Genomics 99 (2012) 189–194

Contents lists available at SciVerse ScienceDirect

Genomics

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

MiRANN: A reliable approach for improved classification of precursor microRNA using Artificial Neural Network model

Md. Eamin Rahman [a,1], Rashedul Islam [b,1], Shahidul Islam [a,*], Shakhinur Islam Mondal [b], Md. Ruhul Amin [a]

[a] Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh
[b] Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh

ARTICLE INFO

Article history:
Received 28 September 2011
Accepted 5 February 2012
Available online 10 February 2012

Keywords:
MiRANN
Precursor microRNA
Reliable approach
Prediction
Artificial neural network

ABSTRACT

MicroRNA (miRNA) is a special class of short noncoding RNA that serves pivotal function of regulating gene expression. The computational prediction of new miRNA candidates involves various methods such as learning methods and methods using expression data. This article has proposed a reliable model — miRANN which is a supervised machine learning approach. MiRANN used known pre-miRNAs as positive set and a novel negative set from human CDS regions. The number of known miRNAs is now huge and diversified that could cover almost all characteristics of unknown miRNAs which increases the quality of the result (99.9% accuracy, 99.8% sensitivity, 100% specificity) and provides a more reliable prediction. MiRANN performs better than other state-of-the-art approaches and declares to be the most potential tool to predict novel miRNAs. We have also tested our result using a previous negative set. MiRANN, opens new ground using ANN for predicting pre-miRNAs with a promise of better performance.

1. Introduction

1.1. Biogenesis of miRNA

MicroRNAs (miRNAs) are single-stranded small noncoding RNA molecules of about ~22 nucleotides which regulate gene expression at the post-transcriptional level by inhibiting the expression of mRNAs bearing fully or partly complementary sequences. The biogenesis of miRNA is a multistep process; in the first step miRNAs are primarily transcribed by RNA polymerase II to produce primary miRNAs (pri-miRNAs) as part of a long primary miRNA transcript. The pri-miRNAs are processed by an RNases III enzyme named Drosha in the nucleus to yield ~70-nt long hairpin structure. This structure is called precursor miRNAs (pre-miRNAs) and both of its termini bear the signature of Drosha processing i.e., a 5′-phosphate and 2-nt overhangs at the 3′-hydroxylated end [1–3].

The exportin-5 protein is involved in active transport of pre-miRNAs hairpin into the cytoplasm through nuclear pore [4]. In the cytoplasm the pre-miRNAs are sliced at the base of the loop by a second RNase III enzyme, called Dicer, which can recognize the 2-nt overhang at the 3′-end of pre-miRNAs and cuts it ~22 nucleotide away from the overhang to produce the miRNA:miRNA* duplex [5,6]. Usually one of the two strands of the duplex disappears quickly, whereas other strand remains as a mature miRNA of ~22-nt length. The final products are incorporated into miRNA-containing RNA-induced silencing complex (miRISC) to function as sequence-specific guide molecules by complementary binding in translational control or cleavage of certain mRNAs [7–9]. The miRNAs recognize their targets mainly through limited base-pairing interactions between the 5′-end of the miRNA (i.e. 2–8 nt length, called the seed region) and the 3′ untranslated regions (3′-UTRs) of the target mRNAs to up or down regulate the expression of the genes [10,11]. The gene regulation differs invariably in plant and animal because of their variable size of stem loop and also for base pairing differences, positional preferences of U [12] etc.

The importance of the study of miRNAs is rapidly increasing due to its impact in gene regulations. Approximately 30% of human genes are regulated by miRNAs [13]. The miRNAs are associated with the onset of cancer and tumor, neurodegenerative disorders, diabetes and other diseases along with the crucial functions in developmental timing, differentiation, proliferation, cell death [14–17] etc. Different strategies for miRNA therapeutics have been discovered using the related miRNAs to cure diseases [18]. Hence identification of novel miRNA is necessary in order to prevent these diseases. Early approaches for miRNA discovery were limited to directional cloning and sequencing. Recently high-throughput sequencing is able to sequence gigabase of nucleotides in a single run with a very low cost. This technology advanced greatly the strategies of discovering novel miRNAs. However, the identification of novel miRNAs from the large pool of small RNA sequences and analysis of such large volume of data is still a great challenge. To reduce the setbacks and support the wet lab approaches *in silico* prediction play an important role in the identification of new miRNAs. In this paper we have presented a

* Corresponding author.
E-mail addresses: eamin_sust@yahoo.com (M.E. Rahman), rashedul.gen@gmail.com (R. Islam), si_sumon@yahoo.com (S. Islam), shakhin2000@yahoo.com (S.I. Mondal), shajib-cse@sust.edu (M.R. Amin).
[1] These authors contributed equally to the study.

