REVIEW ARTICLE

Applications of Machine Learning in miRNA Discovery and Target Prediction

Alisha Parveen[1], Syed H. Mustafa[2], Pankaj Yadav[3] and Abhishek Kumar[4,5,*]

[1]*Institute of Medical Bioinformatics and Systems Medicine Medical Center, Faculty of Medicine, Albert-Ludwigs University of Freiburg, 79110 Freiburg, Germany;* [2]*Department of Computer Engineering, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India;* [3]*Department of Bioscience and Bioengineering, Indian Institute of Technology, Jodhpur, India;* [4]*Institute of Bioinformatics, International Technology Park, Bangalore, 560066, India;* [5]*Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India*

Abstract: MicroRNA (miRNA) is a small non-coding molecule that is involved in gene regulation and RNA silencing by complementary on their targets. Experimental methods for target prediction can be time-consuming and expensive. Thus, the application of the computational approach is implicated to enlighten these complications with experimental studies. However, there is still a need for an optimized approach in miRNA biology. Therefore, machine learning (ML) would initiate a new era of research in miRNA biology towards potential diseases biomarker. In this article, we described the application of ML approaches in miRNA discovery and target prediction with functions and future prospective. The implementation of a new era of computational methodologies in this direction would initiate further advanced levels of discoveries in miRNA.

1. INTRODUCTION

MicroRNAs (miRNAs) are endogenous, short length (around 22 base pairs) non-coding RNA molecules that play an important role in the gene regulation in animals, plants and viruses. miRNA regulates the post-transcriptional expression of a gene by aligning to a different region that directs to translational repression or endonucleolytic cleavage of the coding gene [1]. MiRNA is known to be involved in several biological processes such as muscle development, hematopoiesis, apoptosis, immune system, aging and signal transduction depending upon target regulation (Fig. **1A**). Moreover, it plays a central role during the embryogenesis period that leads to tissue identity and differentiation. It acts as a biomarker for several fatal diseases including cancer, heart disease and neurological disorders [2].

During the 1990s, Lee *et al.* discovered miRNA in the nematode *Caenorhabditis elegans* as lin-4 [3, 4]. About two decades later, let-7 gene reported in the same nematode species with similarity to lin-4 [5]. In the biogenesis, RNA polymerase II (RNA pol II) along with Dorsha enzyme transcribed miRNA gene into long primary transcript (pre-miRNAs) in the nucleus. Exportin V protein transports pre-miRNA into the cytoplasm for further processing of the hydrolysis of Ran-GTP complex. During the

maturation phase, ATP dependent protein Dicer recognizes pre-miRNA and processed into the duplex structure of miRNA-miRNA*. In the duplex structure, one strand is an antisense strand consisting of G: U wobble base pair, other matches, mismatches and unpaired at the 5' end and whereas the other strand is sense strand [2]. Previous studies found that double-stranded load on the argonaute-1 (AGO-1) protein along with an RNA-induced silencing complex (RISC) makes this complex guide to target mRNA that leads to the post-transcriptional expression of the gene as described in Fig. (**1B**).

There are two different processes of the post-transcriptional phases as the endonucleolytic cleavage phase and the repression phase. In the endonucleolytic cleavage process, miRNA sequence is extensively aligned to their target gene by removing poly (A) tail, which is being silenced by AGO protein leading to target fragmentation [6]. In translational repression, miRNA sequence aligns partially to a binding region of their target gene that resists the binding of ribosomes to the target gene resulting in the inhibition of synthesis of the polypeptide [6, 7].

In the past, significant progress has been made in the discovery of miRNAs and their target predictions. This includes the ascertainment of several physical and functional characteristics of miRNA that are indicative of the miRNA functions and targets. These characteristics include information such as folding patterns, thermodynamic properties, and sequence conservation [8]. Machine-learning (ML) methods are very useful for the performance of real-time

