Available online at **www.ijmrhs.com**

ISSN No: 2319-5886

International Journal of Medical Research & Health Sciences, 2019, 8(10): 132-142

Structure Prediction of ncRNA genes: A review on Computational Intelligence Algorithms

Qaisar Abbas*

Al-Imam Mohammad Ibn Saud Islamic, University (IMSIU), Riyadh, Saudi Arabia

**Corresponding e-mail: qaabbas@imamu.edu.sa*

ABSTRACT

Noncoding RNAs (ncRNAs) are an important part of genes and having an important role in human cellular activities and serious diseases. To predict ncRNAs structure, there are many computational intelligence algorithms (CIAs) that are developed in past studies. However, many studies suggested that there were still many structures that are still unpredictable by researchers. In this paper, CIAs algorithms were comprehensively reviewed to predict ncRNAs structures. The advantages and disadvantages of CIA algorithms are briefly mentioned related to ncRNA genes. Moreover, the latest software tools are also compared and reviewed to identify the structure of ncRNAs for mining deep sequencing data. In this study, conventional machine learning algorithms are mainly focused and future trends are also described to predict ncRNAs structure. This paper concludes that there is a need for improving CIA algorithms by using deep learning architectures in terms of layers and computational complexity to predict ncRNAs structures.

Keywords: Genes, Non-coding RNAs, Prediction of structures, Functional annotation, Computational methods, Deep learning

INTRODUCTION

Noncoding RNA (ncRNA) is a typical type of RNA that does not encode a protein. However, they played a vital role in disease cells [1]. There are many types of ncRNAs such as transfer RNA (tRNA), ribosomal RNA (rRNA), microRNA (miRNA), and long ncRNA (lncRNA) [2]. Last decades, there are many non-coding RNAs (ncRNAs) discovered by scientists and still, they are performing extensive experiments on a human body to determine new structure or function of ncRNAs genes in different organisms [3]. The researchers noticed that there are ncRNAs genes involved in the modification of RNA stability, translation and even protein degradation [4]. A wide variety of genomic sequences of ncRNAs genes are publicly available but still, many structures or functions are unknown.

Structure prediction of ncRNAs is nowadays a very hot research topic among many scientists in the domain of bioinformatics. To determine structural information of ncRNA genes, the authors are applying many Computational intelligence algorithms (CIAs). Such as, the authors are trying to classify RNA structure based on real and pseudo-MicRNA and then reconstructing evolution patterns of ncRNA genes [5]. Accordingly, it is very important for any CIAs algorithms to determine accurately structure of ncRNA genes. However recently, this computational intelligence for searching ncRNA gene's structure field has been neglected compared to protein-coding genes.

According to study, it is very much difficult to predict and classify the structure of many RNA genes. So discovering a function of RNAs it is essential to have a good model of its structure. But it is very much challenging task for determining the complete structure ('tertiary structure') of an RNA. In practice, this task required many efforts and it is a time-consuming job [6]. Around seventies, the researchers are trying to develop many computational models for predicting RNA secondary structure. In Figure 1, a broad overview of the type of structures associate programs and resources are visually represented to analyze them in many different aspects.

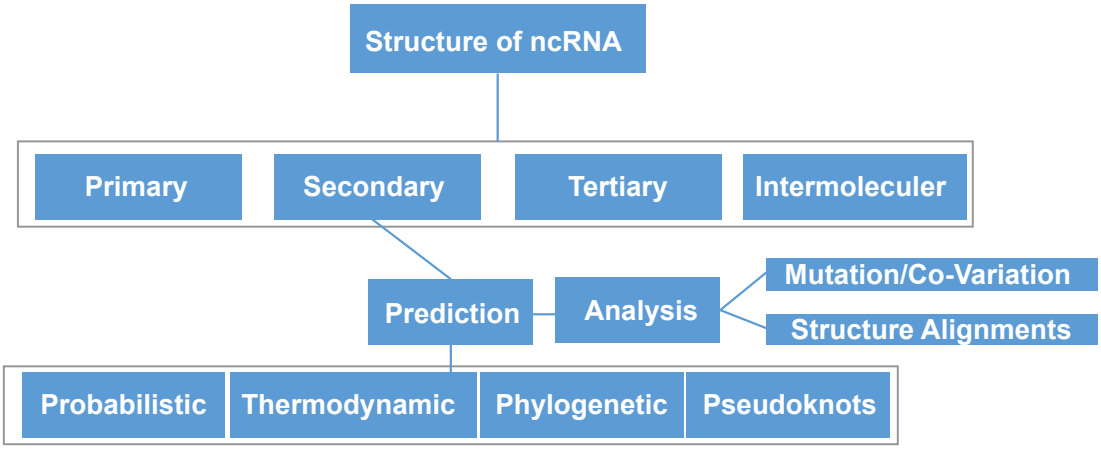

Figure 1 An example of a basic structure of Non-coding RNA (ncRNA)

Good modeling of RNA structure is critical that provides some of the evidence or information towards a clear explanation of its function. Determining the tertiary structure of RNA is a difficult job. Therefore, computational intelligence algorithms (CIAs) and methods are proposed in past studies by many authors to predict three basic structures of ncRNA genes. It noticed that the authors are mainly trying to focus on the likelihood of the secondary structure. To predict structure, they are detecting the pattern of intramolecular base pairs (A-U, G-C, and G-U) [7]. An example of ncRNA genes structures is visually displayed in Figure 2.