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2012.02.001

reliable computational approach for improved identification of miRNAs.

1.2. Existing computational approaches for miRNA prediction

The existing major computational methods have been developed using filter-based approaches, homology-based search, mixed approaches, and machine learning approaches. Filter-based approaches identify the conserved miRNAs in more than one genome using the feature characteristics of known miRNAs. miRo is a filter-based approach that utilized several filtering steps for miRNA prediction and was applied to a number of species [19]. Phylogenetic shadowing of pre-miRNAs gives a characteristic conservation profile and it was utilized to predict novel miRNAs based on human–mouse–rat comparisons [20]. But these systems failed to identify nonconserved miRNAs due to their dependence on conservation criteria. Homology-based approaches such as ERPIN [21] are a profile-based pre-miRNA detector for animal genome to generate miRNA gene candidates and miRAlign [22] uses the secondary structure alignment of pre-miRNAs in the genome-wide for prediction of pre-miRNA.

Machine learning approaches are popularly used in most of the pre-miRNA prediction work during the last decade. The first reported machine-learning approach is ProMiR which was based on the paired Hidden Markov Model (HMM) that simultaneously considered the structure and sequence of pre-miRNAs [23]. Several machine learning tools that used Support Vector Machine (SVM) algorithm are triplet-SVM [24], MiRFinder [25], MiPred [26], microPred [27], yasMiR [28], MiRenSVM [29], MiRPara [30] etc. In addition to SVM, Naïve Bayes classifier has also been employed in the field of miRNA prediction [31]. Another type of machine learning approach is Artificial Neural Nerwork (ANN) which is used in other study [32]. They used a feed forward multi layer perceptron Artificial Neural Network (ANN) classifier to predict pre-miRNAs. However, they failed to keep the prediction accuracy in Human, Mouse and Rat genome when compared to other existing approaches. The latest released kernel based machine learning approach is G²DE [33] to identify pre-miRNAs with higher prediction accuracy than existing kernel and logic based classifiers. Most of the kernel based methods used known pre-miRNA sequences as a positive set and the negative set was made by pseudo hairpins. Features employed in the methods are primarily sequence conservation, nucleotide frequency, topological properties, thermodynamic stability, entropy measures etc. But with the advent of deep sequencing technology new computational models are developing using the deep sequencing data. Among them the most commonly used prediction tool is mirDeep that provides a user friendly environment to analyze the expression data [34]. mirDeep follows mixed approaches, uses a probabilistic model of miRNA biogenesis and a scoring system to identify miRNA. The approaches that used deep sequencing data got a low false positive rate but still it is a challenge to increase the computational accuracy in analyzing the data and to reduce technical difficulties, since the methods use large volume of data.

Over the same period of time, advancement in miRNA detection using learning algorithms is going on. Because this type of prediction models are inevitable for their prediction accuracy in identifying non-conserved miRNAs and to continue the development of novel computational models. Machine learning has been used in recent miRNA research [35,36].

In our research, we have established a computational model named miRANN using a supervised machine learning approach — Artificial Neural Network (ANN). ANN is an effective technique that has the advantage of enhancing its performance by expanding the network with more neurons and hidden layers. Also miRANN integrates an exceptional weight initializing equation where closest neurons slightly differ in weights that make the network impartial for any feature.

Sometimes, machine learning algorithms are criticized mainly for their learning method that made them to predict only known sequences. But currently there is a vast set of known miRNAs with great diversity of sequences. Therefore, this could be claimed that such large dataset covers almost all characteristics of unknown miRNAs. So, learning of unknown sequence should not be a threat for the quality of the result in our model and a reliable prediction of miRNA will be obtained in human genome.

2. Result and discussion

In this section we have presented the achievements of miRANN to classify pre-miRNAs. The details of our system are described in the Materials and methods section. The prediction performance of miRANN was evaluated using the current miRBase release 18.0 [37] and using a novel negative set from the latest annotation data of human CDS, since this annotated data has been upgraded enormously toward a complete annotation of human genome during recent times. So, the CDS data obtained is less probable to contain unknown splice events. We also showed that the miRANN gives a better-quality result using a well known negative set [24] and miRBase release 15.0 sequences. Then we made a comparison with other predictors using HU920 and HU424 datasets [33] and we have shown here that miRANN is better than other approaches.