*Address correspondence to this author at the Institute of Bioinformatics, International Technology Park, Bangalore, 560066, India; and Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India; Tel: +91 80-28416140; Fax: +91 80-28416132; E-mail: abhishek@ibioinformatics.org
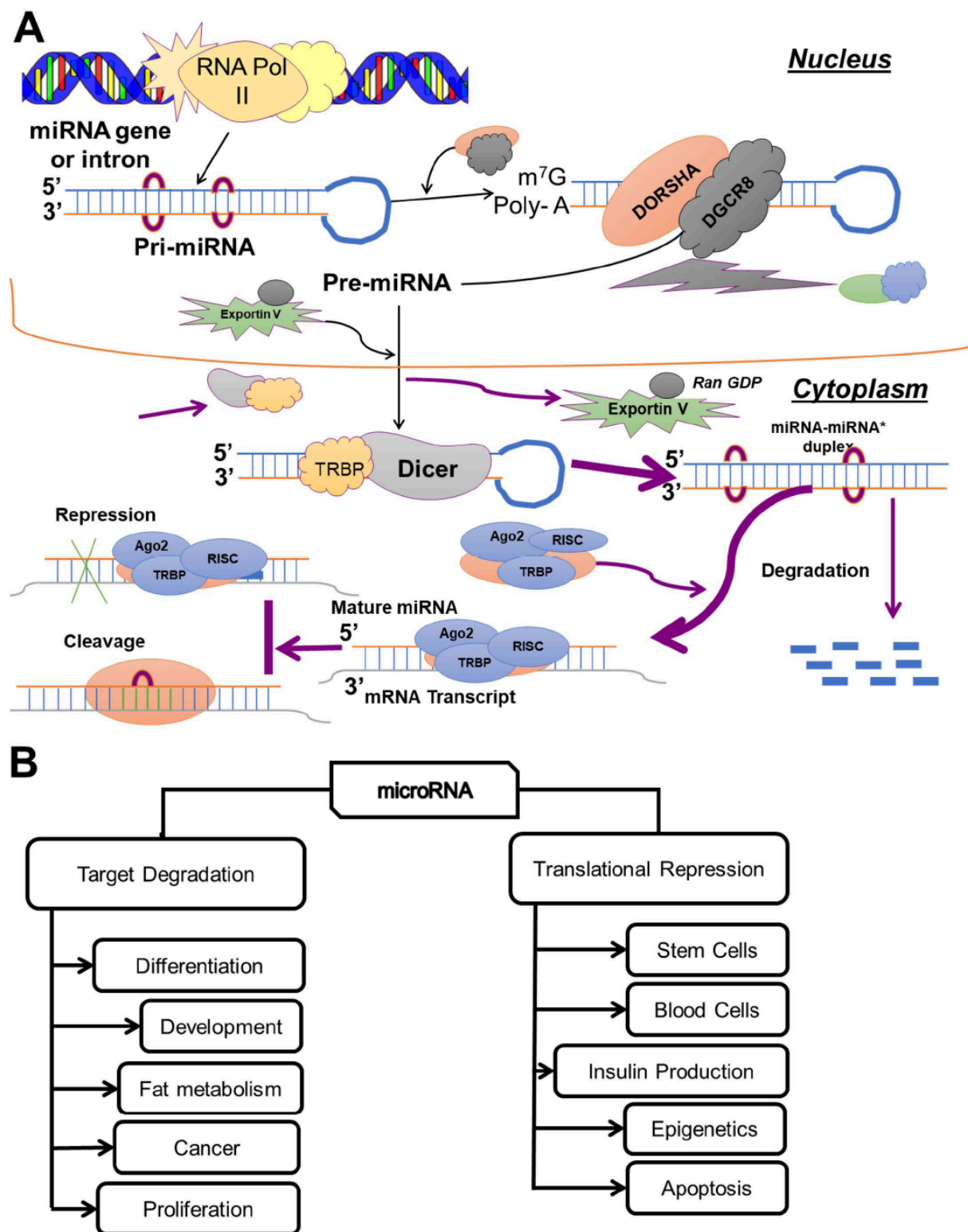
Fig. (1). Overview of microRNA biology. A. Biogenesis of microRNA. Intergenic miRNAs are transcribed by RNA polymerase II or III, that generates a primary miRNA molecule (pri - miRNA), and processed by a microprocessor complex comprised of DGCR8 and Drosha into a precursor miRNA (pre - miRNA). Pre - miRNAs are exported to the cytoplasm in a transporter containing Exportin 5 and Ran - GTP nucle-ocytoplasm. Intronic miRNA is transcribed by RNA polymerase II as part of the precursor mRNA (pre - mRNA). The miRNA sequence is excised by spliceosomal components or the microprocessor from the pre-mRNA in order to release an exported mirtron or pre-miRNA. Alternatively, a primary miRNA (pri - miRNA) is released to generate pre - miRNA by microprocessor cleavage. **B. Post-transcriptional phase.** microRNA induce transcriptional repression and cleavage of the target gene and involve in different biochemical pathways. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

predictive and analytical study in the identification of miR-NA and their target genes which are involved in different diseases [9]. This article presents the recent advancement in the implementation of the ML in the field of microRNA biology and biomedical research. Furthermore, we present the fundamental concept behind various classification methods including supervised and unsupervised approaches.

2. APPLICATION OF MACHINE LEARNING IN MicroRNA IDENTIFICATION

The generalized process for the application of machine learning in miRNA identification is shown in Fig. (**2**) and summarized in Table **1**. The machine learning approach involves several steps in training the miRNA identification classifier as summarized:

a) To train the classifier model, data mining is a crucial step in the extraction and identification of features from a dataset.

b) In the generation of the positive dataset, the hairpin sequence of miRNA extracted from experimentally verified databases [10] undergoes several levels of filtration to improve the high confidence positive set.

c) The negative dataset is equally important to train the classifier, so that it can easily distinguish between negative and positive datasets. The excess amount of negative dataset and positive dataset can create overfitting and underfitting model, respectively. For instance, the positive sample dataset should consist of miRNA duplexes derived from experimentally validated miRNAs. To avoid redundant information, only one miRNA duplex should be included in the positive sample set if both the 5' and 3'

strands of the miRNA duplex are functional. The negative samples should consist of pseudo miRNA duplexes derived from segments randomly selected from pre-miRNA hairpins. The miRNA secondary structure can be predicted using RNAfold package as implemented in Vienna software [11].

d) Learning classifier is trained after the generation of positive and negative datasets. Also, a different algorithm based on the prediction of the mature miRNA sequences can generate features.

e) Different features are measured to train the classifier by different platforms and to build a model for identification of miRNA sequence. For example, Scikit-learn python package and Keras package for deep learning [12].

f) The best classifier model for the identification of miRNA is selected based upon the cross-validation results.

