DOI 10.5113/kjcisv7i1.218

Predicting and Characterizing piRNAs and their Functions Using an Integrated Machine Learning Approach

Anam Umera[a] , Sajid Mahmood[a] , Usman Inayat[a]*

[a] School of Systems and Technology, University of Management and Technology
anamumera333@gmail.com, sajid.mahmood@umt.edu.pk, usman.inayat@umt.edu.pk

*Corresponding Author: Usman Inayat usman.inayat@umt.edu.pk

Abstract

One class of short non-coding RNA molecule that is well recognized is called PIWI-interacting RNA (piRNA). PiRNAs are involved in the creation of novel medications as well as the identification of different kinds of tumors. Additionally, it is associated with stopping transposes, managing gene transcription, and maintaining genomic integrity. The important role that piRNAs play in biological processes has led to a growing body of research in bioinformatics on the discovery of piRNAs and their functionality. In this research, a powerful model is proposed to improve PiRNA prediction and functionality. The suggested model uses four classifiers (Logistic Regression, SVC, Random Forest, and Gradient Boosting Classifier) for classification. Moreover, TNC and DNC are used to acquire features. There are two layers involved in developing the suggested model. A sequence's potential to be piRNA is predicted in the first layer, and its potential to direct target mRNA deadenylation is predicted in the second. In the first layer, the model's accuracy is 98.59%, and in the second layer, it is 94.55%.

Keywords: PIWI-interacting RNAs, Intelligent Model for PiRNAs, Prediction of PiRNAs,

Machine learning based model, Functions of PIWI-interacting RNAs.

List of Abbreviations:

RNA	:	RiboNucleic Acid
PiRNA	:	Piwi-interacting RNAs
RRNAs	:	Ribosomal RNAs
SnoRNAs	:	Small nucleolar RNAs
ncRNA	:	Non-coding RNA
SiRNAs	:	Small interference RNAs
TRNAs	:	Transfer RNAs
MRNA	:	Messenger RNA
LncRNA	:	Long non-coding RNAs
BC	:	Breast Cancer
ORFs	:	Open Reading Frames
OH	:	Hydroxide
PIWI	:	P-element Induced Wimpy
PiRISC	:	PiRNA-Induced Silencing Complex

DNA	:	Deoxyribonucleic Acid
DNMT	:	DNA Methyl Transferase
ShRNA	:	Short hairpin RNA
GC	:	Gastric Cancer
HCC	:	Hepatocellular Carcinoma
TEs	:	Transposable Elements
ISCs	:	Intestine Stems Cells
CLIP	:	Cross-Linking and ImmunoPrecipitation
LFE-GM	:	Luca Fuzzy Entropy and Gaussian Membership
SRA	:	Sparse Recognition Algorithm
FCS	:	Feature Score Criteria
DAC	:	Di-Nucleotide Auto Covariance
DNN	:	Deep Neural Network
GB	:	Gradient boosting
SVM	:	Support Vector Machine
RF	:	Random Forest
LR	:	Logistic Regression
DNC	:	Di-Nucleotide Composition
TNC	:	Tri-Nucleotide Composition

1. Introduction

Through a variety of adjustments, dynamic balance in human health is achieved. Normal daily routines require the internal environment to be in a stable state. Internal environment disturbances lead to mutations and imbalances in genes, which leads to a variety of human illnesses, often to the extent of cancer, heart and brain disorders, claiming millions of lives annually. The regular functional activities of the cells are primarily driven by several biochemical signaling systems and genetic modifications, maintaining the homeostasis of the internal environment [1]. PiRNA is small non-coding RNA molecule, having significant importance in the development of novel medications as well as the identification of different kinds of tumors, heart and brain disorders. Additionally, it is associated with stopping transposes', managing gene transcription, and maintaining genomic integrity [2].

Almost all non-coding RNAs have drawn a great interest for their involvement in cellular processes, and in detection of multiple disorders. All the regulating non-coding RNAs can be roughly split into smaller and large non-coding RNAs depending on the size of the molecule [3]. In prokaryotic cells, PIWI interacting RNA is a documented class of short non-coding RNA molecules having a polymer that is 24 to 31 nucleotides long. PiRNAs carry out a wide range of genetic and biological tasks, such as, controlling the activation of genes, preserving, and forming genetic material, and synthesis of specific protein [4].

2. Background

One class of short non-coding RNA molecule that is well recognized is called PIWI-interacting RNA (piRNA), distinct from other types, including ~~such as~~ microRNAs and siRNAs, which play crucial roles in the control of gene production in animal cells. Initially in early 2000s, piRNAs are predominantly found in the germ line cells of animals, although their presence and functions in

somatic cells have also been recognized [5]. These PiRNA molecules association with PIWI proteins, (a subclade of the Argonaut proteins), forms piRNA complexes, this interaction plays a fundamental role in silencing transposable elements, thereby preventing genomic instability and potential mutagenesis. The primary piRNAs role is to safeguard the genome's stability, achieving this by targeting and regulating transposable elements, which are sequences that can change their positions within the genome, often resulting in mutations and alterations in the cell's genetic material. By the formation of complexes with PIWI proteins, piRNAs facilitate the cleavage of transpose mRNA, thus preventing potentially harmful elements from being transcribed and interfering with normal cellular processes [6].