The primary structure of RNA is a sequence-specific process that determines some functional properties, like mature miRNA and siRNA molecules base-pair to their targets [8]. Compare to the primary structure, the authors are trying to predict many different secondary and tertiary structures of RNA genes. To update the structures of genes, the authors are trying to diagnosis many diseases such as cancer through long ncRNA-binding proteins [9]. As a result, the primary structure prediction through CIA algorithms is very important to save them from many diseases. Moreover, the scientists discovered that the ncRNAs has many different gene expression patterns that are used for cancer diagnosis [10].

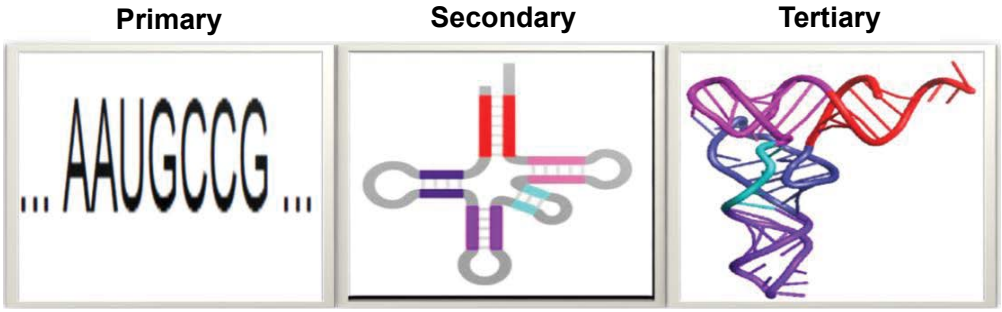

Figure 2 Visual example of the primary, secondary and tertiary structure of non-coding RNA structure

Computational biology domain, the authors are also discovered that the prediction of secondary structures is also an important and time-consuming task. In particular, the secondary structure elements are overall accountable for folding energy that is also available in the tertiary structure of ncRNA genes. As shown in Figure 2, secondary structure poses very important biophysical property and it is easy to predict for CIA algorithms compare to the tertiary structure of ncRNA genes. Therefore, there are many CIA algorithms and methods have been proposed in the literature for predicting secondary structure from the primary sequence [11]. For example, the authors in [11] paper proposed a simple computational intelligence algorithm to detect secondary structure through maximum number of base pairs. In practice, they implemented dynamic programming algorithm to predict RNA secondary structure. To recognize or detect pseudoknot structure is one of the complicated tasks and computationally expensive tasks for CIA algorithms as observed by many scientists. The reason is that the pseudoknot structure is sometimes non-tree like structure. A

pseudoknot is predominant in many ncRNA genes. Still, there are many authors that ignored pseudoknot structure due to complex structure and computational expensive [12].

Compare to pseudoknot structure, the authors in the past studies have developed many computational intelligence algorithms (CIAs) for predicting the secondary and tertiary structure of ncRNA genes. They performed many experiments and high-throughput to determine the previously mentioned ncRNA structures. In human organs, many forms of RNA secondary structures are not having the same biological function [13]. As a result, it is very much difficult to derive the tertiary structure of ncRNA genes in a hierarchic way [14]. This job is very complicated and it is not straight-forward by humans as well as computer experts.

Template-based and de novo prediction algorithms are widely utilized in the past systems to model and predict the tertiary structure of ncRNA genes. In the case of template-based model approach, the authors are not known homology compared to other RNA structures so it can be easy to predict tertiary structure through this approach. On the other hand, the de novo algorithm was totally depended on the 3-D structure by stabilizing non-canonical base pairs and van der Waals interactions [15]. In practice, it is very much important to determine the tertiary structure of ncRNA genes because it poses very important biological processes [16,17]. Compare to primary and secondary structures, the tertiary used different conditions, interactions in proteins [18-21].

The human ncRNA genes are highly limited to functional elements and rules of interactions compared to protein-coding RNA genes [22-25]. In this paper, we are briefly reviewed all computational intelligence algorithms (CIAs) to predict primary, secondary and tertiary structures of ncRNA including long (lncRNA). The long noncoding RNA (lncRNA) is, on the other hand, has been poorly studied in the past papers by authors [26]. Therefore, the ncRNA including lncRNA genes are both studies in this study especially in the field of computational biology. The subsequent section is described in brief all CIA algorithms for predicting structures of ncRNA genes.

Computational Intelligence Algorithms

In gene regulation and other cellular functions, ncRNA genes and RNA structural regulatory motifs play important roles. Their specific secondary structures are often considered critical to their functions and are regularly conserved in phylogenetically or functionally related sequences. It is still a challenge in bioinformatics research to predict common RNA secondary structures in multiple unaligned sequences. One of the most important and challenging problems in computational biology is the computational identification of non-coding RNA (ncRNA) genes. Existing methods mostly dependent on homology information that will limit prediction and identification to ncRNA genes with known homologs.

