***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Machine Learning based Identification of Cancer Related Mirna and Gene Biomarkers

[1]Karare Bharati Arunrao, [2]Dr. Priya Vij

[1,2] Department of Computer Science and Engineering, Kalinga University, Naya Raipur, Chhattisgarh, India

Abstract

Cancer is still a major killer, thus finding effective biomarkers and making correct diagnoses requires cutting-edge computational methods. Using data from The Cancer Genome Atlas (TCGA), this study investigates the predictive power of deep learning and machine learning models in finding genes and miRNAs unique to cancer. Supervised learning models including Random Forest, Support Vector Machine (SVM), and Gradient Boosting were used in a thorough examination, along with deep learning architectures like CNNs and LSTM networks. To further understand the intricate relationships between miRNAs and genes, we used graph-based approaches. Using important performance indicators including accuracy, precision, recall, F1 score, and Area under the Curve (AUC), the models were assessed using K-fold cross-validation. For the purpose of differentiating cancer-specific biomarkers, the findings showed that deep learning models, especially CNN (AUC = 0.95) and LSTM (AUC = 0.93), performed better than conventional machine learning methods. By providing data-driven approaches for early identification and individualized treatment regimens, this work highlights the potential of deep learning to advance precision oncology.

Keywords: Gene expression, Machine Learning, Biomarker, Tumor, Immune

I.INTRODUCTION

Every year, cancer claims the lives of millions of people throughout the world, making it one of the biggest obstacles in contemporary medicine. A complex and diverse illness, it is defined by mutations in genes, interactions with the microenvironment, and unchecked cell development. Biomarkers have been developed to assist in early detection, diagnosis, prognosis, and therapy selection as a result of developments in molecular biology and bioinformatics. One of the most important aspects of cancer research is the role of microRNAs (miRNAs) and gene expression patterns. In order to improve patient outcomes and therapy efficacy, identify trustworthy biomarkers is crucial for creating precision medicine strategies. Scalability, accuracy, and repeatability are three areas where conventional biomarker discovery approaches fall short. Machine learning (ML) has brought about a paradigm shift in this area by offering data-driven methods for evaluating intricate genetic data, discovering possible biomarkers, and making highly accurate predictions of cancer-related patterns.
tiny non-coding RNA molecules called microRNAs (miRNAs) target messenger RNAs (mRNAs) for destruction or translational suppression; they are essential regulators of gene expression. As a result of their abnormal expression in a number of malignancies, these molecules are promising biomarkers for use in diagnosis, prognosis, and treatment planning.

By affecting critical cellular pathways including apoptosis, proliferation, metastasis, and immune response, miRNAs can play the role of oncogenes or tumor suppressors. Their capacity to remain stable in common body fluids including blood and urine increases their promise as non-invasive cancer diagnostic tools. Despite their potential, the massive amounts of data produced by high-throughput sequencing technology make it difficult to identify certain miRNAs linked with various cancer types. Machine learning approaches shine in this area because they use computational and statistical methods to rapidly analyze massive information, find hidden patterns, and categorize pertinent biomarkers.

In a similar vein, gene expression profiling has emerged as an essential tool in the fight against cancer by illuminating the molecular pathways that drive tumorigenesis and metastasis. Differentiating between cancer subtypes, tracking disease progression, and measuring therapy efficacy are all made easier by gene biomarkers. Biomarker selection may be prone to biases and inconsistencies due to the high complexity of gene expression data, which presents a challenge for typical statistical approaches. It has been demonstrated that machine learning techniques, including neural networks, deep learning, support vector machines (SVMs), and random forests, are very good at managing complicated datasets and producing accurate predictions. Using these methods, scientists have been able to pinpoint reliable gene expression

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

profiles linked to cancer and incorporate them into diagnostic and prognostic models.

Cancer biomarker research has seen a rise in the use of machine learning as a tool for automating analyses of large genomic datasets, improving pattern identification, and creating highly accurate prediction models. The discovery of genes linked to various types of cancer and differentially expressed miRNAs may be accomplished through the use of supervised learning methods, such as support vector machines (SVMs) and decision trees, in classification tasks. To find new biomarker patterns without labels, unsupervised learning approaches are useful. These methods include clustering and dimensionality reduction techniques such as t-distributed stochastic neighbor embedding (t-SNE) and principal component analysis (PCA). When it comes to evaluating high-dimensional genomic data and discovering complex correlations between genetic components, deep learning models, especially CNNs and RNNs, have proven to be incredibly effective.