3. SUPPORT VECTOR MACHINE (SVM)

Support vector machine (SVM) is a popular discriminative classifier that has shown to be an efficient classifier model in dealing with classification problems. SVM takes labeled data as input and it generates decision hyperplanes [13]. Fig. (**3**) evaluates a linearly separable 2D axis cartesian plane between two classes with multiple linear lines, which is a solution to the problem. The same concept is applied to problems where more than two categories have to be classified. Therefore, the goal of the SVM is to find the optimal linear line that passes to possible all data points [13]. SVM learning problem is based on generating an optimal solution for multiple existences of lines that generate the largest
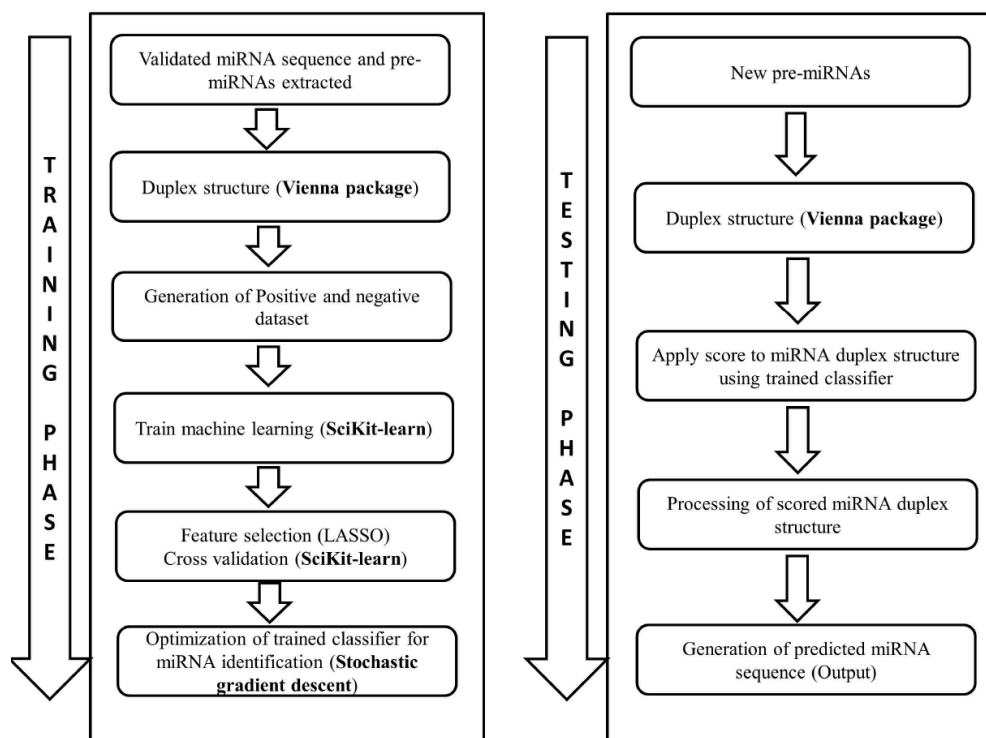

Fig. (2). Generalized steps involved in training machine learning classifier. Learning classifier model involved in 2 different phases: Training stage and testing stage to generate an accurate and robust model.

Table 1. **Summary of different tools of microRNA prediction, target prediction and functional annotation.**