2.1. Performance of miRANN

We trained miRANN for all 33 features using TrS1852 dataset and then tested its prediction performance using TeS1200 data set. The

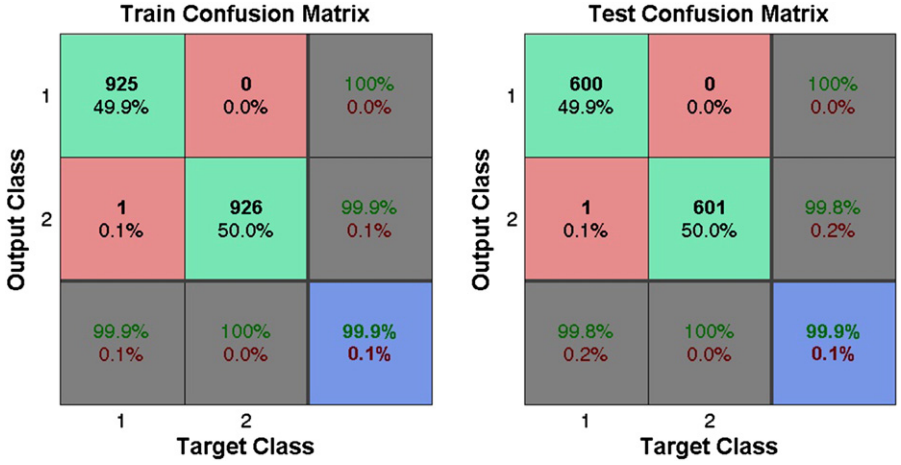

Fig. 1. Confusion Matrix generated for miRANN.

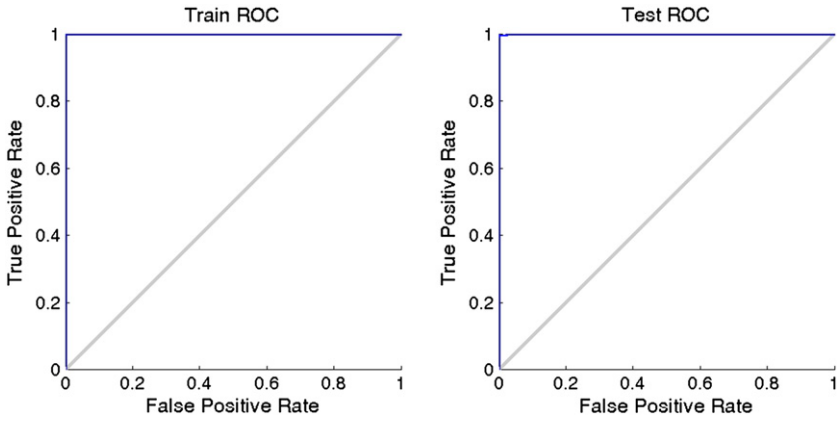

Fig. 2. Receiver Operator Characteristic (ROC) curve generated for miRANN.

prediction gives a high quality result of 99.9% accuracy. The Sensitivity (SE), Specificity (SP) and Accuracy (ACC) of miRANN for training dataset are **99.9% (SE)**, **100% (SP)** and **99.9 (ACC)** and for testing dataset **99.8% (SE)**, **100% (SP)** and **99.9 (ACC)** in due order.

The confusion matrix (Fig. 1) for training dataset shows that number of correctly identified pre-miRNAs (TP) is 925 and pseudo-hairpins (TN) is 926, inaccurately predicted as pre-miRNAs (FP) is 0 and pseudo-hairpins (FN) is 1. On the other hand, the test confusion matrix illustrates that TP, TN, FP and FN are 600, 601, 0 and 1, respectively.

Train Confusion Matrix and Test Confusion Matrix were generated on dataset TrS1852 and TeS1200, respectively. Classes 1 and 2 mean the pre-miRNAs and pseudo hairpins respectively. The blue square displays prediction accuracy 99.9% for TrS1852 and 99.9% for TeS1200.

We demonstrate the performance of miRANN using Receiver Operator Characteristic (ROC) curve which is a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity). Fig. 2 shows Train ROC and Test ROC for dataset TrS1852 and TeS1200 respectively. For both graphs area under the curve (AUC) is almost 1.

Train ROC and Test ROC were generated on dataset TrS1852 and TeS1200, respectively. The AUC of Train ROC and Test ROC almost covers the entire area.

2.2. Evaluation of miRANN predictability using another data set

Triplet-SVM [24] proposed a set of negative hairpins and it was prevalently used in different machine learning algorithms [26–28,33,35,38]. Since this negative set is proven to give dependable results, we used the set to evaluate our performance and still we get a steady performance of miRANN. In this section, we used pre-miRNA sequences from miRBase release 15.0 to construct positive set and negative set was constructed using pseudo hairpins from http://bioinfo.au.tsinghua.edu.cn/mirnasvm/ [24]. Training and testing method of miRANN were identical as described in previous subsection. We used 33 features to construct the data sets (Supplementary material 1). The confusion matrix (Fig. 3) representing the Sensitivity (SE), Specificity (SP) and Accuracy (ACC) of miRANN for training dataset is **96.3%(SE)**, **99.8%(SP)** and **98.1%(ACC)** and for testing dataset **95.3%(SE)**, **98.1%(SP)** and **96.7%(ACC)**.