Additionally, piRNAs play a role in the epigenetic regulation of genes. They influence the modification of chromatin structure, thus aiding in the transcriptional silencing of genes. This regulatory capacity extends beyond transposes to include other genomic elements and is crucial during development, especially in germ cells where it ensures the transmission of genetic information across generations without disruptions from transposable elements. The study of piRNAs holds significant clinical implications, particularly in the field of oncology. Aberrant expression of piRNAs has been associated with various types of cancers, implying that they might act as biomarkers for the detection and prediction of cancer. Moreover, because piRNAs contribute to the maintenance of genomic stability, understanding their mechanisms can provide insights into the progression of cancer, where genomic instability is a hallmark [7]. The potential role of piRNAs in therapeutics is also being explored. As regulators of gene expression, targeting piRNA pathways, could offer new avenues for the treatment of diseases that arise from genetic and epigenetic dysregulations, such as cancer and hereditary disorders. Modulating piRNA activity might help to correct or compensate for pathological gene expression profiles. Recent studies indicate that piRNAs are widely transcribed in a variety of physiological cell types and are associated with a wide range of clinical conditions besides those that have been documented in the germ cells. PiRNAs, for instance, have been found to explain inordinately in many diseases [8].

PiRNAs have also demonstrated promising result as prognostic indicators for a variety of tumors. The dysregulation of piRNAs in cancer suggests that they may be potential targets in cancer therapy. The most prevalent malignancy and the leading reason for tumor related deaths in women is breast tumor. PiRNAs influence a significant impact in breast tumor and may be used as diagnostics and treatment approaches. The excessive transcription of piRNAs is seen in tumors and is linked to proliferation of the cancer cells. Additionally, there is an indication that piRNA-mediated genomic mechanisms contribute to cancer. In breast tumor cells, piR-651 was shown to be substantially amplified. One of the biggest reasons of cancer related deaths worldwide is lung cancer. A possible diagnostic and treatment strategy in lung cancer involves excessive piRNA transcription, therefore targeting it might be a possible tool in limiting the cancer spread and growth. In laboratory and in mammalian lung cancer, progression might be inhibited by piR-55490, which was also inversely correlated with patients' survival. While other research has established that in lung cancer cells piR-651 is enhanced, piR-L-163 was the piRNA that was most frequently down-regulated in long cell lung cancer when opposed to the equivalent non tumor lung cells [9].

3. Literature Review

For recognizing sequences both piRNA and non-piRNA, numerous models have been proposed that are computerized. To predict piRNA, a sequential computational model called "piRNA-CNN" based on convolutional neural network is proposed [11]. DNC and TNC are used for feature extraction. In [12] presented 2L-piRNA, an effective predictor. It's an ensemble model with two layers.-The 1st layer used to determine whether a dataset is a PiRNA or non-PiRNA, and the 2nd layer used to determine whether a piRNA can direct the de-adenylation of a target mRNA or not. For classification, SVM classifier, CNN models are used and pseudo dinucleotide composition (k=2) for feature extraction. "2S-piRCNN", a two-stage deep-learning classifier that makes use of a CNN, is proposed in [13]. In [14] "2L-piRNAPred" model is proposed. 2L-piRNAPred is SVM-based predictor. 2-layer merged Scheme for point out piRNAs in the 1st layer and identifying if piRNA have the task of directing selected mRNA de-adenylation in the 2nd layer. In [17] proposed "piRNN" for identifying piRNA. Convolutional neural network classifiers were used, each of which had been trained using datasets from four organisms. Each sequence was represented by a matrix of k-mer frequencies. In [18] "2L-piRNADNN" model was proposed. To reduce computing complexities via parallel processing, it is suggested to use the DNN model with the Spark computing platform. A feature vector consisting of numerical values that was created from the RNA sequences by the suggested model's use of the dinucleotide auto covariance approach.

In [19] created a quick, reliable, and effective deep learning technique called piRDA for locating the correlations between piRNAs and diseases. Without using any features of engineering, the suggested architecture takes the most important and information that is generically expressed in a piRNA disease pair from the unprocessed sequences. K fold cross validation is used to assess the effectiveness of the suggested method piRDA. In contrast to community methods, the piRDA greatly outperforms them all in terms of quality assessment criteria for the detection of piRNA disease connections. In [20], they provide an integrated strategy for piRNA prediction that considers a range of genetic and epigenetic traits that can be utilized to describe these molecules. They have gathered and examined a sizable variety of piRNA characteristics that have been empirically verified in numerous species. In an object-oriented framework that uses a Various Kernel Learning technique, these properties are expressed by several kernels. The developed tool, known as IpiRId, outperforms all other tools with prediction findings that reach more than 90% accuracy for the three examined species. Additionally, their method enables researchers to examine the applicability of each specified trait in a particular specie.
The proposed method can also be modified to anticipate different types of ncRNAs because it is modular and easily expandable. IpiRId model performance in term of Acc, Sp, Sn is 93.66%, 96.58%, 90.74% respectively. For classification, SVM classifier, CNN models are used and pseudo dinucleotide composition (k=2) for feature extraction. "2S-piRCNN", a two stage deep learning classifier that makes use of a CNN, is proposed in [21]. In this [22] study, they created the 2lpiRNApred a combined algorithm with two layers' approach. In the first layer it recognizes piRNAs and assesses if they are involved in the second layer process of inducing target mRNA deadenylation. To make the attributes' dimensions smaller, a new feature extraction approach depends on Gaussian and Luca fuzzy entropy participation function (LFE-GM) was presented. Two types of classifiers—Sparse Recognition Algorithm (SRA) and Five attribute detection techniques using the Support Vector Machine with Mahalanobis Proximity Dependent Rotational

Basis Process: Extended serial similarity pseudo dinucleotide composition, k-mer, overall sequence similarity pseudo dinucleotide composition, Standardized Moreau Broto auto-correlation, and Geary auto-correlation—were combined to create the unified classifier method with two layers, The outcomes show that 2lpiRNApred outperforms six other available prediction methods by a wide margin. 2lpiRNApred model performance in term of Acc, Sp, Sn, MCC is 88.72%, 85.54%, 91.89%, 0.775 respectively.