A. miRNA Prediction				
Tools	**Algorithm**	**Positive**	**Negative**	**Feature**
miR-abela	SVM	miRBase	Coding region	Structure
Triplet-SVM	SVM	Rfam	Pseudo miRNA hairpin	Structure and sequence feature
miPred	Random forest	Registry database	Pseudo miRNA hairpin	Structure and thermodynamics
ProMir	Probabilistic colearning	Mature pre-miRNA	Randomly extracted stem loop	Structure, thermodynamics, and sequence
MiRRim	Hidden Markov Model	Conserved miRNA	Non-conserved, moderately conserved, and highly conserved	Sequence alignment
HHMMiR	Hidden Markov Model	microRNA registry	Coding region	Structure and thermodynamics
B. Target Prediction				
Tools	**Algorithm**	**Positive**	**Negative**	**Feature**
MBSTAR	Random forest	miRBase	Randomly generated	Sequential and Structural
NbmiRTar	Naïve Bayes	Tarbase	Probability Randomization	Sequence
TargetBoost	Genetic Programming	let-7, lin-4, miR-13a, and bantam	Random string with same frequency	Sequence
TarpmiR	Random forest	CLASH	Reshuffling site of target site	Structure
DeepTarget	Recurrent Neural Network	miRecords and miR-Base	Mocking in alignment	Sequence
TargetMiner	SVM	miRecords	Randomly generated	Seed
C. Functional annotation of miRNA				
Tools	**Algorithm**	**Positive**	**Negative**	**Feature**
GenMiR++	Bayesian learning	Tarbase and miRecords	Negative correlation in expression profiles	Sequence and expression data
Joung *et al.*	Probabilistic learning	Expression profile	Low profile expression data	Parametric adjusted population size, and minimum subset size
Tran *et al.*	Rule based	Expression profile from human cancer	Randomly generated	Alignment
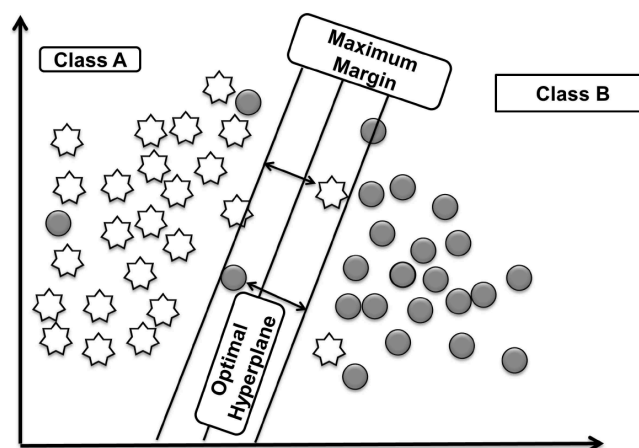
Fig. (3). Support vector model (SVM)-based classification. This shows the basic form of linear classification problem of two data points that generate a linear model (as depicted inline). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

minimum distance to labelled data. Based on the SVM, there are several methods for identification of miRNAs sequence as described below:

a) **miR-abela** is based on the SVM classifier program for predicting mammalian miRNAs and it has shown high specificity for a dataset of 40 pre-miRNAs. The features involved in the pre-miRNA sequence prediction included thermodynamic energy, loop length, conservation and stem length. However, this algorithm has low specificity for the identification of a mature miRNA sequence [14].

b) **Triplet-SVM** has a triple set of nucleotides to generate structural and sequential feature properties that indicate the pairing state of every three adjacent nucleotides in which true miRNAs sequence lies and is separated by pseudo hairpins [15]. It has illustrated more accuracy for genomic data of animals as compared to lower species based on classifier performance [15].

c) **RNAz** is based upon selected features such as thermodynamic stability, conservation and sequential and structural properties in predicting structural noncoding RNAs and cis-acting regulatory elements of mRNAs. It can be used to detect functional RNA structures deployed in genome-wide screens. This algorithm has a high sensitivity, but it also has a high type I error rate [16].

4. HIDDEN MARKOV MODEL (HMM)

The Hidden Markov Model (HMM) is a statistical model using the probabilistic distribution for modeling time series data. HMM can be applied to a stochastic process having unobservable hidden states. HMM-based algorithms are frequently applied for miRNA identification [17].

a) **ProMir** is a web server for the generation of the non-coding miRNA query sequence. ProMir method is trained with loop-based features generated based upon a probabilistic score. ProMir II is an optimized version method used to identify both conserved and non-conserved miRNA sequences. The latest version of ProMir implements miRNA identification score and several filtering criteria that include free energy, GC content and conservation [18].

b) **MiRRim** uses the evolutionary and structural features as a multidimensional vector to train the classifier. In the miRNA structure, the stem region is more conserved than the loop region and their corresponding surrounding regions are also less conserved across different species [19].

c) **HHMMiR** uses thermodynamic energy, similar to RNA-fold program [11], as selected features to train hierarchical HMM classifier to predict hair-loop structure of miRNA, which lacks the evolutionarily conserved feature. This method has high sensitivity and specificity for determining functional roles of the miRNAs [20].

5. NAÏVE BAYES

Naïve Bayes is a simple probabilistic classifier based on applying Bayes' theorem with strong (naïve) independence assumptions. A more descriptive term for the underlying probability model would be the "independent feature model".

In simple terms, a naïve Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Depending on the probability model, Naïve Bayes classifiers can be trained very efficiently in a supervised learning setting [21].

a) **BayesMirFind** was created for the identification of miRNA sequence in *C. elegans* and mice. It uses a comparative post-filtering technique on a large set of sequential and structural features providing >80% of sensitivity and >90% of specificity. However, overall classifier performance is poor as compared to other algorithms [21].

b) **miR-KDE** was developed by Chang *et al.* and is based on the classification problem. For extraction of pre-miRNAs in humans, this method uses the hairpin sequence and structural features collected from previously published work. miR-KDE incorporates a variable kernel density method to classify RNA sequence from a generated set of features. Experimentally verified pre-miRNA is collected from 40 species to evaluate the overall performance of the classifier [22].