An overview of the latest structure prediction computational tools is displayed in Table 1. In this table, there are important methods mentioned with features and machine learning algorithms that have taken from ref [27-46]. In addition to this, the online tools for structure prediction of ncRNAs genes are shown in Table 2. This information is presented to help the computer scientist and biologist to quickly understand the state-of-the-art methods for prediction of structures. Those systems mentioned in Table 1 are explained in the subsequent paragraphs.

Table 1 State-of-the-art computational tools for structure prediction

Reference	Tools	Features	Computational Methods
[27]	RNAz	Predicting structural ncRNAs	[1] MFE RNA folding+[2] SVM
[28]	Foldalign	RNAs Structure and Sequence Alignment	Foldalign+Sankoff
[31]	StrAl	Structural alignment of ncRNAs	Base pairing probabilities
[32]	MASTR	Multiple alignment for strctural RNAs	[3] McMC+[4] SA
[33]	foldalignM	Multiple structures and sequence alignment	foldalign+pairwise scans
[34]	RNA-sampler	Prediction of secondary structure	Base pairing+iterative method
[37]	Denovo	Structure prediction for ncRNAs	Features+[5] NN
[39]	miRanalyzer	Predict the structure of microRNA	[1] MFE+Vienna RNA
[46]	lncRNA_MDFL	lncRNA structure predictor	Multiple features+deep learning

[1] Minimum Free Energy (MFE) RNA; [2] Support vector machine; [3] Markov chain Monte Carlo (McMC); [4] Simulated annealings; [5] Neural network

Table 2 Online tools for structure prediction of Non-coding RNAs

Reference	Name	[1]URL
[27]	RNAz	www.tbi.univie.ac.at/~wash/RNAz
[28]	Foldalign	http://rth.dk/resources/foldalign/
[31]	StrAl	http://www.biophys.uni-duesseldorf.de/stral/download.php
[32]	MASTR	http://mastr.binf.ku.dk/
[33]	foldalignM	http://rth.dk/resources/foldalign/
[34]	RNA-sampler	http://ural.wustl.edu/software.html
[37]	Denovo	http://csbl.bmb.uga.edu/publications/materials/tran/
[39]	miRanalyzer	http://web.bioinformatics.cicbiogune.es/microRNA/
[46]	lncRNA_MDFL	http://compgenomics.utsa.edu/lncRNA_MDFL/
[1] Universal resource locator		

Prediction of secondary structure is performed by a scanning method known as Zuker minimum-energy folding algorithm for detecting of non-coding RNAs (ncRNAs) sequence [27]. In that study, the authors argue that this model is more stable to predict secondary structure that is undoubtedly important in most noncoding RNAs. In [4], the authors combined comparative sequence analysis and structure prediction to determine the appearance of large-scale genomic screens. In that research study, they presented a system based on a measure of RNA secondary structure. The authors claim that this approach is accurate and faster compared to other methods in the past studies. A comparison study was conducted to determine the structural presentation of noncoding RNAs (ncRNAs) [6]. The authors also noticed that 40% of the predicted structure RNAs overlap with the detected sites.

In another study [28], the authors showed that the searching of ncRNAs genes and structural RNA elements are difficult tasks for experts because they often conserved in structure rather than in a sequence. In that study, the pairwise local alignment and the Sankoff algorithm are utilized to determine the structural alignment of multiple sequences. However, the authors introduced a new method based on Genetic programming to predict thousands of ncRNA genes [29]. The authors tested this method on sixteen predictions and show that twelve of these are actual ncRNA transcripts.

The authors suggested that advance computational methods are required to accurately detect ncRNAs because there is an increasing number of genomic sequences [30]. Therefore, they proposed a program known as Dynalign by using support vector machine (SVM) for predicting secondary structures. The authors also compared Dynalign with other two methods such as RNAz and QRNA and obtained significant results. Also in [31], the authors argue that the structure prediction of ncRNAs is very difficult as these sequences may evolve by compensatory mutations. Despite these facts, the proper alignment of multiple structural RNA sequences remains a problem. In that study, the authors presented a system known as StrAl through a heuristic method for the alignment of ncRNA that reduces sequence-structure alignment. The authors also compared StrAl algorithm with ClustalW and reported significantly higher prediction accuracy.

To consider multiple alignment and structure prediction for ncRNAs genes, the authors developed a model by using a Markov chain Monte Carlo (McMC) algorithm in a simulated annealing framework [32]. In this framework, the authors iteratively improved the problem of multiple alignments of structural RNAs by minimizing a cost function. They showed that the proposed system is very effective and efficient compared to other state-of-the-art systems. A different study was conducted about computational RNA structure prediction [33]. The authors presented that multiple alignments are required for the large sequence to predict structure by advanced computational techniques. However, if there are few sequences then it is difficult to predict the structure of ncRNAs genes. Therefore in that study, the authors developed a system known as PMcomp that can cluster sequences based on sequence and structure similarity. The authors developed PMcomp program, in Java and PERL scripts and achieved higher prediction accuracy compared to other past methods.

