zhu et al., 2020). Considering the powerful classification probability in most of the application, cascade forest comes true by assembling amount of decision-tree based random forest. With the importance of diversity in ensemble learning (Zhou, 2012), two random forests are employed here as the basis classifier. In each layer, estimated classifier distribution will be concatenated to the input feature, and worked as the inputs for the next layer.

The framework of DeepForest is illustrated in Fig. 1. With two different sliding windows, 50 and 100, two generated feature vectors will be concatenated into one and as the input for cascade forest. In these two generated feature vectors, one is 604 ($151 \times 2 \times 2$), the other is 404 ($101 \times 2 \times 2$). In the test instances, with the transformed feature representation from a multi-grained scanning procedure, and the cascade will be performed till the last level. The maximum aggregated value of these 4 generated 2-dimensional class vectors will be assumed as the final prediction.

2.3. LMI-DForest: an autoencoder prediction model for lncRNA-miRNA interactions

In the prediction of lncRNA-miRNA interactions, the association between lncRNA and miRNA can be assumed as a heterogeneous bipartite network. Assuming there are N_m miRNAs nodes and N_l lncRNAs nodes, an adjacent matrix $I_{N_m \times N_l}$ will be built to represent the interaction between these miRNAs and lncRNAs. Each element of the matrix $Int_{N_m \times N_l} = (int_{ij})_{N_m \times N_l}$ will represent if there is any interaction between these miRNAs and lncRNAs, in which

$$int_{ij} = f(x) = \begin{cases} 1, & \textit{if there is an interaction between ith miRNA and jth lncRNA} \\ 0, & \textit{Otherwise} \end{cases} \quad (1)$$

The prediction of lncRNA and miRNA interaction can be assumed as referring to the value of unobserved entries in $I_{N_m \times N_l}$ using supervised learning on the observed ones. Feature matrix with expression profiles to miRNAs are $E_{miRNA} = (em_{ij})_{N_m \times E_m}$, and feature matrix to lnRNA are $E_{lncRNA} = (el_{ij})_{N_l \times E_l}$, in which E_m are the feature number to miRNA expression profiles, E_l are the feature number to lncRNA expression profiles.

There are two components in the proposed model LMI-DForest here: 1) the AutoEncoder layer to handle the high-dimension features of lncRNAs and miRNAs on their interaction network; 2) interactions between lncRNA and miRNA will be predicted based on the handled feature in the first component. The whole flow of LMI-DForest is described as below (Fig. 2).

In the AutoEncoder, features to miRNA and lncRNA can be integrated into the matrix

$$Feature = \begin{bmatrix} E_{lncRNA} & 0 \\ 0 & E_{lncRNA} \end{bmatrix} \quad (2)$$

And experimentally verified interaction between lncRNA and miRNA is

$$INTMAT = \begin{bmatrix} 0 & Int_{N_m \times N_l} \\ Int^T_{N_m \times N_l} & 0 \end{bmatrix} \quad (3)$$

3. Experiments and results

3.1. Evaluation of LMI-DForest

With DeepForest in the experiments, we evaluate the effectiveness of LMI-DForest regarding its ability to integrate the raw data of the input feature. With the raw feature, the LMI-DForest is compared on 2-fold, 5-fold, and 10-fold cross-validation (2-CV, 5-CV and 10-CV).

With 5-CV to evaluate prediction models. The lncRNA-miRNA interactions will be randomly split into 5 subsets. One subset is used as the testing data, and the others are used as training data in each fold. In each fold of each prediction model, the following evaluation metrics are calculated, which are widely used by the machine learning-based studies

