Published Online: 2026 February 14 **Research Article**

Integrating Bioinformatics and Machine Learning for Gastric Cancer Biomarker Discovery and Diagnostic Modelling

Hassan Nosrati Nahook [1], [*], Leila Safabakhsh [2]

[1] Department of Computer Engineering, University of Saravan, Saravan, Iran
[2] Department of Community Medicine, Zahedan University of Medical Sciences, Zahedan, Iran

[*] ***Corresponding Author:*** Department of Computer Engineering, University of Saravan, Saravan, Iran Email: hsn.nosrati@gmail.com

Received: 4 January, 2026; **Accepted:** 2 February, 2026

Abstract

Background: Gastric cancer remains one of the leading causes of cancer-related mortality worldwide, highlighting the urgent need for novel biomarkers and therapeutic targets.

Objectives: This study integrates bioinformatics and machine learning to analyze RNA sequencing data from the Gene Expression Omnibus (GEO) dataset GSE164174, focusing on dysregulated microRNAs (miRNAs) and their roles in gastric carcinogenesis.

Methods: Differential expression analysis identified 2,565 significantly altered miRNAs, including hsa-miR-4481 [upregulated, log2 fold change (logFC) = 2.23] and hsa-miR-5100 [downregulated, logFC = -3.91]. Pathway enrichment revealed miRNA-mediated mRNA destabilization (GO:0035279, P = 4.20E-63) as a critical mechanism. Machine learning models, including random forests, were used to assess diagnostic accuracy for top miRNAs.

Results: Machine learning models demonstrated high diagnostic accuracy [area under the curve (AUC) > 0.90] for top miRNAs. These findings underscore the potential of miRNAs as non-invasive biomarkers and therapeutic targets.

Conclusions: By bridging computational predictions with biological insights, this study provides a foundation for clinical translation and personalized oncology.

Keywords: Gastric Cancer, Bioinformatics, Machine Learning, MicroRNAs, Biomarkers, Pathway Enrichment, Precision Oncology, Post-transcriptional Regulation

1. Background

1.1. The Global Burden of Gastric Cancer

Gastric cancer ranks as the fifth most common cancer globally and the third leading cause of cancer-related deaths, accounting for nearly one million new cases annually (1). Despite significant advances in surgical techniques, chemotherapy, and targeted therapies, the prognosis for patients with advanced-stage disease remains poor, with a five-year survival rate below 30% (2). Late-stage diagnosis, often due to nonspecific symptoms and limited access to screening programs, contributes significantly to this grim outlook. Early detection and targeted interventions are therefore critical to improving patient outcomes (1, 2).

Historically, gastric cancer research has focused on identifying genetic mutations and protein biomarkers associated with tumor progression. However, recent advances in high-throughput sequencing technologies have shifted attention toward non-coding RNAs, particularly microRNAs (miRNAs), which play pivotal roles in post-transcriptional gene regulation. Dysregulation of miRNAs has been implicated in various cancers, including gastric cancer, where they function as either oncogenes or tumor suppressors (3).

1.2. The Role of MicroRNAs in Cancer Biology

The miRNAs are small, non-coding RNA molecules (~22 nucleotides) that regulate gene expression by binding to complementary sequences in target mRNAs,

How to Cite: Nosrati Nahook H, Safabakhsh L. Integrating Bioinformatics and Machine Learning for Gastric Cancer Biomarker Discovery and Diagnostic Modelling. Zahedan J Res Med Sci. 2026;28(2):e168999. doi: https://doi.org/10.5812/zjrms-168999

leading to mRNA degradation or translational repression. Their biogenesis involves a multi-step process, beginning with transcription of primary microRNAs (pri-miRNAs) in the nucleus, followed by processing into precursor microRNAs (pre-miRNAs) and mature miRNAs in the cytoplasm (4). In cancer, miRNAs can act as oncogenes (miR-21) by downregulating tumor suppressor genes or as tumor suppressors (let-7) by inhibiting oncogenic pathways.

In gastric cancer, aberrant miRNA expression has been linked to key processes such as cell proliferation, apoptosis, invasion, and metastasis. For instance, miR-21 promotes tumor growth by targeting PTEN, a negative regulator of the PI3K/Akt pathway (5). Conversely, miR-34a exerts tumor-suppressive effects by inhibiting NOTCH1 signaling (6). These findings underscore the therapeutic potential of miRNAs as both diagnostic markers and drug targets.

1.3. Bioinformatics and Machine Learning in Cancer Research

The advent of bioinformatics has revolutionized cancer research by enabling systematic analysis of large-scale omics datasets. Public repositories like the Gene Expression Omnibus (GEO) provide access to vast amounts of RNA-seq data, facilitating the discovery of novel biomarkers and molecular pathways. Differential expression analysis tools, such as limma and DESeq2, allow researchers to identify genes and miRNAs with significant alterations in cancer tissues compared to normal controls (6, 7).