Classes 1 and 2 mean the pre-miRNAs and pseudo hairpins respectively. The training confusion matrix that represents the number of correctly identified pre-miRNAs (TP) is 597 and pseudo-hairpins (TN) is 619, inaccurately predicted as pre-miRNAs (FP) is 1 and pseudo-hairpins (FN) is 23. On the other hand, the test confusion matrix that illustrates TP, TN, FP and FN is 305, 314, 6 and 15, respectively.

Plotting of the true positive rate (sensitivity) against the false positive rate (1-specificity) has been demonstrated using Receiver Operator Characteristic (ROC) curve (Supplementary material 2).

2.3. Prediction performance using HU920 and HU424 dataset and comparison with other predictors

The prediction performance of miRANN has been compared to kernel based classifiers — support vector machine (SVM) [24], relaxed variable kernel density estimator (RVKDE) [39] and generalized Gaussian components based density estimation (G²DE) [33]. We have also compared with logic based classifiers — C4.5 [40] and RIPPER [41]. In both cases we used HU920 and HU424 dataset of G²DE for training and testing, respectively. To make a valid comparison the 33 characteristic features described in the Feature Set section were partitioned into four different sets according to G²DE.

The best performance among the four feature sets and the average best performance are highlighted with bold font. The #kernels indicate number of kernels in average, where in logic based classifiers the numbers indicate the rules they deliver.

Table 1
Comparison of prediction accuracies achieved by SVM, RVKDE, G²DE, miRANN, C4.5 and RIPPER.

Feature set	Number of features	Kernel based classifier					Logic based classifier	
		SVM	RVKDE	G²DE	G²DE-2	miRANN	C4.5	RIPPER
1	17	80.17%	77.59%	80.39%	80.60%	**85.56%**	77.80%	76.72%
2	7	93.32%	92.46%	92.03%	93.10%	**93.97%**	90.95%	90.52%
3	5	91.60%	91.16%	91.60%	92.46%	**92.67%**	91.16%	91.38%
4	4	78.66%	79.53%	78.66%	80.17%	**83.19%**	77.37%	76.72%
Average		85.94%	85.18%	85.67%	86.58%	**88.85%**	84.32%	83.84%
#kernels		361	920	6	36	**2**	10	9

The best performance among the four feature sets and the average best performance are highlighted with bold font. The #kernels indicate number of kernels in average, where in logic based classifiers the numbers indicate the rules they deliver.

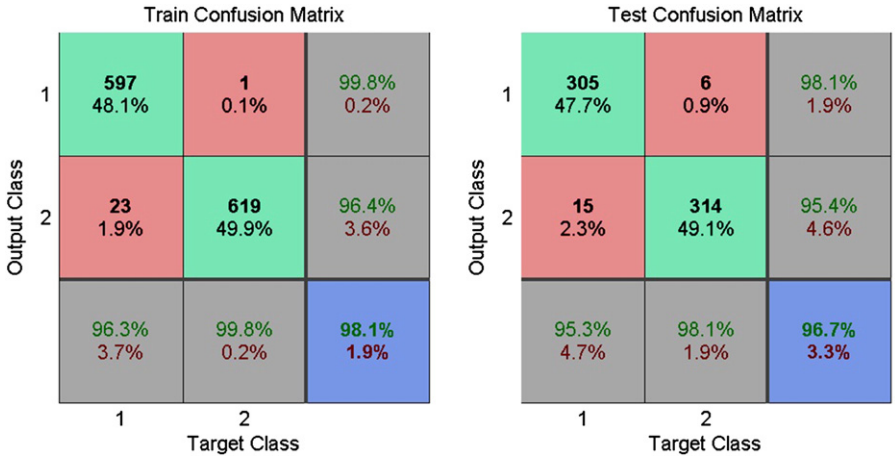

Fig. 3. Confusion Matrix generated for miRANN using another dataset.

Table 1 demonstrates best prediction accuracies of miRANN in all the four feature sets compared to other kernel and logic based classifiers. miRANN gives **85.56%, 93.97%, 92.67%, 83.19%** prediction accuracy in all four feature sets where its performance is **4.96%**, **0.65%**, **0.21%** and **3.02%** (set 1 to 4) better than the existing prediction tools. Average performance of miRANN is **88.85%** and it is **2.27%** higher than G²DE. It is the dazzling improvement of average prediction performance than all other pairwise comparisons of methods using only 2 kernels.