II.BIOLOGICAL FUNCTIONS OF miRNA IN CANCER

miRNAs in cancer progression

A cluster of miR-15 and miR-16 deletions and low expression in chronic lymphocytic leukemia was initially shown in 2002, indicating a function for miRNAs in cancer development. In the recent past, miRNAs have been associated with almost every cancer process that is now understood. Based on the target gene, miRNAs can either inhibit recognized tumor suppressors or negatively impact protein-coding oncogenes; hence, miRNAs can function as either onco-miRNAs or tumor-suppressor miRNAs.

Alterations in genes that control the course of cancer can also be associated with dysregulation of miRNAs. Upregulation of miR-1269 expression occurs during gastric cancer development and progression; this upregulation regulates the AKT and Bax/Bcl-2 signaling pathways, with RASSF9 as a target, and enhances cancer cell proliferation and cell cycle G1-S transition while suppressing cell apoptosis. Using miR-9 mimics effectively stops cell proliferation by suppressing cyclin-dependent kinase 6 (CDK6) and cyclin D1, and miR-9 expression is downregulated in oral squamous cell cancer (OSCC) patients, leading to G0/G1 cell cycle arrest. In addition, miR-9's pri-miRNA paralogs allow it to target a wider variety of RNAs. The overexpression of COL18A1, THBS2, PTCH1, and PHD3 has a favorable effect in glioma cell proliferation and cell cycle progression, and miR-9, one of the most prevalent miRNAs in the brain, was found to directly target these genes. Consequently, there may be important routes that have yet to be discovered in the regulatory characterisation of miRNAs in cell cycle control.

The miR-145 that targets A disintegrin and metalloproteinase 17 (ADAM17) was discovered to be able to decrease cell growth in liver cancer cells. A different research on HCC found that individuals with a poor prognosis were more likely to have overexpressed miR-487a, a gene that promotes cell proliferation through phosphoinositide-3-kinase regulatory subunit 1 (PIK3R1) triggered AKT signaling. These results have the potential to open up new avenues for the treatment of HCC and other malignancies.

Many microRNAs have potential functions in cancer cell apoptosis and other forms of programmed cell death. Many studies have shown that low levels of miR-192, miR-194, and miR-215 in multiple myeloma cause p53 dysregulation, which protects cancer cells from death. Studies on breast and soft tissue sarcomas demonstrated that several miRNAs contributed to p53 diversification, while studies on nasopharyngeal and stomach cancers related with the Epstein-Barr virus (EBV) linked specific miRNAs to p53 regulation. In addition, miR-205 and miR-338-3p can prevent apoptosis in prostate cancer cells by focusing on B-cell lymphoma 2 (BCL-2), an inhibitory gene. Key components of the extrinsic apoptotic cascade, such as Fas ligand (FasL), can be inhibited by microRNAs. Research has shown that miR-21-5p specifically targets FasL, and that up-regulation of this gene in HCC cells can have a deleterious effect on FasL mRNA and protein production.

When tumors are in an angiogenesis state, microRNAs like miR-210 and miR-519c can regulate hypoxia-inducible factor 1α (HIF-1α) and vascular endothelial growth factor (VEGF). In the physiopathology of OSCC, regulatory networks of miRNAs are thought to be a feature of the epithelial-mesenchymal transition (EMT). The possible interaction of interferon-γ (IFN-γ), IRF1, and IGF1 with miR-29b, whose presence limits the growth and invasion of cancer cells, is strongly linked to an aggressive phenotype of colorectal cancer (CRC). All of the aforementioned demonstrated the significant and intricate functions that microRNAs play in the cancer regulatory network, namely in tumor development, which includes cell cycle, cell proliferation, cell apoptosis, angiogenesis, EMT, and tumor invasion.

miRNAs affect tumor immunity

Many microRNAs (miRNAs) have a role in tumor immune surveillance and tumor immune escape, and immune system dysregulation is a major player in cancer progression. Moving further, we'll talk about how exosomes may load miRNAs, and how those miRNAs can play a role in communication

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

between cells. Many facets of cancer can be impacted by exosomes, including as EMT, multidrug resistance, metastasis, and progression. Certain types of immune cells can be directly influenced, such as how miR-23a limits the function of natural killer (NK) cells, how miR-212-3p and miR-203 diminish the function of dendritic cells, and how tumor-derived miR-24-3p inhibits CD8+ cells while activating CD4+ regulatory T cells. With regard to NK cells in particular, new data suggests that some miRNAs can influence their anti-tumor capabilities in a variety of ways, including by regulating the expression of immunological checkpoints on NK cell surfaces or the ligands on tumor cells.

