



# Identifying Cancer-Related lncRNAs Based on a Convolutional Neural Network

Zihao Liu<sup>1,2†</sup>, Ying Zhang<sup>3†</sup>, Xudong Han<sup>4†</sup>, Chenxi Li<sup>2</sup>, Xuhui Yang<sup>1</sup>, Jie Gao<sup>2</sup>, Ganfeng Xie<sup>5\*</sup> and Nan Du<sup>1,2\*</sup>

<sup>1</sup> Department of Oncology, Medical School of Chinese PLA, Chinese PLA General Hospital, Beijing, China, <sup>2</sup> Department of Oncology, The Fourth Medical Center, Chinese PLA General Hospital, Beijing, China, <sup>3</sup> Department of Pharmacy, Heilongjiang Province Land Reclamation Headquarters General Hospital, Harbin, China, <sup>4</sup> College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, <sup>5</sup> Department of Oncology, Southwest Hospital, Army Medical University, Chongqing, China

## OPEN ACCESS

### Edited by:

Lei Deng,  
Central South University, China

### Reviewed by:

Hao Lin,  
University of Electronic Science and  
Technology of China, China  
Juan Wang,  
Inner Mongolia University, China

### \*Correspondence:

Nan Du  
dunan05@aliyun.com  
Ganfeng Xie  
xiegf@aliyun.com

†These authors share first authorship

### Specialty section:

This article was submitted to  
Molecular Medicine,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

Received: 04 June 2020

Accepted: 24 June 2020

Published: 11 August 2020

### Citation:

Liu Z, Zhang Y, Han X, Li C, Yang X,  
Gao J, Xie G and Du N (2020)  
Identifying Cancer-Related lncRNAs  
Based on a Convolutional Neural  
Network. *Front. Cell Dev. Biol.* 8:637.  
doi: 10.3389/fcell.2020.00637

Millions of people are suffering from cancers, but accurate early diagnosis and effective treatment are still tough for all doctors. In recent years, long non-coding RNAs (lncRNAs) have been proven to play an important role in diseases, especially cancers. These lncRNAs execute their functions by regulating gene expression. Therefore, identifying lncRNAs which are related to cancers could help researchers gain a deeper understanding of cancer mechanisms and help them find treatment options. A large number of relationships between lncRNAs and cancers have been verified by biological experiments, which give us a chance to use computational methods to identify cancer-related lncRNAs. In this paper, we applied the convolutional neural network (CNN) to identify cancer-related lncRNAs by lncRNA's target genes and their tissue expression specificity. Since lncRNA regulates target gene expression and it has been reported to have tissue expression specificity, their target genes and expression in different tissues were used as features of lncRNAs. Then, the deep belief network (DBN) was used to unsupervised encode features of lncRNAs. Finally, CNN was used to predict cancer-related lncRNAs based on known relationships between lncRNAs and cancers. For each type of cancer, we built a CNN model to predict its related lncRNAs. We identified more related lncRNAs for 41 kinds of cancers. Ten-cross validation has been used to prove the performance of our method. The results showed that our method is better than several previous methods with area under the curve (AUC) 0.81 and area under the precision-recall curve (AUPR) 0.79. To verify the accuracy of our results, case studies have been done.

**Keywords:** long non-coding RNA (lncRNA), cancer, convolutional neural network (CNN), deep belief network (DBN), machine learning

## INTRODUCTION

Four to nine percent of the sequences' transcription are long non-coding RNAs (lncRNAs) in mammalian genomes (Canzio et al., 2019; Ji et al., 2019). lncRNA was regarded as the noise of genome transcription and did not have biological functions at first. However, an increasing number of studies have reported that lncRNA is widely (Robinson et al., 2019) involved in chromosome

silencing, genomic imprinting, chromatin modification, transcriptional activation, transcriptional interference, and nuclear transport (Cheng et al., 2018a). Recently, it has been proven to be associated with many kinds of cancers.

The secondary structure, spliced form, and subcellular localization of most lncRNAs are conserved (Karner et al., 2020), which is very important for lncRNA to execute functions. However, compared to the functions of microRNAs (miRNAs) and proteins, the function of lncRNA is more difficult to determine. According to the position of lncRNA in the genome relative to protein-coding genes, it can be divided into five types: sense, antisense, bidirectional, intronic, and intergenic.

Many researchers have found lncRNAs play an important role in cancers (Avgeris et al., 2018; Cheng et al., 2018b; Zhao et al., 2020) and neurodegenerative diseases (Peng and Zhao, 2020) as other biological molecules (Zhang T. et al., 2017; Bai et al., 2019; Cheng et al., 2019a; Liang et al., 2019). Although many researchers have verified many associations between lncRNAs and cancers by biological experiments, compared with our knowledge about disease-related genes, we still do not know enough about disease-related lncRNAs. Considering the time and money cost of finding disease-related lncRNAs, more and more researchers tend to use computational methods to identify disease-related lncRNAs. These methods could be divided into three categories: machine learning methods, network methods, and other methods.