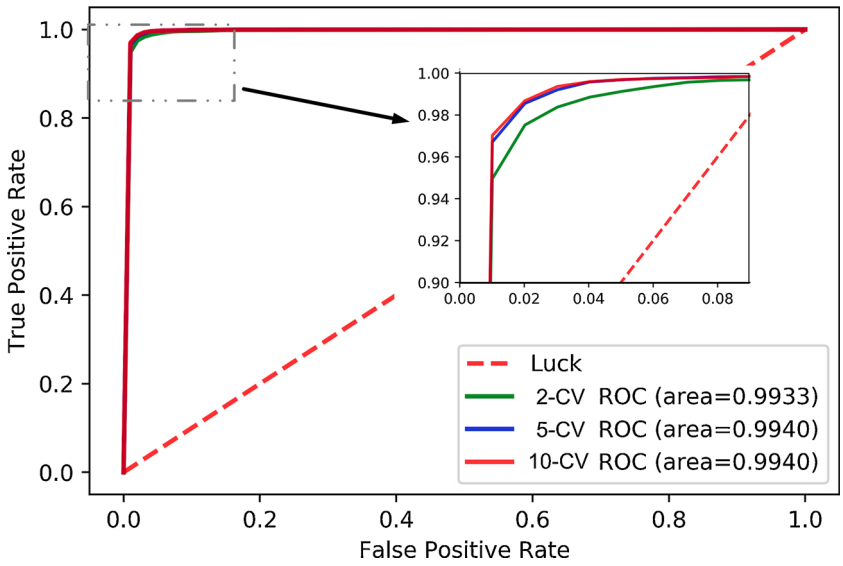

Fig. 3. The ROC curves yielded by LMI-DForest on 2-fold (2-CV), 5-fold (5-CV), and 10-fold (10-CV) cross-validation.

in the bioinformatics field (He et al., 2018; Xiong et al., 2018; Lian et al., 2019; Yang et al., 2019; Zhu et al., 2019; Cheng et al., 2020; Jia et al., 2020; Li et al., 2020b, c; Lissabet et al., 2020; Mu et al., 2020; Yang et al., 2020c; Zhang et al., 2020c).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (7)$$

where TP is true positive, FP is false positive, FN is false negative, and TN is true negative.

4. Results

As a result, LMI-DForest can achieve the best prediction performance with average area under curves (AUCs) of 0.9933, 0.9940, and 0.9940 in the 2-CV, 5-CV, and 10-CV, respectively. receiver operating characteristic curve (ROC) curves on 2-CV, 5-CV, and 10-CV are shown in Fig. 3 as below.

4.1. Performance comparison with other methods

To validate the effectiveness of LMI-DForest, the proposed method was compared with other machine learning models: Random Forest (Tin Kam, 2020), XGBoost (Chen and Guestrin, 2016), support vector machines (SVMs) with rbf kernel (Chang and Lin, 2011), and DeepForest (Zhou and Feng, 2017) on this lncRNA-miRNA interaction data set In the experiments, the data in training and test set are split in the ratio of 80 % vs 20 %, which follows a stratified approach (Hastie et al., 2009). In multi-gained scanning part, there are 500 trees in each forest, and cascade forest are based on 500 trees as the default. In the training and prediction, two completely random forests and two partial random forests were used in both multi-gained scanning and cascade forest. In the two partial random forests, $\sqrt{d}$ (d: number of input features) of features were selected as the candidates and separated with gini values. To overcome of over-fitting, 5-CV is used to evaluate the overall accuracy.

Table 1
Performances of different models based on lncRNA-miRNA interaction dataset.

Methods	Accuracy	F1-measure	Recall	Precision
LMI-DForest	0.9930	0.9445	0.9247	0.9653
DeepForest	0.9387	0.8806	0.9355	0.8318
Random Forest	0.9372	0.7357	0.9024	0.6210
XGBoost	0.9375	0.8279	0. 9332	0.7440
SVM (rbf kernel)	0.9217	0.8145	0.8296	0.8001

In comparison to LMI-DForest with Random Forest, there are 500 trees used here for each random forest, and $\sqrt{d}$ of features were selected as candidate features and separated with gini values. And in DeepForest, similar to the part of DeepForest in LMI-Dforest, both multi-gained-scanning andcascade forest were based on 500 trees in each forest as the default. In the training and prediction period, two completely random forests and two partial random forests were used for both both multi-gained scanning and cascade forest. With Table 1 as below, we can get comparable results when compared to DeepForest from Accuracy, recall, and precision, but reached some better performance than Random Forest, XGBoost and SVM. All these experiments were performed in the operating system of 64-bit Ubuntu 16.04.6 LTS (GNU/Linux 4.15.0−112-generic x86_64) in a machine with Intel(R) Xeon(R) CPU E5−2640 v3 @ 2.60 GHz.