3. Conclusion

In this article, we successfully pioneered a computational approach miRANN as an expanding research field of pre-miRNA prediction with ANN classifier. Extensive evaluation of performance based on human dataset has been done to provide its satisfactory level of prediction performance using miRANN. For human genome our system works as state-of-the-art in pre-miRNA prediction. Established biochemical and secondary structure based features of miRNA were used in miRANN with four biologically significant groupings and also with combined set. Comparison of results with the four different groups proves impressive increase in performance of our proposed system. Moreover, as an ANN classifier, the performance of miRANN can still be improved by increasing size of neuron with more features to achieve a perfect level of prediction.

4. Materials and methods

4.1. Feature set

For our system we used 29 RNA global and intrinsic folding attributes [26] and four "stemloop" features that are based on the miRNA secondary structures [38]. In the 29 features there are 16 dimer and G+C frequency, 6 folding measures — adjusted base pairing propensity (dP), adjusted Minimum Free Energy (MFE) of folding (dG), MFE index 1 (MFEI1), MFE index 2 (MFEI2), adjusted base pair distance (dD), adjusted shannon entropy (dQ), and 1 topological descriptor — degree of compactness (dF), 5 normalized variants of dP, dG, dQ, dD and dF i.e. zP, zG, zQ, zD and zF derived from dimer shuffling. Remaining 4 features are hairpin length, loop length, consecutive base-pairs and ratio of loop length to hairpin length of pre-miRNA secondary structure. Table 2 summarizes all the 33 features.

The table shows the order of a feature within the feature set. Set 1, 2, 3 contains 29 RNA global and intrinsic folding features and set 4 contains stemloop features.

4.2. Dataset

The proposed miRANN system is trained with known human pre-miRNAs as positive dataset and the human pseudo pre-miRNAs hairpins as negative dataset. We retrieved human pre-miRNA sequences from the miRBase release 18.0 to prepare the positive set which

Table 2
Features used in miRANN.

Group	Features	Number (total 33)	Description
Set 1	AA, AC, …, UU	16	Frequencies of dinucleotide pairs
	%G+C	1	Percentage of nitrogenous bases which are either G or C
Set 2	mfe2	1	Ratio of *d*G to the number of stems
	mfe1	1	Ratio of *d*G to %G+C
	P	1	Adjusted base pairing propensity. *d*P is the number of base pairs observed in the secondary structure divided by the sequence length
	*d*G	1	Adjusted minimum free energy of folding. *d*G is the minimum free energy (MFE) divided by the sequence length
	*d*Q	1	Adjusted Shannon entropy. *d*Q measures the entropy of the base pairing probability distribution (BPPD)
	*d*D	1	Adjusted base pair distance. *d*D measures the average distance between all base pairs of structures inferred from the sequence
	*d*F	1	Compactness of the tree-graph representation of the sequence
Set 3	zG, zQ, zD, zP, zF	5	Normalized variants of *d*P, *d*G, *d*Q, *d*D and *d*F
Set 4	lH	1	Hairpin length dangling ends
	lL	1	Loop length
	lC	1	Maximum consecutive base-pairs
	%L	1	Ratio of loop length to hairpin length

The table shows the order of a feature within the feature set. Set 1, 2, 3 contains 29 RNA global and intrinsic folding features and set 4 contains stemloop features.

Table 3
Type, source and number of sequences employed to construct TrS1852 and TeS1200 Data set.

Genome	Type of set	Type of sequence	Source of sequence	Number of available sequence	Number of sequence used	Number of train Sequence	Number of test sequence	% of train sequence	% of test sequence
Human	Positive set	miRNA hairpins	miRBase release 18	1527	1526	926	600	60.68	39.32
	Negative set	Pseudo hairpins	Human CDS hairpins	Unknown	1526	926	600	60.68	39.32
Feature extraction and scaling						TrS1852 Data set	TeS1200 Data set		

contains 1527 reported pre-miRNA entries. We extracted pseudo hairpins randomly from human CDS regions to construct a novel negative set using ScorePin algorithm [42] in human genome from the latest assembly – Feb. 2009 (GRCh37/hg19) – of UCSC genome browser annotations where no experimentally validated splicing event has been reported yet [43]. Although, 0.6–6% miRNAs come from exonic regions of well annotated genomes [44] there is a little chance having CDS hairpins from unknown miRNAs in human genome as we have taken a very small proportion randomly. We tested our negative set to confirm that it does not contain any known miRNA sequence. For all the sequences we predicted secondary structures using RNAfold default parameters at 37 °C [45]. To ensure the similarity of the pseudo hairpins with real pre-miRNAs according to their widely accepted characteristics we selected RNA sequences that are less than 18 base pairs on the stem having the GU wobble pairs, minimum free energy greater than −25 kcal/mol and multiple loops of the predicted secondary structure are taken away.