Machine learning methods build models based on the similarities of diseases or lncRNAs and their biological characteristics (Cheng, 2019; Cheng et al., 2019b; Zeng et al., 2019; Zou et al., 2019). Lan et al. (2017) developed the lncRNA–disease association prediction (LDAP) which is a method based on bagging support vector machine (SVM) to identify lncRNA–disease associations. They used similarities of lncRNAs and diseases as the features. Yu et al. (2019) developed collaborative filtering naive Bayesian classifier (CFNBC) based on naive Bayesian. They integrated miRNA–lncRNA associations, miRNA–disease associations, and lncRNA–disease associations to infer more lncRNA–disease associations. Considering the discriminative contributions of the similarity, association, and interaction relationships among lncRNAs, disease, and miRNAs, Xuan et al. (2019a) developed a dual convolutional neural network (CNN) with attention mechanisms to predict disease-related lncRNAs.

Network methods are the most common way to identify associations between diseases and lncRNAs nowadays (Gu et al., 2017; Yu et al., 2017; Zhang J. et al., 2017; Kuang et al., 2019; Wang L. et al., 2019; Liu et al., 2020). This kind of method would build one or multiple networks to infer new information. Wang L. et al. (2019) built a lncRNA–miRNA–disease interactive network and used their novel method “LDLMD” to predict associations between lncRNAs and diseases. Sumathipala et al. (2019) used a multilevel network topology which includes lncRNA–protein, protein–protein interaction, protein–disease relationship to use network diffusion algorithm to predict disease-related lncRNAs. The graph convolutional network (GCN) and CNN were used on a lncRNA–miRNA–disease network by Xuan et al. (2019b). Deng et al. (2019) built lncRNA similarity network, disease similarity network, miRNA

similarity network, and their associations. Then, they calculated the meta-path and feature vector for each lncRNA–disease pair in the heterogeneous information network.

Other methods may borrow the feature extraction method or similarity conjecture of network methods, but the core of this method is matrix decomposition or matrix completion. Lu et al. (2019) developed the geometric matrix completion lncRNA–disease association (GMCLDA) which is a method based on geometric matrix completion. They calculated disease similarity based on Disease Ontology (DO) and calculated the Gaussian interaction profile kernel similarity for lncRNAs. Then they inferred disease-related lncRNAs based on the association patterns among functionally similar lncRNAs and similar diseases. Wang Y. et al. (2019) proposed a weighted matrix factorization to capture the inter(intra)-associations between different types of nodes. Then, they approximated the lncRNA–disease association matrix using the optimized matrices and weights to predict disease-related lncRNAs. Locality-constrained linear coding label propagation Latent Dirichlet Allocation (LLCLPLDA) was developed by Xie et al. (2019). Firstly, local-constraint features of lncRNAs and diseases were extracted by locality-constrained linear coding (LLC). Then, they predicted disease-related lncRNAs by label propagation (LP) strategy.

However, previous methods did not consider the regulating target gene expression of lncRNA, which is an important function of lncRNA and plays an important role in associations between lncRNAs and diseases. In addition, deep learning methods are an important tool and have shown their power in bioinformatics (Chen et al., 2019; Lv et al., 2019; Wei et al., 2019; Wu et al., 2019; Zhao et al., 2019a,b,c). Therefore, in this paper, we used this information as features of lncRNA. In addition, the expression of lncRNA in different tissues were also used as the features of lncRNA. Then, the deep belief network (DBN) was used to encode, and the CNN was used to classify.

## METHODS

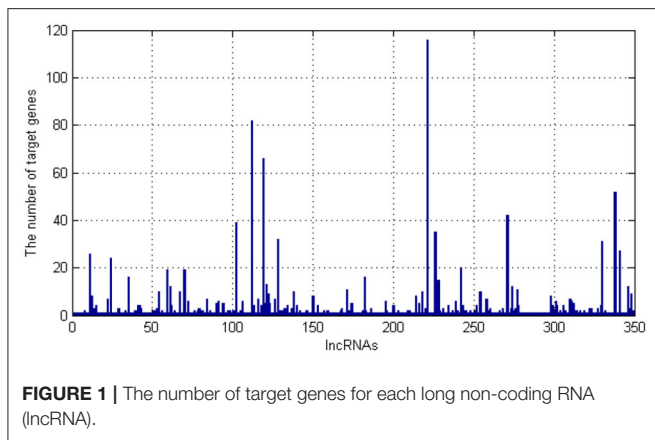
### Feature Extraction

#### Tissue Expression Specificity of Long Non-coding RNA

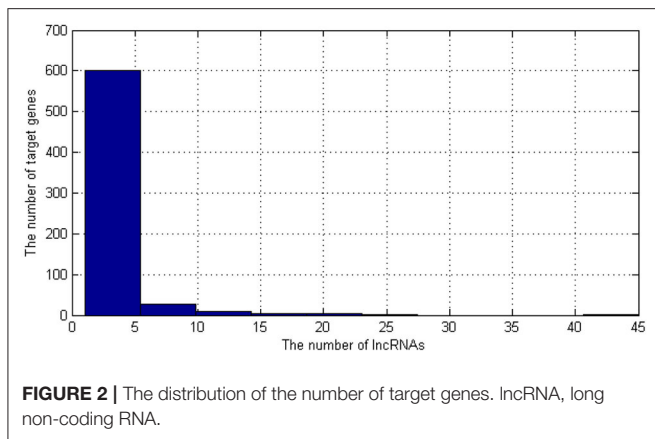
Compared with protein-coding genes, lncRNA shows strong tissue specificity. The specificity of lncRNAs in different kinds of tissues and cell types has been proven by many biological experiments. The different expression also plays an important role in essential cellular processes. Sasaki et al. (2007) tested the expression of lncRNAs in 11 different tissues and found 67% lncRNAs exhibited tissue-specific expression and 29% of lncRNAs were only expressed in one discrete tissue. Therefore, the expression of lncRNAs in different tissues were used as the features.