6. APPLICATION OF LEARNING CLASSIFIER IN THE TARGET PREDICTION

Previous research shows that miRNA sequence has a seed region, which is 6-8 nucleotide in length that aligns at the 5′ end of the mature mRNA [8]. The binding of miRNA to target is an important feature by different pairing sites (Fig. **4**). Thus, linear development in the amount of data in genomics and proteomics needs accurate and precise prediction algorithm unlike rule-based traditional methods [9]. In the rule-based prediction, the algorithm determines whether user-provided miRNA sequences are not manually created; rather generate characteristics trained to learn classifier [23].
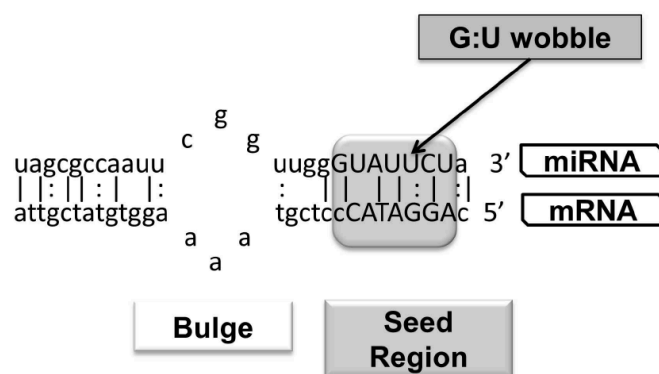
Fig. (4). Features involved in microRNA-target. For a training classifier model, different features generation is involved in microRNA-target interactions such as sequence, structural, thermodynamic and functional feature of alignment. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Several machine learning-based miRNA target prediction algorithms were developed in the last decade. The general process in target prediction has been as follows:

a) For each miRNA, identify the putative binding site from the validated target (as positive) and non-target (as nega-

tive) mRNAs based on seed site complementarity and other features such as thermodynamics energy and global alignments.

b) Extract or identify features from these interactions (irrespective of whether they are functional or not).

c) Train a classifier to distinguish between target and non-targets.

d) For an unknown miRNA-mRNA pair, use the classifier to label it as positive (target) or negative (non-target).

A few of the popular machine learning-based algorithm in target prediction are described below:

a) **MBSTAR** is based on the random forest (RF) algorithm with a combination of 31 structural and 340 sequence features and that are applied to an unsupervised learning algorithm to select 40 putative features for miRNA-mRNA interaction. After 5-fold cross-validation, the top six different multiple instance learning (MIL) techniques are considered to evaluate the classifier. After performing an analysis of different learning classifiers, the RF was shown to have the highest accuracy. For further analysis, MBSTAR shows more putative binding sites in nucleotide conversion in biological sequence. This method is generated from mature miRNA sequences downloaded from miRBase database and 3UTRs sequence of mRNA for feature extraction [24].

b) **TarPmiR** detects miRNA and their corresponding target sites by the application of random forest by the integration of six regular features and seven other newly generated features for a prediction. For the generation of the negative dataset, TarPmiR applies random application in the target site with equal expression level and molecular functions. The selection of proper negative dataset is based on a different parameter per se no overlapping between positive sites and negative sites and low binding energy. To evaluate the learning classifier, TarpmiR shows good results with the PAR-CLIP dataset in a human HEK293 cell line [25].

c) **DeepTarget** is the combination of both supervised and unsupervised learning methods. This method relies on the application of autoencoder to generate target prediction by utilizing a sequence-based interaction feature to train the recurrent neural network (RNN) model. DeepTarget has a high level of accuracy and eliminates the necessity of manually curated features for prediction. The appearance of pattern nucleotide positions in the RNN layer corresponds to interaction. DeepTarget delivers a quantum leap in the longstanding challenge of robust miRNA target prediction (Website, http://data.snu.ac.kr/pub/deep Target/).

d) **NBmiRTar** is a Naïve Bayes classifier learning approach. To train this classifier, features are extracted from the seed and out-seed regions filtered from the output of miRanda tool. In this approach, incorporation of these in seed and out-seed structural and sequential features improves the performance of NBmiRTar. These artificial mature miRNAs consist of a random string of nucleosides A, C, G and U with a probability of 0.34, 0.19, 0.18 and 0.29, respectively, that are not consistent with

the base frequencies in true miRNAs. Several parameters based on free energy and conservation are applied to artificial interaction for the production of a negative dataset, which is further used as input to the classifier [26].