To improve the use of deep learning methods for piRNA and function identification, a two-layer predictor is proposed in this [23] study. The suggested approach uses multiple feature acquisition methods to consider the physical and chemical characteristics of the biological sequences while extracting features. The k-fold cross-validation approach is used to thoroughly assess the suggested strategy output. The evaluation's findings indicate that the suggested model outperformed the current models, with accuracy gains of 7.59 and 2.81 percent at layer 1st and layer 2nd, correspondingly. The suggested paradigm is believed to be useful for precision medicine and cancer diagnostics. In [24], they address current developments in our knowledge of piRNA function as well as prospective testing and treatment uses of piRNAs in a variety of digestion malignancies. In order to predict connections between piRNAs and disorders using information-retrieving technology, they present a unique predictor dubbed iPiDA-LTR. According to research findings, iPiDA-LTR shows promising results in detecting disorders linked to both previously identified and newly discovered piRNAs.

In this study, a powerful model is proposed to improve PiRNA prediction and its functionality. The proposed model uses four classifiers (Logistic Regression, SVC, Random Forest, and Gradient Boosting Classifier) for classification. Moreover, TNC and DNC are used to acquire features. There are two layers involved in developing the suggested model. A sequence's potential to be piRNA is predicted in the first layer, and its potential to direct target mRNA deadenylation is predicted in the second.

This contribution is organized as follows: Section 4 represent material and methods. Section 5 represents the results and discussion of the suggested model. Finally, in Section 6, a conclusion and future directions are offered.

4. Materials and methods

The following five step are used to complete the proposed model as shown in Figure 1: (1) data acquisition and analysis, (2) feature extraction (DNC, TNC), (3) training the three ML algorithms, (4) generating new features on the basis of the three ML model's outputs of step # 3 and the original dataset, and (5) training the one ML classifier. The detail of these stages is given below.
Same model and stages are used for both "Dataset 1" and "Dataset 2" prediction. As Shown in section 4.1 "Dataset 1" represent S and "Dataset 2" represent S+.

4.1. Data Acquisition and Analysis

The same dataset as in [20] is used. There are 709 piRNA sequences that instruct target mRNA deadenylation (denoted as S^{+}_{inst}), 709 piRNA sequences that do not have this function (denoted as $S^{+}_{non-inst}$), and 1418 non-piRNA sequences. PiRNA sequences denoted as S^{+} and non-PiRNA sequences denoted as S^{-}. Consequently, the datasets for this investigation can be described as follows:

$$S = S^{+} \cup S^{-} \ldots\ldots \quad (1)$$
$$S^{+} = PiRNA \ldots\ldots\ldots\ldots \quad (2)$$
$$S^{-} = NonPiRNA \ldots\ldots \quad (3)$$
$$S^{+} = S^{+}_{inst} \cup S^{+}_{non-inst} \quad (4)$$

.

S represents “Dataset 1” and S+ represents “Dataset 2”.

Figure 1. 5 Steps of the Proposed model

4.2. Feature Extraction

The piRNA and non-PiRNA sequence dataset is used to extract features, and two feature encoding techniques are used (i.e., DNC, TNC). In this research, iLearn web server is used for feature extraction. DNC is used to extract 16 features and TNC to extract 64 features in total.

a. ***Di-Nucleotide Composition (DNC)***

Feature extraction approach that uses a set of 2 sequential nucleotides to represent an RNA sequence is called Di-Nucleotide Composition (DNC). The probability of each set, denoted by N1 N2 for the first set, N2 N3 for the second set, and so on, is calculated [12]. 16 features are provided by the di-nucleotide composition. It's described as:

$$\boldsymbol{D(a,b)} = \frac{\boldsymbol{Nab}}{\boldsymbol{n-1}}, \text{ a, b} \in \{A, C, G, (U)\} \ldots\ldots(5)$$

In equation (5) *Nab* is the number of Di-Nucleotide described by nucleic acid types *a* and *b*. For example, *ab* ∈ {AA, AC, . . . L, GT, TT} sequences.

b. ***Tri-Nucleotide Composition (TNC)***

Another feature extraction approach is called Tri-Nucleotide Composition (TNC), which uses three pairs of sequential nucleotides to describe an RNA sequence. Each pair's probability is computed. For instance, the first set in an RNA sequence is N1 N2 N3, the second set is N2 N3 N4, and so on, producing a 4 x 4 x 4 = 64D matching features vector [10]. It's outlined as:

$$\boldsymbol{D(a,b,c)} = \frac{\boldsymbol{N(abc)}}{\boldsymbol{n-2}}, \text{ a, b, c} \in \{A, C, G, (U)\} \ldots\ldots(6)$$

In equation (6) *Nabc* is the number of Tri-Nucleotide described by nucleic acid types *a, b* and *c*. For example, *abc* ∈ *{AAA, AAC, AAG, ..., TTT}* sequences.