5. Discussion

As an alternative to deep learning, DeepForest has been proved to be very powerful in the classification in practice (Zhou and Feng, 2017). As a further implementations and applications of the standard deep forest model (DeepForest), AutoEncoder is employed here to handle high-dimension before the classification with DeepForest. It is shown to be an effective method in the label prediction of lncRNA-miRNA interactions. Our LMI-DForest method is an effective option to investigate label classification by applying deep learning on small-scale biology datasets.

Based on the data constructed in this study, LMI-DForest is compared with other machine learning classification models in the performance metrics. LMI-DForest has achieved comparable performance on the dataset with original features. In further studies, our proposed approach will be tested on more strictly experimental settings and applied on other more similar bioinformatics problems, such as various types of

associations between ncRNAs, ncRNAs and disease, ncRNA and drug targets, small molecules and ncRNAs, genome analysis applications, etc (Ling et al., 2013; Chen et al., 2018; Bai et al., 2019; Wang et al., 2019b).

Funding

This work was supported by the grants from the Key Research Area Grant2016YFA0501703 of the Ministry of Science and Technology of China, theNational Science Foundation of China (Grant Nos.32070662,61832019,32030063), theScience and Technology Commission of Shanghai Municipality (Grant No.:19430750600), the-Natural Science Foundation of Henan Province (162300410060), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research atShanghai Jiao Tong University (YG2017ZD14).

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

The computations were partially performed at the Center for High-Performance Computing, Shanghai Jiao Tong University.

References

Bai, Y., Dai, X., Ye, T., Zhang, P., Yan, X., Gong, X., et al., 2019. PlncRNADB: a repository of plant lncRNAs and lncRNA-RBP protein interactions. Curr. Bioinform. 14 (7), 621–627. https://doi.org/10.2174/1574893614666190131161002.

Berezikov, E., Cuppen, E., Plasterk, R.H., 2006. Approaches to microRNA discovery. Nat. Genet. 38 (Suppl), S2–7. https://doi.org/10.1038/ng1794.

Betel, D., Wilson, M., Gabow, A., Marks, D.S., Sander, C., 2008. The microRNA.oRg resource: targets and expression. Nucleic Acids Res. 36, D149–153. https://doi.org/10.1093/nar/gkm995 (Database issue).

Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27. https://doi.org/10.1145/1961189.1961199.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chen, X., Guan, N.N., Sun, Y.Z., Li, J.Q., Qu, J., 2018. MicroRNA-small molecule association identification: from experimental results to computational models. Brief Bioinform. https://doi.org/10.1093/bib/bby098.

Cheng, N., Li, M., Zhao, L., Zhang, B., Yang, Y., Zheng, C.H., et al., 2020. Comparison and integration of computational methods for deleterious synonymous mutation prediction. Brief Bioinform 21 (3), 970–981. https://doi.org/10.1093/bib/bbz047.

Chu, Y., Kaushik, A.C., Wang, X., Wang, W., Zhang, Y., Shan, X., et al., 2019. DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. Brief Bioinform. https://doi.org/10.1093/bib/bbz152.

Conte, F., Fiscon, G., Chiara, M., Colombo, T., Farina, L., Paci, P., 2017. Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer. PLoS One 12 (2). https://doi.org/10.1371/journal.pone.0171661 e0171661.

Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W., Liu, S., 2020. A multimodal deep learning framework for predicting drug-drug interaction events. Bioinformatics. https://doi.org/10.1093/bioinformatics/btaa501.

Fan, Y.X., Cui, J., Zhu, Q.Q., 2020. Heterogeneous graph inference based on similarity network fusion for predicting lncRNA-miRNA interaction. RSC Adv. 10 (20), 11634–11642. https://doi.org/10.1039/c9ra11043g.