We obtained the expression of lncRNAs in 13 different tissues which included adipose, adrenal, breast, colon, heart, kidney, liver, lung, lymph node, ovary placenta, prostate, testis, and thyroid.

Therefore, the dimension of each lncRNA's expression feature is  $1 * 13$ .



**FIGURE 1 |** The number of target genes for each long non-coding RNA (lncRNA).



**FIGURE 2 |** The distribution of the number of target genes. lncRNA, long non-coding RNA.

### Target Gene of Long Non-coding RNA

Quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) and Western blot were used to test the different expression genes after knocking down or overexpressing lncRNAs.

We obtained target genes of lncRNA from lncRNA2Target (Jiang et al., 2015).

As we can see in **Figure 1**, there are 349 kinds of lncRNAs. One lncRNA has more than 100 target genes. Then, we draw the distribution of the number of target genes corresponding to lncRNA.

As shown in **Figure 2**, most of the target genes are corresponding to less than five lncRNAs. Therefore, if we used them to be the features of lncRNAs, the features would be sparse. Therefore, we only select the most common target genes to be the features. The genes which are corresponding to more than five lncRNAs were selected as the features of lncRNAs. There are 45 kinds of genes. Then, we need to encode these genes.

$$F = [G_1, G_2, \dots, G_{45}] \tag{1}$$

where  $G_1$  denotes the first gene of these 45 genes, and  $F$  denotes the feature of lncRNA. For each lncRNA, if  $G_1$  is the target gene of it, then  $G_1 = 1$ , otherwise  $G_1 = 0$ .

Therefore, the dimension of each lncRNA's target gene feature is  $1 * 45$ .

### Deep Belief Network

The DBN can effectively learn complex dependencies between variables (Zhao et al., 2019d). The DBN contains many layers of hidden variables, which can effectively learn the internal feature representation of the data and can also be used as an effective non-linear dimensionality reduction method.

When the observable variables are known, the joint posterior probabilities of the hidden variables are no longer independent of each other, so it is difficult to accurately estimate the posterior probabilities of all hidden variables. The posterior probability of early DBN is generally approximated by Monte Carlo method, but its efficiency is relatively low, which makes its parameter learning difficult. In order to effectively train the DBN, we convert the sigmoid belief network of each layer to a restricted Boltzmann machine (RBM). The advantage of this is that the posterior probabilities of the hidden variables are independent of each other, which makes it easy to sample. In this way, the DBN can be regarded as being stacked from top to bottom by multiple RBMs, and the hidden layer of the  $L$ th RBM is used as the observable layer of the  $L + 1$ th RBM. Further, the DBN can be trained quickly by layer-by-layer training, that is, starting from the bottom layer and training only one layer at a time until the last layer. The specific layer-by-layer training process is to train the RBM of each layer in turn from bottom to top. Assuming we have trained the RBM in the first  $L-1$  layer, we can calculate the conditional probability of the bottom-up hidden variables:

$$p(h^{(i)}|h^{(i-1)}) = \sigma(b^{(i)} + W^{(i)}h^{(i-1)}) \tag{2}$$

where  $b^{(i)}$  is the bias of  $i$ th layer of RBM.  $W^{(i)}$  is the connection weight.  $h^{(i)}$  is the  $i$ th layer of RBM.

The process of training DBN is as follows:

```

Input : train dataset  $\hat{V}^{(n)}$ , learning rate  $\lambda$ 
Output: weight matrix  $W^{(l)}$ , bias  $a^{(l)}$  and  $b^{(l)}$ 

For l = 1:L
  Initialization:  $W^{(l)}, a^{(l)}, b^{(l)} = 0$ 
  Sample from train dataset  $\hat{h}^{(l)}$ 
  For i = 1:  $L-l$ 
    Sample  $h^{(l)}$  based on  $p(h^{(l)}|\hat{h}^{(l-1)})$ 
  End
  Set  $h^{(l-1)}$  as the train sample to train  $l$ th layer of RBM
End
    
```

Since the dimension of expression feature and target gene feature are different, we should reduce the dimension of target gene feature and make it the same as the expression feature's. Therefore, in this paper, two layers of RBM were used to build a DBN model.

The number of nodes of the two layers was 32 and 12, respectively. Sigmoid function was used as the

activation function.

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

Therefore, the dimension of final features is  $2 * 13$ .

$$F = \begin{bmatrix} G_1, G_2, \dots, G_{13} \\ E_1, E_2, \dots, E_{13} \end{bmatrix} \quad (4)$$

where  $G_1, G_2, \dots, G_{13}$  denotes target gene feature after DBN, and  $E_1, E_2, \dots, E_{13}$  denotes the expression of lncRNAs in 13 different tissues.

