Laryngeal Cancer Diagnosis via miRNA-based Decision Tree Model

Aarav Arora[a], Igor F Tsigelny[bcde*], Valentina L Kouznetsova[bce]

[a]REHS Program, San Diego Supercomputer Center, UC San Diego, La Jolla, CA, USA

[b]San Diego Supercomputer Center, UC San Diego, La Jolla, CA, USA

[c]BiAna, La Jolla, CA, USA

[d]Department of Neurosciences, UC San Diego, La Jolla, CA, USA

[e]CureScience, San Diego, CA, USA

***Corresponding Author Email: itsigeln@ucsd.edu**

Acknowledgements

The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga

Abstract

Purpose Laryngeal cancer (LC) is the most common head and neck cancer, which often goes undiagnosed due to the expensiveness and inaccessible nature of current diagnosis methods. Many recent studies have shown that microRNAs (miRNAs) are crucial biomarkers for a variety of cancers.

Methods In this study, we create a decision tree model for the diagnosis of laryngeal cancer using a calculated miRNAs' attributes, such as sequence-based characteristics, predicted miRNA target genes, and gene pathways. This series of attributes is extracted from both differentially expressed blood-based miRNAs in laryngeal cancer and random, non-associated with cancer miRNAs.

Results Several machine-learning (ML) algorithms were tested in the ML model, and the Hoeffding Tree (HT) classifier yields the highest accuracy (86.8%) in miRNAs-based recognition of laryngeal cancer. Furthermore, HT-based model is validated with the independent laryngeal cancer datasets and can

accurately diagnose laryngeal cancer with 86% accuracy. We also explored the biological relationships of the attributes used in HT-based model to understand their relationship with cancer proliferation or suppression pathways.

Conclusion Our study demonstrates that the proposed model and an inexpensive miRNA testing strategy have the potential to serve as a cost-effective and accessible method for diagnosing laryngeal cancer.

Keywords: microRNA; laryngeal cancer; decision tree; machine learning, Hoeffding Tree classifier

1 Introduction

Laryngeal cancer is the presence of a tumor in the larynx, or the voice box within a human's throat. According to the American Cancer Association, laryngeal cancers represent one-third of head and neck cancers [1]. In a study of thousands of laryngeal cancer patients in the United States, results showed that 57.7% of the subjects died from the disease [2].

The current gold standard for laryngeal cancer diagnosis is either physical exam of throat and neck, a biopsy of the throat, or medical imaging [3], all of which are relatively expensive and are inaccessible for many areas of the world. Moreover, these forms of diagnosis are more invasive and have a high propensity of false negative results. That is why most diagnoses of laryngeal cancer are made in the late stages of the disease, which makes it very difficult to cure. The late-stage disease is associated with worse outcomes, warrants multimodal therapy, and is less likely to allow for the preservation of the larynx [1]. Non-cancerous voice box diseases, such as laryngitis, voice box paralysis, and vocal leukoplakia, can also cause similar symptoms and may be misdiagnosed as laryngeal cancer [4]. However, the early detection of laryngeal cancer significantly increases survival rate and reduces the length of hospital stay and health care costs for patients [5]. Thus, the need for a reliable and accurate diagnosis method exists and can serve as a potential solution for decreasing laryngeal cancer mortality rates.

MicroRNAs (miRNAs) are small noncoding RNAs that typically have around 22 nucleotides in length [6]. MiRNAs have shown potential to serve as biomarkers for diagnosing cancer, since the expression of several miRNAs varies between normal and tumorous tissue [7]. In a study of 10,841 laryngeal cancer cases, Broseghini and co-workers found that 69 miRNAs are upregulated in the tumor and 95 miRNAs are downregulated in tumor compared to normal mucosa—the downregulated miRNAs are putative tumor suppressor miRNAs [8]. MiRNAs are found to be linked to several cancers because they can affect mRNA translation through the repression of specific mRNAs during translation or accelerating the destruction of mRNAs, which can trigger cancers, and it has been shown that specific miRNAs are linked to the occurrence of tumors [9, 10]. Since miRNAs are stable in bodily fluids like blood, urine, and saliva, they offer a minimally invasive means of assessing disease states and therapeutic efficacy [11]. Studies analyzing mice with miRNA overexpression or ablation have shown causal links between miRNAs and cancer development [12, 13]. Thus, these studies show the importance and relevance in studying miRNA as biomarkers for laryngeal cancer diagnosis.

Over the past few years, many different studies have attempted to utilize the miRNA-disease association to identify novel miRNAs that show association with different types of cancer. For example, L. Li and co-authors developed a miRNA-cancer association database, miCancerna, which used text-mining to find

thousands of miRNA-cancer associations from current studies published [14]. Another study used a Gaussian Bayesian network to analyze relationships between genes and miRNAs, which provided information of significant genes in cancer-related pathways [15].

A study by Sultan and co-workers used a decision tree model to compare relative miRNA expressions in regular and cancer tissues to identify significant miRNAs [16]. Our study goes a step further by analyzing specific attributes of miRNAs that contribute to miRNA being a biomarker of laryngeal cancer, which also enables us to identify previously unassociated miRNAs.

There were several attempts to diagnose laryngeal cancer using machine learning (ML). A study by Aicha [17] extracted acoustic features from recorded speech and performed feature selection to determine which speech features were most important for diagnosis. Another study by Z. Li and colleagues. Analyzed the Raman spectra from normal larynxes and cancerous larynxes and used a feature extraction approach to identify which parts of the Raman spectra indicated presence of a tumor [18]. Finally, a study by Singh and Maurya proposed a computer-aided-diagnosis system which can analyze patches of endoscopic videos, perform feature extraction, and determine which parts of an endoscopy are most crucial for diagnosis [19]. However, there is no existing research that uses biological attributes of microRNA for laryngeal cancer diagnosis. Thus, we propose a novel ML-based approach that can use biomarkers of cancer-associated miRNAs. This approach is far less invasive and more accessible, and it has the potential to improve diagnostics.