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., et al., 2018. NONCODEV5: a comprehensive annotation database for long non-coding RNAs. Nucleic Acids Res. 46 (D1), D308–D314. https://doi.org/10.1093/nar/gkx1107.

Fatica, A., Bozzoni, I., 2014. Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. 15 (1), 7–21. https://doi.org/10.1038/nrg3606.

Hastie, T., Tibshirani, R., Friedman, J., 2009. Model assessment and Selection. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer New York, New York, NY, pp. 219–259.

He, J., Fang, T., Zhang, Z., Huang, B., Zhu, X., Xiong, Y., 2018. PseUI: pseudouridine sites identification based on RNA sequence information. BMC Bioinformatics 19 (1), 306. https://doi.org/10.1186/s12859-018-2321-0.

Hu, P.W., Huang, Y.A., Chan, K.C.C., You, Z.H., 2018. Discovering an integrated network in heterogeneous data for predicting lncRNA-miRNA interactions. Intell. Comput. Theories Appl. Pt I 10954, 539–545. https://doi.org/10.1007/978-3-319-95930-6_51.

Hu, P., Huang, Y.A., Chan, K.C.C., You, Z.H., 2020. Learning multimodal networks from heterogeneous data for prediction of lncRNA-miRNA interactions. IEEEACM Trans. Comput. Biol. Bioinform. 17 (5), 1516–1524. https://doi.org/10.1109/TCBB.2019.2957094.

Huang, Y.A., Chan, K.C.C., You, Z.H., 2018a. Constructing prediction models from expression profiles for large scale lncRNA-miRNA interaction profiling. Bioinformatics 34 (5), 812–819. https://doi.org/10.1093/bioinformatics/btx672.

Huang, Z.A., Huang, Y.A., You, Z.H., Zhu, Z., Sun, Y., 2018b. Novel link prediction for large-scale miRNA-lncRNA interaction network in a bipartite graph. BMC Med. Genomics 11 (Suppl 6), 113. https://doi.org/10.1186/s12920-018-0429-8.

Huang, Y.A., Huang, Z.A., You, Z.H., Zhu, Z., Huang, W.Z., Guo, J.X., et al., 2019. Predicting lncRNA-miRNA interaction via graph convolution auto-encoder. Front. Genet. 10, 758. https://doi.org/10.3389/fgene.2019.00758.

Huang, F., Yue, X., Xiong, Z., Yu, Z., Liu, S., Zhang, W., 2020. Tensor decomposition with relational constraints for predicting multiple types of microRNA-disease associations. Brief Bioinform. https://doi.org/10.1093/bib/bbaa140.

Hung, T., Chang, H.Y., 2010. Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol. 7 (5), 582–585. https://doi.org/10.4161/rna.7.5.13216.

Ismalia, B., Qiang, K., Luan, Y.S., Jun, M., 2019. Predicting miRNA-lncRNA interactions and recognizing their regulatory roles in stress response of plants. Math. Biosci. 312, 67–76. https://doi.org/10.1016/j.mbs.2019.04.006.

Jalali, S., Bhartiya, D., Lalwani, M.K., Sivasubbu, S., Scaria, V., 2013. Systematic transcriptome wide analysis of lncRNA-miRNA interactions. PLoS One 8 (2). https://doi.org/10.1371/journal.pone.0053823 e53823.

Ji, J., Tang, J., Xia, K.-j., Jiang, R., 2019. LncRNA in Tumorigenesis Microenvironment. Curr. Bioinform. 14 (7), 640–641. https://doi.org/10.2174/1574893614071909171654.

Jia, C., Bi, Y., Chen, J., Leier, A., Li, F., Song, J., 2020. PASSION: an ensemble neural network approach for identifying the binding sites of RBPs on circRNAs. Bioinformatics 36 (15), 4276–4282. https://doi.org/10.1093/bioinformatics/btaa522.