**TABLE 1** | The structure of convolutional neural network (CNN).

Layers	Parameter
Convolutional layer	Filter = 64 kernel size = (1,4) Activation function = tanh
Pooling layer	pool size = (2,2) Activation function = tanh
Convolutional layer	Filter = 128 kernel size = (1,2) Activation function = tanh
Pooling layer	pool size = (1,2) Activation function = tanh
Fully connected layer	Units = 512 Activation function = tanh
Output	Units = 2 Activation function = sigmoid

## Convolutional Neural Network

The power of CNN in dealing with bioinformatic problems has been proven by many researchers. We selected CNN as the classifier based on two reasons. (1) The dimension of features is  $2 * 13$ , which can be regarded as an image. (2) The outstanding performance of CNN in image classification.

There are five layers in our CNN model. The structure of CNN is shown as **Table 1**.

## Work Frame

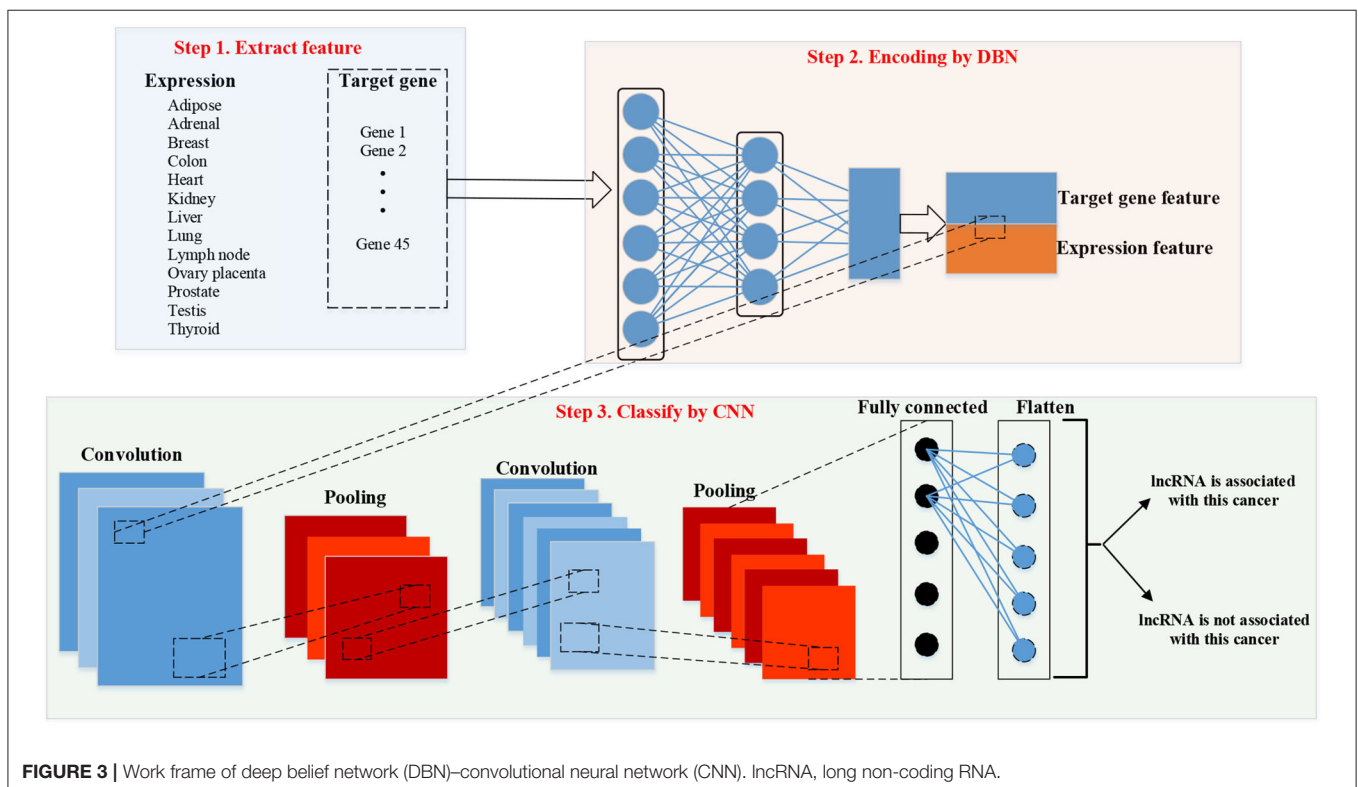
**Figure 3** shows the work frame of our method “DBN–CNN.” There are three steps of our methods. Firstly, we should extract features of lncRNAs. There are two parts of features: expression feature and target gene feature. Then, DBN was used to encode the target gene feature. After encoding, the two kinds of features were combined together. Finally, CNN was used to classify.

## RESULTS

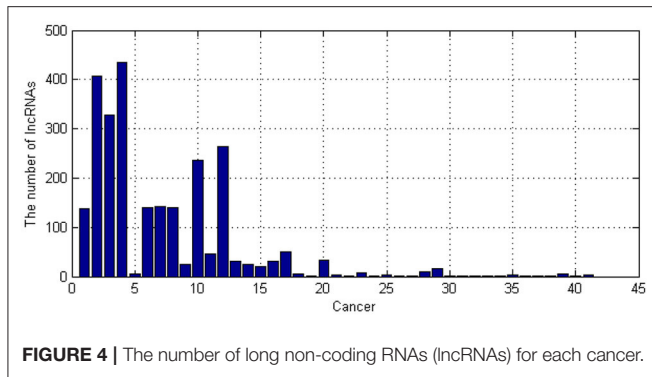
### Data Description

The known associations between lncRNA and diseases were obtained from LncRNADisease database (Bao et al., 2019). We totally obtained 41 kinds of cancer-related lncRNAs. The number of their corresponding lncRNAs is shown as **Figure 4**.