4.3. Train Machine Learning Classifiers

We merge DNC and TNS features and make a dataset. Before experiment, we check the sub features contribution in overall dataset. So, we select best top 10 sub features from dataset using Random Forest Regressor. Figure 2 presents the important sub features. All sub features range is 0.00 to 0.05, but CG sub feature range is 0.35. So, we normalize the CG sub feature with 0.15.

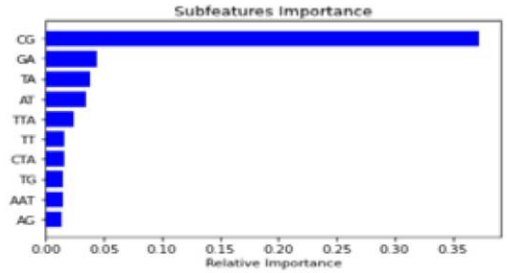

Figure 2. Top 10 important Sub-features with their relative importance

Figure 3 represents a workflow for creating a machine learning model pipeline using a combination of classifiers and a meta-classifier. Figure 4 represents the same workflow as Figure 3, but the difference between Figure 3 and 4 is that Figure 3 is used for "Dataset 1" and Figure 4 is used for "Dataset 2". "Dataset 1" as shown in session 2.1 contain PiRNA and non-PiRNA data sequences. And "Dataset 2" contains samples having the function of instructing target RNA deadenylation and samples without this function.

The process begins with loading the initial dataset, referred to as "Dataset 1" and "Dataset 2". Three classifiers (Logistic Regression, SVM, and Random Forest) are used for prediction. The predictions from the three classifiers are combined with the original features from the test set to create a new dataset. The new dataset, which includ s both the original features and the predictions from the three classifiers, is used to train a meta-classifier. In this case, the meta-classifier is a Gradient Boosting Classifier. The Gradient Boosting Classifier is trained on the new dataset, and its accuracy is evaluated to assess the achievement of the overall model.The proposed model has been developed in Python using Google Colab. We employed the Holdout Method to divide the dataset into two parts: 80% for training and 20% for testing.

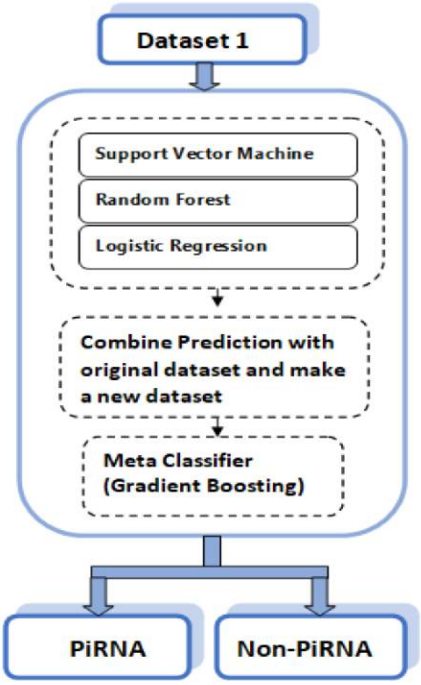

Figure 3. Proposed Model for Dataset 1

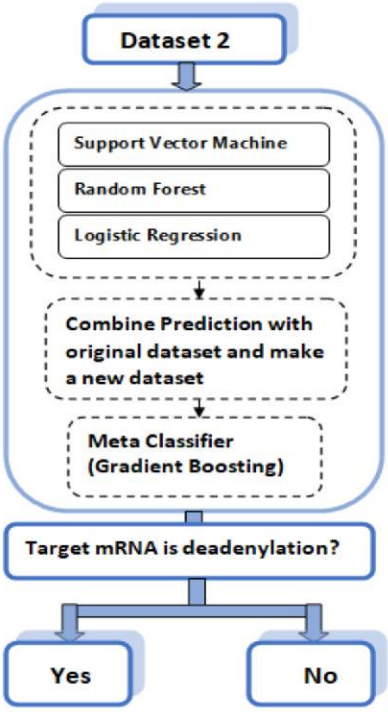

Figure 4. Proposed Model for Dataset 2

5. Results & Discussion

In this section, the experimental results of the suggested model for both Dataset 1 and Dataset 2 are discussed. As Figure 3 indicates, when Dataset 1 is applied to the suggested model, the results are obtained in the form of PiRNA or Non-PiRNA. Whereas, when Dataset 2 is applied, the results were obtained in the form of target mRNA deadenylation or no deadenylation.

5.1. Results of the Proposed Model for Dataset 1

Figure 5 provides a visual summary of the model's performance across various metrics: Precision, AUC (Area Under the Curve), MCC (Matthews Correlation Coefficient), F1-Score, Balanced Accuracy, Specificity, Sensitivity (Recall), and Accuracy.