The authors presented a solution to predict common RNA secondary structures based on base-pairing probabilities and iterative techniques for multiple unaligned sequences [34]. This algorithm can be also suitable for multiple sequence alignment problems as noticed by the authors due to introduce of iteratively convergence technique. The

structural alignments were more efficient as compared to other programs in sequences of a wide range of identities, and a more accurate representation of RNA secondary structure conservations was achieved. Whereas, the authors presented a new version of the RNAalifold method for the prediction of a consensus structure through improving the evaluation of sequence covariance, structure centroid, and stochastic backtracking approaches [35]. In the past, the authors are trying to compute suboptimal structures based on Boltzmann weights to determine the statistical features of the ncRNA structure. The author concluded that the new version of the RNAalifold method is more accurate compared to other computational methods while maintaining the same complexity time.

The latest computational technique such as deep learning algorithms is developed to predict the secondary structure of miRNA from the large pool of sequenced transcripts from a single deep sequencing run remains a major challenge [36]. The authors named this model as miRDeep that is applied to the latest dataset of RNA sequence and achieved a higher prediction rate compared to other systems. A de novo prediction algorithm is developed for prediction of ncRNA genes using features derived from the structures and sequences [37]. The authors extracted the set of features from known ncRNA genes and then performed training based on these features by a neural network-based classifier. The authors achieved an average prediction sensitivity of 68% while specificity was 70% for identifying potential ncRNA genes. There is another improvement done for predicting secondary structures of RNAs based on decoding the posterior probabilities of the base-pairing algorithm to overcome the problem of minimum free energy methods known as improved CentroidFold algorithm [38].

Whereas, a new machine learning algorithms were developed to predict the new microRNAs structure and achieved an area under the curve values of 97.9% and recall values of up to 75% on unseen data based on random forest classifier known as miRanalyzer [39]. The authors utilized minimum free energy (MFE) and the Vienna RNA package for predicting the secondary structure. To understand RNA sequence structure, the authors developed a program based on existing design of RNA structure [40]. The authors observed that when designing a small interfering RNAs through the set of nearest neighbor parameters consideration RNA structure is very important. It includes methods for secondary turner group. Moreover, the authors also suggested that structure prediction is often important for analyzing the deep sequence of ncRNAs [41]. As a result, a genome-wide algorithm was developed to analyze the structure of MicroRNAs and compared with other systems too [42]. The authors obtained high sensitivity and specificity using some basic properties to recognize the structure of microRNAs.

The prediction of the structure of ncRNAs is also important to diagnose the cancer metastasis [43]. As in this study, the authors reported that the prediction of the long ncRNAs (lncRNAs) structure is also important. Therefore, the authors developed a computational model known as IPKnot to predict the secondary structures based on maximizing excepted accuracy [44]. Also, the authors detected the structure of lncRNAs, related to the development, epigenetics, cancer, brain function and hereditary disease [45]. It was also noticed that the long noncoding RNAs (lncRNAs) are emerging class of ncRNA genes plays an important role in cellular functions [46]. In practice, the lncRNAs are carefully connected with the development of some diseases. As a result, the authors developed a lncRNA-MFDL system to predict and identify lncRNAs through fusion of variant features and multi-layer deep learning algorithms. By using the 10-fold cross-validation test, the authors reported 97.1% accuracy, which is very high as compared to other systems.

The advantage and limitations of some of the above-mentioned computational tools are mentioned in Table 3. This table is clearly represented that the previously developed methods for structure predictions are suffered from computational complexity and focused only on certain structures for the finding of genes. As described in this table, the RNAz [4] computational tool is better for prediction of secondary structures but having high computational cost due to use to support vector machine (SVM) algorithm. It noticed that the SVM algorithm needs a lot of training data set by making it difficult for the prediction of complex structures. Instead of using RNAz, there is also another MASTR [32] developed to predict multiple sequence structures. However, the MASTR system is developed on the particular sequence structures. Moreover, the RNA-Sampler [34], Denovo [37] and lncRNA_MDFL [46] tools are required higher computational time and therefore these are not suitable for determining the structures of next-generation-sequences (NGS).

Table 3 Various comparisons of State-of-the-art computational tools for structure prediction

Reference Methods	Advantage	Limitations
[27] RNAz	Better accuracy of prediction	High computational cost due to [1]SVM so not suitable for [2]NGS
[32] MASTR	It can predict multiple sequence structures	Does not consider sequence-depend structures
[34] RNA-sampler	Suitable for predicting secondary structures	Computational complexity
[37] Denovo	Not suitable for ncRNAs structures	High complexity time
[46] lncRNA_MDFL	Suitable for long ncRNAs structure	Better computational complexity
[1] Support vector machine; [2] Next generation sequence		

DISCUSSION

Structure prediction of non-coding RNAs (ncRNAs) is still a challenging task for biological experts and computational intelligence person. Although, there are many computational intelligence algorithms (CIAs) [27-46] have been developed in the past to predict the primary, secondary and tertiary structure of non-coding RNA structure. Researchers are trying to develop computational tools for prediction the previously mentioned structure for ncRNA including long ncRNA (lncRNA) genes. Since the ncRNA genes are transcripts that function directly as RNA molecule structure. These ncRNA genes are not being translated into protein. As a result, the prediction of ncRNA genes structure is stilling a challenging and time-consuming task because there are still hidden structures in recent genomes. In the bioinformatics domain, the de novo prediction ncRNA genes is still a difficult task due to lack of statistical significant properties in primary genome sequences.

