KICS The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express xxx (xxxx) xxx

www.elsevier.com/locate/icte

One-class models for validation of miRNAs and ERBB2 gene interactions based on sequence features for breast cancer scenarios

Juan Gutiérrez-Cárdenas[a,b], Zenghui Wang[a,*]

[a] *University of South Africa, Florida 1710, South Africa*
[b] *Universidad de Lima, Lima, Peru*

Received 10 November 2020; received in revised form 28 February 2021; accepted 8 March 2021
Available online xxxx

Abstract

One challenge in miRNA–genes–diseases interaction studies is that it is challenging to find labeled data that indicate a positive or negative relationship between miRNA and genes. The use of one-class classification methods shows a promising path for validating them. We have applied two one-class classification methods, Isolation Forest and One-class SVM, to validate miRNAs interactions with the ERBB2 gene present in breast cancer scenarios using features extracted via sequence-binding. We found that the One-class SVM outperforms the Isolation Forest model, with values of sensitivity of 80.49% and a specificity of 86.49% showing results that are comparable to previous studies. Additionally, we have demonstrated that the use of features extracted from a sequence-based approach (considering miRNA and gene sequence binding characteristics) and one-class models have proven to be a feasible method for validating these genetic molecule interactions.

Keywords: MiRNAs; Breast cancer; One-class models; Unsupervised learning

1. Introduction

MicroRNAs (miRNAs) are small molecules that belong to the group of non-coding RNA. Their importance is that they bind to genes for regulating their expression or degrade them. In some instances, this process is linked to the outcome of certain diseases, such as tumor growth or certain forms of cancers [1–5]. Moreover, the study of these units is essential because they act as biomarkers, which, if targeted and recognized, could help in the treatment and diagnosis of several diseases, see Fig. 1. Their study's complexity lies in that multiple miRNAs could interact with a targeted gene [3], and the task of finding or predicting miRNA target genes is not straightforward [6].

In past years, researchers found that by using Machine Learning techniques, it was probable to predict these interactions or to classify groups of miRNAs that are likely to increase or decrease the expression of specific genes [7]. The use of these computational models has been less time and resource consuming than their in-vitro experimentations [8]. However, in some models, e.g., supervised learning for classification, a set of labeled data should be used to train the model to discretize when one sample belongs to one class or another. Although, on many occasions, it is impossible to find labeled data; or the data from one class is too scarce [7,9–11] that might lead to a scenario of unbalanced data.

There have been numerous studies of miRNA and RNA interactions that point to the scarcity of validated data. For example, [10] described the problem related to predicting miRNA hairpins derived from mRNA hairpins structures. The theoretical basis was that miRNA hairpins, with a length of 21 to 25 nucleotides, are obtained from RNA hairpins of 60 to 90 nucleotides long. The difficulty in this scenario was that the dataset of available miRNA hairpins was moderately limited; therefore, using a two-class classifier was not possible, being a feasible option utilizing a one-class model. In this case, one can find that there existed two main issues, the first one related to the difficulty of finding labeled data due to the presence of not validated or weak miRNAs and mRNAs interactions, and second, we might end up with an unbalanced dataset. In either way, there is a limiting factor in the direct application of supervised classifier models.

* Corresponding author.
E-mail address: wangzengh@gmail.com (Z. Wang).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2021.03.001

Please cite this article as: J. Gutiérrez-Cárdenas and Z. Wang, One-class models for validation of miRNAs and ERBB2 gene interactions based on sequence features for breast cancer scenarios, ICT Express (2021), https://doi.org/10.1016/j.icte.2021.03.001.

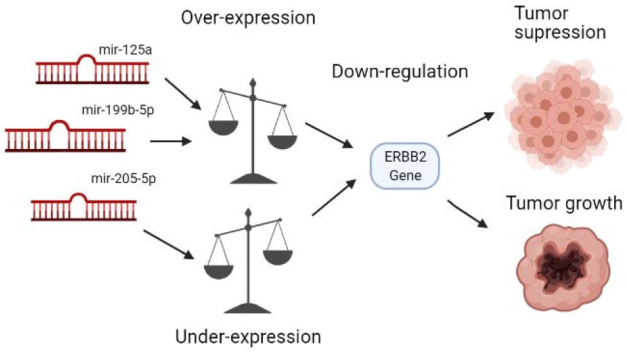

Fig. 1. Schematic of the influence of miRNAs when interacting with the ERBB2 gene, depending on if they are over-expressed or under-expressed, they could suppress or proliferate malignant tumoral cells [12–14].
Source: Image created with Biorender.com.