2 Methods

2.1 Data Pre-Processing and Mining

To elucidate which miRNA characteristics are most important for cancer association, we explored existing studies where miRNAs shown an association with laryngeal cancer [20–27].

2.2 Generating Attributes for miRNAs

To prepare the ML model using miRNA descriptors for diagnostics of laryngeal cancer (LC) we prepared two sets of miRNAs. The first set was selected from public sources of miRNA-related to LC, and the second set included random miRNA extracted from the miRBase [28]. Then we created a strategy to generate attributes for these miRNAs. Our strategy was three-fold: (1) generate sequence-based attributes for each miRNA's sequence, (2) generate a list of predicted target genes for each miRNA, (3) generate a list of predicted pathway processes in which are involved genes that are targets of selected miRNA. Such a list of attributes was generated for each miRNA. This data was used for training of the ML model. Before generating any other attributes, we created a separate attribute that categorized classification—we assigned a 1 to all LC associated miRNA and a 0 to all randomly associated miRNA. Our model was trained to classify miRNA based on this attribute.

2.2.1 Sequence-Based Attributes

Our strategy for generating sequence-based attributes is based on a strategy developed by Kang et al. [29], where the authors analyze quantity of bases, frequency, mean mass, hydrogen bonds, and motifs. We analyzed the sequence of all miRNAs because it is important in complimentary binding to the target genes mRNAs to inhibit the mRNA's expression [30].

Table 1 Sequence-Based Attributes and their Descriptions.

Sequence-Based Attribute	**Description**
Bases in miRNA sequence	N
Frequency of each base	N_A / N, N_C / N, N_G / N, N_U / N
Mean Mass of bases	135.1(N_A) + 111.1(N_C) + 151.1(N_G) +112.1(N_U) / N
Number of hydrogen bonds	2(N_A + N_U) + 3(N_C + N_G)
2 base motifs	Checks the sequence for each occurrence of 2 base motifs. Each motif is a separate attribute, and the miRNA is assigned 1 if it has the motif, 0 if not.
3 base motifs	Checks the sequence for each occurrence of 3 base motifs. Each motif is a separate attribute, and the miRNA is assigned 1 if it has the motif, 0 if not.
4 base motifs	Checks the sequence for each occurrence of 4 base motifs. Each motif is a separate attribute, and the miRNA is assigned 1 if it has the motif, 0 if not.

Note: 2 hydrogen bonds form between adenine and uracil and three between cytosine and guanine

We downloaded a full database of miRNA sequences from miRBase (https://mirbase.org/) [28], which contained the human miRNA names and their corresponding sequences. Using this information, we created a script, which parsed through the database and prepared a list of all the miRNA sequences that we were analyzing.

With the list of sequences, we generated sequence-based attributes for all the individual miRNAs. Table 1 shows the types of attributes that we had for each sequence. N_A is the amount of Adenine bases, N_C is the amount of Cytosine bases, N_G is the amount of Guanine bases, N_U is the amount of Uracil bases and N is the total amount of bases.

2.2.2 miRNA Target Genes

We chose to use target genes as descriptors because through them miRNAs affect cancer development. Dysregulated miRNAs can affect cancer growth, and depending on the target genes of miRNAs, they can function as either oncogenes or tumor suppressor genes [31].

In order to find the predicted target genes for all the miRNA, we used miRDB 6.0 (http://www.mirdb.org/) [32, 33]. miRDB provides a prediction score for all suggested target genes, so we removed all target genes that had a score below 98. We created a Python script, which found the individual target genes from all associated miRNA. These genes were added to the list of attributes. Then we created a script, which iterated

through all miRNAs and their target genes. For each predicted gene target in the file, the miRNA would be assigned "1" for that attribute. All target genes that a particular miRNA was not predicted to have relation with were assigned a "0" for the attribute.

2.2.3 miRNA Pathways

Finally, we used different pathway processes of the miRNAs as attributes. This was important to study because a single miRNA can target hundreds of mRNAs which can affect transcription and subsequently influence signaling pathways for cancer processes, which can either trigger or suppress the growth of cancer [34].

To find the predicted pathway processes for different miRNA, we downloaded data from NcPath (http://ncpath.pianlab.cn/) [35] to get a list of all experimentally validated miRNA-target mRNA-pathway lists. The pathways, which had a weak association, as determined by NcPath's verification methods like Luciferase reporter assay, were removed from the list. Similar to the target genes, we created a script, which elucidated all the individual pathways associated with each selected miRNA and created a list of them, to be added to the list of attributes. We also created a script, which iterated through all the pathways and assigned a "1" attribute to the miRNAs, which had those pathway processes, and a 0 to the miRNAs, which did not.

Once we added the miRNA pathway attributes, the list of attributes was finalized for all the LC-associated miRNAs. Figure 1 shows an example of an LC-associated pathway, Central Carbon Metabolism in Cancer. This pathway and one of its target genes, RPS6KA5, were chosen by attribute selection and are areas of future study for biological association between miRNAs targeting related pathways that trigger proliferation.