According to the literature review, the secondary structure prediction through CIA algorithms has obtained significantly better results compared to other structures. Those CIA algorithms included traditional or modern multi-layer deep learning architecture techniques. Table 1 described those CIA algorithms that are utilized in the past to predict structures of ncRNA genes. Some authors have also developed online tools for other researchers to assist them in automatically identify the structures of ncRNA genes. Those online tools or resources are mentioned in Table 2.

There are many systems developed in the past to predict primary or secondary structures. Those computational intelligent systems are described in Table 3. However, most of them are not accurate and they recognize the structure in a large amount of time. Therefore, those algorithms are not suitable for screening large genome sequence. Recently, authors were trying to propose computational intelligence algorithms (CIAs) to analyze the structural [28-50] information and functional annotation [51-73] of ncRNA genes. Due to limit the scope, the structure prediction algorithms are only described in this paper. If someone author is interested to study computational intelligence methods for functional annotation then he or she can study this article [74].

According to the literature review, it concludes that there is not a single study to study computational intelligence algorithms (CIAs) for predicting all three structures of ncRNA genes including both traditional and state-of-the-art deep learning architectures. For predicting structures of ncRNA including lncRNA genes, some authors are trying to use the convolutional neural network (CNN) model and softmax classifiers compared to traditional algorithms such as support vector machine (SVM), neural network (NN) or decision-tree. After doing experiments, the authors conclude that the advanced CNN models are very powerful to identify the structure of ncRNA genes because these models can be used to train large genome sequences and predicted response time is very fast [67]. As a result, these advanced models are very fast once they trained on a large sequence. Unluckily, at present, there are few CIA algorithms proposed through advance deep learning algorithms in the past systems to predict structure.

To highlight these issues, let us consider a study of Washietl, et al., [4], the authors developed RNAz system to predict the structure of ncRNAs but it requires high computational cost due to support vector machine (SVM) classifiers. Therefore, it is unsuitable for the next-generation sequence (NGS) or where there is a requirement for big data. Also, the authors developed an RNA-sample system to predict secondary structure but this system required high computational complexity [34]. This information is briefly described in Table 3 about structure prediction. From this table, it noticed that the high-throughput sequences are not possible to extract from the large-sequence ncRNA genes. Therefore, these approaches must require deep-learning algorithms to effectively detect large sequence structures without focusing on domain expert knowledge about machine learning or data mining algorithms.

A similar trend can be seen in the case of functional annotation of ncRNAs or lncRNAs genes. Many authors utilized

old fashion machine learning algorithms as mentioned above. In past studies, the authors developed many different computational approaches to annotate the function of thousands of ncRNAs or lncRNAs genes. In fact, the functional annotation of lncRNAs genes is very important and vital step to detect human diseases because it has diverse biological processes. However, the authors developed many computational techniques to provide valuable important insides to functionally annotate the ncRNAs or lncRNAs genes. In short, the coming advance in the study of lncRNAs, especially at a large genome-wide scale, poses an exciting chance to examine the lncRNA function in the future.

Current Computational Challenges

Computational intelligent techniques had many limitations that developed in the past systems for functional annotation of noncoding RNAs (ncRNAs). It was mentioned before that the ncRNAs genes are a very much important function for the operation of the cell. Therefore, the authors focused more on developing methodological techniques for understanding the different functions of ncRNAs. The authors applied many computational intelligent models to determine functional divergence and resemblance ncRNAs genes in the cell along with limited data sources.

There are many varieties of ncRNAs such as micro RNAs or long noncoding RNAs. Therefore to get accurate accuracy for determining the functions of ncRNAs is still a challenging task for computational intelligent methods. Compare to functional annotation of ncRNAs, the authors have also developed structure prediction algorithms that obtained higher accuracy. Currently, many authors proposed computational methods to analyze the structural [28-46] and functional annotation [53-72] of ncRNAs.

CONCLUSION

Advances in bioinformatics have made enormous progress toward the development of computational techniques for structure and functions of ncRNA or lncRNA. In this review article, the advanced computational methods are focused to annotate the function of ncRNAs. This paper described the computational methods to predict the structure or to functionally annotate the ncRNAs genes. To increase the accuracy of functional annotation, there is a dire need for domain expert knowledge. In fact, the state-of-the-art computational techniques that have been used in the past required for pre-processing of raw input data, feature selection and to fine-tune parameters for getting high accuracy. However, the authors put many efforts to do pre- or post-processing on data instead of developing effective computational intelligent techniques. It was noticed that the deep learning algorithms are classifying the data without doing pre- or post-processing steps. These algorithms are advanced in the domain of CIA techniques. The methods were developed using deep learning algorithms get higher accuracy for data classification. In practice, the deep learning algorithms have outperformed compared to conventional machine learning algorithms for functional annotation of ncRNAs but still lack of expert knowledge is required to define multilayer architecture. As reported in the past studies, the techniques developed using deep learning algorithms must be further investigated in terms of performance.