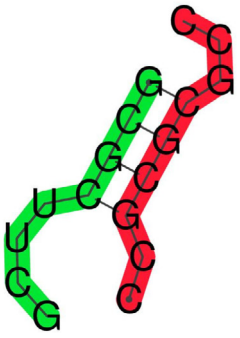

Fig. 2. Potential miRNA and gene sequences binding features like energy released, complementarity score, and coupling region can be used as input features for the proven models.

Novelty detection or one-class classification is the implementation of computational models that try to find a complementary class by using a dataset in which only one class is present. T. Spinosa and de Carvalho's work used this type of novelty detection in the field of Bioinformatics [15]. In their work, they tested a One-class SVM for detecting ALL-B leukemia samples in a dataset composed of non-differentiable classes of the disease, such as ALL-B, ALL-T, and AML. Their dataset chosen consisted of limited records, ranging from 17, 27, or 30 registers per each leukemia class, but with a vast number of attributes of approximately 7000 features. Considering their results, for the AML type, the obtained accuracy was approximately 85% for the regular class and 60% for the class containing most outliers. We can also mention the work of [11] and Yousef et al. [7], where the authors tried different types of one-class models to predict the presence of miRNAs by using features such as the secondary structure or gene sequence information. The authors supported the use of one class model because it is usually a complicated and a biased procedure to obtain negative data based on the positive miRNAs class. For validating their proposal, they predicted a set of miRNAs that were related to the Epstein Barr Virus. They obtained, by using One-class SVM, values of sensitivity of 72% and 99% of specificity using the secondary structure features in Human data. However, information about the hyperparameters tuning of this model was not explicitly mentioned in the research.

Concerning the use of Machine Learning classifiers to validate miRNAs involved in Breast Cancer scenarios, we can mention the work of [5]. This study used a dataset obtained from the National Cancer Institute's Genomic Data Commons Data Portal [16], which contained samples from 1207 patients with 1881 miRNA features. The samples contained 1103 tumoral samples, seven metastatic, and 104 healthy ones. In this research, we observed a disbalance between the number of patient records and the number of features, being the latter one which outperforms in quantity to the patients' records. For that reason, the authors proposed the use of feature selection techniques such as Information Gain, Chi-Squared, or Least Absolute Shrinkage and Selection Operator (Lasso) for choosing the most relevant miRNAs which served as features for an SVM and Random Forest classifiers [5].

In the study of miRNA and mRNA interactions, two significant approaches exist: one relates to the study of their characteristics of the sequences involved in the binding, such as pairing sites, accessibility, or evolutionary conservation data; and the second approach considers the negative correlation present in the expression levels of miRNA and mRNAs [17]. In this research, we will be using the sequence-based technique, see Fig. 2. We found a limited set of studies regarding the validation of miRNAs–mRNA interactions in cancer scenarios using features obtained from the sequence interactions. For example, we found the use of unsupervised models in the studies of [11] and Yousef et al. [7]. Additionally, we encountered the research of [9], where they proposed a mix of supervised and unsupervised techniques for miRNA target prediction, but the final results obtained were advocated towards the use of SVM supervised binary classifiers.

Additionally, a drawback in the study of miRNA and gene interaction is that it is not straightforward to obtain samples from the positive (or negative) class in the right proportion [9, 18]; having the risk of ending up with an imbalanced dataset classification scenario. Still, these situations influence the use of binary classification models by utilizing a mix of techniques to obtain the negative interactions class.

In this research, we used two well-known techniques: Isolation Forest [19] and One-class SVM [20], for our one class model classification. Concerning the miRNA and gene interactions' sequence features present in these associations, we will use the data available from mirWalk [21,22], in contraposition to the use of gene expression levels as in [7,9–11], which could lead to cases of unbalanced data. mirWalk is based on the sequence-based approach for retrieving miRNAs interactions being the number of features available to be manageable for our purposes. For validation purposes, by using metrics such as precision or specificity, we had to manipulate our dataset for obtaining a small subset of the negative class. To accomplish this, we decided to test an approach that consisted of using those interactions between miRNAs and mRNAs, which present weak interactions supported by the available studies or do not have supporting evidence, e.g., obtained by prediction methods, but not validated by wet-lab experiments. For experimentation purposes, the miRNAs we selected interact with the