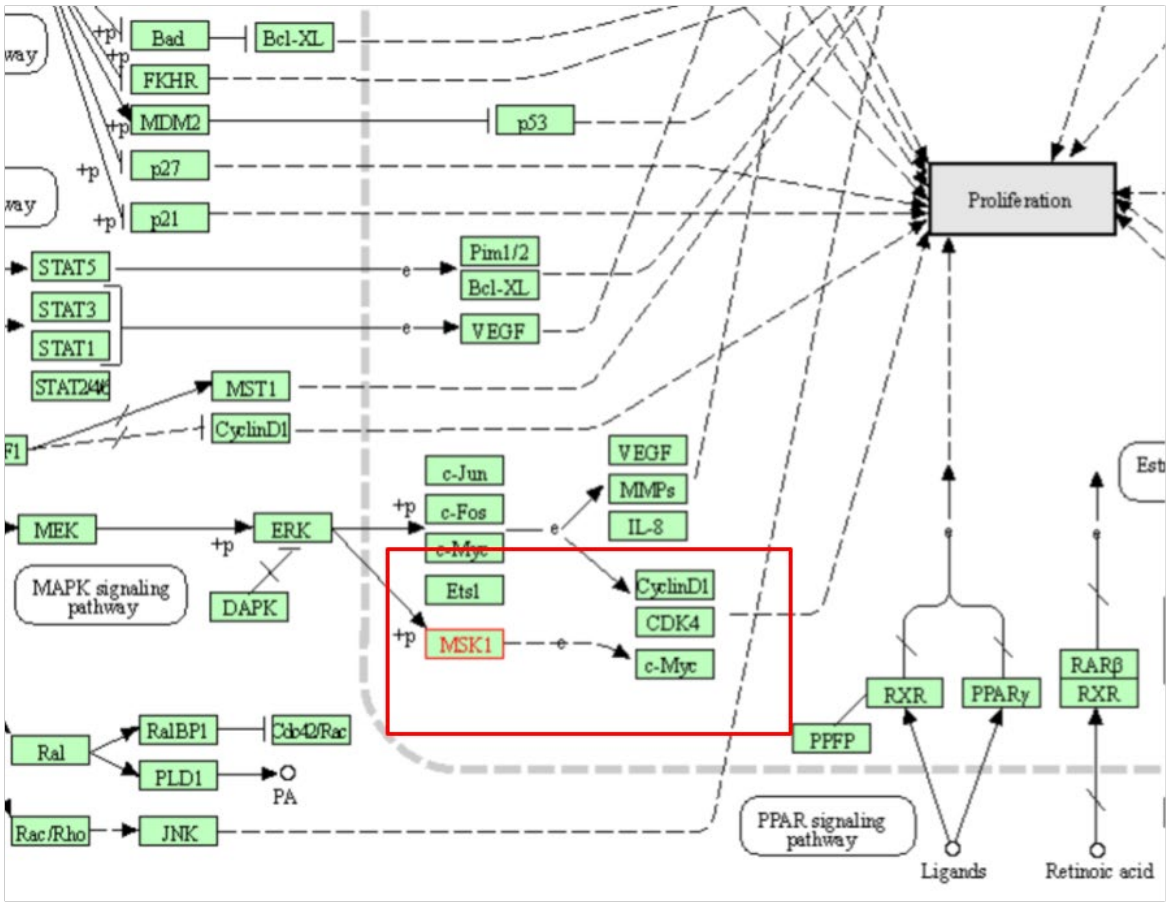

Fig. 1 *Central Carbon Metabolism in Cancer* Pathway section from KEGG, the area boxed shows the function of MSK1, (RPS6KA5), which was a chosen target gene by our model and triggers cancer cell proliferation.

2.3 Adding Random miRNAs

We also used miRBase [28] to get a list of randomly generated miRNAs that have not shown association with laryngeal cancer. In order to do this, we generated a list of miRNAs that have no known association with laryngeal cancer and randomly chose 57 miRNAs. We then repeated all the steps, which were previously described for generating attributes of the random miRNAs.

2.4 Attribute Selection

Once we had a final list of miRNAs and their attributes, we input our data into Waikato Environment for Knowledge Analysis (WEKA) [36]. Since we had many attributes and many of them may add noise to the model, we used WEKA's InfoGainAttributeEval feature to perform attribute selection on the dataset. We used the Ranker functionality to choose the attributes that were most significant in classification of the data. WEKA's attribute selection narrowed our list of attributes from 876 attributes to the 25 most important for classification. The selected target pathways are shown in Table 2. These target pathways are more common in the miRNAs that were used in training our model and already have known association with LC. We also analyzed the features chosen from the attribute selection. For the predicted target genes and pathway

processes, we visualized the relation of the different target genes in more complex cancer pathways using KEGG (https://www.genome.jp/kegg/) [37–39].

2.5 Training Decision Tree Model

The set of attributes we created was used as input to train our model. We tested different classification algorithms of WEKA to see which produced the highest classification accuracy. Our goal was also to optimize the area under the receiver-operating characteristic (ROC) curve, accuracy, precision, true-positive rate, false-positive rate, and area under the precision-recall (AUC-PR) curve.

We tested several classifiers from which the best were Random Forest, Naïve Bayes, Multilayer Perceptron, Logistic Model Tree, Hoeffding Tree (HT), and Logistic Regression. Each algorithm was tested using cross-validation, which divides the dataset into a given number of folds and tests each fold on the remaining folds, so each fold is tested multiple times. WEKA then averaged the results for each of these tests to return the accuracy, which reduces variance in the model. After testing models with different numbers of folds for the cross-validation on different algorithms, we found that the HT with five folds performed the highest. This model had 86.8% accuracy and was also ranked among the highest in the other metrics, when compared against the other models. We exported this model from WEKA and used it for testing data from other independent datasets.