From the positive set randomly 926 pre-miRNAs and from negative set 926 pseudo hairpins, total 1852 sequences have been selected for training data set and then all 33 feature values that are described in the Feature set are generated for these sequences to construct TrS1852. The remaining 600 pre-miRNAs and 600 pseudo hairpins, total 1200 sequences are used as testing data set and again those 33 feature values have been extracted from these sequences to construct TeS1200 (Table 3). Both TrS1852 and TeS1200 are scaled by SVM scale program [46] to the interval of [−1.0, 1.0]. For generating dataset HU920 and HU464 we used the same procedure as G²DE for reasonable comparison [33].

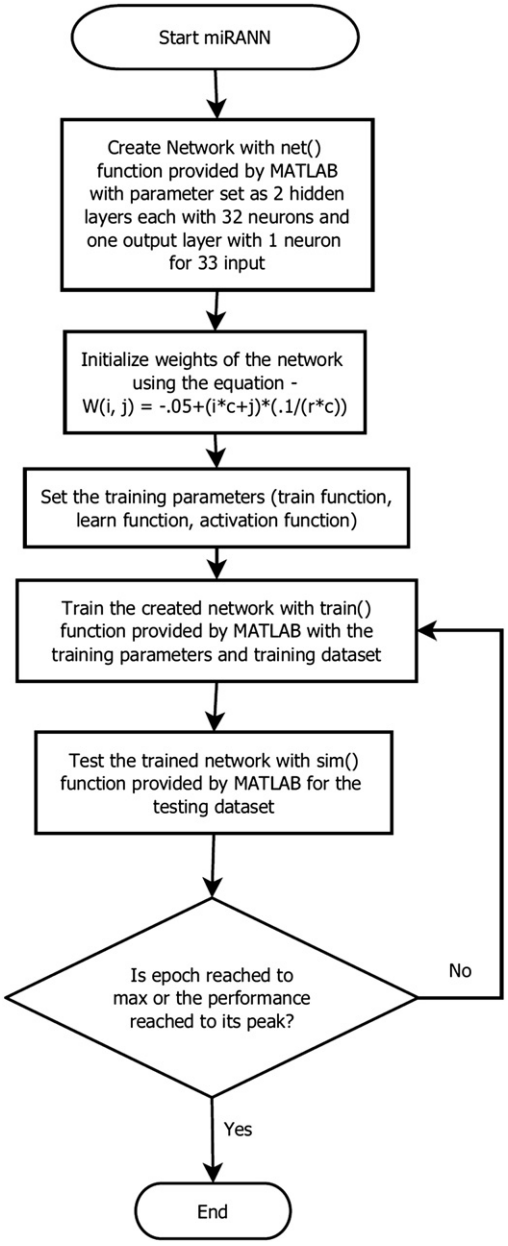

Fig. 4. Shows step by step architectural design of miRANN.

4.3. ANN classifier

Artificial neural networks (ANNs) are computer-based interconnected artificial neurons which simulate the biological neurons of the human brain that can be trained to distinguish and classify complex patterns [47]. In miRANN, multilayer feed-forward backpropagation neural networks has been used as the classifier for pre-miRNA prediction.

We choose three layer backpropagation network with two hidden layers of size 32 neurons and one output layer of 1 neuron for the classification of pre-miRNAs. To implement miRANN we used Quasi-Newton BFGS Algorithm (trainbfg) for training and Radial basis transfer function (radbas) and weight learning function learnlv1; provided by MATLAB Neural Network Toolbox [48]. For weight initialization we applied an exceptional equation −

$$W(i,j) = -.05 + (i * c + j) * (.1/(r * c))$$

where W(i, j) is the weight of j'th neuron of i'th layer, r is the total number of neurons in that layer and c is the total number of inputs. This weight equation distributes weights among neurons uniformly within the range of −.05 to +.05 where closest neurons differ slightly in weights. Optimization was done by repeating the process and miRANN needed 230 epochs.

For inputs we used total 33 features described in the Feature set section. For outputs '1' is set if the sequence is pre-miRNA and '0' is set if the sequence is pseudo hairpin. The complete flow diagram of miRANN containing step by step architectural design is shown in Fig. 4.

Authors' contributions

MER, RI, SI and SIM formed the group and initiated the project. Selection of the research field and forming of the inventive ideas have been done by MER, RI and SI. MER developed the model and wrote the source code and implemented most of the experiments under the guidance of SI. RI assisted in collecting and pre-processing the

data and provided significant insight in developing the biological background under the guidance of SIM. RI and MER wrote the manuscript. MRA conceptualized the computational aspects and helped to improve the manuscript. All authors have read and approved the final manuscript.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2012.02.001.