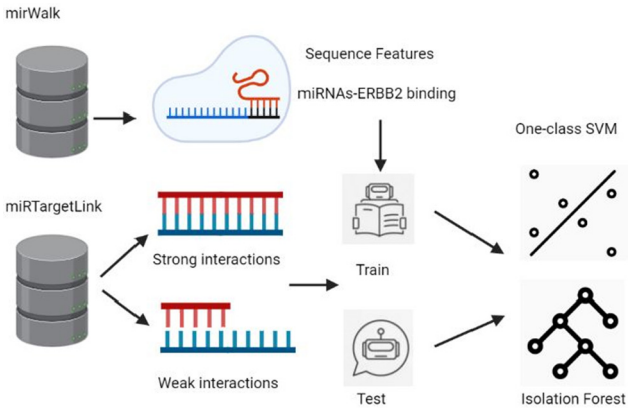

Fig. 3. Depiction of the methodology followed. The binding sequence features are extracted from mirWalk [36], and the evidence of strong and weak interactions is extracted from miRTargetLink [23], both serve as inputs for the one-class models.

ERBB2 gene, which is a gene that could lead to breast cancer scenarios when altered in its expression. The data obtained for our experiments was validated from miRNAs gene interaction tools, such as mirTargetLink [23].

2. Materials and methods

2.1. Methodology

The steps we have followed entail two main parts. The first one is related to the miRNA–mRNA sequence binding characteristics extracted from mirWalk [36], and the second step is to gather a set of miRNA that interacts with the ERBB2 gene; we obtained this set from miRTargetLink [23]. From this set, we extracted a subset of them to form a small validation subset (25% of the total miRNAs extracted) that will serve as a negative class, see Fig. 3. The criteria for choosing the elements of this validation subset was to consider those miRNAs that had weak or no interactions validated by the literature. This schema will allow us to obtain the necessary metrics, such as sensitivity and specificity, to validate the interactions between miRNAs and the ERBB2 gene.

After we got these subsets, we will apply an Isolation forest model for checking the presence of outliers in our dataset considering all the samples. These outliers would represent mRNA and miRNA interactions that are weak among these components and match the miRNAs interactions we chose for the validation subset. For comparison purposes, we will apply a One-class SVM classifier also to check the presence of outliers. It is valuable to mention that we will consider these outliers are weak interactions between mRNA and miRNA components. With the metrics obtained from both one-class models, we will compare them using a confusion matrix and the metrics of precision, specificity, sensitivity, and F1-score. Also, it is valuable to keep in mind that these models only need one class for training, and any other data found (outliers) that it is not enclosed in the boundaries could be considered anomalies. The creation of validation or test subset was only to apply the metrics mentioned above. A similar methodology can be found in the work of Eude and Chang [37].

As an additional detail, it is valuable to mention that both one-class models will be fitted or trained in the strong interaction datasets and then tested in the created validation subset to find outliers. We hypothesize that there should be a scarce presence of outliers in the training set, while in the second dataset that contains weak interactions, we hope to find more than half of the presence of outliers. We will validate our results by analyzing a generated confusion matrix with their metrics of accuracy, precision, recall, and F1-score and check some of the outliers found via a literature review.

2.2. Dataset extraction and samples division in classes

We will use the data downloaded from the mirWalk [36] web page for our experimental part. mirWalk is a dataset that enables to download miRNA–mRNA interactions, both predicted and validated by wet-lab experiments. The format of their downloaded data is in CSV, and it gives us a set of attributes that appear in sequence interactions between miRNA and mRNA as described in the research works of Stitch et al. [38] and [21,22,39]. Most of the features that appear in mirWalk were extracted from the TarPmirR software [40].

As a sample gene for our experimentation purposes, we decided to choose the ERBB2 gene, a molecule that appears in distinct types of breast cancer scenarios. Regarding the miRNAs that interact with the ERBB2, we have downloaded a list of miRNA with strong or weak evidence and predicted interactions from miRTargetLink Human [23]. In some cases, we could not find the miRNA–mRNA interaction because the gene name ERBB2 was not present in the file. In this case, we searched for the ERBB2 gene's aliases by using the GeneCards database (https://www.genecards.org/). The complete list of miRNAs, evidence support type according to [23], gene name or alias, and literature reference that points to its relationship with the gene of study is in Table 1.

From this point, we proceed to analyze if there were outliers in our dataset using a boxplot diagram; the results from this examination confirmed the presence of these anomalies. At this point, it is worthy of mentioning that those miRNA and mRNA interactions with weak evidence would be considered as our subset of artificial data. This data is sometimes be generated using the real data and checking if it is possible to use a one-class classifier to detect those data points that would not belong to our main class, and per se could be considered outliers.