As shown in **Figure 4**, People’s understanding of cancer-related lncRNAs varies widely. We have known more than 100 lncRNAs for some cancers, but few lncRNAs are known for some cancers. To better build our model, we only selected cancers which have more than 20 related lncRNAs. Therefore, 16 kinds of cancers were selected.







**TABLE 2** | The performance of deep belief network (DBN)–convolutional neural network (CNN) in 16 cancers.

Cancer	Area under curve (AUC)	Area under precision curve (AUPR)
Cervical cancer	0.87	0.82
Breast cancer	0.81	0.78
Colorectal cancer	0.81	0.75
Stomach cancer	0.88	0.83
Urinary bladder cancer	0.93	0.85
Lung cancer	0.79	0.82
Ovarian cancer	0.93	0.85
Thyroid cancer	0.94	0.86
Prostate cancer	0.84	0.72
Liver cancer	0.84	0.74
Pancreatic cancer	0.77	0.69
Ovarian epithelial cancer	0.90	0.87
Gallbladder cancer	0.96	0.88
Endometrial cancer	0.76	0.72
Colon cancer	0.88	0.82
Esophageal cancer	0.91	0.87

The target genes of lncRNAs were obtained from lncRNA2Target database. We have discussed about this in section *Target Gene of Long Non-coding RNA*.

The expression of lncRNAs in 13 different tissues was obtained from NON-CODEV5 (Zhao et al., 2016). We only used human data.

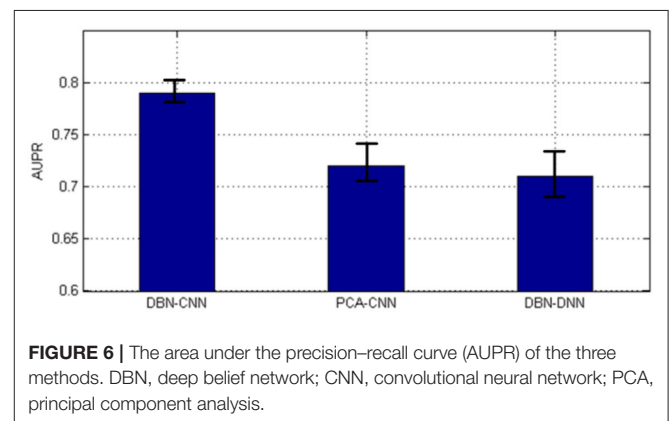
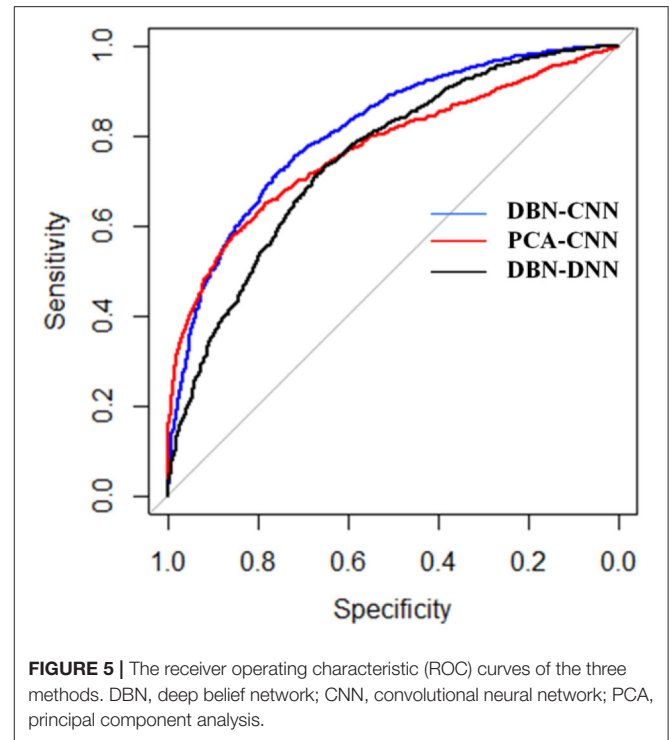
## The Performance of Deep Belief Network–Convolutional Neural Network

We did 10-cross validation on each cancer. Area under the curve (AUC) (Cheng, 2019; Dao et al., 2020; Zhang et al., 2020) and area under the precision–recall curve (AUPR) were used to evaluate the performance of DBN–CNN. The results are shown in **Table 2**.

As we can see in **Table 2**, the performance of DBN–CNN is quite different in different cancers. This may be caused by the different sample sizes. The average AUC is 0.86 and AUPR is 0.80.

## Comparison Experiments

To verify the superior of DBN–CNN, we compared it with similar methods. Since the main function of DBN is to reduce dimension, principal component analysis (PCA) has the same function.