7. MACHINE LEARNING IN FUNCTIONAL CHARACTERIZATION OF miRNA

In an early section, we have provided a brief overview of several different machine learning-based methods in the identification of miRNA and the target prediction. In a broad range of genomics, machine learning is a useful tool for interpretation of a large amount of genomics data [27]. Functional annotation helps miRNA research. In this section, we provide functional information and involvement of miRNA in cancer and other disease pathways using machine-learning techniques. Functional analysis that reveals the involvement of miRNA in several physiological processes are essential for determining the association with diseases [28]. miRNA is an emerging therapeutic agent against diseases. Therefore, several databases have been developed in past few decades that provide functional information of miRNA such as DAVID [29] which is a knowledge-based visualization of functional annotation of miRNA. Another online database is miRDB [30] that is based on functional information of miRNA target prediction sequence. There are two statistical tools for evaluation of miRNA functions in different diseases, namely MAGIA (miRNA and genes integrated analysis) [31] and FAME (Functional Assignment of miRNAs *via* Enrichment) [32]. These methods have several limitations due to the dependence on target prediction. The main disadvantage of these approaches is that they are unable to predict whether miRNA binds outside the binding region. Due to this, these approaches are not able to predict functional analysis beyond the binding region. GenMiR++ is a generative Bayesian interface that has high sensitivity and it is used to evaluate the expression profiles. GenMiR++ provides a balanced score between predicted and other generated miRNA target interactions and identifies their functional annotation. To test the performance of this approach, the dataset of biological process (BP) annotations from the gene ontology (GO) annotation database is collected to test the performance of the classifier. GenMiR++ is more consistent with high score confidence on the set of predicted functional targets from the sequence-based predictions [33].

8. FURTHER DIRECTIONS

With the extensive increment in novel discoveries of therapeutic, an identified biomarker must undergo trained validation procedure before undergoing clinical trials. For example, in recent studies, scientists executed the application of advanced learning techniques to determine the stability of Tarcolimus (immunosuppressive drug) in renal transplantation [34].

Recently, deep learning is receiving more attention in the domain of genomics. In the past, pharma companies have been trying to develop new learning technology that resulted in Deep genomics costing approx. $13 million and iCarbonX. Due to this high-end learning technology, scientists have advanced knowledge in understanding of human genome in a much faster and accurate way.

Chen *et al.* carried out an investigation for the improvement of the accuracy of metabolism defects using the advanced learning model [35]. This study was primarily covered with chemical compounds that involve in specified conditions reduced false positive rates for *e.g.*, phenyl ketonuria decline false positive rates from 21 to 2, hypermethioninemia from 30 to 10 data point, and 2-methyl crotonyl-CoA-carboxylase reduced from 209 to 46 deficiency [35].

The applications of ML have diverse importance in genomics and personalized medicine that generate fast and accurate models before undergoing further clinical trials, more than traditional medicine. Companies are expecting to adopt advanced learning technology to know more about unsolved queries in their related field.

CONCLUSION

Discovery in miRNA has brought forth a new era in the field of molecular biology across the world. This discovery excited researchers and there have been aggressively taken a peak in the development of computational biology approaches in miRNA biology. An enormous amount of genomic data has been generated from molecular biology technique like next-generation sequencing, microarray, *etc*. This application has overcome the difficulties of experimental procedures involved in miRNA discovery and target predictions but also the limitation of conservation-based computational approaches. Several ML approaches like SVM HMM, naïve bayes, and deep learning (advanced artificial neural networks), which is an optimized version of neural network classifier have been an efficient framework to identify novel miRNA and their target prediction with high accuracy and a less false positive rate. In deep learning, the recurrent neural network is naturally capable of temporary remodeling of naturally fitting biological molecules like DNA, RNA, proteins, and miRNA.

Therefore, the advancement in ML approaches is needed in miRNA research for novel identification, target prediction and functional annotation for clinical biomarkers.

LIST OF ABBREVIATIONS

AGO-1	=	Argonaute-1
BP	=	Biological Process
FAME	=	Functional Assignment of miRNAs *via* Enrichment
GC	=	Guanine-Cytosine
GO	=	Gene Ontology
HMM	=	Hidden Markov Model
MAGIA	=	miRNA and Genes Integrated Analysis
MIL	=	Multiple Instance Learning
miRNAs	=	microRNAs
ML	=	Machine Learning
RF	=	Random Forest
RISC	=	RNA Induced Silencing Complex
RNA pol II	=	RNA polymerase II
RNN	=	Recurrent Neural Network
SVM	=	Support Vector Machine

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

REFERENCES

[1]　Kaeuferle, T.; Bartel, S.; Dehmel, S.; Krauss-Etschmann, S. MicroRNA methodology: advances in miRNA technologies. *Methods Mol. Biol.,* **2014**, *1169*, 121-130.
[http://dx.doi.org/10.1007/978-1-4939-0882-0_12]　[PMID: 24957235]

[2]　Felekkis, K.; Touvana, E.; Stefanou, Ch.; Deltas, C. MicroRNAs: A newly described class of encoded molecules that play a role in health and disease. *Hippokratia,* **2010**, *14*(4), 236-240.
[PMID: 21311629]

[3]　Lee, R.C.; Feinbaum, R.L.; Ambros, V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* **1993**, *75*(5), 843-854.
[http://dx.doi.org/10.1016/0092-8674(93)90529-Y]　[PMID: 8252621]

[4]　Bhaskaran, M.; Mohan, M. MicroRNAs: History, biogenesis, and their evolving role in animal development and disease. *Vet. Pathol.,* **2014**, *51*(4), 759-774.
[http://dx.doi.org/10.1177/0300985813502820] [PMID: 24045890]