DECLARATIONS

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] Zhang, Tao Jiang Michael Q. *Current topics in computational molecular biology.* MIT Press, 2002.

[2] Jiang, Peng, et al. "MiPred: Classification of real and pseudo microRNA precursors using random forest prediction model with combined features." *Nucleic Acids Research,* Vol. 35, No. 2, 2007, pp. W339-W44.

[3] Lowe, Todd M., and Sean R. Eddy. "tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence." *Nucleic Acids Research,* Vol. 25, No. 5, 1997, pp. 955-64.

[4] Washietl, Stefan, Ivo L. Hofacker, and Peter F. Stadler. "Fast and reliable prediction of noncoding RNAs." *Proceedings of the National Academy of Sciences,* Vol. 102, No. 7, 2005, pp. 2454-59.

[5] Gardner, Paul P., Andreas Wilm, and Stefan Washietl. "A benchmark of multiple sequence alignment programs upon structural RNAs." *Nucleic Acids Research,* Vol. 33, No. 8, 2005, pp. 2433-39.

[6] Washietl, Stefan, et al. "Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome." *Nature Biotechnology,* Vol. 23, No. 11, 2005, p. 1383.

[7] Tinoco, Ignacio, Olke C. Uhlenbeck, and Mark D. Levine. "Estimation of secondary structure in ribonucleic acids." *Nature,* Vol. 230, No. 5293, 1971, pp. 362-67.

[8] Carthew, R. W., and E. J. Sontheimer. "Origins and mechanisms of miRNAs and siRNAs." *Cell,* Vol. 136, pp. 642-55.

[9] Wapinski, Orly, and Howard Y. Chang. "Corrigendum: Long noncoding RNAs and human disease." *Trends in Cell Biology,* Vol. 21, 2011, pp. 354-61.

[10] Tsai, Miao-Chih, Robert C. Spitale, and Howard Y. Chang. "Long intergenic noncoding RNAs: new links in cancer progression." *Cancer Research,* Vol. 71, No. 1, 2011, pp. 3-7.

[11] Nussinov, Ruth, and Ann B. Jacobson. "Fast algorithm for predicting the secondary structure of single-stranded RNA." *Proceedings of the National Academy of Sciences,* Vol. 77, No. 11, 1980, pp. 6309-13.

[12] Washietl, Stefan, et al. "Computational analysis of noncoding RNAs." *Wiley Interdisciplinary Reviews: RNA,* Vol. 3, No. 6, 2012, pp. 759-78.

[13] Novikova, Irina V., Scott P. Hennelly, and Karissa Y. Sanbonmatsu. "Tackling structures of long noncoding RNAs." *International Journal of Molecular Sciences,* Vol. 14, No. 12, 2013, pp. 23672-84.

[14] Tsai, Miao-Chih, Robert C. Spitale, and Howard Y. Chang. "Long intergenic noncoding RNAs: New links in cancer progression." *Cancer Research,* Vol. 71, No. 1, 2011, pp. 3-7.

[15] Gautheret, Daniel, Francois Major, and Robert Cedergren. "Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA." *Bioinformatics,* Vol. 6, No. 4, 1990, pp. 325-31.

[16] Pedersen, Jakob Skou, et al. "Identification and classification of conserved RNA secondary structures in the human genome." *PLoS Computational Biology,* Vol. 2, No. 4, 2006, p. e33.

[17] Dann III, Charles E., et al. "Structure and mechanism of a metal-sensing regulatory RNA." *Cell,* Vol. 130, No. 5, 2007, pp. 878-92.

[18] Wan, Yue, et al. "Understanding the transcriptome through RNA structure." *Nature Reviews Genetics,* Vol. 12, No. 9, 2011, pp. 641-55.

[19] Brion, Philippe, and Eric Westhof. "Hierarchy and dynamics of RNA folding." *Annual Review Of Biophysics and Biomolecular Structure,* Vol. 26, No. 1, 1997, pp. 113-37.

[20] Yan, Kun, et al. "Structure prediction: New insights into decrypting long noncoding RNAs." *International Journal of Molecular Sciences,* Vol. 17, No. 1, 2016, p. 132.

[21] Zhao, Yi, et al. "Large-scale study of long non-coding RNA functions based on structure and expression features." *Science China Life Sciences,* Vol. 56, No. 10, 2013, pp. 953-59.

[22] Guo, Xingli, et al. "Advances in long noncoding RNAs: Identification, structure prediction and function annotation." *Briefings in Functional Genomics,* Vol. 15, No. 1, 2015, pp. 38-46.

[23] Harrow, Jennifer, et al. "GENCODE: The reference human genome annotation for The ENCODE Project." *Genome Research,* Vol. 22, No. 9, 2012, pp. 1760-74.