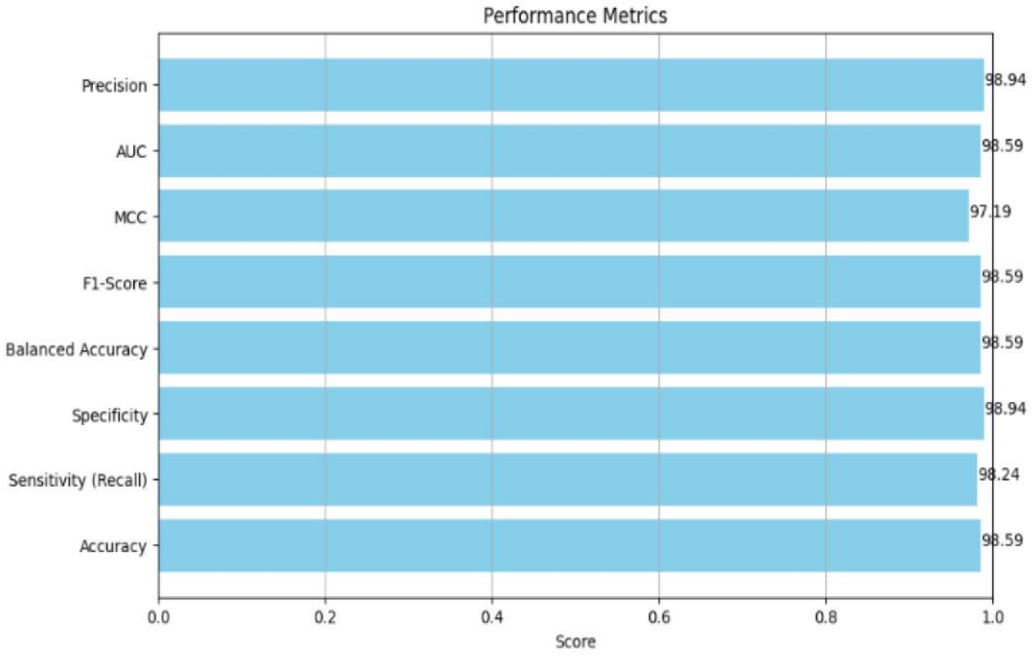

Figure 5. Results of the Proposed Model for Dataset 1

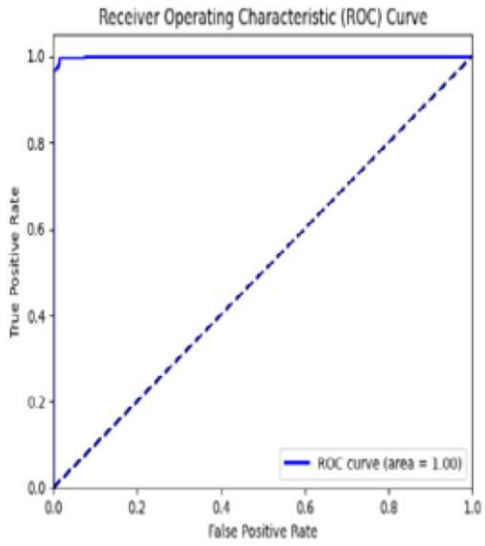

Figure 6. ROC Curve for Dataset 1

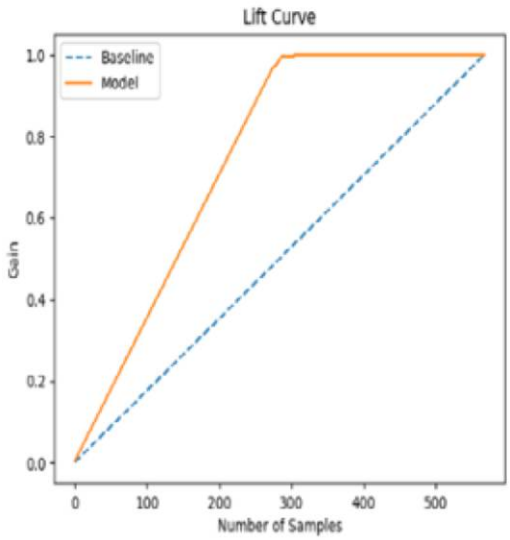

Figure 7. Lift Curve for Dataset 1

Figure 6 shows, a Receiver Operating Characteristic (ROC) Curve, which plots the True Positive Rate (TPR, also known as Sensitivity or Recall) against the False Positive Rate (FPR) for different threshold settings of a binary classification model. In the figure.6, performance of the model shows using blue line. In this case: the ROC curve is almost touching the top-left corner, indicating excellent model performance. The Area Under the Curve (AUC) is 1.00, suggesting a perfect classification ability of the proposed model. The lift curve, depicted in Figure. 7, compares the outcomes produced with and without a predictive model to determine how effective the model is. The performance of a random model is represented by the horizontal line, which is the baseline. The lift, or the ratio of the model-assisted outcomes to the model-off outcomes, is represented by the vertical axis.

5.2. Results of the Proposed Model for Dataset 2

Figure 8 shows, a visual summary of the model's achievement for Dataset 2 across various metrics. These metrics collectively deliver a comprehensive overview of the model's achievement, ensuring it is robust across different aspects of prediction quality. Figure 8 illustrates these metrics, emphasizing the model's capability to effectively classify the data in Dataset 2, which involves distinguishing between target mRNA deadenylation and not.