Considering the features obtained from mirWalk, we decided to work with the quantitative features and drop the irrelevant features. The features not selected were: mirnaid, refseqid, genesymbol, seed (because all the obtained values were set up as one), position (it gave us the position of the longest consecutive pairs [40]. The values could be 3 UTR (Untranslated region), 5 UTR or CDS (Coding sequence), it was a categorical value), validated (It contains all validated interactions that are present in mirTarBase [41], some data was missing), TargetScan, and miRDB (both of them pointed out if

Table 1
miRNA and ERBB2 interactions with type of evidence supporting according to miRTargetLink [23]. The data is divided in strong and weak interactions.

miRNA	Evidence	miRNA	Evidence	miRNA	Evidence
hsa-miR-125a-5p	Strong [14,24]	hsa-miR-323b-5p	Strong [25]	hsa-miR-124-3p	Weak [26]
hsa-miR-125b-5p	Strong [27]	hsa-miR-331-3p	Strong [28]	hsa-miR-326	NA [29]
hsa-miR-134-5p	Strong [30]	hsa-miR-375-3p	Strong [31]	hsa-miR-4326	Weak [32]
hsa-miR-193a-5p	Strong [33]	hsa-miR-375-5p	Strong [31]	hsa-miR-670-3p	NA
hsa-miR-199b-5p	Strong [13]	hsa-miR-498-3p	Strong [34]	hsa-miR-6739-3p	Weak
hsa-miR-205-5p	Strong [12]	hsa-miR-498-5p	Strong [34]		
hsa-miR-25-3p	Strong [1]	hsa-miR-541-3p	Strong [35]		
hsa-miR-552-3p	Strong [4]				

Table 2
Confusion matrix for the Isolation Forest and One-Class SVM models.

	Isolation Forest		One-Class SVM	
	True positive	True negative	True positive	True negative
Predicted positive	87	18	99	5
Predicted negative	36	19	24	32

Table 3
Metrics obtained from the Isolation Forest and One Class SVM.

Model	Accuracy	Sensitivity	Specificity	F1-Score
Isolation Forest	66.25%	70.73%	51.35%	76.32%
One Class SVM	81.88%	80.49%	86.49%	87.22%

Table 4
Selected Hyperparameters for the Isolation forest and One Class SVM.

Hyperparameter	Isolation Forest	One Class SVM
Number of trees	20	
Number of features	70%	
Number of samples	30	–
Bootstrap	True	–
Contamination	True	–
Kernel	–	RBF
ν	–	0.17163
γ	–	0.1

the data was validated with some of these databases). We will perform tests using all the remaining features and a subset of the features for validation purposes, as mentioned, for example, in the work of [7,11]. Additionally, we had to normalize our data, and for this purpose, we use a standard scalar which empirically is the same as the z-score normalization with zero degrees of freedom.

2.3. One-class models application and hyperparameters tuning

We decided to apply first the model of Isolation forest to check the presence of outliers. The metric used for hyperparameter tuning was the weighted F1-score. After testing a list of probable hyperparameters, we ended up with the results showed in Table 2. We set it up in 30% regarding the contamination level, which we knew beforehand that it was approximately the number of miRNAs and ERBB2 interactions with weak support from the literature. We followed a similar procedure for the use of One-class SVM but using the Grid Search algorithm in this case with cross-validation of ten folds.

We had to perform two additional modifications before using the Grid Search algorithm. First, we needed to have a sort of labeling for our outputs in the training and testing sets for fitting our model. For that reason, we put a value of +1 to those samples that presented a strong verified miRNA and mRNA interaction and a value of -1 for those that presented a weak verified interaction. At this point, it is valuable to remember that One class SVM works with only one dataset that has all the elements of the same class. If we would have an exact division of knowing which class is, for example, positive and which one is negative, then we could have reduced our solution to a two-class classifier and use a supervised technique. The second modification we made was choosing the best metric for the Grid Search algorithm's scoring function. In this case, we were not able to use precision or accuracy because we were dealing with an unsupervised model, so we decided to choose a model based on the F1-score, which relates to the precision and recall metrics, with a weighted average of the obtained results from each of the outputs derived from each one of the cross-validation cases. For the scoring function to pass to our Grid Search algorithm, we have used the F1-score [42], which is suitable for binary classification with imbalanced data. After applying the Grid Search algorithm and validating via a manual selection, we found a list of the best hyperparameters; see Table 4.

About our dataset division, we divide it into a training dataset, with 123 miRNA and mRNA interactions, and a validation or test dataset with 37 interactions. In a first moment, we apply the One-class SVM to the dataset that contained only the train set for outliers detection to detect novelties within it, and then in a second scenario, we train our model with approximately 70% of samples of all the data (positive validated training data), and then applied this fitted model to the validation or test set. This last procedure allowed us to obtain a confusion matrix for validation purposes and calculate our model's metrics.