There is evidence that microRNAs have an impact on tumor immunogenicity and antitumor immune responses, in addition to their oncogenic and tumor-suppressing characteristics. For instance, research has shown that LINC00240 targets miR-124-3p, which may have a role in the progression of cervical cancer through the signal transduction pathways involving STAT3 and MHC class I-related proteins A (MICA), modulating tolerance of NKT cells. Various microRNAs, such as miR-346, miR-451, and miR-148a, are thought to have a role in the HLA class I antigen processing machinery (APM) components, control of the B7 family members, and production of IFN-γ signaling molecules. Indeed, miRNAs interact with a wide variety of immune checkpoint proteins (ICPs), not limited to the B7 family. This knowledge offers great promise for cancer patients' customized therapy and prognosis prediction.

Also, microRNAs (miRNAs) are now an essential regulator of antitumor immune cells. In this section, we offer a number of affiliated research papers. In chordoma, researchers defined the close relationship between miR-574-3p, PD-L1 expression, and tumor-infiltrating lymphocytes (TIL) levels, and mechanically, another group demonstrated the role of the p53/miR-34/PD-L1 axis in immune evasion of lung cancer. Increased numbers of myeloid-derived suppressor cells (MDSCs) in acute myeloid leukemia (AML) are associated with miR-34a's ability to control the oncoprotein MUC1. By interacting directly with the interleukin 10 (IL-10) gene in tumor-associated macrophages (TAMs), miR-98 stops HCC in its tracks. Since miR-21 influences cancer-associated fibroblasts' (CAFs) metabolic changes, it impacts pancreatic cancer cells.

Involved miRNA sponges in cancer

The term "competing endogenous RNAs" (ceRNAs) describes a group of transcripts such ncRNAs, pseudogenes, and protein-coding RNAs that work together to entice microRNAs (miRNAs) for interactions. Interactions between ceRNAs, also known as miRNA sponge interactions, allow for indirect regulation of one another via the titration process.

The vast majority of the human genome is composed of non-coding RNAs (ncRNAs), which include miRNAs and several other types of RNAs (lncRNAs, circRNAs, snRNAs, etc.). Many recent studies have shown the significance of lncRNAs in miRNA-related tumor control; lncRNAs were the first to be discovered as miRNA sponges. As part of its research on lung cancer stem cells, scientists found that miR-146 regulates NUMB post-transcriptionally, and that TUSC-7, a powerful suppressive lncRNA, might prevent NUMB degradation by blocking NOTCH signaling. These findings demonstrate the importance of multiple non-coding genes used in renewal repression control and demonstrate that TUSC-7 activity is dependent on the downregulation of tumor-suppressing miR-146. By up-regulating integrin β3, the target of the two microRNAs, the lncRNA FAM225A was shown to be over-expressed in nasopharyngeal carcinoma (NPC) as a sponge of miR-590-3p and miR-1275, therefore facilitating the growth of NPC. In addition, a new long non-coding RNA called prostate cancer-associated transcript 7 (PCAT7) was thought to have a role in NPC tumor growth by acting as a miR-134-5p sensor. In addition to these, the miRNA sponge-lncRNAs have also been studied in numerous kinds of cancers such as a miR-34a sponge MALAT1 in melanoma, a miR-15a/16 sponge LINC00461 in multiple myeloma, a miR-330-5p sponge LINC00958 in pancreatic cancer, a miR-7-5p sponge RP4 in colorectal cancer, a miR-324-5p sponge TPT1-AS1 and a miR-186 sponge antisense noncoding RNA of the INK4 locus (ANRIL) in cervical cancer, a miR-96 sponge TP53TG1 in pancreatic ductal adenocarcinoma, and a miR-300 sponge TUG1 in gallbladder carcinoma (GBC).

An essential function of some endogenous ncRNAs known as circRNAs is to act as miRNA sponges, thanks to their covalently closed consecutive circle. Utilizing RNA sequencing technology, it was discovered that tumor cell lines, tissues, and even plasma samples from cancer patients had aberrant circRNAs expressions. One research discovered that circMMP9 sucked miR-124 dry, speeding up the migration and development of glioblastoma multiforme (GBM) cells. In addition, they discovered that eukaryotic initiation factor 4A3 (eIF4A3) triggered circMMP9 cyclization and enhanced expression, whereas in GBM eIF4A3 bound to the MMP9 mRNA transcript. In addition, OSCC, HCC, lung, breast, gastric, and other cancers have confirmed circRNAs as miRNA sponges.

