

Determination of Colorectal Cancer and Lung Cancer Related LncRNAs based on Deep Autoencoder and Deep Neural Network

Ahmet TOPRAK*

Selcuk University, Department of Electricity and Energy, Konya-Turkiye

* Corresponding Author Email: atoprak@selcuk.edu.tr - ORCID: 0000-0003-3337-4917

Article Info:

DOI: 10.22399/ijcesen.636

Received : 15 November 2024

Accepted : 29 December 2024

Keywords :

lncRNA,
disease,
lncRNA-disease association,
Autoencoder,
Deep Learning.

Abstract:

Until recently, non-coding RNAs were considered junk RNA and were always ignored, but studies have revealed that many non-coding RNAs such as miRNA, lncRNA, and circRNAs play important roles in biological processes. A subclass of non-coding RNAs with transcripts longer than 200 nucleotides, called lncRNAs, play important roles in many cellular processes such as gene regulation. For this reason, since wet experimental studies to identify disease-related lncRNA are time-consuming, computational methods are used. Many researchers have applied similarity-based and machine learning-based computational methods and achieved very successful results. Due to its high success rate, the deep learning technique is applied to many fields today. In this study, we used the Deep Autoencoder and Deep Neural Network method to predict disease related lncRNAs. As input data of Deep Autoencoder, the concatenated feature vector obtained from integrated disease similarity and integrated lncRNA similarity was used. To train the deep neural network for predicting relationships between lncRNAs and diseases, the features extracted from the autoencoder's output were utilized. The prediction performance of our method was evaluated with the commonly used 5-fold cross validation and an AUC value of 0.9575 was obtained. It can be seen that the method we proposed is more successful than other compared methods. Additionally, case studies on colorectal cancer and lung cancer were conducted and confirmed with the literature. As a result, the Deep Autoencoder and Deep Neural Network method can be used reliably to identify candidate disease-related lncRNAs.

1. Introduction

Long non-coding RNAs (lncRNAs) are a subtype of non-coding RNA (ncRNA) characterized by a length exceeding 200 nucleotides [1]. Thought to be noise when first discovered, lncRNAs constitute the largest portion of the mammalian non-coding transcriptome [2]. Thanks to the development of bioinformatics technology, researchers have found that lncRNAs are involved in important biological processes such as cell growth, cell development, and gene expression regulation [3]. Studies have revealed that dysregulation of lncRNAs is related with many diseases, such as various cancers, cardiovascular diseases, and neurodegenerative disorders. For example, lncRNA has been discovered to be a tumor suppressor in prostate cancer cells [4]. Moreover, numerous lncRNAs have been identified in Alzheimer's disease, Huntington's

disease, Parkinson's disease, and amyotrophic lateral sclerosis [5].

Accurate identification of potential disease-associated lncRNAs helps diagnose, treatment, and prevent diseases. It even contributes to the development of personalized medicines. However, revealing these relationships with traditional biological methods is a very expensive and long process. Therefore, many computational methods have been developed to identify lncRNAs associated with diseases. These developed methods are generally divided into five categories: network propagation-based methods, matrix factorization-based methods, machine learning-based methods, deep learning-based methods, and graph neural network-based methods.

- Network propagation methods: The network-based approach aimed to reveal possible new associations at the network level by creating networks to represent lncRNA-disease

correlations. Network-based methods often do not require negative samples.

- **Matrix completion and factorization methods:** The lncRNA-disease relationship can also be considered as a recommendation system. With the matrix factorization method, two low-dimensional matrices that can approximate the original input matrix and fill in the missing relationships are obtained. Computational methods based on matrix completion and factorization do not require negative samples to train the model.
- **Machine learning methods:** Today, machine learning methods are widely used to predict disease-related lncRNAs, as in every field. In this method, experimentally confirmed lncRNA-disease associations are labeled as positive, while unknown ones are labeled as negative. Afterwards, classification methods are applied to predict disease-associated lncRNAs.
- **Deep learning methods:** In recent decades, deep learning methods have been widely used for data mining and pattern recognition research. Additionally, deep learning is widely used in areas such as lncRNA-miRNA interaction prediction, lncRNA-protein interaction prediction, miRNA-disease relationship prediction, and drug repositioning.

In this study, to estimate the relationship between lncRNA and disease, Gaussian interaction profile Kernel similarities of lncRNAs and diseases, similarities of lncRNA functional and disease semantic were used to create lncRNA-disease association features. To evaluate the success performance of the method we applied, the commonly used 5-fold cross-validation technique was applied, and the ROC curve was plotted. An AUC value of 0.9575 was obtained in the 5-fold cross-validation technique. Then, we compared our method with eight previous computational methods such as SIMCLDA [6], BRWLDA [7], DMFCDA [8], LLCLPLDA [9], NIMCGCN [10], VGAELDA [11], MDA-SKF [12], and GDCL-NcDA [13]. We also evaluated the success performance of our method by conducting a case study on colorectal cancer and lung cancer. When the results were examined, it was seen that the method we used could successfully predict potential lncRNA-disease relationships.

2. Materials and Methods

2.1. Datasets

The datasets we used in this study are known human lncRNA-disease associations, lncRNA functional

similarity and disease semantic similarity, respectively. Obtaining and calculating this data is explained in detail below.