[24] Bhartiya, Deeksha, and Vinod Scaria. "Genomic variations in non-coding RNAs: Structure, function and regulation." *Genomics,* Vol. 107, No. 2-3, 2016, pp. 59-68.

[25] Derrien, Thomas, et al. "The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression." *Genome Research,* Vol. 22, No. 9, 2012, pp. 1775-89.

[26] Mercer, Tim R., Marcel E. Dinger, and John S. Mattick. "Long non-coding RNAs: Insights into functions." *Nature Reviews Genetics,* Vol. 10, No. 3, 2009, pp. 155-59.

[27] Rivas, Elena, and Sean R. Eddy. "Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs." *Bioinformatics,* Vol. 16, No. 7, 2000, pp. 583-605.

[28] Havgaard, Jakob Hull, et al. "Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%." *Bioinformatics,* Vol. 21, No. 9, 2005, pp. 1815-24.

[29] Sætrom, Pål, et al. "Predicting non-coding RNA genes in *Escherichia coli* with boosted genetic programming." *Nucleic Acids Research,* Vol. 33, No. 10, 2005, pp. 3263-70.

[30] Uzilov, Andrew V., Joshua M. Keegan, and David H. Mathews. "Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change." *BMC Bioinformatics,* Vol. 7, No. 1, 2006, p. 173.

[31] Dalli, Deniz, et al. "STRAL: Progressive alignment of non-coding RNA using base pairing probability vectors in quadratic time." *Bioinformatics,* Vol. 22, No. 13, 2006, pp. 1593-99.

[32] Lindgreen, Stinus, Paul P. Gardner, and Anders Krogh. "MASTR: Multiple alignment and structure prediction of non-coding RNAs using simulated annealing." *Bioinformatics,* Vol. 23, No. 24, 2007, pp. 3304-11.

[33] Torarinsson, Elfar, Jakob H. Havgaard, and Jan Gorodkin. "Multiple structural alignment and clustering of RNA sequences." *Bioinformatics,* Vol. 23, No. 8, 2007, pp. 926-32.

[34] Xu, Xing, Yongmei Ji, and Gary D. Stormo. "RNA Sampler: A new sampling-based algorithm for common RNA secondary structure prediction and structural alignment." *Bioinformatics,* Vol. 23, No. 15, 2007, pp. 1883-91.

[35] Bernhart, Stephan H., et al. "RNAalifold: Improved consensus structure prediction for RNA alignments." *BMC Bioinformatics,* Vol. 9, No. 1, 2008, p. 474.

[36] Friedländer, Marc R., et al. "Discovering microRNAs from deep sequencing data using miRDeep." *Nature Biotechnology,* Vol. 26, No. 4, 2008, pp. 407-15.

[37] Tran, Thao T., et al. "De novo computational prediction of non-coding RNA genes in prokaryotic genomes." *Bioinformatics,* Vol. 25, No. 22, 2009, pp. 2897-905.

[38] Hamada, Michiaki, et al. "Prediction of RNA secondary structure using generalized centroid estimators." *Bioinformatics,* Vol. 25, No. 4, 2008, pp. 465-73.

[39] Hackenberg, Michael, et al. "miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments." *Nucleic Acids Research,* Vol. 37, No. 2, 2009, pp. W68-W76.

[40] Reuter, Jessica S., and David H. Mathews. "RNAstructure: Software for RNA secondary structure prediction and analysis." *BMC Bioinformatics,* Vol. 11, No. 1, 2010, p. 129.

[41] Kertesz, Michael, et al. "Genome-wide measurement of RNA secondary structure in yeast." *Nature,* Vol. 467, No. 7311, 2010, pp. 103-07.

[42] Mathelier, Anthony, and Alessandra Carbone. "MIReNA: Finding microRNAs with high accuracy and no learning at the genome-scale and from deep sequencing data." *Bioinformatics,* Vol. 26, No. 18, 2010, pp. 2226-34.

[43] Tsai, Miao-Chih, Robert C. Spitale, and Howard Y. Chang. "Long intergenic noncoding RNAs: New links in cancer progression." *Cancer Research,* Vol. 71, No. 1, 2011, pp. 3-7.

[44] Sato, Kengo, et al. "IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming." *Bioinformatics,* Vol. 27, No. 13, 2011, pp. i85-i93.

[45] Novikova, Irina V., et al. "Rise of the RNA machines: Exploring the structure of long non-coding RNAs." *Journal of Molecular Biology,* Vol. 425, No. 19, 2013, pp. 3731-46.

[46] Fan, Xiao-Nan, and Shao-Wu Zhang. "lncRNA-MFDL: Identification of human long non-coding RNAs by fusing multiple features and using deep learning." *Molecular BioSystems,* Vol. 11, No. 3, 2015, pp. 892-97.

[47] Khachane, Amit N., and Paul M. Harrison. "Mining mammalian transcript data for functional long non-coding RNAs." *PLoS One,* Vol. 5, No. 4, 2010, p. e10316.

[48] Sacco, Letizia da, Antonella Baldassarre, and Andrea Masotti. "Bioinformatics tools and novel challenges in long non-coding RNAs (lncRNAs) functional analysis." *International Journal of Molecular Sciences,* Vol. 13, No. 1, 2012, pp. 97-114.