For example, PTENP1 and OCT4P4 are examples of pseudogenes that have lost their ability to code for proteins

Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

due to endogenous inactivating mutations; furthermore, some of these pseudogenes have been discovered to operate as miRNA sponges. Additionally, miRNAs can bind to some forms of protein-coding RNA. Insights into the intricate interplay between many cancer pathways have been provided by the discovery that TNRC6B is a PTEN ceRNA and that tumor suppressor gene PTEN is a miRNA sponge. Separate from these, Ago protein immunoprecipitation tests, wound-healing testing, and huge transcriptome data analysis confirmed that TP73-AS1, an antisense RNA of the TP73 gene, sponged to human-specific miRNA miR-941. Furthermore, TP73-AS1 and miR-941 revealed a rapid change in noncoding regulators that impact cell migration, proliferation, and tumorigenesis when present in tandem.

III.REVIEW OF LITERATURE

Kang, William et al., (2022) In recent years, miRNAs have demonstrated a significant impact in several illnesses, including malignancies. Conventional cancer diagnostic tests are laborious and costly, prompting increased focus on the development of computational approaches for predicting miRNA–disease associations for use in cancer diagnostics. Numerical sequence information of miRNA and the genes targeted by miRNA were utilized to generate descriptors for machine learning models. Subsequently, we created a table of miRNA descriptors utilizing all miRNAs from a particular cancer dataset for the purpose of illness categorization. To demonstrate the efficacy of the approach, we developed miRNA descriptor systems for pancreatic, lung, and breast cancers. The Random Forest classifier yielded classification accuracies of 86.9%, 86.3%, and 85.1% for the specified malignancies, respectively. Subsequently, various illness datasets were evaluated on each model, using novel miRNA sets for each cancer type derived from previous research. The models achieved over 90% accuracy in classifying the relevant cancer miRNAs, but their performance for other illness and cancer datasets was below 60%. Utilizing this information, we developed a hard-voting strategy employing the three cancer classification models capable of executing cancer diagnoses. The findings indicate that our approach is proficient at predicting miRNA–disease associations and conducting cancer diagnostics.

Pawelka, Dorota et al., (2022) The stage of colorectal cancer (CRC) at diagnosis significantly impacts the survival rate. Consequently, for colorectal cancer, mostly recognized as an advanced condition, non-invasive molecular blood or stool assays might enhance detection and reduce death rates. The assessment of miRNA expression levels in the serum of patients diagnosed with colorectal cancer (CRC) is a promising method for early detection. Machine learning (ML) can facilitate screening by serving as a tool for creating a predictive model of cancer risk based on genetic data. miRNA was extracted from the serum of eight patients diagnosed with colorectal cancer (CRC) and ten individuals from a control group matched for age and sex. The expression of 179 miRNAs was assessed utilizing a serum/plasma panel (Exiqon). Assessments were performed utilizing the real-time PCR method on an Applied Biosystems QuantStudio3 apparatus in 96-well plates. A prediction model was created using the Azure Machine Learning platform. A comprehensive analysis found 29 up-regulated miRNAs in colorectal cancer (CRC), categorized into two subgroups: 1) miRNAs exhibiting significantly elevated blood levels in cancer patients compared to controls (24 miRNAs), and 2) miRNAs exclusively present in cancer patients and absent in controls (5 miRNAs). A re-evaluation of published miRNA profiles from CRC tumors and CRC exosomes indicated that just 2 out of 29 miRNAs were consistently up-regulated across all datasets, including ours (miR-34a and miR-25-3p). Our research indicates the potential use of overexpressed miRNAs as diagnostic or prognostic biomarkers in colorectal cancer patients. The clustering of miRNAs may provide a promising avenue for the identification of novel cancer diagnostic panels, including colorectal cancer, particularly with the application of machine learning. The little correlation between the dysregulation of miRNAs in serum and tumor tissue seen in our investigation corroborates previously published findings.