Human lncRNA–disease associations

Experimentally validated lncRNA-disease associations data were downloaded from the LncRNADisease [14] (v3.0) database (<http://www.rnaut.net/lncrnadisease/>). In this database, there are 13191 relationships between 6066 lncRNAs and 566 diseases. The dataset interaction density is about 0.3842%. Matrix A is created according to the relationship between lncRNAs and diseases. Here, $A(i, j)$ set to 1, if there is a relationship between lncRNA(l_i) and disease(d_j), otherwise set to 0.

$$\begin{cases} A(l_i, d_j) = 1 & \text{if lncRNA } l_i \text{ has relation with disease } d_j \\ A(l_i, d_j) = 0 & \text{if lncRNA } l_i \text{ has no relation with disease } d_j \end{cases} \quad (1)$$

lncRNA functional similarity (LFS)

Functionally similar lncRNAs are generally considered to be associated with similar diseases, according to assumption this manner, we calculated the LFS. For example, if lncRNA l_i and lncRNA l_j are associated with m diseases and n diseases, respectively. Firstly, the similarity value between a disease d and a disease set D is calculated as follows and represented by SV .

$$SV(d, D) = \max_{1 \leq i \leq k} (SV(d, d_i)) \quad (2)$$

where the disease set associated with lncRNA l_i is represented by D . The calculation of LFS scores between the two lncRNAs is carried out as follows:

$$LFS(l_i, l_j) = \frac{\sum_{1 \leq i \leq m} SV(d, D(l_i)) + \sum_{1 \leq j \leq n} SV(d, D(l_j))}{m+n} \quad (3)$$

Construction of disease-semantic similarity (SS)

In addition to containing disease-specific information, the MeSH database [15] also contains directed acyclic graphs (DAGs) created from relationships between diseases. DAG structure is used when calculating the semantic similarities of diseases. As an example, the DAG structure of two diseases is shown in Figure 1. Semantic similarity scores of diseases were computed using the methodology proposed by Toprak [16].

The following equation is utilized to compute the semantic value (V) of disease d .

$$V(d) = \sum_{t \in N_d} D_d(t) \quad (4)$$

In the DAG_d structure, the contribution of disease t to d can be explained by the following equation.

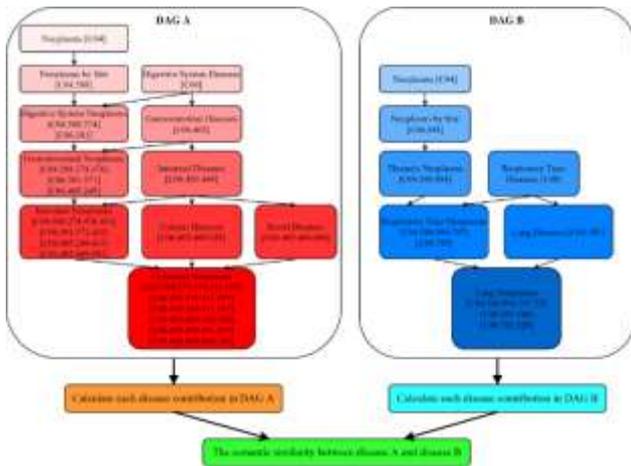
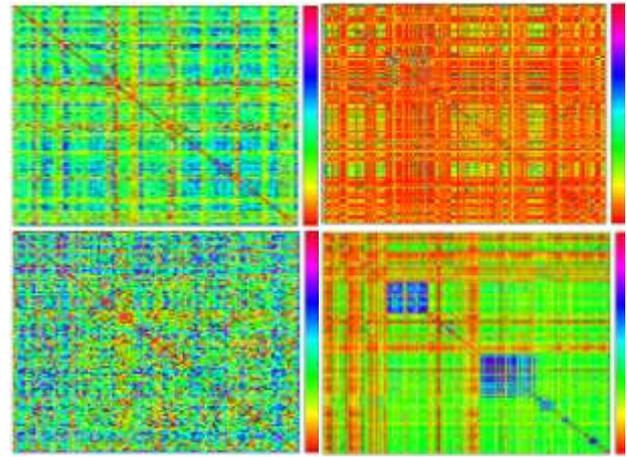


Figure 1. Colorectal and Lung cancers' DAG structures.



a) lncRNA matrices b) Disease matrices
Figure 3. Visualization of lncRNA matrices and disease matrices

$$\begin{cases} D_d(t) = 1, & \text{if } t = d \\ D_d(t) = \max\{\Delta * D_d(t') | t' \in \text{children of } t\}, & \text{if } t \neq d \end{cases} \quad (5)$$

The calculation of the semantic similarity between diseases d_1 and d_2 is as follows.

$$SS(d_1, d_2) = \frac{\sum_{t \in N_{d_1} \cap N_{d_2}} (d_1(t) + d_2(t))}{V(d_1) + V(d_2)} \quad (6)$$

The semantic value of colorectal cancer in the DAG A structure was calculated as 3.406250 and semantic value of lung neoplasms in the DAG B structure was calculated as 2.687500. Afterwards, the semantic similarity between colorectal neoplasms and lung neoplasms was calculated as 0.046153846. Figure 2 shows an example disease semantic network.