[49] Guttman, Mitchell, and John L. Rinn. "Modular regulatory principles of large non-coding RNAs." *Nature,* Vol. 482, No. 7385, 2012, pp. 339-46.

[50] Weinberg, Zasha, and Walter L. Ruzzo. "Sequence-based heuristics for faster annotation of non-coding RNA families." *Bioinformatics,* Vol. 22, No. 1, 2005, pp. 35-39.

[51] Xie, Fuliang, et al. "miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs." *Plant Molecular Biology,* Vol. 80, No. 1, 2012, pp. 75-84.

[52] Boerner, Susan, and Karen M. McGinnis. "Computational identification and functional predictions of long non-coding RNA in Zea mays." *PLoS ONE,* Vol. 7, No. 8, 2012, p. e43047.

[53] Sun, Lei, et al. "lncRScan-SVM: a tool for predicting long non-coding RNAs using support vector machine." *PLoS ONE,* Vol. 10, No. 10, 2015, p. e0139654.

[54] Ma, Lina, Vladimir B. Bajic, and Zhang Zhang. "On the classification of long non-coding RNAs." *RNA Biology,* Vol. 10, No. 6, 2013, pp. 924-33.

[55] Liu, Jinfeng, Julian Gough, and Burkhard Rost. "Distinguishing protein-coding from non-coding RNAs through support vector machines." *PLoS Genetics,* Vol. 2, No. 4, 2006, p. e29.

[56] Li, Aimin, Junying Zhang, and Zhongyin Zhou. "PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme." *BMC Bioinformatics,* Vol. 15, No. 1, 2014, p. 311.

[57] Achawanantakun, Rujira, et al. "LncRNA-ID: Long non-coding RNA IDentification using balanced random forests." *Bioinformatics,* Vol. 31, No. 24, 2015, pp. 3897-905.

[58] Cunningham, Fiona, et al. "Ensembl 2015." *Nucleic Acids Research,* Vol. 43, No. D1, 2014, pp. D662-D69.

[59] Li, Jun, et al. "TANRIC: An interactive open platform to explore the function of lncRNAs in cancer." *Cancer Research,* Vol. 75, No. 18, 2015, pp. 3728-37.

[60] Jiang, Qinghua, et al. "LncRNA2Function: A comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data." *BMC Genomics,* Vol. 16. No. 3, 2015.

[61] Chen, Xing. "KATZLDA: KATZ measure for the lncRNA-disease association prediction." *Scientific Reports,* Vol. 5, 2015, p. 16840.

[62] Pan, Xiaoyong, and Kai Xiong. "PredcircRNA: computational classification of circular RNA from other long non-coding RNA using hybrid features." *Molecular Biosystems,* Vol. 11, No. 8, 2015, pp. 2219-26.

[63] Liu, Ming-Xi, et al. "A computational framework to infer human disease-associated long noncoding RNAs." *PLoS ONE,* Vol. 9, No. 1, 2014, p. e84408.

[64] Fan, Xiao-Nan, and Shao-Wu Zhang. "lncRNA-MFDL: identification of human long non-coding RNAs by fusing multiple features and using deep learning." *Molecular BioSystems,* Vol. 11, No. 3, 2015, pp. 892-97.

[65] Chen, Xing. "Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA." *Scientific Reports,* Vol. 5, 2015, p. 13186.

[66] Chen, Xing, et al. "Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity." *Scientific Reports,* Vol. 5, 2015, p. 11338.

[67] Li, Yifeng, Wenqiang Shi, and Wyeth W. Wasserman. "Genome-wide prediction of cis-regulatory regions using supervised deep learning methods." *BioRxiv,* 2016, pp. 1-17.

[68] Al-Tobasei, Rafet, Bam Paneru, and Mohamed Salem. "Genome-wide discovery of long non-coding RNAs in rainbow trout." *PLoS ONE,* Vol. 11, No. 2, 2016, p. e0148940.

[69] Liang, Cheng, Yue Li, and Jiawei Luo. "A novel method to detect functional microRNA regulatory modules by bicliques merging." *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB),* Vol. 13, No. 3, 2016, pp. 549-56.

[70] Arruda, Wosley, et al. "ncRNA-Agents: A multiagent system for non-coding RNA annotation." *Brazilian Symposium on Bioinformatics,* Vol. 8213, 2013, pp. 136-47.

[71] McLaren, William, et al. "The ensemble variant effect predictor." *Genome Biology,* Vol. 17, No. 1, 2016, p. 122.

[72] Milligan, Michael J., et al. "Global intersection of long non-coding RNAs with processed and unprocessed pseudogenes in the human genome." *Frontiers in Genetics,* Vol. 7, 2016, p. 26.

[73] Huang, Yu-An, et al. "ILNCSIM: improved lncRNA functional similarity calculation model." *Oncotarget,* Vol. 7, No. 18, 2016, p. 25902.

[74] Abbas, Qaisar. "Current challenges of computational intelligence techniques for functional annotation of ncRNA genes." *Health Sciences,* Vol. 8, No. 6, 2019, pp. 54-63.