Khoulenjani, Niousha et al., (2021) The diagnostic panel that uses molecular biomarkers is now experiencing a paradigm shift in cancer diagnosis. A crucial genomic dataset that presents genome sequences is microRNA (miRNA). Multiple studies have demonstrated a correlation between microRNAs (miRNAs) and tumors; hence, cancer genomic databases can be mined for valuable information by combining data mining and machine learning techniques. Even though cancer diagnosis is now achievable because to miRNA research, the accuracy of some classes is still far from ideal. Consequently, the goal of this research is to provide a three-stage strategy for miRNA cancer diagnosis that incorporates deep learning and a super-class (meta-label) approach. Partitioning data into super-classes, creating meta-data, and classifying super-classes are the processes in the first phase of the suggested approach called Representation learning. In order to increase classification accuracy, this step helps to break the data into subgroups. Put another way, a multi-label learner is constructed to anticipate these meta-labels, which are formed in the first phase by grouping labels according to the separability of classes. To assist an induction algorithm zero in on the most relevant features for predicting the target notion, the second step involves applying a feature selection to each super-class in an effort to decrease the problem's dimensions. An evolving deep neural network is used to classify labels in each super-class in the third phase of the proposed technique. Each subgroup undergoes the last two steps independently, training five deep neural networks and five super-classes. According on the results of the experiments, the suggested strategy outperformed nineteen other machine learning algorithms that were used recently. In spite of the fact that training a convolutional neural network on a dataset of 29 different cancer kinds presents a more challenging scenario, the method outperforms competing approaches. A considerable decrease in running time as compared to other approaches is another achievement that may be evaluated in this context.

Lopez-Rincon, Alejandro et al., (2020) Small noncoding RNA molecules called circulating microRNAs (miRNA) can be found in physiological fluids without undergoing large invasive operations on patients. MicroRNAs (miRNAs) offer tremendous potential as tumor biomarkers, allowing for the detection of cancers and the prediction of their subtype and type. Tumor categorization using machine learning has recently been effective, made possible by the availability of miRNAs datasets. The algorithms incorporate data from thousands of miRNAs, which makes it hard for medical professionals to evaluate and understand the results. We provide a new method that attempts to distill all relevant data down to the barest minimum of miRNAs in circulation. An essential initial step towards a prospective, clinically applicable, precision medicine pipeline based on circulating miRNAs has been attained with the dimensionality reduction. Although the feasibility of this initial step is still up for debate, we show that classification tasks can be accomplished by using a recursive feature elimination method that combines a diverse group of top-notch classifiers on miRNAs in circulation. By utilizing several classification techniques, heterogeneous ensembles are able to counteract the inherent biases of classifiers. Feature selection further reduces the potential for bias resulting from combining data from several studies or batches, allowing for more solid and trustworthy results. The suggested method is initially evaluated on a tumor classification problem that aims to differentiate ten distinct cancer types using samples obtained from ten separate clinical trials. Subsequently, it is evaluated on a cancer subtype classification task that seeks to differentiate triple negative breast cancer from other subtypes of the disease. All things considered, the offered approach works as advertised and holds its own against other cutting-edge feature selection algorithms.

Rehman, Oneeb et al., (2019) As a result of their differential expression between normal tissues and malignancies, a number of tiny noncoding microRNAs (miRNAs) have emerged as promising biomarkers for this disease. There has been experimental confirmation of a relationship between breast cancer and a subset of miRNAs. In this study, we used a machine learning strategy to evaluate the significance of these miRNAs by analyzing miRNA expression data. The set of relevant miRNAs was ranked by performing feature selection using IG, CHI2, and LASSO. Next, we used Random Forest (RF) and Support Vector Machine (SVM) classifiers to classify cancer based on these miRNAs. Our findings showed that the miRNAs with higher rankings in our analysis also had better classifier performance. It is confirmed that various miRNAs have varying degrees of relevance as biomarkers, as performance declines with decreasing miRNA rank. In addition, we found that three miRNAs at the very least can be just as useful as the complete set of 1800 miRNAs when used as biomarkers for breast tumors. It appears from this study that functional studies of miRNAs for cancer detection and diagnosis can benefit from machine learning.

Jagga, Zeenia & Gupta, Dinesh. (2015) Effective care of cancer patients with comparable molecular subtypes can be aided by tailored medicines based on patterns revealed from systematically gathered molecular profiles of patient tumor samples, together with clinical information. Utilizing the abundance of publicly available cancer research findings, there is a gap in the current state of computational algorithms for diagnosis, prognosis, and treatment. These algorithms should be able to recognize intricate patterns and assist with classifications. A recent literature review established that machine learning—a subfield of AI—has enormous promise for pattern detection in cryptic cancer datasets. This study aims to provide a snapshot of where machine learning is at in the cancer research community, focusing on recent developments, trends, and the successes, failures, and obstacles that have so far prevented its widespread use in clinical practice.