Kallen, A.N., Zhou, X.B., Xu, J., Qiao, C., Ma, J., Yan, L., et al., 2013. The imprinted H19 lncRNA antagonizes let-7 microRNAs. Mol. Cell 52 (1), 101–112. https://doi.org/10.1016/j.molcel.2013.08.027.

Kang, Q., Meng, J., Cui, J., Luan, Y., Chen, M., 2020. PmliPred: a method based on hybrid model and fuzzy decision for plant miRNA-lncRNA interaction prediction. Bioinformatics 36 (10), 2986–2992. https://doi.org/10.1093/bioinformatics/btaa074.

Kuang, L., Zhao, H., Wang, L., Xuan, Z., Pei, T., 2019. A novel approach based on point cut set to predict associations of diseases and LncRNAs. Curr. Bioinform. 14 (4), 333–343. https://doi.org/10.2174/1574893613666181026122045.

Li, J.H., Liu, S., Zhou, H., Qu, L.H., Yang, J.H., 2014. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 42, D92–97. https://doi.org/10.1093/nar/gkt1248 (Database issue).

Li, F., Chen, J., Leier, A., Marquez-Lago, T., Liu, Q., Wang, Y., et al., 2020a. DeepCleave: a deep learning predictor for caspase and matrix metalloprotease substrates and cleavage sites. Bioinformatics 36 (4), 1057–1065. https://doi.org/10.1093/bioinformatics/btz721.

Li, F., Leier, A., Liu, Q., Wang, Y., Xiang, D., Akutsu, T., et al., 2020b. Procleave: predicting protease-specific substrate cleavage sites by combining sequence and structural information. Genom. Proteom. Bioinf. https://doi.org/10.1016/j.gpb.2019.08.002.

Li, K., Zhang, S., Yan, D., Bin, Y., Xia, J., 2020c. Prediction of hot spots in protein-DNA binding interfaces based on supervised isometric feature mapping and extreme gradient boosting. BMC Bioinf. 21 (Suppl 13), 381. https://doi.org/10.1186/s12859-020-03683-3.

Li, M., Wang, Y., Li, F., Zhao, Y., Liu, M., Zhang, S., et al., 2020d. A deep learning-based method for identification of bacteriophage-host interaction. IEEEACM Trans. Comput. Biol. Bioinform. https://doi.org/10.1109/TCBB.2020.3017386. PP.

Lian, X., Yang, S., Li, H., Fu, C., Zhang, Z., 2019. Machine-learning-based predictor of human-bacteria protein-Protein interactions by incorporating comprehensive host-network properties. J. Proteome Res. 18 (5), 2195–2205. https://doi.org/10.1021/acs.jproteome.9b00074.

Ling, H., Fabbri, M., Calin, G.A., 2013. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. Nat. Rev. Drug Discov. 12 (11), 847–865. https://doi.org/10.1038/nrd4140.

Lissabet, J.F.B., Belen, L.H., Farias, J.G., 2020. PPLK(+)C: a bioinformatics tool for predicting peptide ligands of potassium channels based on primary structure information. Interdiscip. Sci. 12 (3), 258–263. https://doi.org/10.1007/s12539-019-00356-5.

Liu, H.S., Ren, G.F., Chen, H.Y., Liu, Q., Yang, Y.J., Zhao, Q., 2020. Predicting lncRNA-miRNA interactions based on logistic matrix factorization with neighborhood regularized. Knowledge Based Syst. 191. https://doi.org/10.1016/j.knosys.2019.105261.

Miao, Y.R., Liu, W., Zhang, Q., Guo, A.Y., 2018. lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. Nucleic Acids Res. 46 (D1), D276–D280. https://doi.org/10.1093/nar/gkx1004.

Mu, Y., Zhang, R., Wang, L., Liu, X., 2020. iPseU-layer: identifying RNA pseudouridine sites using layered ensemble model. Interdiscip. Sci. 12 (2), 193–203. https://doi.org/10.1007/s12539-020-00362-y.

Su, R., Liu, X., Wei, L., Zou, Q., 2019. Deep-Resp-Forest: a deep forest model to predict anti-cancer drug response. Methods 166, 91–102. https://doi.org/10.1016/j.ymeth.2019.02.009.