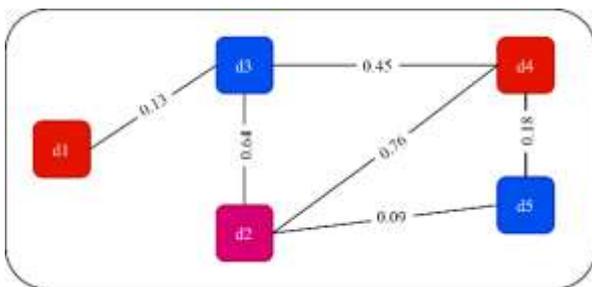


Figure 2. Disease semantic network

A heat map of the lncRNA functional matrix and the fusion of lncRNA matrix are seen in Figure 3(a) and heat map of the disease semantic matrix and the fusion of disease matrix are seen in Figure 3(b). As seen in the figure, both matrices are denser after fusion.

Gaussian (GIP) Kernel methods for lncRNAs and diseases

According to van Laarhoven et al., “lncRNAs with similar functions are likely to be associated with diseases with similar phenotypes (and vice versa)” [17]. Based on this assumption, the GIP kernel

methods for lncRNAs and diseases are computed in the following manner.

$$GIP_lnc(l_i, l_j) = \exp(-\lambda_l \|M(l_i) - M(l_j)\|^2) \quad (7)$$

$$GIP_disease(d_i, d_j) = \exp(-\lambda_d \|M(d_i) - M(d_j)\|^2) \quad (8)$$

Here, parameters λ_l and λ_d , which are the normalized kernel bandwidths, they can be calculated as follows.

$$\lambda_l = \frac{1}{n_l} \sum_{i=1}^{n_l} \|M(l_i)\|^2 \quad (9)$$

$$\lambda_d = \frac{1}{n_d} \sum_{i=1}^{n_d} \|M(d_i)\|^2 \quad (10)$$

where n_l and n_d are number of all lncRNAs all diseases, respectively.

Fusion of multi-source similarity

We combined two similarities, such as lncRNA functional and GIP Kernel, according to previous studies [16, 18-20]. The similarity between two lncRNAs is calculated as follows, and integrated new lncRNA similarity matrix is represented by $LS(l_i, l_j)$.

$$LS(l_i, l_j) = \begin{cases} \frac{LFS(l_i, l_j) + GIP_lnc(l_i, l_j)}{2} & l(i) \text{ and } l(j) \text{ have functional similarity} \\ GIP_lnc(l_i, l_j) & \text{other} \end{cases} \quad (11)$$

As the same way, the similarity between two diseases is calculated as follows and integrated new disease similarity matrix is represented by $DS(d_i, d_j)$.

$$DS(d_i, d_j) = \begin{cases} \frac{SD(d_i, d_j) + GIP_disease(d_i, d_j)}{2} & d(i) \text{ and } d(j) \text{ have semantic similarity} \\ GIP_disease(d_i, d_j) & \text{other} \end{cases} \quad (12)$$

2.2. Method

This method we apply comprises two components: an Autoencoder and a Deep Neural Network. The data in the feature vector, which is the input of the Autoencoder, is the combination of lncRNA and disease vectors. That is, the integrated lncRNAs similarity matrix has dimensions $n_l \times n_l$, and the integrated disease similarity matrix has dimensions $n_d \times n_d$. By combining integrated lncRNAs similarity matrix and integrated disease similarity matrix, we derived a feature vector of size $n_l + n_d$ for each lncRNA-disease pair. The labels for these feature vectors were acquired from the lncRNA-disease relationship matrix.

However, the lncRNA-disease association matrix consists of only 1 (positive) and zeros. Here, we randomly selected the same number of samples from unconfirmed lncRNA-disease associations as the number of positive samples and assigned them as -1 (negative).

The labels for each feature vector were obtained from the known lncRNA-disease relationships. The feature vectors of positive and negative labelled samples were determined as input data of a deep autoencoder.

The resulting features from the autoencoder's output were utilized for training a deep neural network. Afterwards, the association probability of unconfirmed lncRNA-disease pairs was predicted with the trained deep neural network.

Feature selection based on Autoencoder

An autoencoder is an unsupervised neural network comprising two components called the encoder and the decoder, that reproduces its input in its output. Given a vector x at the input of an autoencoder, using a mapping function f , the encoder maps input sample to vector y with the following equation.

$$y = f(Wx + b) \tag{13}$$

Here, f , W , and b represent the encoder's transfer function, weight matrix, and bias vector, respectively. Then, using a similar mapping function, the encoded representation y is decoded into the vector x' by the decoder with following equation.

$$x' = f(W'y + b') \tag{14}$$

Here, W' and b' represent the reconstructed weight matrix and reconstructed bias vector, respectively. Mean squared error were used as the loss function and Adam algorithm were used as the optimizer.