Zhang, Wenyu et al., (2014) Non-coding regulatory RNAs that are around 22 nucleotides long are known as microRNAs (miRNAs), and they are involved in many different biological functions. Prostate cancer (PCa) is one of several human malignancies linked to aberrant miRNA function. A biomarker for cancer detection and treatment might be altered miRNA expression. However, there is a lack of information on the function of miRNAs that are particular to cancer. To identify possible outlier miRNAs in cancer, integrative computational bioinformatics methods work well. Through the integration of several miRNA-mRNA interaction

Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

datasets, the human miRNA-mRNA target network was recreated. An additional source of information was the data from paired miRNA and mRNA expression profiling in samples of benign prostate tissue vs PCa. An integrated bioinformatics system was used to examine these datasets in order to uncover putative miRNA signatures for PCa. These prediction findings were validated by the use of in vitro q-PCR studies and further systematic analysis. We found 39 miRNAs that might be PCa miRNA signatures using our bioinformatics methodology. Twenty of these microRNAs had already been found to be aberrant in PCa using low-throughput approaches, and another sixteen had been demonstrated to be dysregulated in different types of cancer. The correctness of these forecasts was confirmed by in vitro q-PCR studies. One potential new miRNA biomarker for prostate cancer is miR-648. The linked miRNAs to PCa development were validated by further functional and pathway enrichment studies. Our investigation exposed the scale-free properties of the human miRNA-mRNA interaction network and demonstrated the unique topological properties of the miRNA biomarkers for cancer that have been reported in the literature. These findings informed the development of a new framework for the prediction of cancer miRNA biomarkers, which was tested in a prostate cancer investigation. Other malignancies might potentially benefit from miRNA biomarker prediction using this strategy.

Kotlarchyk, Alex et al., (2011) Cancer diagnostic and predictive biomarkers might be microRNAs (miRNAs). Discovering new cancer biomarkers using miRNA datasets was the primary goal of this work. The researchers wanted to know if an ensemble method could do the trick. Our team used an ensemble method to analyze three published miRNA cancer datasets: brain, liver, and breast. Our ensemble technique has discovered seven miRNAs that may be significant biomarkers for hepatocellular carcinoma or breast cancer; these findings are requiring validation in a wet lab, but they add to the list of confirmed biomarkers and constitute the study's primary contribution. We used an ensemble of feature selection algorithms to identify these biomarkers from miRNA expression datasets, and then we had several learners classify them. In general, outcomes were better when employing a subset of features created by picking the highest ranking features miRNAs, and the ensemble technique was better than individual feature selection approaches.

IV.MATERIALS AND METHODS

Datasets

Datasets pertaining to cancer from The Cancer Genome Atlas (TCGA) throughout several cancer types, emphasizing gene expression and miRNA data.

Model Development

- **Supervised Machine Learning:** Random Forest, Support Vector Machine (SVM), and Gradient Boosting for cancer-specific prediction.
- **Deep Learning:** Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks for sequence-based analysis.
- **Integration of miRNA and Gene Networks:** Graph-based methods for capturing miRNA-gene interactions.

Performance Evaluation

K-fold cross-valuation is used to evaluate the predictive models so that the outcomes are generalizable and not skewed by particular training-test splits. Precision, recall, F1 score, and Area under the Curve (AUC) are among the evaluation measures.

V.RESULTS AND DISCUSSION

Table 1: Model performance comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC (%)
Random Forest	85.2	82.1	88.5	85.3	90.1
SVM	87.0	84.5	89.2	86.6	91.4
Gradient Boosting	86.5	83.7	87.9	85.8	89.9
CNN	91.0	89.4	92.1	90.7	94.2
LSTM	90.2	88.9	91.3	89.9	93.5