Tang, W., Wan, S., Yang, Z., Teschendorff, A.E., Zou, Q., 2018. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. Bioinformatics 34 (3), 398–406. https://doi.org/10.1093/bioinformatics/btx622.

Tin Kam, H., 2020. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, 271, pp. 278–282.
Turner, M., Galloway, A., Vigorito, E., 2014. Noncoding RNA and its associated proteins as regulatory elements of the immune system. Nat. Immunol. 15 (6), 484–491. https://doi.org/10.1038/ni.2887.
Veneziano, D., Marceca, G.P., Di Bella, S., Nigita, G., Distefano, R., Croce, C.M., 2019. Investigating miRNA-lncRNA interactions: computational tools and resources. Methods Mol. Biol. 1970, 251–277. https://doi.org/10.1007/978-1-4939-9207-2_14.
Wang, L., Xuan, Z., Zhou, S., Kuang, L., Pei, T., 2019a. A novel model for predicting LncRNA-disease associations based on the LncRNA-MiRNA-disease interactive network. Curr. Bioinform. 14 (3), 269–278. https://doi.org/10.2174/1574893613666180703105258.
Wang, W.T., Han, C., Sun, Y.M., Chen, T.Q., Chen, Y.Q., 2019b. Noncoding RNAs in cancer therapy resistance and targeted drug development. J. Hematol. Oncol. 12 (1), 55. https://doi.org/10.1186/s13045-019-0748-z.
Wang, M.N., You, Z.H., Li, L.P., Wong, L., Chen, Z.H., Gan, C.Z., 2020a. GNMFLMI: graph regularized nonnegative matrix factorization for predicting LncRNA-MiRNA interactions. IEEE Access 8, 37578–37588. https://doi.org/10.1109/Access.2020.2974349.
Wang, W., Dai, Q., Li, F., Xiong, Y., Wei, D.Q., 2020b. MLCDForest: multi-label classification with deep forest in disease prediction for long non-coding RNAs. Brief Bioinform. https://doi.org/10.1093/bib/bbaa104.
Wong, L., Huang, Y.A., You, Z.H., Chen, Z.H., Cao, M.Y., 2020. LNRLMI: linear neighbour representation for predicting lncRNA-miRNA interactions. J. Cell. Mol. Med. 24 (1), 79–87. https://doi.org/10.1111/jcmm.14583.
Xie, G.B., Wu, C.M., Sun, Y.P., Fan, Z.L., Liu, J.H., 2019. LPI-IBNRA: long non-coding RNA-Protein interaction prediction based on improved bipartite network recommender algorithm. Front. Genet. 10, 10. https://doi.org/10.3389/fgene.2019.00343.
Xiong, Y., Wang, Q., Yang, J., Zhu, X., Wei, D.Q., 2018. PredT4SE-stack: prediction of bacterial type IV secreted effectors from protein sequences using a stacked ensemble method. Front. Microbiol. 9, 2571. https://doi.org/10.3389/fmicb.2018.02571.
Yang, S., Li, H., He, H., Zhou, Y., Zhang, Z., 2019. Critical assessment and performance improvement of plant-pathogen protein-protein interaction prediction methods. Brief Bioinform 20 (1), 274–287. https://doi.org/10.1093/bib/bbx123.
Yang, Q., Wu, J., Zhao, J., Xu, T., Han, P., Song, X., 2020a. The expression profiles of lncRNAs and their regulatory network during Smek1/2 knockout mouse neural stem cells differentiation. Curr. Bioinform. 15 (1), 77–88. https://doi.org/10.2174/1574893614666190308160507.
Yang, S., Wang, Y., Lin, Y., Shao, D., He, K., Huang, L., 2020b. LncMirNet: predicting LncRNA-miRNA interaction based on deep learning of ribonucleic acid sequences. Molecules 25 (19). https://doi.org/10.3390/molecules25194372.
Yang, X., Yang, S., Li, Q., Wuchty, S., Zhang, Z., 2020c. Prediction of human-virus protein-protein interactions through a sequence embedding-based machine learning method. Comput. Struct. Biotechnol. J. 18, 153–161. https://doi.org/10.1016/j.csbj.2019.12.005.