3. Results

3.1. Comparison of isolation forest vs one-class SVM

In the confusion matrices in Table 2, corresponding to Isolation Forest and One-class SVM models, the True Positives

represented those miRNAs that interact with the ERBB2 gene, while the True Negatives are those miRNAs in which there is no strong evidence of their interaction with the gene ERBB2. Concerning the results of the metrics shown in Table 3, we believe that the Accuracy use as a sole metric for comparing these models is relatively inaccurate. The reason is that an analysis of the rate of true positive and false negatives should also be considered for obvious reasons, i.e., medical systems. Additionally, we considered the F1-score as a well-suited metric in cases like the described when we can have an imbalanced dataset. By comparing our both one class models, Isolation and One-class SVM, we found that the SVM model ruled out the Isolation Forest, with values of 81.88% in terms of Accuracy, and an F1-score of 87.22% compared with the values obtained of 66.25% 76.32% for the mentioned metrics to name a few.

4. Discussion and conclusions

In this paper, we have used a One-class SVM for finding miRNA and mRNA interactions when one has only a unique set of data to extract these relationships, and it is not possible to find a set of genes that could act as a second class to be used as in regularly supervised classifiers. Until the moment of writing the present article, we could not find evidence of the use of one-class classifiers to study miRNA and mRNA interactions by using features of the mRNA–miRNA sequences for breast cancer scenarios. Tran et al. [21] and Yousef et al. [11], made a closer proposal, where one-class classifiers were used to predict miRNA hairpins or miRNAs prediction by using sequence characteristics in contraposition to gene expression data.

We also found that, even though the one-class models are oriented to anomaly detection in imbalanced datasets and for unsupervised learning, some authors like [21] also use a subset of the training data to be converted into test data. In the end, this trick is useful for applying metrics such as precision or F1-score for validating our results; and it seems helpful when there in not a straightforward method to generate this information Yousef et al. [7,11]. A similar approach was used in our proposal because we used as testing data samples that had weak verified breast cancer correlations considered the literature reviewed. In the end, this was considered our negative samples dataset.

The results obtained in similar scenarios [11] found that the one-class models tested have a higher sensitivity and low specificity compared to the common two-class supervised models. The experiments they performed was to detect miRNAs in the Epstein Barr Virus. The results of the sensitivity criteria for the different one-class models were in the range of 82% approximately, with no information about the specificity metric. However, our one-class SVM model obtained values of 80.49% in sensitivity and 86.49% in specificity, giving more stable results.

We have shown that it is possible to obtain relatively good accuracy and F1-scores of 81.88% and 87.22%. Respectively, that allowed us to find interesting relationships between miRNAs and an ERBB oncogene, which could be the initial point for further studies. When we examined the results after applying the One-class SVM to the training set, we found approximately 19.51% of outliers; even though they had strong supporting evidence, we decided to find in the literature what was occurring with these samples. For example, we found that the hsa-miR-25-3p in our dataset has nine different ways of interacting with the ERBB2 gene, but only one of them was a false negative. A similar situation occurred with the hsa-miR-125a-5p in which one interaction out of ten available, was marked as a false negative. One point to consider is that a miRNA could bind to different sections of the mRNA or even given other values of some features as free energy, stem-loop, or flanking conservation; that could have influenced the final classification of these miRNAs. Maybe the generation of a voting system, like the one used in a KNN model, would help obtain the miRNA's final classification as an outlier.

Concerning the results obtained by using Isolation Forest compared with One-class SVMs, we could argue that the former's relatively low performance is related to how Random Forest performs its classification. As we know, Random Forests tend to divide the space in sort of rectangular sections, while an SVM model could have smooth separating spaces via the use of different types of kernels. This situation in which an SVM model performs better than RF in genomic data was mentioned in [43], and it would be interesting to do more research in this direction.

As a concluding remark, this study's importance is to motivate the study and the use of unsupervised learning techniques along with datasets such as mirWalk [36] to find interesting miRNA and mRNA interactions. In contraposition to supervised techniques in those cases where labeled data is not feasible or the process of distinction between classes is a hard one.

CRediT authorship contribution statement

Juan Gutiérrez-Cárdenas: Conceptualization, Methodology, Data analysis, Writing - original draft. **Zenghui Wang:** Confirmed results, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported partially by South African National Research Foundation Grants (No. 112108 & 112142) and Tertiary Education Support Programme (TESP) of South African ESKOM.

References

[1] H. Chen, H. Pan, Y. Qian, W. Zhou, X. Liu, MiR-25-3p promotes the proliferation of triple negative breast cancer by targeting BTG2, Mol Cancer. 17 (2018) 4, http://dx.doi.org/10.1186/s12943-017-0754-0.