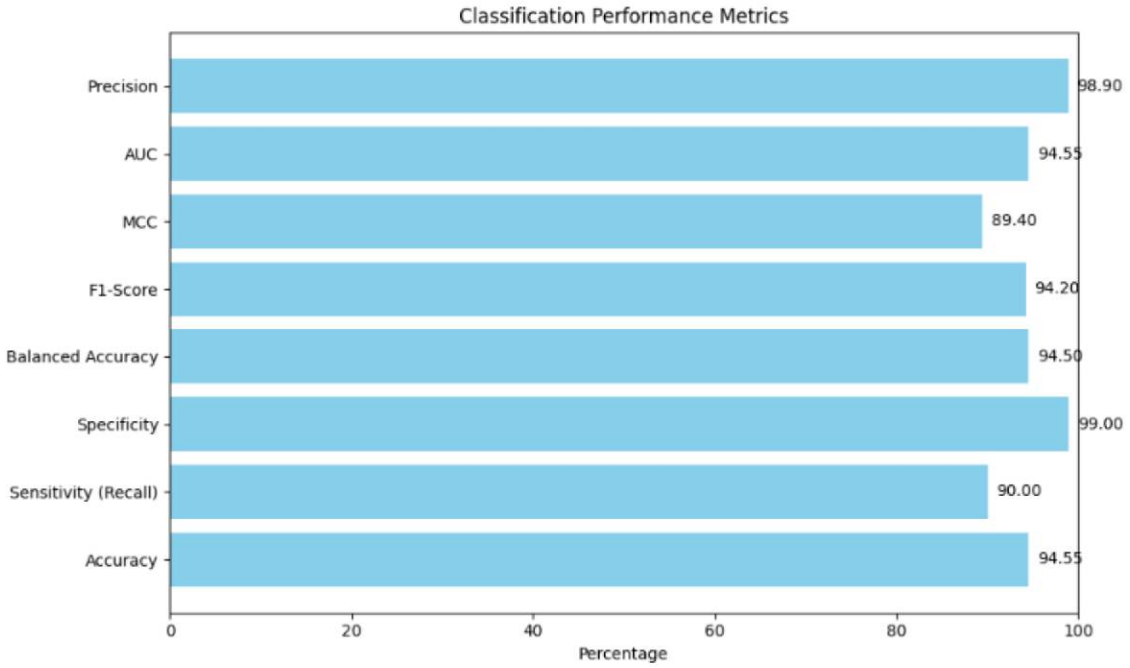

Figure 8. Results of the Proposed Model for Dataset 2

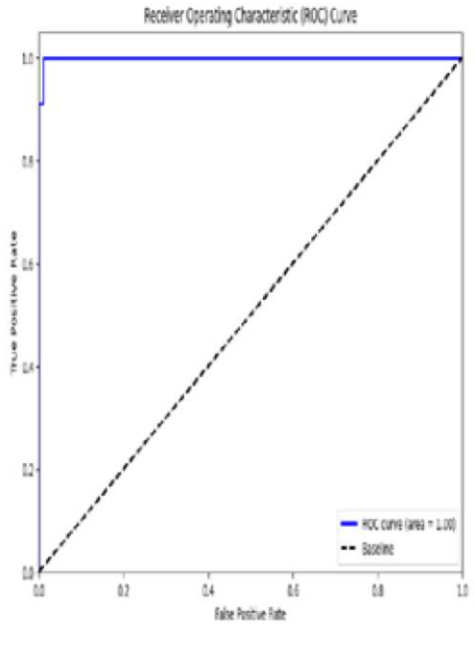

Figure 9. ROC Curve for Dataset 2

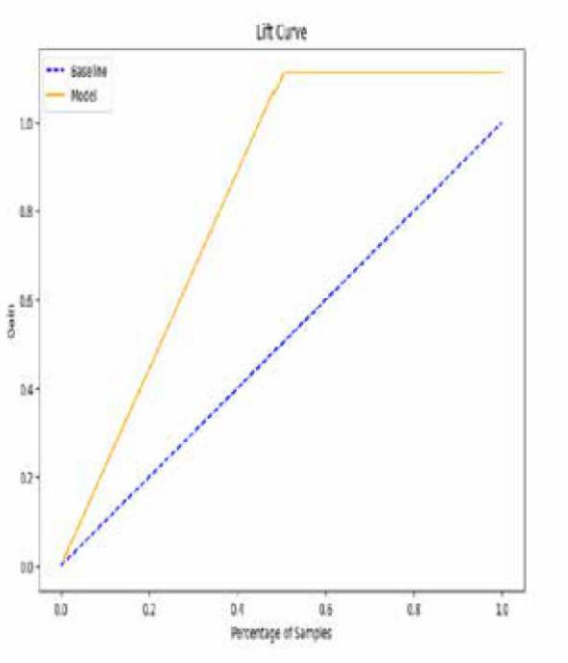

Figure 10. Lift Curve for Dataset 2

Figure 9 shows, a Receiver Operating Characteristic (ROC) Curve for Dataset 2. Similarly, figure 10 shows, Lift Curve for Dataset 2.

The proposed model demonstrates significant efficacy in classifying piRNA and non-piRNA sequences, as well as in distinguishing piRNA samples based on their function in target mRNA deadenylation. The high accuracy and other performance metrics obtained for both Dataset 1 and Dataset 2 indicate that the combined feature extraction techniques (DNC and TNC) and the machine learning approach are effective in capturing the relevant features and patterns within the data. To check the model performance, we use various metrics: Accuracy, Balance Accuracy, Precision, AUC (Area Under the Curve), MCC (Matthews Correlation Coefficient), F1-Score, Balanced Accuracy, Specificity, and Sensitivity (Recall).