Therefore, instead of using DBN to encode, we used PCA this time and CNN was used to classify the features after PCA. We call this method PCA–CNN. In addition, we also used the deep neural network (DNN) to replace CNN so this comparison method was called DBN–DNN.

We used these three methods to test on 16 cancers and summarized the results to get a final AUC and AUPR for each method. The receiver operating characteristic (ROC) curves are shown in **Figure 4**.

As shown in **Figure 5**, the blue curve denotes the results of DBN–CNN. The red and black curves denote PCA–CNN and DBN–DNN, respectively. As we can see, DBN–CNN performed best among these three methods. The AUC of DBN–CNN is 0.81, which is better than 0.77 and 0.75 for PCA–CNN and DBN–DNN, respectively.

As shown in **Figure 6**, the AUPR of DBN–CNN is the highest with the least standard error.

## Case Study

Liu et al. (2002) found down syndrome cell adhesion molecule - antisense RNA 1 (DSCAM-AS1) is associated with breast cancer by constructing two suppression subtracted cDNA libraries.

Martens-Uzunova et al. (2014) reported the association between H19 and bladder cancer. They also pointed out that H19 could be the biomarker of bladder cancer.

Shi et al. (2014) measured the expression level of lncRNAs-Loc554202 in breast cancer tissues and found that Loc554202 was significantly increased compared with normal control and associated with advanced pathologic stage and tumor size.

## CONCLUSIONS

Increasing evidence has shown the relationship between lncRNAs and cancers. lncRNAs could be the biomarkers to help diagnose cancer and also help researchers understand the mechanism of cancers. Compared with people's knowledge of disease-related protein coding genes, we knew few about disease-related lncRNAs. However, the biological experiments for finding disease-related lncRNAs are time-consuming and expensive.

Therefore, in this paper, we proposed a novel method for identifying cancer-related lncRNAs. We called this method "DBN–CNN," which is a fusion of DBN and CNN. Two kinds of features were used based on the biological background. Since lncRNAs have tissue-specific expression and the expression of cancer tissues is different from normal tissues, the expression of lncRNAs in different tissues could provide important

information for us to identify cancer-related lncRNAs. In addition, lncRNAs execute their regulation function by interacting with their target genes. Therefore, the target genes of lncRNAs can also be the features of lncRNAs. To encode the features, DBN was used to reduce the dimension. Finally, CNN was used to identify real cancer-related lncRNAs based on the final feature.

To verify the effectiveness of our method, we compared DBN–CNN with PCA–CNN and DBN–DNN since PCA can also reduce the dimension of features and DNN can also do classification. The results showed that DBN–CNN performed best. Finally, case studies have been done to verify the accuracy of our results. We found potential lncRNAs for 16 kinds of cancers, which can be a kind of guidance for researchers finding novel cancer-related lncRNAs.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

## AUTHOR CONTRIBUTIONS

ND and GX designed the research. ZL performed the research and wrote the manuscript. YZ and XH acquired the data and reviewed and edited the manuscript. CL, XY, and JG analyzed the data. All authors reviewed the manuscript and provided comments.

## REFERENCES

- Avgeris, M., Tsilimantou, A., Levis, P. K., Tokas, T., Sideris, D. C., Stravodimos, K., et al. (2018). Loss of GAS5 tumour suppressor lncRNA: an independent molecular cancer biomarker for short-term relapse and progression in bladder cancer patients. *Br. J. Cancer* 119, 1477–1486. doi: 10.1038/s41416-018-0320-6
- 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, 621–627. doi: 10.2174/1574893614666190131161002
- Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). lncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* 47, D1034–D1037. doi: 10.1093/nar/gky905
- Canzio, D., Nwabeke, C. L., Horta, A., Rajkumar, S. M., Coffey, E. L., Duffy, E. E., et al. (2019). Antisense lncRNA transcription mediates DNA demethylation to drive stochastic protocadherin  $\alpha$  promoter choice. *Cell* 177, 639–653. doi: 10.1016/j.cell.2019.03.008
- Chen, X., Shi, W., and Deng, L. (2019). Prediction of disease comorbidity using hetsim scores based on multiple heterogeneous networks. *Curr. Gene Ther.* 19, 232–241. doi: 10.2174/1566523219666190917155959
- Cheng, L. (2019). Computational and biological methods for gene therapy. *Curr. Gene Ther.* 19, 210. doi: 10.2174/156652321904191022113307
- Cheng, L., Hu, Y., Sun, J., Zhou, M., and Jiang, Q. (2018a). DincRNA: a comprehensive web-based bioinformatics toolkit for exploring disease associations and ncRNA function. *Bioinformatics* 34, 1953–1956. doi: 10.1093/bioinformatics/bty002
- Cheng, L., Jiang, Y., Ju, H., Sun, J., Peng, J., Zhou, M., et al. (2018b). InfAcrOnt: calculating cross-ontology term similarities using information flow by a random walk. *BMC Genomics* 19(Suppl. 1):919. doi: 10.1186/s12864-017-4338-6
- Cheng, L., Yang, H., Zhao, H., Pei, X., Shi, H., Sun, J., et al. (2019a). MetSigDis: a manually curated resource for the metabolic signatures of diseases. *Brief Bioinform.* 20, 203–209. doi: 10.1093/bib/bbx103
- Cheng, L., Zhao, H., Wang, P., Zhou, W., Luo, M., Li, T., et al. (2019b). Computational Methods for identifying similar diseases. *molecular therapy. Nucleic Acids* 18, 590–604. doi: 10.1016/j.omtn.2019.09.019
- Dao, F. Y., Lv, H., Zulfiqar, H., Yang, H., Su, W., Gao, H., et al. (2020). A computational platform to identify origins of replication sites in eukaryotes. *Brief Bioinform.* doi: 10.1093/bib/bbaa017. [Epub ahead of print].
- Deng, L., Li, W., and Zhang, J. (2019). LDAH2V: Exploring meta-paths across multiple networks for lncRNA-disease association prediction. *IEEE/ACM Transac. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2019.2946257. [Epub ahead of print].
- Gu, C., Liao, B., Li, X., Cai, L., Li, Z., Li, K., et al. (2017). Global network random walk for predicting potential human lncRNA-disease associations. *Sci. Rep.* 7:12442. doi: 10.1038/s41598-017-12763-z
- Ji, J., Tang, J., Xia, K. J., and Jiang, R. (2019). lncRNA in tumorigenesis microenvironment. *Curr. Bioinform.* 14, 640–641. doi: 10.2174/157489361407190917161654
- Jiang, Q., Wang, J., Wu, X., Ma, R., Zhang, T., Jin, S., et al. (2015). lncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res.* 43, D193–D196. doi: 10.1093/nar/gku1173
- Karner, H., Webb, C.-H., Carmona, S., Liu, Y., Lin, B., Erhard, M., et al. (2020). Functional conservation of lncRNA JPX despite sequence and structural divergence. *J. Mol. Biol.* 432, 283–300. doi: 10.1016/j.jmb.2019.09.002