Yu, B., Chen, C., Yu, Z., Ma, A., Liu, B., Ma, Q., 2020. Prediction of protein-protein interactions based on elastic net and deep forest. bioRxiv. https://doi.org/10.1101/2020.04.23.058644, 2020.2004.2023.058644.
Zeng, X., Zhong, Y., Lin, W., Zou, Q., 2020. Predicting disease-associated circular RNAs using deep forests combined with positive-unlabeled learning methods. Brief. Bioinformatics 21 (4), 1425–1436.
Zhang, H., Cai, K., Wang, J., Wang, X., Cheng, K., Shi, F., et al., 2014. MiR-7, inhibited indirectly by lincRNA HOTAIR, directly inhibits SETDB1 and reverses the EMT of breast cancer stem cells by downregulating the STAT3 pathway. Stem Cells 32 (11), 2858–2868. https://doi.org/10.1002/stem.1795.
Zhang, W., Li, Z., Guo, W., Yang, W., Huang, F., 2019a. A fast linear neighborhood similarity-based network link inference method to predict microRNA-disease associations. IEEEACM Trans. Comput. Biol. Bioinform. https://doi.org/10.1109/TCBB.2019.2931546.
Zhang, W., Tang, G., Zhou, S., Niu, Y., 2019b. LncRNA-miRNA interaction prediction through sequence-derived linear neighborhood propagation method with information combination. BMC Genomics 20 (Suppl 11), 946. https://doi.org/10.1186/s12864-019-6284-y.
Zhang, H., Saravanan, K.M., Yang, Y., Hossain, M.T., Li, J., Ren, X., et al., 2020a. Deep learning based drug screening for novel coronavirus 2019-nCov. Interdiscip. Sci. 12 (3), 368–376. https://doi.org/10.1007/s12539-020-00376-6.
Zhang, P., Meng, J., Luan, Y., Liu, C., 2020b. Plant miRNA-lncRNA interaction prediction with the ensemble of CNN and IndRNN. Interdiscip. Sci. 12 (1), 82–89. https://doi.org/10.1007/s12539-019-00351-w.
Zhang, S., Zhao, L., Zheng, C.H., Xia, J., 2020c. A feature-based approach to predict hot spots in protein-DNA binding interfaces. Brief Bioinform 21 (3), 1038–1046. https://doi.org/10.1093/bib/bbz037.
Zhao, Q., Yu, H., Ming, Z., Hu, H., Ren, G., Liu, H., 2018. The bipartite network projection-recommended algorithm for predicting long non-coding RNA-Protein interactions. Mol. Ther. Nucleic Acids 13, 464–471. https://doi.org/10.1016/j.omtn.2018.09.020.
Zhou, Z.H., 2012. Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC, New York. https://doi.org/10.1201/b12207.
Zhou, Z.H., Feng, J., 2017. Deep forest: towards an alternative to deep neural networks. arXiv preprint arXiv:1702.08835.
Zhou, S., Yue, X., Xu, X., Liu, S., Zhang, W., Niu, Y., 2019. LncRNA-miRNA interaction prediction from the heterogeneous network through graph embedding ensemble learning. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 622–627. https://doi.org/10.1109/BIBM47256.2019.8983044.
Zhu, X., He, J., Zhao, S., Tao, W., Xiong, Y., Bi, S., 2019. A comprehensive comparison and analysis of computational predictors for RNA N6-methyladenosine sites of Saccharomyces cerevisiae. Brief. Funct. Genomics 18 (6), 367–376. https://doi.org/10.1093/bfgp/elz018.
Zhu, Y.H., Hu, J., Ge, F., Li, F., Song, J., Zhang, Y., et al., 2020. Accurate multistage prediction of protein crystallization propensity using deep-cascade forest with sequence-based features. Brief Bioinform. https://doi.org/10.1093/bib/bbaa076.