[5]　Shi, Z.; Hayes, G.; Ruvkun, G. Dual regulation of the lin-14 target mRNA by the lin-4 miRNA. *PLoS One,* **2013**, *8*(9), e75475.
[http://dx.doi.org/10.1371/journal.pone.0075475]　[PMID: 24058689]

[6]　Obernosterer, G.; Leuschner, P.J.; Alenius, M.; Martinez, J. Post-transcriptional regulation of microRNA expression. *RNA,* **2006**, *12*(7), 1161-1167.
[http://dx.doi.org/10.1261/rna.2322506] [PMID: 16738409]

[7]　Ha, M.; Kim, V.N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.,* **2014**, *15*(8), 509-524.
[http://dx.doi.org/10.1038/nrm3838] [PMID: 25027649]

[8]　Parveen, A.; Gretz, N.; Dweep, H. Obtaining miRNA-target interaction information from miRWalk2.0. *Curr. Protoc. Bioinformatics,* **2016**, *55*(1), 12.15.1-12.15.27.

[9]　Allmer, J.; Yousef, M. Computational miRNomics. *J. Integr. Bioinform.,* **2016**, *13*(5), 1-2.
[http://dx.doi.org/10.1515/jib-2016-302] [PMID: 29216003]

[10]　Kozomara, A.; Griffiths-Jones, S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.,* **2014**, *42*(Database issue), D68-D73.
[http://dx.doi.org/10.1093/nar/gkt1181] [PMID: 24275495]

[11]　Lorenz, R.; Bernhart, S.H.; Höner Zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.,* **2011**, *6*, 26.
[http://dx.doi.org/10.1186/1748-7188-6-26] [PMID: 22115189]

[12]　Medved, D.; Nugues, P.; Nilsson, J. Predicting the outcome for patients in a heart transplantation queue using deep learning. *Conf. Proc. IEEE Eng. Med. Biol. Soc.,* **2017**, *2017*, 74-77.
[http://dx.doi.org/10.1109/EMBC.2017.8036766]　[PMID: 29059814]

[13]　Chen, Q.; Cao, F. Distributed support vector machine in master-slave mode. *Neural Netw.,* **2018**, *101*, 94-100.
[http://dx.doi.org/10.1016/j.neunet.2018.02.006]　[PMID: 29494875]

[14]　Sebastian, B.; Aggrey, S.E. Specificity and sensitivity of PROMIR, ERPIN and MIR-ABELA in predicting pre-microRNAs in the chicken genome. *In Silico Biol. (Gedrukt),* **2008**, *8*(5-6), 377-381.
[PMID: 19374126]

[15]　Xue, C.; Li, F.; He, T.; Liu, G.P.; Li, Y.; Zhang, X. Classification

of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics,* **2005**, *6*, 310.
[http://dx.doi.org/10.1186/1471-2105-6-310] [PMID: 16381612]

[16]   Lindow, M.; Gorodkin, J. Principles and limitations of computational microRNA gene and target finding. *DNA Cell Biol.,* **2007**, *26*(5), 339-351.
[http://dx.doi.org/10.1089/dna.2006.0551] [PMID: 17504029]

[17]   Muthusamy, S.K.; Dalal, M.; Chinnusamy, V.; Bansal, K.C. Genome-wide identification and analysis of biotic and abiotic stress regulation of small heat shock protein (HSP20) family genes in bread wheat. *J. Plant Physiol.,* **2017**, *211*, 100-113.
[http://dx.doi.org/10.1016/j.jplph.2017.01.004] [PMID: 28178571]

[18]   Nam, J. W.; Kim, J.; Kim, S.K.; Zhang, B.T. ProMiR II: A web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. *Nucleic Acids Res.,* **2006**, *34*(Web Server issue), W455-W458.
[http://dx.doi.org/10.1093/nar/gkl321]

[19]   Terai, G.; Komori, T.; Asai, K.; Kin, T. miRRim: A novel system to find conserved miRNAs with high sensitivity and specificity. *RNA,* **2007**, *13*(12), 2081-2090.
[http://dx.doi.org/10.1261/rna.655107] [PMID: 17959929]

[20]   Kadri, S.; Hinman, V.; Benos, P.V. HHMMiR: Efficient *de novo* prediction of microRNAs using hierarchical hidden Markov models. *BMC Bioinformatics,* **2009**, *10*(Suppl. 1), S35.
[http://dx.doi.org/10.1186/1471-2105-10-S1-S35]   [PMID: 19208136]

[21]   Yousef, M.; Nebozhyn, M.; Shatkay, H.; Kanterakis, S.; Showe, L.C.; Showe, M.K. Combining multi-species genomic data for microRNA identification using a Naive Bayes classifier. *Bioinformatics,* **2006**, *22*(11), 1325-1334.
[http://dx.doi.org/10.1093/bioinformatics/btl094]   [PMID: 16543277]