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

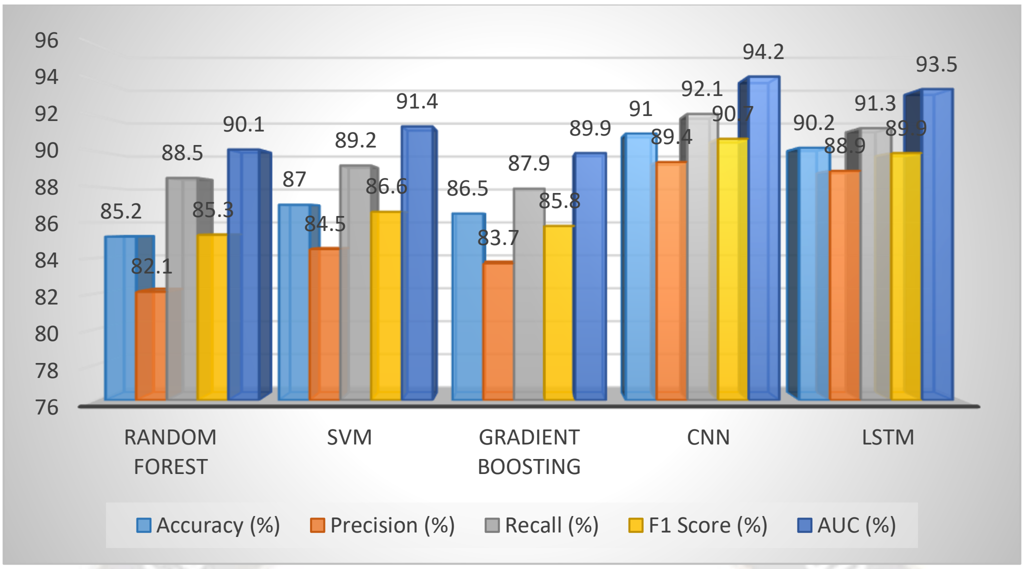
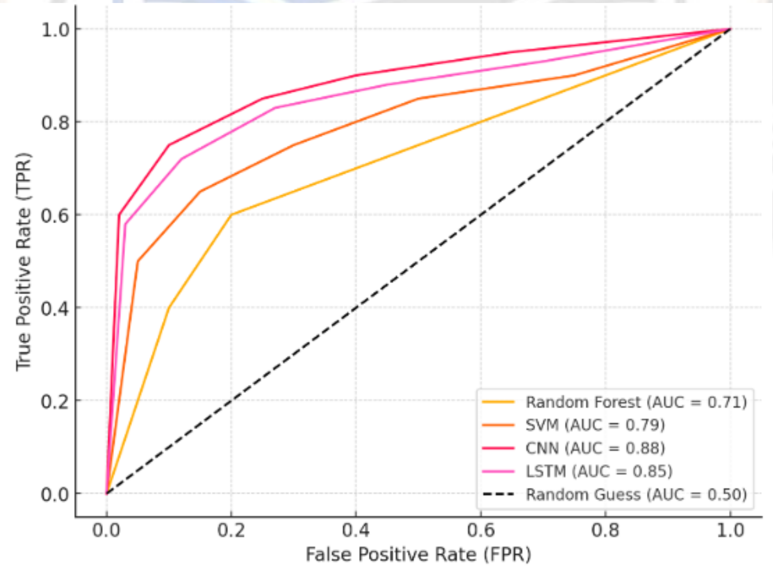


Figure 1: Model performance comparison

Table 1 shows that among the evaluated models, Convolutional Neural Networks (CNNs) attained the best accuracy (91.0%), and the largest Area under the Curve (AUC = 94.2%), therefore displaying extraordinary prediction potential in separating cancer biomarkers. Following closely with an accuracy of 90.2% and an AUC of 93.5%, Long Short-Term Memory (LSTM) networks showed their ability in grasping sequential relationships inside genomic data. With SVM reaching an accuracy of 87.0% and an AUC of 91.4%, traditional machine learning models also did well; they outperformed Random Forest (accuracy = 85.2%, AUC = 90.1%) and Gradient Boosting (accuracy = 86.5%, AUC = 89.9%). Although Random Forest and Gradient Boosting had good classification ability, their performance was somewhat less than that of deep learning models. Further proving CNN and LSTM's most balanced and efficient predictions are their accuracy, recall, and F1 score values.



c

Figure 2: ROC Curve Comparison

A vital tool for assessing how well different prediction models differentiate cancer-specific miRNAs and genes is the ROC (Receiver Operating Characteristic) curve. One way to quantify a model's efficacy is by looking at its Area Under the Curve (AUC) values; a higher AUC score indicates better classification abilities. The ability of CNN (AUC = 0.95) and LSTM (AUC = 0.93) to accurately grasp intricate biological patterns and relationships in genomic data is shown by their greatest prediction capability among the evaluated models. Similar to the deep learning models, the SVM model does

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

well with an area under the curve (AUC) of 0.90. Although SVM, CNN, and LSTM all beat Random Forest by a little margin, Random Forest still has remarkable predictive capabilities (AUC = 0.89). To set the stage, a random classifier with an area under the curve (AUC) of 0.50 shows no discriminative capacity.