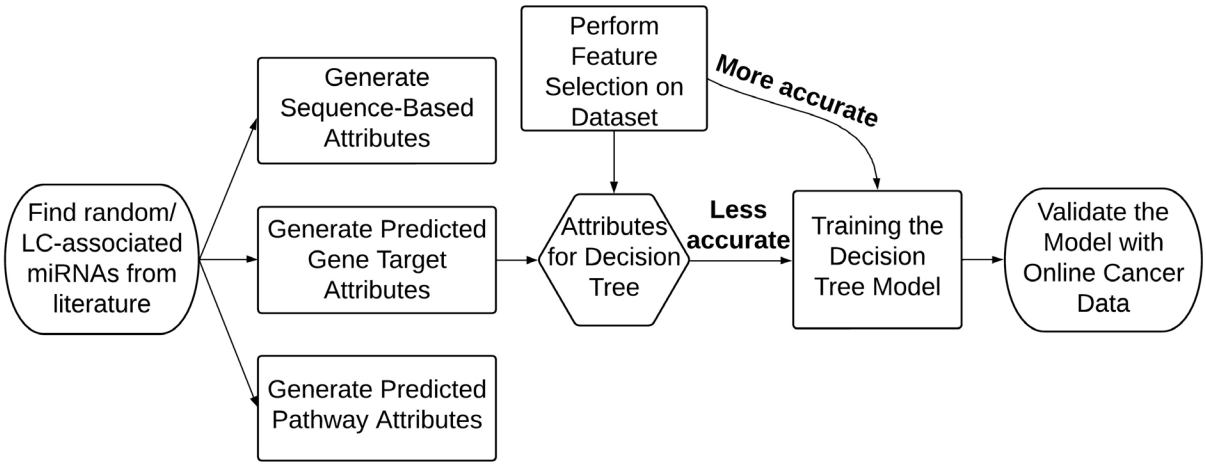

Fig 2 Workflow of Methods. First, obtain a list of random and LC-associated miRNAs from existing studies and generate sequence-based, gene target, and predicted pathway attributes. Once this is done, the model can be trained, or feature selection can be done before training. Performing feature selection increases the accuracy of the model when it is trained. Finally, the model is validated using independent cancer datasets. Created using LucidChart (https://www.lucidchart.com/).

2.6 Testing Model with the Independent Dataset

Once we finalized the model, we downloaded data from the Cancer Genome Atlas (https://portal.gdc.cancer.gov/) [40], which provided patient data and their concentration of specific

miRNAs. We only included miRNAs in testing that were more commonly expressed in patients who had LC ($p < 0.05$). We removed all miRNAs that were used in the training of the model. This finalized 29 miRNAs, which were used for independent testing. For each of these 29 miRNAs, we used our developed Python script to generate values for the attributes selected during the attribute selection process.

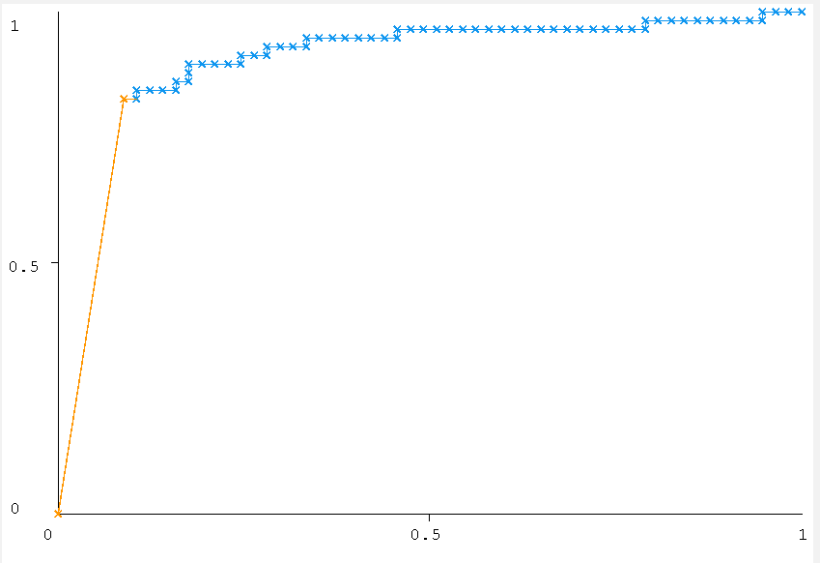

Fig. 3 ROC Curve for the Developed Hoeffding-Tree-Based Model.

We then input all this data into the previously created HT-based model in WEKA and the associated miRNAs were classified with 82% accuracy and the randomly selected miRNAs had <55% accuracy. The full workflow is visualized in Figure 2.

Table 2 Cancer-Associated Pathways Chosen by Attribute Selection of HT-based ML Model.

Cancer-Associated Pathways	Central carbon metabolism in cancer, Transcriptional misregulation in cancer, Melanoma, Small cell lung cancer, Hepatocellular carcinoma
Other Pathways	Hypertrophic cardiomyopathy, Pathways of neurodegeneration, Chagas disease, Toxoplasmosis

3 Results

3.1 Model Performance

Our final model, which is based on Hoeffding Tree (HT) classifier, showed high performance across multiple statistics. It had 0.868 accuracy, 0.871 precision, 0.868 recall, 0.868 F-measure, 0.740 Mathews correlation coefficient (MCC), 0.902 Area under ROC curve (AUROC), and 0.860 Area under precision-recall curve (AUC-PR).