Prediction of lncRNA–disease associations based on Deep Neural Network

In this study, we used a feed-forward deep neural network model consisting of fully connected layers to predict novel relationships between lncRNAs and diseases. The deep neural network model we implemented consists of an input layer, three hidden layers and an output layer. We used the high-level feature vectors obtained from the autoencoder output as the input data of our feed-forward deep neural network model. We used the Adaptive Moment Estimation (Adam) optimizer for optimization, the Rectified Linear Unit (ReLU) activation function in the hidden layers, and the sigmoid function in the output layer.

The flowchart of our proposed method to estimate lncRNA-disease relationships is shown in Figure 4. The integrated lncRNA and disease features were determined as the input of the deep autoencoder. The resulting features obtained at the autoencoder's output were employed to train a deep neural network to predict disease-related lncRNAs.

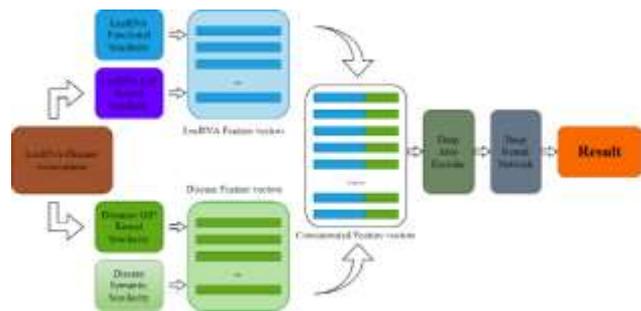


Figure 4. The flowchart of our proposed method

3. Results and Discussion

3.1. Performance Evaluation

In the cross-validation method, the data is divided into k subsets, 80% of the data is used to train the model, while 20% is used to test the model. For example, in 5-fold cross validation, the data is divided into five equal subsets. While the model is trained with four subsets, the model is tested with the remaining one subset and the average is calculated by repeating for each subset. We used the receiver operation characteristic (ROC) [21] curve to evaluate the prediction performance of our applied method. ROC curves are obtained by plotting true-positive rate (TPR, Sensitivity) versus false positive rate (FPR, 1-Specificity).

$$TPR, Sensitivity = \frac{TP}{TP+FN} \tag{15}$$

$$FPR, 1 - Sensitivity = \frac{TN}{TN+FP} \tag{16}$$

where TP, TN, FP, and FN represent correctly classified positive samples, correctly classified negative samples, incorrectly classified positive samples, and incorrectly classified negative samples, respectively.

Moreover, we also calculated the area under ROC curve (AUC) value. While an AUC value of 0.5 indicates a random result, an AUC value close to 1 indicates that the model's predictive performance is successful. In five-fold cross validation, an AUC value of 0.9575 was obtained.

In order to assess the accuracy of our proposed method's predictions, we conducted a comparison with eight other calculation methods such as SIMCLDA, BRWLDA, DMFCDA, LLCLPLDA, NIMCGCN, VGAELDA, MDA-SKF, and GDCL-NcDA. Table 1 shows the sorted AUC values of the other compared methods. When the table is examined, it is seen that the method we used obtains a better AUC value than the other eight compared methods. Figure 5 shows the ROC curve of our proposed method and the other eight compared methods in five-fold cross-validation technique.

Table 1. AUC values

SIMCLDA	0.7455
BRWLDA	0.7661
DMFCDA	0.8044
LLCLPLDA	0.8237
NIMCGCN	0.8612
VGAELDA	0.9126
MDA-SKF	0.9222
GDCL-NcDA	0.9382

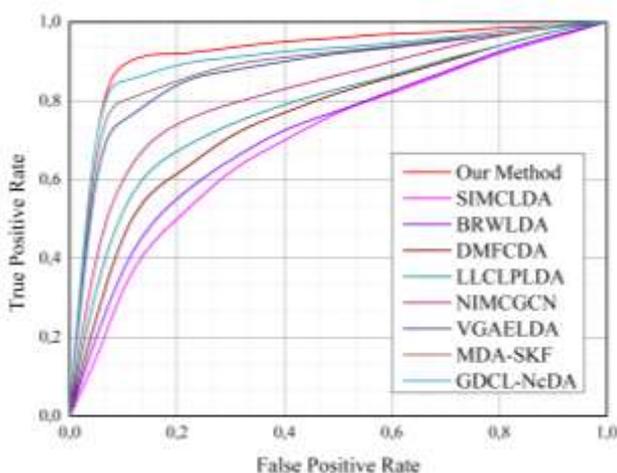


Figure 5. ROC curve

3.1. Case Studies

Colorectal cancer and lung cancer are the most common cancer types worldwide and have a high mortality rate. For this reason, we conducted a case

study on these two cancer diseases to demonstrate the prediction accuracy of our proposed method. The lncRNA-disease association data we obtained from the LncRNADisease database was used as the training set. In our training set, we assigned a value of zero to the known associations between lncRNAs and each disease. This means that we did not consider any prior knowledge of lncRNA-disease associations in our training process. This approach allows us to assess the accuracy of our proposed method in predicting novel associations between lncRNAs and diseases. After training our model and obtaining the top 20 predicted lncRNAs for both diseases, we verified these predictions using information from databases.