[2] H.-Y. Loh, B.P. Norman, K.-S. Lai, N.M.A.N.Abd. Rahman, N.B.M. Alitheen, M.A. Osman, The regulatory role of microRNAs in breast cancer, IJMS 20 (2019) 4940, http://dx.doi.org/10.3390/ijms20194940.

[3] P. Paul, A. Chakraborty, D. Sarkar, M. Langthasa, M. Rahman, M. Bari, RK.S. Singha, A.K. Malakar, S. Chakraborty, Interplay between miRNAs and human diseases, J. Cell Physiol. 233 (2018) 2007–2018, http://dx.doi.org/10.1002/jcp.25854.

[4] A. Penyige, É. Márton, B. Soltész, M. Szilágyi-Bónizs, R. Póka, J. Lukács, L. Széles, B. Nagy, Circulating miRNA profiling in plasma samples of ovarian cancer patients, IJMS 20 (2019) 4533, http://dx.doi.org/10.3390/ijms20184533.

[5] O. Rehman, H. Zhuang, A. Muhamed Ali, A. Ibrahim, Z. Li, Validation of miRNAs as breast cancer biomarkers with a machine learning approach, Cancers. 11 (2019) 431, http://dx.doi.org/10.3390/cancers11030431.

[6] X. Yan, T. Chao, K. Tu, Y. Zhang, L. Xie, Y. Gong, J. Yuan, B. Qiang, X. Peng, Improving the prediction of human microRNA target genes by using ensemble algorithm, FEBS Lett. 581 (2007) 1587–1593, http://dx.doi.org/10.1016/j.febslet.2007.03.022.

[7] M. Yousef, N. Najami, W. Khalifav, A comparison study between one-class and two-class machine learning for MicroRNA target detection, JBiSE 03 (2010) 247–252, http://dx.doi.org/10.4236/jbise.2010.33033.

[8] K. Zheng, Z.-H. You, L. Wang, Y. Zhou, L.-P. Li, Z.-W. Li, MLMDA: a machine learning approach to predict and validate MicroRNA–disease associations by integrating of heterogenous information sources, J. Transl. Med. 17 (2019) 260, http://dx.doi.org/10.1186/s12967-019-2009-x.

[9] N. Sedaghat, M. Fathy, M.H. Modarressi, A. Shojaie, Combining supervised and unsupervised learning for improved miRNA target prediction, IEEE/ACM Trans. Comput. Biol. Bioinf. (2018) 1, http://dx.doi.org/10.1109/TCBB.2017.2727042.

[10] D.H. Tran, T.H. Pham, K. Satou, T.B. Ho, Prediction of microRNA hairpins using one-class support vector machines, in: 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, IEEE, Shanghai, China, 2008, pp. 33–36, http://dx.doi.org/10.1109/ICBBE.2008.15.

[11] M. Yousef, S. Jung, L.C. Showe, M.K. Showe, Learning from positive examples when the negative class is undetermined- microRNA gene identification, Algorithms Mol. Biol. 3 (2008) 2, http://dx.doi.org/10.1186/1748-7188-3-2.

[12] A. De Cola, S. Volpe, M.C. Budani, M. Ferracin, R. Lattanzio, A. Turdo, D. D'Agostino, E. Capone, G. Stassi, M. Todaro, C. Di Ilio, G. Sala, M. Piantelli, M. Negrini, A. Veronese, V. De Laurenzi, miR-205-5p-mediated downregulation of ErbB/HER receptors in breast cancer stem cells results in targeted therapy resistance, Cell Death Dis. 6 (2015) e1823, http://dx.doi.org/10.1038/cddis.2015.192.

[13] C. Fang, Y. Zhao, B. Guo, MiR-199b-5p targets HER2 in breast cancer cells, J. Cell. Biochem. 114 (2013) 1457–1463, http://dx.doi.org/10.1002/jcb.24487.

[14] L. Ninio-Many, E. Hikri, T. Burg-Golani, S.M. Stemmer, R. Shalgi, I. Ben-Aharon, miR-125a induces HER2 expression and sensitivity to trastuzumab in triple-negative breast cancer lines, Front. Oncol. 10 (2020) 191, http://dx.doi.org/10.3389/fonc.2020.00191.

[15] E.J. Spinosa, Andre de Carvalho, SVMs for novel class detection in Bioinformatics, in: Brazilian Workshop on Bioinformatics, 2004, pp. 81–88.

[16] The website of the national cancer institute, 2020, https://www.cancer.gov, (accessed 20 September 2020).