The high true-positive rate shows that the model is successful is accurately classifying miRNAs as cancer-associated. The high precision shows that the model consistently performs well. In addition, the AUROC being over 0.9 indicates that the model performs well in both classification of associated and random miRNAs, so it is a reliable method for classification. The ROC curve and AUROC are shown in Figure 3.

Other algorithms used on this dataset also performed at high accuracies, but we found that the HT-based model was most optimal across all statistics. Table 3 illustrates the performance of other models, all with cross-validation of five folds. The HT algorithm shown at the top of Table 3 is the model we chose. Figure 4 shows the accuracy of all the different models, including those that were less optimal.

Table 3 Performance of ML Algorithms on Dataset with Five-Fold Cross-Validation.

	Accuracy	**Precision**	**Recall**	**F-Measure**	**AUROC**	**AUC-PR**
Hoeffding Tree	0.868	0.871	0.868	0.868	0.902	0.860
Random Forest	0.789	0.790	0.789	0.789	0.910	0.906
Multilayer Perceptron	0.816	0.818	0.816	0.815	0.844	0.824
Logistic Regression	0.807	0.813	0.807	0.806	0.723	0.693

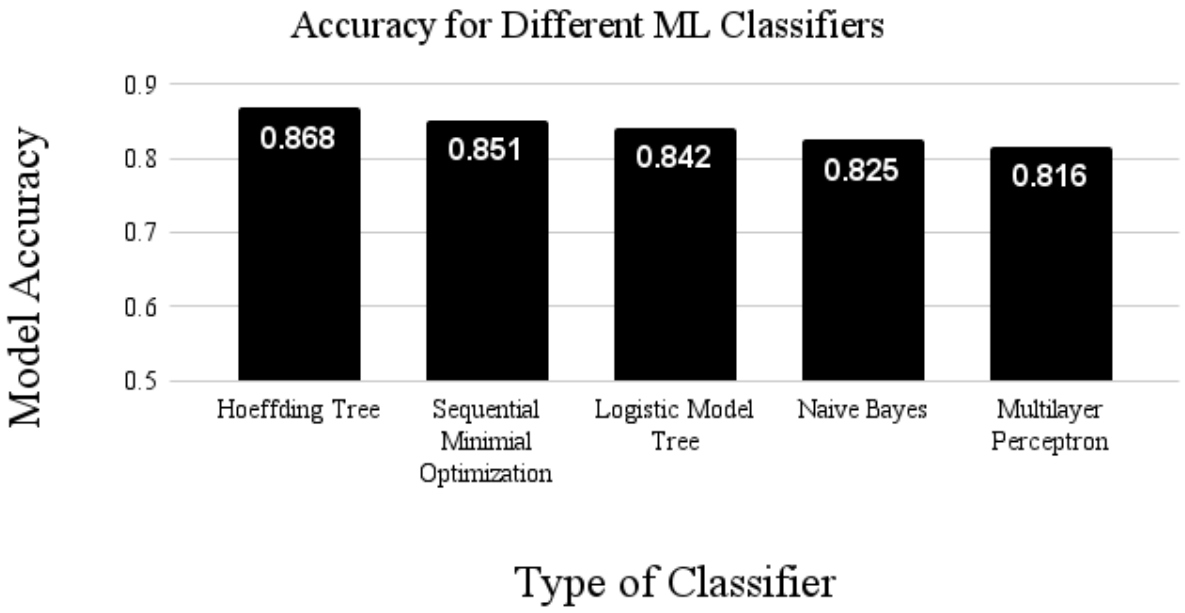

Fig. 4 Comparison of the Accuracies of Five Different ML Classifiers on the Cross-Validation of Training Dataset.

In addition, we verified that HT-based ML model was most optimal based on the number of miRNAs that were provided for training and the number of attributes that were in the model after attribute selection. Both factors affected the accuracy of classification, since the number of miRNAs defined how much data the

model had for training, and the number of attributes affected which biological characteristics were determined to be most important in classification. Figure 5 shows a comparison of the highest-performing model accuracy with different numbers of miRNA being used in training and different numbers of attributes being used for classification.

HT-based model is the most optimal based on the number of attributes and the number of associated miRNAs. The model performed the highest across multiple metrics, and it was the best to use for validation with new independent data. Figure 5a shows that HT-based model performed better with increasing number of miRNAs included in training. The model was trained with 57 miRNAs, which includes the majority of available studies on biologically associated miRNAs. The accuracy of predictions after some threshold is inversely proportional to the number of attributes (Figure 5b).

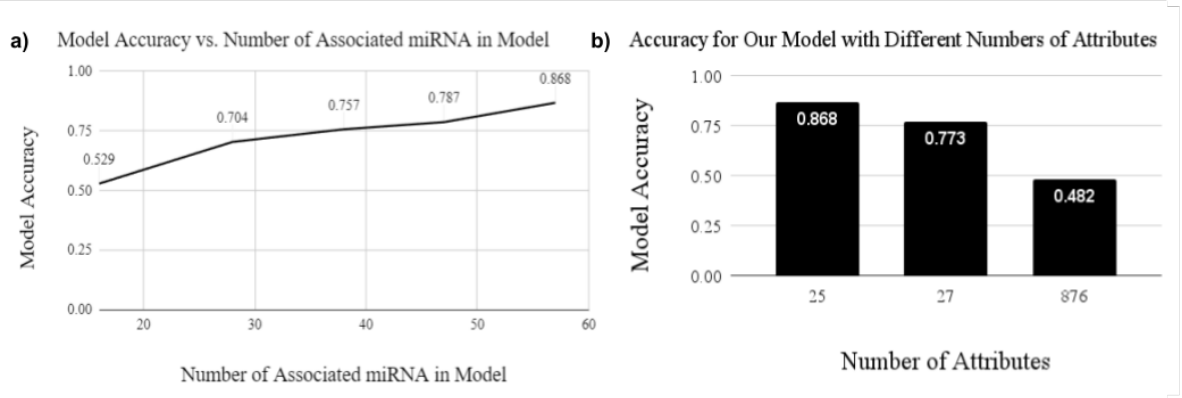

Fig. 5 Comparison of the accuracies of the highest-performing ML models. (**a**) Accuracies of Model with Different Numbers of miRNA in Model Training. (**b**) Accuracies of Model with Different Numbers of Attributes in Model Training.