Colorectal cancer, which is a major threat to human health, is the second most common type of cancer in females and the third in males [22]. Approximately more than 1 million new cases of colorectal cancer are diagnosed each year. As seen in Table 2, 19 of the top 20 predicted lncRNAs were validated in the literature. For example, it has been observed that PVT1 is overexpressed in colorectal cancer samples, and the expression level of PVT1 is higher in colorectal cancer cells than in normal cells [23]. Yan *et al.* demonstrated that XIST is highly expressed in colorectal cancer tissues and cells [24]. BANCR, also known as LINC00586, has been noted to exhibit overexpression in colorectal cancer tissues and has been linked to patient survival outcomes [25]. Low expression of CAS5 in colorectal cancer patients increased tumor metastasis and reduced survival rate [26]. Upregulation of HULC has been observed in colorectal cancer tissues [27]. NPTN-IT, on the other hand, could not be validated in the literature. Therefore, we think that lncRNA NPTN-IT could be a potential biomarker for colorectal cancer.

Lung cancer stands as one of the most prevalent cancer types and is a primary contributor to cancer-related fatalities globally [28]. There are two primary forms of lung cancer: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC is less common but tends to grow and spread quickly than NSCLC. On the other hand, NSCLC is the more prevalent type, comprising approximately 80% of all cases [29]. As seen in Table 3, 17 of the top 20 predicted lncRNAs were validated in the literature. For example, overexpression of PVT1 has been observed to regulate proliferation in lung cancer patients [30]. Li *et al.* stated that XIST was upregulated in lung cancer tissues [31]. In most lung cancer patients, HULC expression level was observed to be increased in cancer tissues compared to normal tissues [32]. lncRNA HAR1A has been observed to be downregulated in lung adenocarcinoma and negatively correlates with prognosis [33]. BCAR4 expression level is higher

than normal level in NSCLC tissues and cells [34]. In both lung cancer tissues and cancer cell lines, there was observed an increased expression of NEAT1 [35]. Therefore, CCDC26, LINC00032, and PCA3 are likely to be associated with lung cancer. As a result, we think that these lncRNAs may be a potential biomarker for lung cancer. Besides cancer research deep learning has been applied in different applications [36-46].

Table 2. Colorectal Cancer

Rank	lncRNA	Evidence (PMID and Article)	
1	PVT1	PMID: 34418232	DOI: 10.1111/oca.15113
2	XIST	PMID: 35725976	DOI: 10.1007/s10142-022-00673-5
3	BANCR	PMID: 35586041	DOI: 10.3389/ijbar.2022.887822
4	GAS5	PMID: 36062165	DOI: 10.1155/2022/3298939
5	HULC	PMID: 27496341	DOI: 10.1016/j.gen.2016.08.002
6	MALAT1	PMID: 34290314	DOI: 10.3390/jma22116147
7	SPRY4-IT1	PMID: 33029299	DOI: 10.1089/10768354.2020.1784274
8	HCP5	PMID: 32606965	DOI: 10.2147/CMAR.S255866
9	HNF1A-AS1	PMID: 36230970	DOI: 10.3390/cells11193008
10	CDKN2B-AS1	PMID: 34365874	DOI: 10.3892/or.2021.8164
11	CCAT2	PMID: 34668956	DOI: 10.3389/fonc.2021.751903
12	TUG1	PMID: 35508523	DOI: 10.1038/s41419-022-04805-w
13	HOTTIP	PMID: 32587834	DOI: 10.22088/JPACM.BRUMS.8.4.240
14	NEAT1	PMID: 36262350	DOI: 10.1155/2022/8130132
15	UCA1	PMID: 34212174	DOI: 10.1042/BSR20211115
16	NFTN-IT1	Unknown	Unknown
17	MIR318G	PMID: 35733512	DOI: 10.2147/CMAR.S251928
18	SNHG3	PMID: 35923404	DOI: 10.1038/s41420-022-01116-z
19	ZFAS1	PMID: 35036950	DOI: 10.1016/j.annonc.2021.12.010
20	H19	PMID: 35220220	DOI: 10.21873/anticanres.15597

Table 3. Lung Cancer

Rank	lncRNA	Evidence (PMID and Article)	
1	PVT1	PMID: 36376862	DOI: 10.1186/s12935-022-02770-0
2	XIST	PMID: 35067156	DOI: 10.1089/10799893.2021.2019274
3	HULC	PMID: 30575912	DOI: 10.26355/annrev.201812.16637
4	HAR1A	PMID: 35740511	DOI: 10.3390/cancers14122845
5	BCAR4	PMID: 33204025	DOI: 10.1038/s41416-020-01146-3
6	NEAT1	PMID: 32296457	DOI: 10.3389/fonc.2020.00250
7	CCDC26	Unknown	Unknown
8	SPRY4-IT1	PMID: 29484753	DOI: 10.1111/cpe.12446
9	BANCR	PMID: 29651883	DOI: 10.1177/1073274818769840
10	LINC00032	Unknown	Unknown
11	TUG1	PMID: 33073961	DOI: 10.7754/Clin.Lab.2020.200217
12	SNHG16	PMID: 35646120	DOI: 10.1155/2022/2411642
13	TUSC7	PMID: 30542753	DOI: 10.3892/or.2018.6839
14	CRNDE	PMID: 35611803	DOI: 10.1002/ajm2.12558
15	HIF1A-AS1	PMID: 26339353	Unknown
16	TDRG1	PMID: 36444960	DOI: 10.1002/ox.23714
17	KCNQ1OT1	PMID: 33705625	DOI: 10.4143/cr.2020.1208
18	PCA3	Unknown	Unknown
19	GHET1	PMID: 29286919	DOI: 10.5233/CHM-170431
20	PTENP1	PMID: 32698750	Unknown