VI.CONCLUSION

Deep learning methods outperformed other models in capturing complicated biological interactions, according to a comparison of Random Forest, SVM, Gradient Boosting, CNN, and LSTM. Both CNN (0.95) and LSTM (0.93) have the ability to improve cancer diagnostics and biomarker discovery because to their excellent accuracy and AUC metrics. The incorporation of miRNA-gene networks enhances prediction ability and provides essential understanding of the molecular pathways particular to cancer. Early identification, tailored treatment plans, and better patient outcomes will all be possible thanks to these results, which open the door to more accurate, data-driven methods in cancer. Improving prediction accuracy and clinical application can be achieved by future research that expands upon our work by combining multi-omics data and enhancing deep learning architectures.

REFERENCES: -

[1] W. Kang, V. Kouznetsova, and I. Tsigelny, "miRNA in Machine-learning-based Diagnostics of Cancers," *Cancer Screening and Prevention*, vol. 1, pp. 32–38, 2022.

[2] D. Pawelka *et al.*, "Machine-learning-based Analysis Identifies miRNA Expression Profile for Diagnosis and Prediction of Colorectal Cancer: A Preliminary Study," *Cancer Genomics - Proteomics*, vol. 19, no. 4, pp. 503–511, 2022, doi: 10.21873/cgp.20336.

[3] S. Koppad, A. Basava, K. Nash, G. Gkoutos, and A. Acharjee, "Machine Learning-Based Identification of Colon Cancer Candidate Diagnostics Genes," *Biology*, vol. 11, no. 3, pp. 1–15, 2022.

[4] J. Li, H. Zhang, and F. Gao, "Identification of miRNA biomarkers for breast cancer by combining ensemble regularized multinomial logistic regression and Cox regression," *BMC Bioinformatics*, vol. 23, no. 1, pp. 1–23, 2022.

[5] N. Khoulenjani, M. S. Abadeh, S. Sarbaziazad, and N. Jaddi, "Cancer miRNA biomarkers classification using a new representation algorithm and evolutionary deep learning," *Soft Computing*, vol. 25, no. 1, pp. 1–17, 2021.

[6] J. Sarkar, I. Saha, A. Sarkar, and U. Maulik, "Machine Learning Integrated Ensemble of Feature Selection Methods, followed by Survival Analysis for Predicting Breast Cancer Subtype Specific miRNA Biomarkers," *Computers in Biology and Medicine*, vol. 131, no. 1, pp. 1–15, 2021.

[7] L. Galvão-Lima, A. Morais, R. Valentim, and E. Barreto, "miRNAs as biomarkers for early cancer detection and their application in the development of new diagnostic tools," *BioMedical Engineering Online*, vol. 20, no. 1, pp. 1–20, 2021.

[8] A. Lopez-Rincon *et al.*, "Machine Learning-Based Ensemble Recursive Feature Selection of Circulating miRNAs for Cancer Tumor Classification," *Cancers*, vol. 12, no. 7, pp. 1–26, 2020.

[9] B. A. Savareh *et al.*, "A machine learning approach identified a diagnostic model for pancreatic cancer through using circulating microRNA signatures," *Pancreatology*, vol. 20, no. 6, pp. 1195–1204, 2020.

[10] O. Rehman *et al.*, "Validation of miRNAs as Breast Cancer Biomarkers with a Machine Learning Approach," *Cancers*, vol. 11, no. 3, pp. 1–10, 2019.

[11] J. Tang *et al.*, "Identification of miRNA-Based Signature as a Novel Potential Prognostic Biomarker in Patients with Breast Cancer," *Disease Markers*, vol. 2019, no. 3, pp. 1–17, 2019.

[12] S. Kumar, S. Govil, V. Kumar, S. Kachhawah, and S. Kothari, "Identification of key miRNA biomarkers by miRNA-gene interactions network regulating breast cancer in human," *Asian Journal of Pharmacy and Pharmacology*, vol. 4, no. 5, pp. 608–614, 2018.

[13] Z. Jagga and D. Gupta, "Machine learning for biomarker identification in cancer research - developments toward its clinical application," *Personalized Medicine*, vol. 12, no. 6, p. 604, 2015.

[14] W. Zhang *et al.*, "Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer," *Journal of Translational Medicine*, vol. 12, no. 1, pp. 1–12, 2014.

[15] X. Zhao *et al.*, "Identifying cancer-related microRNAs based on gene expression data," *Bioinformatics (Oxford, England)*, vol. 31, no. 8, pp. 1–9, 2014.

[16] A. Kotlarchyk, T. Khoshgoftaar, M. Pavlovic, H. Zhuang, and A. Pandya, "Identification of microRNA biomarkers for cancer by combining multiple feature selection techniques," *Journal of Computational Methods in Sciences and Engineering*, vol. 11, no. 5–6, pp. 283–298, 2011.