- Kuang, L., Zhao, H., Wang, L., Xuan, Z., and Pei, T. (2019). A novel approach based on point cut set to predict associations of diseases and lncRNAs. *Curr. Bioinform.* 14, 333–343. doi: 10.2174/1574893613666181026122045
- Lan, W., Li, M., Zhao, K., Liu, J., Wu, F.-X., Pan, Y., et al. (2017). LDAP: a web server for lncRNA-disease association prediction. *Bioinformatics* 33, 458–460. doi: 10.1093/bioinformatics/btw639
- Liang, C., Changlu, Q., He, Z., Tongze, F., and Xue, Z. (2019). gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Res.* 48:7603.
- Liu, D., Rudland, P., Sibson, D., and Barraclough, R. (2002). Identification of mRNAs differentially-expressed between benign and malignant breast tumour cells. *Br. J. Cancer* 87, 423–431. doi: 10.1038/sj.bjc.6600456
- Liu, X., Hong, Z., Liu, J., Lin, Y., Alfonso, R.-P., Zou, Q., et al. (2020). Computational methods for identifying the critical nodes in biological networks. *Brief. Bioinform.* 21, 486–497. doi: 10.1093/bib/bbz011
- Lu, C., Yang, M., Li, M., Li, Y., Wu, F., and Wang, J. (2019). Predicting human lncRNA-disease associations based on geometric matrix completion. *IEEE J. Biomed. Health Inform.* doi: 10.1109/JBHI.2019.2958389. [Epub ahead of print].
- Lv, Z. B., Ao, C. Y., and Zou, Q. (2019). Protein function prediction: from traditional classifier to deep learning. *Proteomics* 19:e1900119. doi: 10.1002/pmic.201900119
- Martens-Uzunova, E. S., Böttcher, R., Croce, C. M., Jenster, G., Visakorpi, T., and Calin, G. A. (2014). Long noncoding RNA in prostate, bladder, and kidney cancer. *Eur. Urol.* 65, 1140–1151. doi: 10.1016/j.eururo.2013.12.003
- Peng, J., and Zhao, T. (2020). Reduction in TOM1 expression exacerbates Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* 117, 3915–3916. doi: 10.1073/pnas.1917589117
- Robinson, E. K., Covarrubias, S., and Carpenter, S. (2019). The how and why of lncRNA function: an innate immune perspective. *Biochim. Biophys. Acta Gene Regul. Mech.* 1863:194419. doi: 10.1016/j.bbagr.2019.194419
- Sasaki, Y. T., Sano, M., Ideue, T., Kin, T., Asai, K., and Hirose, T. (2007). Identification and characterization of human non-coding RNAs with tissue-specific expression. *Biochem. Biophys. Res. Commun.* 357, 991–996. doi: 10.1016/j.bbrc.2007.04.034
- Shi, Y., Lu, J., Zhou, J., Tan, X., He, Y., Ding, J., et al. (2014). Long non-coding RNA Lnc554202 regulates proliferation and migration in breast cancer cells. *Biochem. Biophys. Res. Commun.* 446, 448–453. doi: 10.1016/j.bbrc.2014.02.144
- Sumathipala, M., Maiorino, E., Weiss, S. T., and Sharma, A. (2019). Network diffusion approach to predict lncRNA disease associations using multi-type biological networks: LION. *Front. Physiol.* 10:888. doi: 10.3389/fphys.2019.00888
- Wang, L., Xuan, Z., Zhou, S., Kuang, L., and Pei, T. (2019). A novel model for predicting lncRNA-disease associations based on the lncRNA-MiRNA-disease interactive network. *Curr. Bioinform.* 14, 269–278. doi: 10.2174/1574893613666180703105258
- Wang, Y., Yu, G., Wang, J., Fu, G., Guo, M., and Domeniconi, C. (2019). Weighted matrix factorization on multi-relational data for lncRNA-disease association prediction. *Methods* 173, 32–43. doi: 10.1016/j.ymeth.2019.06.015
- Wei, L., Su, R., Wang, B., Li, X., Zou, Q., and Gao, X. (2019). Integration of deep feature representations and handcrafted features to improve the prediction of N 6-methyladenosine sites. *Neurocomputing* 324, 3–9. doi: 10.1016/j.neucom.2018.04.082