[17] V.V. Pham, J. Zhang, L. Liu, B. Truong, T. Xu, T.T. Nguyen, J. Li, T.D. Le, Identifying miRNA-mRNA regulatory relationships in breast cancer with invariant causal prediction, BMC Bioinformatics 20 (2019) 143, http://dx.doi.org/10.1186/s12859-019-2668-x.

[18] I. Irigoien, B. Sierra, C. Arenas, Towards application of one-class classification methods to medical data, Sci. World J. 2014 (2014) 1–7, http://dx.doi.org/10.1155/2014/730712.

[19] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation forest, in: 2008 Eighth IEEE International Conference on Data Mining, IEEE, Pisa, Italy, 2008, pp. 413–422, http://dx.doi.org/10.1109/ICDM.2008.17.

[20] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution, Neural Comput. 13 (2001) 1443–1471, http://dx.doi.org/10.1162/089976601750264965.

[21] H. Dweep, C. Sticht, P. Pandey, N. Gretz, miRwalk – database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes, J. Biomed. Inform. 44 (2011) 839–847, http://dx.doi.org/10.1016/j.jbi.2011.05.002.

[22] H. Dweep, C. Sticht, N. Gretz, In-silico algorithms for the screening of possible microRNA binding sites and their interactions, CG 14 (2013) 127–136, http://dx.doi.org/10.2174/1389202911314020005.

[23] M. Hamberg, C. Backes, T. Fehlmann, M. Hart, B. Meder, E. Meese, A. Keller, miRtargetlink—miRNAs, genes and interaction networks, IJMS 17 (2016) 564, http://dx.doi.org/10.3390/ijms17040564.

[24] D.T. Vo, N.K. Karanam, L. Ding, D. Saha, J.S. Yordy, U. Giri, J.V. Heymach, M.D. Story, miR-125a-5p functions as tumor suppressor microRNA and is a marker of locoregional recurrence and poor prognosis in head and neck cancer, Neoplasia. 21 (2019) 849–862, http://dx.doi.org/10.1016/j.neo.2019.06.004.

[25] B.M. Sugita, S.R. Pereira, R.C. de Almeida, M. Gill, A. Mahajan, A. Duttargi, S. Kirolikar, P. Fadda, R.S. de Lima, C.A. Urban, K. Makambi, S. Madhavan, S.M. Boca, Y. Gusev, I.J. Cavalli, E.M.S.F. Ribeiro, L.R. Cavalli, Integrated copy number and miRNA expression analysis in triple negative breast cancer of latin American patients, Oncotarget 10 (2019) 6184–6203, http://dx.doi.org/10.18632/oncotarget.27250.

[26] Y. Wang, L. Chen, Z. Wu, M. Wang, F. Jin, N. Wang, X. Hu, Z. Liu, C.-Y. Zhang, K. Zen, J. Chen, H. Liang, Y. Zhang, X. Chen, miR-124-3p functions as a tumor suppressor in breast cancer by targeting CBL, BMC Cancer. 16 (2016) 826, http://dx.doi.org/10.1186/s12885-016-2862-4.

[27] M. Ferracin, C. Bassi, M. Pedriali, S. Pagotto, L. D'Abundo, B. Zagatti, F. Corrà, G. Musa, E. Callegari, L. Lupini, S. Volpato, P. Querzoli, M. Negrini, miR-125b targets erythropoietin and its receptor and their expression correlates with metastatic potential and ERBB2/HER2 expression, Mol Cancer. 12 (2013) 130, http://dx.doi.org/10.1186/1476-4598-12-130.

[28] D. Zhao, Y. Sui, X. Zheng, miR-331-3p inhibits proliferation and promotes apoptosis by targeting HER2 through the PI3K/Akt and ERK1/2 pathways in colorectal cancer, Oncol. Rep. 35 (2016) 1075–1082, http://dx.doi.org/10.3892/or.2015.4450.

[29] Z. Ghaemi, B.M. Soltani, S.J. Mowla, MicroRNA-326 functions as a tumor suppressor in breast cancer by targeting ErbB/PI3k signaling pathway, Front. Oncol. 9 (2019) 653, http://dx.doi.org/10.3389/fonc.2019.00653.

[30] J.-Y. Pan, F. Zhang, C.-C. Sun, S.-J. Li, G. Li, F.-Y. Gong, T. Bo, J. He, R.-X. Hua, W.-D. Hu, Z.-P. Yuan, X. Wang, Q.-Q. He, D.-J. Li, miR-134: A human cancer suppressor? Mol. Ther.- Nucleic Acids 6 (2017) 140–149, http://dx.doi.org/10.1016/j.omtn.2016.11.003.