4. Conclusion

Experimental studies have revealed that lncRNAs have an impact on numerous biological processes. Nevertheless, the identification of these interactions can be both expensive and time-consuming. Therefore, it is imperative to employ computational methods in order to identify lncRNAs associated with diseases. In this study, we utilized Deep Autoencoder and Deep Neural Network techniques to predict potential associations between lncRNAs and diseases. The effectiveness of our method was evaluated using five-fold cross-validation and case

studies. Through the implementation of five different cross-validation techniques, we obtained an AUC value of 0.9575. Furthermore, we compared this AUC value with those obtained from previous studies such as SIMCLDA, BRWLDA, DMFCDA, LLCLPLDA, NIMCGCN, VGAELDA, MDA-SKF, and GDCL-NcDA to assess the performance of our proposed method. Additionally, we conducted case studies specifically focused on colorectal cancer and lung cancer in order to further validate the predictive capabilities of our method. The top 20 predicted lncRNAs for each cancer type were carefully compared with existing literature findings and are presented in table 2 and table 3 accordingly. Upon close examination of these results, it becomes increasingly apparent that our method can be relied upon as a dependable and efficient technique for identifying potential disease-associated lncRNAs.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- [1] P. Kapranov *et al.*, (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription, *Science*, 316(5830);1484-8, doi: <https://doi.org/10.1126/science.1138341>.
- [2] M. Esteller, (2011). Non-coding RNAs in human disease, *Nat Rev Genet*, 12(12);861-874, doi: <https://doi.org/10.1038/nrg3074>.
- [3] J. K. DiStefano, (2018). The Emerging Role of Long Noncoding RNAs in Human Disease, *Methods Mol Biol*, 1706;91-110, doi: https://doi.org/10.1007/978-1-4939-7471-9_6.
- [4] K. Sakurai *et al.*, (2024). Chromogenic in situ hybridization reveals specific expression pattern of long non-coding RNA DRAIC in formalin-fixed paraffin-embedded specimen, *Noncoding RNA Res*,