For Dataset 1, which involves the classification of piRNA and non-piRNA sequences, the success rate of the suggested model is expressed in terms of Acc, Sn, Sp, BalanceAcc, F1-Score, MCC, AUC, and Precision are 98.59%, 98.24%, 98.94%, 98.59%, 98.59%, 97.19%, 98.59%, 98.94% respectively. The results suggest that the proposed feature extraction techniques and model architecture can effectively differentiate between piRNA and non-piRNA sequences.

For Dataset 2, which classifies piRNA samples based on their ability to instruct, target mRNA deadenylation. For Dataset 2, the suggested model success rate is expressed in the form of ACC, Sn, Sp, BalanceAcc, F1-Score, MCC, AUC, and Precision are 94.55%, 90.00%, 90.00%, 94.50%, 94.20%, 89.40%, 94.55%, 98.90% respectively.

Table 1. The suggested model is compared to the benchmark model [20].

Layers	Models	Acc (%)	Sp (%)	Sn (%)	Mcc (%)
1st Layer	**Proposed Model**	99.59	98.94	98.24	97.19
	2L-piRNAPred	89	87.5	90.4	0.779
2nd Layer	**Proposed Model**	94.55	90.00	90.00	89.40
	2L-piRNAPred	84.0	83.6	84.3	0.680

While the proposed model achieves excellent performance, several limitations should be acknowledged. First, the model's training and testing were conducted on a specific dataset, this could restrict the applicability of the findings to other piRNA datasets or different organisms. Second, the feature extraction methods used, while effective, are computationally intensive. This may pose challenges for scaling the model to very large datasets or real-time applications. Third, the current model focuses solely on sequence-based features. Incorporating additional types of data, such as secondary structure information or interaction networks, could potentially improve the model's performance.

6. Conclusions

This study introduces a robust and innovative model, aimed at enhancing the prediction and functional analysis of PIWI-interacting RNA (piRNA) sequences. By employing a combination of four advanced classifiers (Logistic Regression, Support Vector Classifier (SVC), Random Forest, and Gradient Boosting Classifier) the model effectively discriminates between piRNA and non-piRNA sequences, achieving a remarkable accuracy of 98.59% in this primary classification task.

Additionally, by utilizing features extracted through Dinucleotide Composition (DNC) and Trinucleotide Composition (TNC), the model further predicts the functional role of identified piRNA sequences, specifically their ability to instruct target mRNA deadenylation, with an accuracy of 94.55%. The high accuracy rates at both classification steps highlight the potential of the proposed model to contribute meaningfully to bioinformatics research and medical applications involving piRNAs.

Future research should aim to address the limitations mentioned in section 4 by validating the model on a wider range of datasets and exploring more efficient feature extraction techniques. Additionally, integrating multi-omics data, such as transcriptomics and proteomics, could enhance the understanding of piRNA function and their regulatory mechanisms.

References

1. Shoji, K., & Tomari, Y. (2024). A potential role of inefficient and non-specific piRNA production from the whole transcriptome. https://doi.org/10.1101/2024.01.24.577019.

2. Taverna, S., Masucci, A., & Cammarata, G. (2023). PIWI-RNAs Small Noncoding RNAs with Smart Functions: Potential Theranostic Applications in Cancer. Cancers, 15(15), 3912. https://doi.org/10.3390/cancers15153912.

3. Allikka Parambil, S., Li, D., Zelko, M., Poulet, A., & van Wolfswinkel, J. C. (2024). piRNA generation is associated with the pioneer round of translation in stem cells. *Nucleic Acids Research*, *52*(5), 2590–2608. https://doi.org/10.1093/nar/gkad1212.

4. Huang, X., Wang, C., Zhang, T., Li, R., Chen, L., Leung, K. L., Lakso, M., Zhou, Q., Zhang, H., & Wong, G. (2023). PIWI-interacting RNA expression regulates pathogenesis in a Caenorhabditis elegans model of Lewy body disease. Nature Communications, 14(1), 6137. https://doi.org/10.1038/s41467-023-41881-8.

5. Liu, L., Li, L., Zu, W., Jing, J., Liu, G., Sun, T., & Xie, Q. (2023). PIWI-interacting RNA-17458 is on cogenic and a potential therapeutic target in cervical cancer. Journal of Cancer, 14(9), 1648–1659. https://doi.org/10.7150/jca.83446.

6. Wang, J., Zhu, S., Meng, N., He, Y., Lu, R., & Yan, G.-R. (2019). NcRNA-Encoded Peptides or Proteins and Cancer. Molecular Therapy, 27(10), 1718–1725. https://doi.org/10.1016/j.ymthe.2019.09.001.

7. López-Jiménez, E., & Andrés-León, E. (2021). The Implications of ncRNAs in the Development of Human Diseases. Non-Coding RNA, 7(1), 17. https://doi.org/10.3390/ncrna7010017.

8. Wang, K., Wang, T., Gao, X., Chen, X., Wang, F., & Zhou, L. (2021). Emerging functions of piwi-interacting RNAs in diseases. Journal of Cellular and Molecular Medicine, 25(11), 4893–4901. https://doi.org/10.1111/jcmm.16466.