[31] Z.-Y. Shen, Z.-Z. Zhang, H. Liu, E.-H. Zhao, H. Cao, miR-375 inhibits the proliferation of gastric cancer cells by repressing ERBB2 expression, Exp. Therapeutic Med. 7 (2014) 1757–1761, http://dx.doi.org/10.3892/etm.2014.1627.

[32] A.D. Martinez-Gutierrez, D. Cantú de León, O. Millan-Catalan, J. Coronel-Hernandez, A.D. Campos-Parra, F. Porras-Reyes, A. Exayana-Alderete, C. López-Camarillo, N.J. Jacobo-Herrera, R. Ramos-Payan, C. Pérez-Plasencia, Identification of miRNA master regulators in breast cancer, Cells 9 (2020) 1610, http://dx.doi.org/10.3390/cells9071610.

[33] F. Xie, S. Hosany, S. Zhong, Y. Jiang, F. Zhang, L. Lin, X. Wang, S. Gao, X. Hu, MicroRNA-193a inhibits breast cancer proliferation and metastasis by downregulating WT1, PLoS One 12 (2017) e0185565, http://dx.doi.org/10.1371/journal.pone.0185565.

[34] N. Matamala, M.T. Vargas, R. González-Cámpora, J.I. Arias, P. Menéndez, E. Andrés-León, K. Yanowsky, A. Llaneza-Folgueras, R. Miñambres, B. Martínez-Delgado, J. Benítez, MicroRNA deregulation in triple negative breast cancer reveals a role of miR-498 in regulating BRCA1 expression, Oncotarget 7 (2016) 20068–20079, http://dx.doi.org/10.18632/oncotarget.7705.

[35] R.M. Sareyeldin, I. Gupta, I. Al-Hashimi, H.A. Al-Thawadi, H.F. Al Farsi, S. Vranic, A.-E. Al Moustafa, Gene expression and miRNAs profiling: Function and regulation in human epidermal growth factor receptor 2 (HER2)-positive breast cancer, Cancers 11 (2019) 646, http://dx.doi.org/10.3390/cancers11050646.

[36] mirWalk database, 2020, http://mirwalk.umm.uni-heidelberg.de/resources/, (accessed 20 September 2020).

[37] T. Eude, C. Chang, One-class SVM for biometric authentication by keystroke dynamics for remote evaluation: One-class SVM for biometric authentication by keystroke dynamics for remote evaluation, Comput. Intell. 34 (2018) 145–160, http://dx.doi.org/10.1111/coin.12122.

[38] C. Sticht, C. De La Torre, A. Parveen, N. Gretz, miRwalk: An online resource for prediction of microRNA binding sites, PLoS One 13 (2018) e0206239, http://dx.doi.org/10.1371/journal.pone.0206239.

[39] H. Dweep, N. Gretz, C. Sticht, miRwalk database for miRNA–target interactions, in: M.L. Alvarez, M. Nourbakhsh (Eds.), RNA Mapping, Springer New York, New York, NY, 2014, pp. 289–305, http://dx.doi.org/10.1007/978-1-4939-1062-5_25.

[40] J. Ding, X. Li, H. Hu, TarPmiR: a new approach for microRNA target site prediction, Bioinformatics. 32 (2016) 2768–2775, http://dx.doi.org/10.1093/bioinformatics/btw318.

[41] H.-Y. Huang, Y.-C.-D. Lin, J. Li, K.-Y. Huang, S. Shrestha, H.-C. Hong, Y. Tang, Y.-G. Chen, C.-N. Jin, Y. Yu, J.-T. Xu, Y.-M. Li, X.-X. Cai, Z.-Y. Zhou, X.-H. Chen, Y.-Y. Pei, L. Hu, J.-J. Su, S.-D. Cui, F. Wang, Y.-Y. Xie, S.-Y. Ding, M.-F. Luo, C.-H. Chou, N.-W. Chang, K.-W. Chen, Y.-H. Cheng, X.-H. Wan, W.-L. Hsu, T.-Y. Lee, F.-X. Wei, H.-D. Huang, miRtarbase 2020: updates to the experimentally validated microRNA–target interaction database, Nucleic Acids Res. (2019) http://dx.doi.org/10.1093/nar/gkz896, gkz896.

[42] C.C. Aggarwal, Outlier Analysis, Springer International Publishing, Cham, 2017, http://dx.doi.org/10.1007/978-3-319-47578-3.

[43] A. Statnikov, C.F. Aliferis, Are random forests better than support vector machines for microarray-based cancer classification? in: AMIA Annual Symposium Proceedings, n.d., pp. 686–690.