- 9(1);76-83, doi: <https://doi.org/10.1016/j.ncrna.2023.11.004>.
- [5] W. Xiong, L. Lu, and J. Li, (2024). Long non-coding RNAs with essential roles in neurodegenerative disorders, *Neural Regen Res*, 19(6); 1212-1220, doi: <https://doi.org/10.4103/1673-5374.385850>.
- [6] C. Lu *et al.*, (2018). Prediction of lncRNA-disease associations based on inductive matrix completion, *Bioinformatics*, 34(19);3357-3364, doi: <https://doi.org/10.1093/bioinformatics/bty327>.
- [7] G. Yu, G. Fu, C. Lu, Y. Ren, and J. Wang, (2017). BRWLDA: bi-random walks for predicting lncRNA-disease associations, *Oncotarget*, 8(36);60429-60446, doi: <https://doi.org/10.18632/oncotarget.19588>.
- [8] C. Lu, M. Zeng, F. Zhang, F. X. Wu, M. Li, and J. Wang, (2021). Deep Matrix Factorization Improves Prediction of Human CircRNA-Disease Associations, *IEEE J Biomed Health Inform*, 25(3);891-899, doi: <https://doi.org/10.1109/JBHI.2020.2999638>.
- [9] G. Xie, S. Huang, Y. Luo, L. Ma, Z. Lin, and Y. Sun, (2019). LLCLPLDA: a novel model for predicting lncRNA-disease associations, *Mol Genet Genomics*, 294(6);1477-1486, doi: <https://doi.org/10.1007/s00438-019-01590-8>.
- [10] J. Li, S. Zhang, T. Liu, C. Ning, Z. Zhang, and W. Zhou, (2020). Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction, *Bioinformatics*, 36(8);2538-2546, doi: <https://doi.org/10.1093/bioinformatics/btz965>.
- [11] Z. Shi, H. Zhang, C. Jin, X. Quan, and Y. Yin, (2021). A representation learning model based on variational inference and graph autoencoder for predicting lncRNA-disease associations, *BMC Bioinformatics*, 22(1);136, doi: <https://doi.org/10.1186/s12859-021-04073-z>.
- [12] L. Jiang, Y. Ding, J. Tang, and F. Guo, (2018) MDA-SKF: Similarity Kernel Fusion for Accurately Discovering miRNA-Disease Association, *Front Genet*, 9;618,doi: <https://doi.org/10.3389/fgene.2018.00618>.
- [13] N. Ai, Y. Liang, H. Yuan, D. Ouyang, S. Xie, and X. Liu, (2023). GDCL-NcDA: identifying non-coding RNA-disease associations via contrastive learning between deep graph learning and deep matrix factorization, *BMC Genomics*, 24(1);424, doi: <https://doi.org/10.1186/s12864-023-09501-3>.
- [14] Z. Bao, Z. Yang, Z. Huang, Y. Zhou, Q. Cui, and D. Dong, (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases, *Nucleic Acids Res*, 47; D1034-D1037, doi: <https://doi.org/10.1093/nar/gky905>.
- [15] C. E. Lipscomb, (2000). Medical Subject Headings (MeSH), *Bull Med Libr Assoc*, 88(3);265-6, doi: <https://www.ncbi.nlm.nih.gov/pubmed/10928714>
- [16] A. Toprak, (2023). Identification of disease-related miRNAs based on weighted k-nearest known neighbours and inductive matrix completion, *International Journal of Data Mining and Bioinformatics*, 27(4);231-251, doi: <https://doi.org/10.1504/ijdmb.2023.134297>.
- [17] T. Van Laarhoven, S. B. Nabuurs, and E. Marchiori, (2011). Gaussian interaction profile kernels for predicting drug-target interaction, *Bioinformatics*, 27(21);3036-43, doi: <https://doi.org/10.1093/bioinformatics/btr500>.
- [18] A. Toprak and E. Eryilmaz, (2021). Prediction of miRNA-disease associations based on Weighted k-Nearest known neighbors and network consistency projection, *J Bioinform Comput Biol*, 19(1);2050041, doi: <https://doi.org/10.1142/S0219720020500419>.
- [19] A. Toprak and E. Eryilmaz Dogan, (2021). Prediction of Potential MicroRNA-Disease Association Using Kernelized Bayesian Matrix Factorization, *Interdiscip Sci*, 13(4);595-602, doi: <https://doi.org/10.1007/s12539-021-00469-w>.
- [20] A. Toprak, (2024). circRNA-disease association prediction with an improved unbalanced Bi-Random walk, *Journal of Radiation Research and Applied Sciences*, 17(2);100858, doi: <https://doi.org/10.1016/j.jrras.2024.100858>.
- [21] F. S. Nahm, (2022). Receiver operating characteristic curve: overview and practical use for clinicians, *Korean J Anesthesiol*, 75(1);25-36, doi: 10.4097/kja.21209.
- [22] P. Favoriti, G. Carbone, M. Greco, F. Pirozzi, R. E. Pirozzi, and F. Corcione, (2016). Worldwide burden of colorectal cancer: a review, *Updates Surg*, 68(1);7-11, doi: <https://doi.org/10.1007/s13304-016-0359-y>.
- [23] X. Liu *et al.*, (2022). Long noncoding RNA plasmacytoma variant translocation 1 promotes progression of colorectal cancer by sponging microRNA-152-3p and regulating E2F3/MAPK8 signaling, *Cancer Sci*, 113(1);109-119, doi: <https://doi.org/10.1111/cas.15113>.
- [24] Z. Yan, J. Li, J. Guo, R. He, and J. Xing, (2022). LncRNA XIST sponges microRNA-448 to promote malignant behaviors of colorectal cancer cells via regulating GRHL2, *Funct Integr Genomics*, 22(5);977-988, doi: <https://doi.org/10.1007/s10142-022-00873-5>.
- [25] F. Liu, X. Ma, X. Bian, C. Zhang, X. Liu, and Q. Liu, (2022). LINC00586 Represses ASXL1 Expression Thus Inducing Epithelial-To-Mesenchymal Transition of Colorectal Cancer Cells Through LSD1-Mediated H3K4me2 Demethylation, (in English), *Front Pharmacol*, 13;887822, doi: <https://doi.org/10.3389/fphar.2022.887822>.
- [26] J. Xie, J. J. Wang, Y. J. Li, J. Wu, X. J. Gu, and X. R. Yang, (2022). LncRNA GAS5 Suppresses Colorectal Cancer Progress by Target miR-21/LIFR Axis, *Evid Based Complement Alternat Med*, 2022;3298939, doi: <https://doi.org/10.1155/2022/3298939>.
- [27] X. J. Yang, C. Q. Huan, C. W. Peng, J. X. Hou, and J. Y. Liu, (2016). Long noncoding RNA HULC promotes colorectal carcinoma progression through epigenetically repressing NKD2 expression, (in English), *Gene*, 592(1);172-178, doi: <https://doi.org/10.1016/j.gene.2016.08.002>.