Acknowledgments

We acknowledge Professor Dr. Muhammed Zafar Iqbal and Professor Dr.Yasmeen Haque for their cooperation and supportive hands throughout the research. We would like to thank Md. Nazmul Alam for the implementation of ScorePin algorithm to construct negative hairpin set. We also thank the Department of Computer Science and Engineering and the Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh for supporting this research.

References

[1] Y. Lee, M. Kim, J. Han, K.H. Yeom, S. Lee, S.H. Baek, V.N. Kim, MicroRNA genes are transcribed by RNA polymerase II, EMBO J. 23 (20) (2004) 4051–4060.
[2] Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, V.N. Kim, The nuclear RNase III Drosha initiates microRNA processing, Nature 425 (6956) (2003) 415–419.
[3] E. Basyuk, F. Suavet, A. Doglio, R. Bordonné, E. Bertrand, Human let-7 stem-loop precursors harbor features of RNase III cleavage products, Nucleic Acids Res. 31 (22) (2003) 6593–6597.
[4] M.T. Bohnsack, K. Czaplinski, D. Görlich, Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs, RNA 10 (2004) 185–191.
[5] G. Hutvágner, J. McLachlan, A.E. Pasquinelli, É. Bálint, T. Tuschl, P.D. Zamore, A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA, Science 293 (5531) (2001) 834–838.
[6] R.F. Ketting, S.E. Fischer, E. Bernstein, T. Sijen, G.J. Hannon, R.H. Plasterk, Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*, Genes Dev. 15 (20) (2001) 2654–2659.
[7] S.M. Elbashir, W. Leneckel, T. Tuschl, RNA interference is mediated by 21- and 22-nucleotide RNAs, Genes Dev. 15 (2001) 188–200.
[8] V.N. Kim, MicroRNA biogenesis: coordinated cropping and dicing, Nat. Rev. Mol. Cell Biol. 6 (2005) 376–385.
[9] S.M. Elbashir, J. Martinez, A. Patkaniowska, W. Lendeckel, T. Tuschl, Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate, EMBO J. 20 (2001) 6877–6888.
[10] K.K. Farh, A. Grimson, C. Jan, B.P. Lewis, W.K. Johnston, L.P. Lim, C.B. Burge, D.P. Bartel, The widespread impact of mammalian microRNAs on mRNA repression and evolution, Science 310 (2005) 1817–1821.
[11] A. Stark, J. Brennecke, N. Bushati, R.B. Russell, S.M. Cohen, Animal microRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution, Cell 123 (2005) 1133–1146.
[12] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.
[13] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (2005) 15–20.
[14] D. Wang, C. Qiu, H. Zhang, J. Wang, Q. Cui, Y. Yin, Human microRNA oncogenes and tumor suppressors show significantly different biological patterns: from functions to targets, PLoS One 5 (9) (Sep 30 2010).
[15] L.H. Kasper, J. Shoemaker, Multiple sclerosis immunology: the healthy immune system vs the MS immune system, Neurology 74 (Supplement 1) (2010) S2–S8.
[16] V. Ambros, MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing, Cell 114 (2) (2003) 269.
[17] D. Anglicheau, T. Muthukumar, M. Suthanthiran, microRNAs: small RNAs with big effects, Transplantation 90 (2) (2010) 105–112.
[18] L. Chunsheng, F. Yi, C. George, Z. Lin, Therapeutic microRNA strategies in human cancer, AAPS J. 11 (4) (2009) 747–757.
[19] P. May, G. Klau, M. Bauer, T. Steinke, Accelerated microRNA-precursor detection using the Smith–Waterman algorithm on FPGAs, Proceedings of GCCB 2006, LNBI, 4360, 2007, pp. 19–32.
[20] E. Berezikov, V. Guryev, J. van de Belt, E. Wienholds, R.H. Plasterk, E. Cuppen, Phylogenetic shadowing and computational identification of human microRNA genes, Cell 120 (2005) 21–24.
[21] M. Legendre, A. Lambert, D. Gautheret, Profile-based detection of microRNA precursors in animal genomes, Bioinformatics 21 (2005) 841–845.
[22] X. Wang, J. Zhang, F. Li, J. Gu, T. He, X. Zhang, Y. Li, MicroRNA identification based on sequence and structure alignment, Bioinformatics 21 (2005) 3610–3614.
[23] A. Sewer, N. Paul, P. Landgraf, A. Aravin, S. Pfeffer, M.J. Brownstein, T. Tuschl, E. van Nimwegen, M. Zavolan, Identification of clustered microRNAs using an ab initio prediction method, BMC Bioinformatics 6 (2005) 267.