3.2 Validation of Trained Model

The Hoeffding-Tree-based ML model, which we created, was exported and used for validation. Using the independent data of patients with LC from the Cancer Genome Atlas, we inputted 29 miRNAs into HT-based model, and it classified 25 miRNAs as associated with LC. The classification accuracy was 86.2%.

Since these miRNAs were shown to have biological association based on both a patient's clinical data and HT-based model, they are especially important to investigate for future diagnostic of laryngeal cancer.

3.3 Analysis of Model Attributes

We also investigated the model attributes that were chosen after attribute selection. InfoGainAttributeEval evaluated all the different attributes based on the effect of the attribute on reducing the overall entropy when

performing classification. Thus, the attributes that are selected have the greatest relevance for which biological characteristics of miRNA cause the miRNA to be associated to laryngeal cancer.

We found which pathways and target genes were more/less associated with LC by comparing the proportion of associated miRNAs with the target/pathway process to the proportion of random miRNAs with it.

Elucidated gene targets of miRNAs related to LC likely have a clear relation to cancer. For example, Figure 1 shows how one of selected target genes, RPS6KA5 triggers cancer proliferation. RPS6KA5 (MSK1), which is shown in the figure. This is an area of future study because it shows that the RPS6KA5 target gene of miRNA and the pathway it operates in promotes cancer growth, which is the central carbon metabolism in cancer, or hsa05230. We found that this pathway was more linked with laryngeal cancer-associated miRNA, so we can look further into these pathways for treating cancer in the future. Thus, the presence of this target gene and its pathway give an indication of cancer presence and are also a potential target for controlling cancer growth. Similarly, the HT-based model identified the MDM1 target gene of miRNA to be associated with LC. This gene deactivates p53, which causes decreased cancer proliferation. This gene's interaction can be visualized in the hsa05206 pathway, which is for microRNAs in cancer. In addition, we found that the TET2 target gene of miRNA was less expressed in LC-associated miRNA, indicating that LC-associated miRNAs may downregulate TET2. Downregulation of TET2 can prevent its ability to suppress tumors, which is an important area to investigate for cancer proliferation.

4 Discussion

Our model based on the Hoeffding Tree (HT) classifier has high accuracy, AUROC, and precision when compared to other classifiers. HT-based model is also the highest performing based on the number of miRNAs and attributes used in training. Thus, the created model is most effective for the analysis performed.

The HT algorithm was especially efficient because it was able to deal with noisy data. The tree structure makes local decisions based on sufficient statistics, which reduced the impact of noise on the final classification performance. Cross-validation showed that this model performed with high accuracy. The AUROC indicates the robustness of the model since it can optimize the classification rate without creating any false positives. This shows the promise of HT-based model because it can distinguish which attributes are most important in miRNA-cancer association.

In addition, we found that the model can effectively classify any given miRNA as associated or non-associated with laryngeal cancer. The model has the potential to be applied to any given miRNA, by running a Python script to generate a list of attributes and checking whether the miRNA is classified as cancer-associated or not.

The system described in this study can also be applied to other forms of cancer or other disease diagnosis when they have an association with miRNA and there is existing data about it. The same methodology as described in Figure 2 can be applied. In the future, we can generate attributes for a given set of miRNAs that has already shown experimental association with the disease. Our approach has the potential to elucidate unknown molecular mechanisms of various diseases, since miRNAs participate in many of them.

In future studies using this approach would also help to add attributes such as age, gender, and ethnic groups of patients. These factors have been shown to be associated with higher mortality rates [41] and may be confounding variables that made certain miRNAs or attributes more important in the model, even if they are not as significant in laryngeal cancer patients.

We would also suggest studying the biological associations of the newly discovered miRNAs to verify if they are associated with laryngeal cancer. This can advance our current understanding of the relevant miRNA characteristics, which are associated with laryngeal cancer.

We would also suggest studying cancers or diseases, which have more existing data about associated miRNAs, since LC has historically been under-researched.

5 Conclusions

We selected miRNAs that have already shown association with laryngeal cancer. For each miRNA, we generated sequence-based attributes, predicted target genes, and predicted related pathways and used attribute selection to narrow which characteristics are most important for cancer-related miRNAs. We then used the data to train different ML models and found that the HT classifier performed the best. The HT-based model can predict LC using miRNAs analysis data with 86.8% accuracy and was also validated with the independent patients' data showing 83% accuracy. We identified new miRNAs that had not previously shown biological relationship with laryngeal cancer, and we found out which pathways and target genes are predicted to trigger cancer proliferation. Our methodology can be applied to other diseases in the future and the data found can be crucial in providing diagnosis or identifying targets for cancer proliferation.

Statements and Declarations

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

Funding Statement

None.

Competing Interests

The authors have no conflicts of interest related to this publication.