- [28] W. International Agency for Research on Cancer, "GLOBOCAN 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012," ed, 2012.
- [29] K. D. Sutherland and A. Berns, (2010). Cell of origin of lung cancer, *Mol Oncol*, 4(5); 397-403, doi: <https://doi.org/10.1016/j.molonc.2010.05.002>.
- [30] W. Shen *et al.*, (2022). The RNA demethylase ALKBH5 promotes the progression and angiogenesis of lung cancer by regulating the stability of the LncRNA PVT1, *Cancer Cell Int*, 22(1);353, doi: <https://doi.org/10.1186/s12935-022-02770-0>.
- [31] J. Li *et al.*, (2022). XIST/miR-34a-5p/PDL1 axis regulated the development of lung cancer cells and the immune function of CD8(+) T cells, *J Recept Signal Transduct Res*, 42(5);469-478, doi: <https://doi.org/10.1080/10799893.2021.2019274>.
- [32] L. Liu *et al.*, (2018). LncRNA HULC promotes non-small cell lung cancer cell proliferation and inhibits the apoptosis by up-regulating sphingosine kinase 1 (SPHK1) and its downstream PI3K/Akt pathway, *Eur Rev Med Pharmacol Sci*, 22(24);8722-8730, doi: <https://doi.org/10.26355/eurev.201812.16637>.
- [33] J. Ma, K. Cao, X. Ling, P. Zhang, and J. Zhu, (2022). LncRNA HAR1A Suppresses the Development of Non-Small Cell Lung Cancer by Inactivating the STAT3 Pathway, *Cancers (Basel)*, 14(12);2845, doi: <https://doi.org/10.3390/cancers14122845>.
- [34] H. Yang *et al.*, (2019). lncRNA BCAR4 Increases Viability, Invasion, and Migration of Non-Small Cell Lung Cancer Cells by Targeting Glioma-Associated Oncogene 2 (GLI2), *Oncol Res*, 27(3);359-369, doi: <https://doi.org/10.3727/096504018X15220594629967>.
- [35] F. Ma, Y. Y. Lei, M. G. Ding, L. H. Luo, Y. C. Xie, and X. L. Liu, (2020). LncRNA NEAT1 Interacted With DNMT1 to Regulate Malignant Phenotype of Cancer Cell and Cytotoxic T Cell Infiltration via Epigenetic Inhibition of p53, cGAS, and STING in Lung Cancer, *Front Genet*, 11;250, doi: <https://doi.org/10.3389/fgene.2020.00250>.
- [36] J. Jeysudha, K. Deiwakumari, C.A. Arun, R. Pushpavalli, Ponmurugan Panneer Selvam, & S.D. Govardhan. (2024). Hybrid Computational Intelligence Models for Robust Pattern Recognition and Data Analysis . *International Journal of Computational and Experimental Science and Engineering*, 10(4);1032-1040. <https://doi.org/10.22399/ijcesen.624>
- [37] PATHAPATI, S., N. J. NALINI, & Mahesh GADIRAJU. (2024). Comparative Evaluation of EEG signals for Mild Cognitive Impairment using Scalograms and Spectrograms with Deep Learning Models. *International Journal of Computational and Experimental Science and Engineering*, 10(4);859-866. <https://doi.org/10.22399/ijcesen.534>
- [38] M. Devika, & S. Maflin Shaby. (2024). Optimizing Wireless Sensor Networks: A Deep Reinforcement Learning-Assisted Butterfly Optimization Algorithm in MOD-LEACH Routing for Enhanced Energy Efficiency. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1329-1336. <https://doi.org/10.22399/ijcesen.708>
- [39] Rakesh Jha, & Singh, M. K. (2024). Analysing the Impact of Social Influence on Electric Vehicle Adoption: A Deep Learning-Based Simulation Study in Jharkhand, India. *International Journal of Computational and Experimental Science and Engineering*, 10(4);639-644. <https://doi.org/10.22399/ijcesen.371>
- [40] L. Smitha, Maddala Vijayalakshmi, Sunitha Tappari, N. Srinivas, G. Kalpana, & Shaik Abdul Nabi. (2024). Plant Disease Detection Using CNN with The Optimization Called Beluga Whale Optimization Mechanism. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1300-1310. <https://doi.org/10.22399/ijcesen.697>
- [41] S.D.Govardhan, Pushpavalli, R., Tatiraju.V.Rajani Kanth, & Ponmurugan Panneer Selvam. (2024). Advanced Computational Intelligence Techniques for Real-Time Decision-Making in Autonomous Systems. *International Journal of Computational and Experimental Science and Engineering*, 10(4);928-937. <https://doi.org/10.22399/ijcesen.591>
- [42] Machireddy, C., & Chella, S. (2024). Reconfigurable Acceleration of Neural Networks: A Comprehensive Study of FPGA-based Systems. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1007-1014. <https://doi.org/10.22399/ijcesen.559>
- [43] Priti Parag Gaikwad, & Mithra Venkatesan. (2024). KWHO-CNN: A Hybrid Metaheuristic Algorithm Based Optimized Attention-Driven CNN for Automatic Clinical Depression Recognition . *International Journal of Computational and Experimental Science and Engineering*, 10(3);491-506. <https://doi.org/10.22399/ijcesen.359>
- [44] Nagalapuram, J., & S. Samundeeswari. (2024). Genetic-Based Neural Network for Enhanced Soil Texture Analysis: Integrating Soil Sensor Data for Optimized Agricultural Management. *International Journal of Computational and Experimental Science and Engineering*, 10(4);962-970. <https://doi.org/10.22399/ijcesen.572>
- [45] GUNDA, P., & Thirupathi Rao KOMATI. (2024). Integrating Self-Attention Mechanisms For Contextually Relevant Information In Product Management. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1361-1371. <https://doi.org/10.22399/ijcesen.651>
- [46] AY, S. (2024). Vehicle Detection And Vehicle Tracking Applications On Traffic Video Surveillance Systems: A systematic literature review. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1059-1068. <https://doi.org/10.22399/ijcesen.629>