[24] C. Xue, F. Li, T. He, G.P. Liu, Y. Li, X. Zhang, Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine, BMC Bioinformatics 6 (2005) 310.
[25] T.H. Huang, B. Fan, M.F. Rothschild, Z.L. Hu, K. Li, S.H. Zhao, MiRFinder: an improved approach and software implementation for genome-wide fast microRNA precursor scans, BMC Bioinformatics 8 (2007) 341.
[26] S. Kwang Loong, S.K. Mishra, De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures, Bioinformatics 23 (11) (2007) 1321–1330.
[27] R. Batuwita, V. Palade, microPred: effective classification of pre-miRNAs for human miRNA gene prediction, Bioinformatics 25 (8) (2009) 989–995.
[28] D. Pasailă, I. Mohorianu, A. Sucilă, Ş. Panţiru, L. Ciortuz, Yet another SVM for miRNA recognition: yasMiR, Technical Report, Alexandru Ioan Cuza, 2010 TR 10-01.
[29] J. Ding, S. Zhou, J. Guan, MiRenSVM: towards better prediction of microRNA precursors using an ensemble SVM classifier with multi-loop features, BMC Bioinformatics 11 (Suppl 11) (Dec 14 2010) S11.
[30] Y. Wu, B. Wei, H. Liu, T. Li, S. Rayner, MiRPara: a SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences, BMC Bioinformatics 12 (2011) 107.
[31] M. Yousef, M. Nebozhyn, H. Shatkay, S. Kanterakis, L. Showe, M. Showe, Combining multi-species genomic data for microRNA identification using a Naïve Bayes classifier, Bioinformatics 22 (11) (2006) 1325–1334.
[32] C. Vinod, G. Reshmi, A pre-microRNA classifier by structural and thermodynamic motifs, IEEE-NaBIC 2009, 2009, pp. 9–11.
[33] C.H. Hsieh, D.T.H. Chang, C.H. Hsueh, C.Y. Wu, Y.J. Oyang, Predicting microRNA precursors with a generalized Gaussian components based density estimation algorithm, BMC Bioinformatics 11 (Suppl 1) (2010) S52.
[34] M.R. Friedländer, W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel, N. Rajewsky, Discovering microRNAs from deep sequencing data using miRDeep, Nat. Biotechnol. 26 (2008) 407–415.
[35] J. Xiao, X. Tang, Y. Li, Z. Fang, D. Ma, Y. He, M. Li, Identification of microRNA precursors based on random forest with network-level representation method of stem-loop structure, BMC Bioinformatics 12 (2011 May 17) 165.
[36] P. Xuan, M. Guo, Y. Huang, W. Li, Y. Huang, MaturePred: efficient identification of MicroRNAs within novel plant pre-miRNAs, PLoS One 6 (11) (2011) e27422.
[37] S. Griffiths-Jones, H.K. Saini, S. van Dongen, A.J. Enright, miRBase: tools for microRNA genomics, Nucleic Acids Res. 36 (2008) D154–D158.
[38] D.T.H. Chang, C.C. Wang, J.W. Chen, Using a kernel density estimation based classifier to predict species-specific micro-RNA precursors, BMC Bioinformatics 9 (Suppl 12) (2008) S2.
[39] Y.J. Oyang, S.C. Hwang, Y.Y. Ou, C.Y. Chen, Z.W. Chen, Data classification with radial basis function networks based on a novel kernel density estimation algorithm, IEEE Trans. Neural Netw. 16 (1) (2005) 225–236.
[40] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Francisco, 1993.
[41] W.W. Cohen, Fast effective rule induction, International Conference on Machine Learning 1995, 1995, pp. 115–123.
[42] S.A. Helvik, O. Snøve Jr., P. Saetrom, Reliable prediction of Drosha processing sites improves microRNA gene prediction, Bioinformatics 23 (2) (2007 Jan 15) 142–149.
[43] D. Karolchik, R. Baertsch, M. Diekhans, T.S. Furey, A. Hinrichs, Y.T. Lu, K.M. Roskin, M. Schwartz, C.W. Sugnet, D.J. Thomas, R.J. Weber, D. Haussler, W.J. Kent, The UCSC Genome Browser Database, Nucleic Acids Res. 31 (1) (2003) 51–54.
[44] L.C.G. Hinske, P.A.F. Galante, W.P. Kuo, L.O. Machado, A potential role for intragenic miRNAs on their hosts' interactome, BMC Genomics 11 (2010) 533.
[45] I.L. Hofacker, Vienna RNA secondary structure server, Nucleic Acids Res. 31 (13) (2003) 3429–3431.
[46] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm. 2001.
[47] C.M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.
[48] The MathWorks Online Documentation: Neural Network Toolbox, http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/function.html. 2011.