Author Contributions

VK and IFT proposed the development of a machine-learning system that uses miRNAs, gene targets, and miRNA pathway processes to diagnose cancer. AA found and processed the data, trained the model, and wrote the article. VK and IFT assisted with development of the machine-learning system and edited the article.

Supplementary Files

All codes used for this study is available upon request. All other data is publicly accessible.

References

1. Koroulakis A, Agarwal M. Laryngeal cancer. (2022) National Library of Medicine, 2022. StatPearls [Internet]: https://www.ncbi.nlm.nih.gov/books/NBK526076/ (accessed 9 July 2023).
2. Afify,AY, Ashry MH, Sadeq MA, Elsaid M.(2023) Causes of death after laryngeal cancer diagnosis: A US population-based study. Eur Arch Otorhinolaryngol, 280(4):1855–1864. https://doi.org/10.1007/s00405-022-07730-y.
3. Laryngeal Cancer Treatment (PDQ®)–Patient Version. (2023) National Cancer Institute https://www.cancer.gov/types/head-and-neck/patient/adult/laryngeal-treatment-pdq. (accessed 16 July 2023).
4. Jovanović MB. Diagnosis of laryngeal carcinoma. (2008) Med Preg 61(11–12):591–595. https://doi.org/0.2298/mpns0812591j.
5. Cohen SM, Kim J, Roy N, Asche C, Courey M. (2012) Direct health care costs of laryngeal diseases and disorders. Laryngoscope 122(7):1582–1588. https://doi.org/10.1002/lary.23189.
6. Esquela-Kerscher A, Slack FJ. (2006) Oncomirs—microRNAs with a role in cancer. Nat Rev Cancer 6(4):259–269. https://doi.org/10.1038/nrc1840.
7. Broseghini E, Filippini DM, Fabbri L, Leonardi R, Abeshi A, Dal Molin D, Fermi M, Ferracin M, Fernandez IJ. (2023) Diagnostic and Prognostic Value of microRNAs in Patients with Laryngeal Cancer: A Systematic Review. Noncoding RNA 9(1):9. https://doi.org/10.3390/ncrna9010009.
8. Widera C, Gupta SK, Lorenzen JM, Bang C, Bauersachs J, Bethmann K, Kempf T, Wollert KC, Thum T. (2011) Diagnostic and prognostic impact of six circulating microRNAs in acute coronary syndrome. J Mol Cell Cardiol 51(5):872–875. https://doi.org/10.1016/j.yjmcc.2011.07.011.
9. Hegazy M, Elkady MA, Yehia AM, Elsakka EG, Abulsoud AI, Abdelmak-soud NM, Elshafei A, Abdelghany TM, Elkhawaga SY, Ismail A, et al. (2023) The role of miRNAs in laryngeal cancer pathogenesis and therapeutic resistance-a focus on signaling pathways interplay. Pathol Res Pract 246:154510. https://doi.org/10.1016/j.prp.2023.154510.
10. Soifer HS, Rossi JJ, Sætrom P. (2007) MicroRNAs in disease and potential therapeutic applications. Mol Ther 15(12):2070–2079. https://doi.org/10.1038/sj.mt.6300311.
11. Gayosso-Gomez LV, Ortiz-Quintero B. (2021) Circulating microRNAs in blood and other body fluids as biomarkers for diagnosis, prognosis, and therapy response in lung cancer. Diagnostics (Basel) 11(3):421. https://doi.org/10.3390/diagnostics11030421.
12. Iorio MV, Croce CM. (2012) MicroRNA dysregulation in cancer: Diagnostics, monitoring and therapeutics. Acomprehensive review. EMBO Mol Med 4(3):143–159. https://doi.org/10.1002/emmm.201100209.
13. Izumchenko E, Chang X, Michailidi C, Kagohara L, Ravi R, Paz K, Brait M, Hoque MO, Ling S, Bedi A, et al. (2014) The TGFβ–miR200–MIG6 pathway orchestrates the EMT-associated kinase switch that induces resistance to EGRF inhibitors. Cancer Res 74(14):3995–4005. https://doi.org/10.1158/0008-5472.CAN-14-0110.
14. Li L, Hu X, Yang Z, Jia Z, Fang M, Zhang L, Zhou Y. et al. (2014) Establishing reliable miRNA-cancer association network based on text-mining method. Comput Math Meth Med 2014:746979. https://doi.org/10.1155/2014/746979.
15. Jin D, Lee H. (2015) A computational approach to identifying gene-microrna modules in cancer. PLoS Comput Biol 11(1):1004042. https://doi.org/10.1371/journal.pcbi.1004042.
16. Sultan A, Asa AA, Guimbangunan TM, Serapio ED, Fellizar A, Albano PM, Tomas RC. (2023) Machine learning-based prediction of the likelihood of colorectal cancer using mirna expression. Philipp J Sci 152(4):1413–1432. https://doi.org/10.56899/152.04.12.