- Wu, B., Zhang, H., Lin, L., Wang, H., Gao, Y., Zhao, L., et al. (2019). A similarity searching system for biological phenotype images using deep convolutional encoder-decoder architecture. *Curr. Bioinform.* 14, 628–639. doi: 10.2174/1574893614666190204150109
- Xie, G., Huang, S., Luo, Y., Ma, L., Lin, Z., and Sun, Y. (2019). LLCLPLDA: a novel model for predicting lncRNA-disease associations. *Mol. Genet. Genomics* 294, 1477–1486. doi: 10.1007/s00438-019-01590-8
- Xuan, P., Cao, Y., Zhang, T., Kong, R., and Zhang, Z. (2019a). Dual convolutional neural networks with attention mechanisms based method for predicting disease-related lncRNA genes. *Front. Genet.* 10:416. doi: 10.3389/fgene.2019.00416
- Xuan, P., Pan, S., Zhang, T., Liu, Y., and Sun, H. (2019b). Graph convolutional network and convolutional neural network based method for predicting lncRNA-disease associations. *Cells* 8, 1012. doi: 10.3390/cells8091012
- Yu, G., Fu, G., Lu, C., Ren, Y., and Wang, J. (2017). BRWLDA: bi-random walks for predicting lncRNA-disease associations. *Oncotarget* 8, 60429–60446. doi: 10.18632/oncotarget.19588
- Yu, J., Xuan, Z., Feng, X., Zou, Q., and Wang, L. (2019). A novel collaborative filtering model for lncRNA-disease association prediction based on the Naïve Bayesian classifier. *BMC Bioinform.* 20:396. doi: 10.1186/s12859-019-2985-0
- Zeng, X. X., Wang, W., Deng, G. S., Bing, J. X., and Zou, Q. (2019). Prediction of potential disease-associated microRNAs by using neural networks. *Mol. Ther. Nucleic Acids* 16, 566–575. doi: 10.1016/j.omtn.2019.04.010
- Zhang, J., Zhang, Z., Chen, Z., and Deng, L. (2017). Integrating multiple heterogeneous networks for novel lncRNA-disease association inference. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 16, 396–406. doi: 10.1109/TCBB.2017.2701379
- Zhang, T., Tan, P., Wang, L., Jin, N., Li, Y., Zhang, L., et al. (2017). RNAlocate: a resource for RNA subcellular localizations. *Nucleic Acids Res.* 45, D135–D138. doi: 10.1093/nar/gkw728
- Zhang, Z. M., Tan, J. X., Wang, F., Dao, F. Y., Zhang, Z. Y., and Lin, H. (2020). Early diagnosis of hepatocellular carcinoma using machine learning method. *Front. Bioeng. Biotechnol.* 8:254. doi: 10.3389/fbioe.2020.00254
- Zhao, T., Cheng, L., Zang, T., and Hu, Y. (2019a). Peptide-major histocompatibility complex class I binding prediction based on deep learning with novel feature. *Front. Genet.* 10:1191. doi: 10.3389/fgene.2019.01191
- Zhao, T., Hu, Y., Zang, T., and Cheng, L. (2019b). Identifying Alzheimer's disease-related proteins by LRRGD. *BMC Bioinform.* 20:570. doi: 10.1186/s12859-019-3124-7
- Zhao, T., Hu, Y., Zang, T., and Cheng, L. (2020). MRTFB regulates the expression of NOMO1 in colon. *Proc. Natl. Acad. Sci. U.S.A.* doi: 10.1073/pnas.2000499117
- Zhao, T., Hu, Y., Zang, T., and Wang, Y. (2019c). Integrate GWAS, eQTL, and mQTL Data to Identify Alzheimer's disease-related genes. *Front. Genet.* 10:1021. doi: 10.3389/fgene.2019.01021
- Zhao, T., Wang, D., Hu, Y., Zhang, N., Zang, T., and Wang, Y. (2019d). Identifying Alzheimer's disease-related miRNA based on semi-clustering. *Curr. Gene Ther.* 19, 216–223. doi: 10.2174/1566523219666190924113737
- Zhao, Y., Li, H., Fang, S., Kang, Y., Wu, W., Hao, Y., et al. (2016). NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.* 44, D203–D208. doi: 10.1093/nar/gkv1252
- Zou, Q., Xing, P., Wei, L., and Liu, B. (2019). Gene2vec: gene subsequence embedding for prediction of mammalian N6-methyladenosine sites from mRNA. *RNA* 25, 205–218. doi: 10.1261/rna.069112.118

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu, Zhang, Han, Li, Yang, Gao, Xie and Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.