[22]   Chang, D.T.; Wang, C.C.; Chen, J.W. Using a kernel density estimation-based classifier to predict species-specific microRNA precursors. *BMC Bioinformatics,* **2008**, *9*(Suppl. 12), S2.
[http://dx.doi.org/10.1186/1471-2105-9-S12-S2]   [PMID: 19091019]

[23]   Kleftogiannis, D.; Korfiati, A.; Theofilatos, K.; Likothanassis, S.; Tsakalidis, A.; Mavroudi, S. Where we stand, where we are moving: Surveying computational techniques for identifying miRNA genes and uncovering their regulatory role. *J. Biomed. Inform.,* **2013**, *46*(3), 563-573.
[http://dx.doi.org/10.1016/j.jbi.2013.02.002] [PMID: 23501016]

[24]   Bandyopadhyay, S.; Ghosh, D.; Mitra, R.; Zhao, Z. MBSTAR: Multiple instance learning for predicting specific functional binding sites in microRNA targets. *Sci. Rep.,* **2015**, *5*, 8004.
[http://dx.doi.org/10.1038/srep08004] [PMID: 25614300]

[25]   Ding, J.; Li, X.; Hu, H. TarPmiR: A new approach for microRNA target site prediction. *Bioinformatics,* **2016**, *32*(18), 2768-2775.
[http://dx.doi.org/10.1093/bioinformatics/btw318]   [PMID: 27207945]

[26]   Yousef, M.; Jung, S.; Kossenkov, A.V.; Showe, L.C.; Showe, M.K. Naïve Bayes for microRNA target predictions-machine learning for microRNA targets. *Bioinformatics,* **2007**, *23*(22), 2987-2992.
http://dx.doi.org/10.1093/bioinformatics/btm484 PMID: 17925304

[27]   Williams, A.M.; Liu, Y.; Regner, K.R.; Jotterand, F.; Liu, P.; Liang, M. Artificial intelligence, physiological genomics, and precision medicine. *Physiol. Genomics,* **2018**, *50*(4), 237-243.
[http://dx.doi.org/10.1152/physiolgenomics.00119.2017]   [PMID: 29373082]

[28]   Orange, D.E.; Agius, P.; DiCarlo, E.F.; Robine, N.; Geiger, H.; Szymonifka, J.; McNamara, M.; Cummings, R.; Andersen, K.M.; Mirza, S.; Figgie, M.; Ivashkiv, L.; Pernis, A.B.; Jiang, C.; Frank, M.; Darnell, R.; Lingampali, N.; William, R.; Gravallese, E.; Bykerk, V.P.; Goodman, S.M.; Donlin, L.T. Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. *Arthritis Rheumatol.,* **2018**, *70*(5), 690-701.
[http://dx.doi.org/10.1002/art.40428]

[29]   Huang, W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.,* **2009**, *4*(1), 44-57.
[http://dx.doi.org/10.1038/nprot.2008.211] [PMID: 19131956]

[30]   Wong, N.; Wang, X. miRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.,* **2015**, *43*(Database issue), D146-D152.
[http://dx.doi.org/10.1093/nar/gku1104] [PMID: 25378301]

[31]   Sales, G.; Coppe, A.; Bisognin, A.; Biasiolo, M.; Bortoluzzi, S.; Romualdi, C. MAGIA, a web-based tool for miRNA and genes integrated analysis. *Nucleic Acids Res,* **2010**, *38*(Web Server issue), W352-W359.
[http://dx.doi.org/10.1093/nar/gkq423]

[32]   Ulitsky, I.; Laurent, L.C.; Shamir, R. Towards computational prediction of microRNA function and activity. *Nucleic Acids Res.,* **2010**, *38*(15), e160.
[http://dx.doi.org/10.1093/nar/gkq570] [PMID: 20576699]

[33]   Huang, J.C.; Morris, Q.D.; Frey, B.J. Bayesian inference of MicroRNA targets from sequence and expression data. *J. Comput. Biol.,* **2007**, *14*(5), 550-563.
[http://dx.doi.org/10.1089/cmb.2007.R002] [PMID: 17683260]

[34]   Tang, J.; Liu, R.; Zhang, Y.L.; Liu, M.Z.; Hu, Y.F.; Shao, M.J.; Zhu, L.J.; Xin, H.W.; Feng, G.W.; Shang, W.J.; Meng, X.G.; Zhang, L.R.; Ming, Y.Z.; Zhang, W. Application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients. *Sci. Rep.,* **2017**, *7*, 42192.
[http://dx.doi.org/10.1038/srep42192] [PMID: 28176850]

[35]   Chen, W.H.; Hsieh, S.L.; Hsu, K.P.; Chen, H.P.; Su, X.Y.; Tseng, Y.J.; Chien, Y.H.; Hwu, W.L.; Lai, F. Web-based newborn screening system for metabolic diseases: machine learning *versus* clinicians. *J. Med. Internet Res.,* **2013**, *15*(5), e98.
[http://dx.doi.org/10.2196/jmir.2495] [PMID: 23702487]