17. Aicha AB. (2020) Conventional machine learning techniques with features engineering for preventive larynx cancer detection. In: Proc. IEEE 2020 5th Int Conf on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 1–5.
18. Li Z, Li Z, Chen Q, Zhang J, Dunham ME, McWhorter AJ, Feng JM, Li Y, Yao S, Xu J. (2022) Machine-learning-assisted spontaneous Raman spectroscopy classification and feature extraction for the diagnosis of human laryngeal cancer. Comput Biol Med 146:105617. https://doi.org/10.1016/j.compbiomed.2022.105617.
19. Singh,VP, Maurya AK. (2021) Role of machine learning and texture features for the diagnosis of laryngeal cancer. In: Mohanti SN, Nalinipriya G, Jena OP, Sarkar A. (Eds.) *Machine Learning for Healthcare Applications*. John Wiley & Sons: Hoboken, NJ, USA, 2021. Chapter 23, pp. 353–367. https://doi.org/10.1002/9781119792611.ch23.
20. Ayaz L, Görür A, Yaroğlu HY, Ozcan C, Tamer L. (2013) Differential expression of microRNAs in plasma of patients with laryngeal squamous cell carcinoma: Potential early-detection markers for laryngeal squamous cell carcinoma. J Cancer Res Clin Oncol 139:1499–1506. https://doi.org/10.1007/s00432-013-1469-2.
21. Cao P, Zhou L, Zhang J, Zheng F, Wang H, Ma D, Tian J. (2013) Comprehensive expression profiling of microRNAs in laryngeal squamous cell carcinoma. Head Neck 35(5):720–728. https://doi.org/10.1002/hed.23011.
22. Chen L, Sun DZ, Fu YG, Yang PZ, Lv HQ, Gao Y, Zhang XY.. (2020) Upregulation of microRNA-141 suppresses epithelial-mesenchymal transition and lymph node metastasis in laryngeal cancer through HOXC6-dependent TGF-β signaling pathway. CellSignal 66:109444. https://doi.org/10.1016/j.cellsig.2019.109444.
23. Guo L, Cai X, Hu W, Hua W, Yan W, Lin Y, Yin S, Chen Y. (2019) Expression and clinical significance of miRNA-145 and miRNA-218 in laryngeal cancer. Oncol Lett 18(1):764–770. https://doi.org/ 10.3892/ol.2019.10353.
24. Hu Y, Liu H. (2015) MicroRNA-10a-5p and microRNA-34c-5p in laryngeal epithelial premalignant lesions: Differential expression and clinicopathological correlation. Eur Arch Otorhinolaryngol. 272:391–399.
25. Huang Y, Gu,M, Tang Y, Sun Z, Luo J, Li Z. (2021) Systematic review and meta-analysis of prognostic microRNA biomarkers for survival outcome in laryngeal squamous cell cancer. Cancer Cell Int 21(1):1–14. https://doi.org/10.1186/s12935-021-02021-8.
26. Li P, Liu H, Wang Z, He F, Wang H, Shi Z, Yang A, Ye J. (2016) MicroRNAs in laryngeal cancer: implications for diagnosis, prognosis and therapy. Am J Transl Res 8(5):1935.
27. Takeuchi T, Kawasaki H, Luce A, Cossu AM, Misso G, Scrima M, Bocchetti M, Ricciardiello F, Caraglia M, Zappavigna S. (2020) Insight toward the microRNA profiling of laryngeal cancers: Biological role and clinical impact. Int J Mol Sci 21(10):3693. https://doi.org/10.3390/ijms21103693.
28. Kozomara A, Birgaoanu M, Griffiths-Jones S. (2019) miRBase: From microRNA sequences to function. Nucleic Acids Res 47(D1):155–162. https://doi.org/10.1093/nar/gky1141.
29. Kang W, Kouznetsova VL, Tsigelny,IF. (2022) miRNA in machine-learning-based diagnostics of cancers. Cancer Screen Prevent 1(1):32–38. https://doi.org/10.14218/CSP.2021.00001
30. Macfarlane LA, Murphy, PR. (2010) MicroRNA: Biogenesis, function and role in cancer. Curr Genomics 11(7):537–561. https://doi.org/10.2174/138920210793175895.
31. Peng Y, Croce CM. (2016) The role of microRNAs in human cancer. Signal Transd Target Ther 1(1):1–9. https://doi.org/0.1038/sigtrans.2015.4.
32. Chen,Y, Wang X. (2020) miRDB: An online database for prediction of functional microRNA targets. Nucleic Acids Res 48(D1):D127–D131. https://doi.org/10.1093/nar/gkz757.

33. Liu W, Wang X. (2019) Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. Genome Biol 20:1–10. https://doi.org/10.1186/s13059-019-1629-z.
34. Reddy KB. (2015) MicroRNA (miRNA) in cancer. Cancer Cell Int 15(1):1–6. https://doi.org/10.1186/s12935-015-0185-1.
35. Li Z, Zhang Y, Fang J, Xu Z, Zhang H, Mao M, Chen Y, Zhang L, Pian C. (2023) NcPath: A novel platform for visualization and enrichment analysis of human non-coding RNA and KEGG signaling pathways. Bioinformatics 39(1):812. https://doi.org/10.1093/bioinformatics/btac812.
36. Frank E, Hall MA, and Witten IH. (2016) Appendix B: The WEKA Workbench. In Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Morgan Kaufmann Cambridge, Mass., USA. Appendix B, pp. 553–572.
37. Kanehisa, M. (2019) Toward understanding the origin and evolution of cellular organisms. Protein Sci 28(11):1947–1951. https://doi.org/10.1002/pro.3715.
38. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe,M. (2021) KEGG: Integrating viruses and cellular organisms. Nucleic Acids Res 49(D1):D545–D551. https://doi.org/10.1093/nar/gkaa970.
39. Kanehisa M, Goto S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28(1):27–30. https://doi.org/10.1093/nar/28.1.27.
40. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart,JM. (2013) The Cancer Genome Atlas Pan—Cancer analysis project. Nat Genetics 45(10):1113–1120. https://doi.org/10.1038/ng.2764.
41. Hu K, Lin R, Zhang Z, Chen H, Rao X. (2020) Impact of prior cancer history on the survival of patients with larynx cancer. BMC Cancer 20(1):1137. https://doi.org/10.1186/s12885-020-07634-2.