9. Weng, W., Li, H., & Goel, A. (2019). Piwi-interacting RNAs (piRNAs) and cancer: Emerging biological concepts and potential clinical implications. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1871(1), 160–169. https://doi.org/10.1016/j.bbcan.20.

10. Yu, Y., Xiao, J., & Hann, S. S. (2019). The emerging roles of PIWI-interacting RNA in human cancers. Cancer Management and Research, Volume 11, 5895–5909.

https://doi.org/10.2147/CMAR.S209300.

11. Fathizadeh, H., &Asemi, Z. (2019). Epigenetic roles of PIWI proteins and piRNAs in lung cancer. Cell & Bioscience, 9(1), 102. https://doi.org/10.1186/s13578-019-0368-x.

12. Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C., &Lordick, F. (2020). Gastric cancer. The Lancet, 396(10251), 635–648. https://doi.org/10.1016/S0140-6736(20)31288-5.

13. Feng, J. X., & Riddle, N. C. (2020). Epigenetics and genome stability. Mammalian Genome, 31(5–6), 181–195. https://doi.org/10.1007/s00335-020-09836-2.

14. M. Tahir, M. Hayat, S. Khan, and K. to Chong, "Prediction of Piwi-Interacting RNAs and Their Functions via Convolutional Neural Network," IEEE Access, vol. 9, pp. 54233–54240, 2021, doi: 10.1109/ACCESS.2021.3070083.

15. S. D. Ali, W. Alam, H. Tayara, and K. Chong, "Identification of Functional piRNAs Using a Convolutional Neural Network," IEEE/ACM Trans. Comput. Biol. Bioinform., pp. 1–1, 2020, doi: 10.1109/TCBB.2020.3034313.

16. T. Li, M. Gao, R. Song, Q. Yin, and Y. Chen, "Support Vector Machine Classifier for Accurate Identification of piRNA," Appl. Sci., vol. 8, no. 11, p. 2204, Nov. 2018, doi: 10.3390/app8112204.

17. Khan, S., Khan, M., Iqbal, N., Hussain, T., Khan, S. A., & Chou, K.-C. (2020). A Two-Level Computation Model Based on Deep Learning Algorithm for Identification of piRNA and Their Functions via Chou's 5-Steps Rule. International Journal of Peptide Research and Therapeutics, 26(2), 795–809. https://doi.org/10.1007/s10989-019-09887-3.

18. Khan, S., Khan, M., Iqbal, N., Li, M., & Khan, D. M. (2020). Spark-Based Parallel Deep Neural Network Model for Classification of Large Scale RNAs into piRNAs and Non-piRNAs. IEEE Access, 8, 136978–136991. https://doi.org/10.1109/ACCESS.2020.301150.

19. Ali, S. D., Tayara, H., & Chong, K. T. (2022). Identification of piRNA disease associations using deep learning. Computational and Structural Biotechnology Journal, 20, 1208–1217. https://doi.org/10.1016/j.csbj.2022.02.026.

20. Ali, S. D., Alam, W., Tayara, H., & Chong, K. (2020). Identification of Functional piRNAs Using a Convolutional Neural Network. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 1–1. https://doi.org/10.1109/TCBB.2020.3034313.

21. Zuo, Y., Zou, Q., Lin, J., Jiang, M., & Liu, X. (2020). 2lpiRNApred: A two-layered integrated algorithm for identifying piRNAs and their functions based on LFE-GM feature selection. RNA Biology, 17(6), 892–902. https://doi.org/10.1080/15476286.2020.1734382.

22. Wu, J., Liu, X., Han, L., Nie, H., Tang, Y., Tang, Y., Song, G., Zheng, L., & Qin, W. (2022). Small RNA sequencing revealed aberrant piRNA expression profiles in deciduas of recurrent spontaneous abortion patients. BIOCELL, 46(4), 1013–1023. https://doi.org/10.32604/biocell.2022.016744.

23. Khan, S., Khan, M., Iqbal, N., Amiruddin Abd Rahman, M., &Khalis Abdul Karim, M. (2022). Deep-piRNA: Bi-Layered Prediction Model for PIWI-Interacting RNA Using Discriminative Features.

Computers, Materials & Continua, 72(2), 2243–2258. https://doi.org/10.32604/cmc.2022.022901.

24. Cai, A., Hu, Y., Zhou, Z., Qi, Q., Wu, Y., Dong, P., Chen, L., & Wang, F. (2022). PIWI-Interacting RNAs (piRNAs): Promising Applications as Emerging Biomarkers for Digestive System Cancer. Frontiers in Molecular Biosciences, 9, 848105. https://doi.org/10.3389/fmolb.2022.848105.

KARACHI INSTITUTE OF ECONOMICS AND TECHNOLOGY

PUBLISHED BY
College of Computing and Information Sciences, KIET
cocis.kiet.edu.pk
kjcis.kiet.edu.pk
kjcis@kiet.edu.pk

Main Campus
PAF Airmen Academy
Korangi Creek
Tel: 35091114 - 7
Cell: 0336-2508284 - 85

City Campus
Shahra-e-Faisal Site
28-D, Block 6, P.E.C.H.S.
Tel: 34546872, 34532182
Cell: 0336-2508286 - 87

City Campus
North Nazimabad Site
F-103, Block F, Nazimabad
Tel: 36628381, 36679314
Cell: 0336-2444191-92

www.kiet.edu.pk