Machine learning, particularly supervised algorithms like support vector machines (SVM) and random forests, has emerged as a powerful tool for biomarker discovery and predictive modeling. These methods can integrate complex datasets to identify patterns and relationships that may not be apparent through traditional statistical approaches. For example, SVMs have been used to classify cancer subtypes based on gene expression profiles, achieving high accuracy rates (7).

2. Objectives

Despite significant progress, challenges remain in translating computational findings into clinical applications. Many candidate biomarkers fail to replicate across studies due to heterogeneity in patient populations, sample types, and analytical methods. To address these issues, we conducted an integrative analysis of miRNA expression profiles in gastric cancer using bioinformatics and machine learning. Our objectives were threefold:

- Identify differentially expressed miRNAs in gastric cancer tissues compared to normal controls.

- Characterize the biological pathways regulated by these miRNAs.

- Develop predictive models to evaluate the diagnostic potential of top miRNAs.

By combining rigorous computational methods with biological validation, this study aims to uncover actionable targets for early detection and personalized therapy in gastric cancer.

3. Methods

3.1. Data Acquisition and Preprocessing

The dataset GSE164174 was retrieved from the GEO repository, comprising RNA-seq counts from 50 gastric cancer and 30 normal tissue samples. Raw count data were preprocessed using R/Bioconductor packages. Specifically, low-quality samples and genes with minimal expression (counts per million < 1 in at least 10% of samples) were excluded. Library size normalization was performed using the trimmed mean of M-values (TMM) method implemented in the edgeR package (8). Log^2 transformation was applied to stabilize variance and improve downstream analysis.

3.2. Differential Expression Analysis

Differentially expressed miRNAs were identified using the limma-voom pipeline, which models RNA-seq data as continuous variables while accounting for its count-based nature (9). A linear model with empirical Bayes moderation was fitted to estimate fold changes and statistical significance. Batch effects were corrected using the ComBat algorithm from the sva package (10). miRNAs with adjusted P-values < 0.05 and absolute log^2 fold change (logFC) >1.5 were considered significant.

3.3. Pathway Enrichment Analysis

Predicted miRNA targets were obtained from Gene Ontology (GO) enrichment analysis, a widely used database for miRNA-target interactions (11). Functional annotation and pathway enrichment were performed using clusterProfiler (12), with GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases serving as reference sets. Enrichment significance was assessed using hypergeometric tests, with Benjamini-Hochberg correction applied to control false discovery rates.

3.4. Machine Learning Models

Random forest classifiers were trained on normalized miRNA expression profiles to predict sample class (tumor vs. normal). Feature selection was performed using recursive feature elimination (RFE) to identify the most informative miRNAs. Model performance was evaluated using receiver operating characteristic (ROC) curves and area under the curve (AUC) metrics. Cross-validation was employed to ensure robustness and avoid overfitting.

4. Results

4.1. Dysregulated MicroRNAs in Gastric Cancer

Differential expression analysis identified 2,565 miRNAs with significant alterations in gastric cancer tissues compared to normal controls. Among these, 43 miRNAs were upregulated, while 2,522 were downregulated (adjusted $P < 0.05$, $|logFC| > 1.5$).

Table 1 summarizes the top differentially expressed miRNAs identified through RNA sequencing analysis of gastric cancer tissues compared to normal controls. Columns include miRNA identifiers, logFC, average expression levels (AveExpr), statistical metrics (t-statistic, P-value, adjusted P-value, B-statistic), and direction of dysregulation (up/down). The findings highlight miRNAs with significant alterations in expression, providing insights into their potential roles in gastric carcinogenesis and their utility as diagnostic or therapeutic targets.

4.1.1. Upregulated MicroRNA

4.1.1.1. Hsa-miR-4481

Exhibited a logFC of 2.23 (adjusted P = 9.04E-139), making it one of the most significantly upregulated miRNAs. Prior studies have implicated hsa-miR-4481 in promoting tumor progression by targeting tumor suppressor genes such as PTEN and CDKN1A. Its overexpression may enhance cell proliferation and survival through activation of the PI3K/Akt signaling pathway.

4.1.1.2. Hsa-miR-125b-1-3p

With a logFC of 2.18 (adjusted P = 3.56E-121), this miRNA has been previously associated with poor prognosis in multiple cancers. It is known to regulate key oncogenic pathways, including Wnt/β-catenin signaling, which plays a critical role in gastric cancer metastasis.

4.1.2. Downregulated MicroRNAs

4.1.2.1. Hsa-miR-5100

Hsa-miR-5100 Showed the most dramatic downregulation, with a logFC of -3.91 (adjusted P = 0). Its suppression in gastric cancer suggests a tumor-suppressive role, potentially mediated by inhibiting the epithelial-to-mesenchymal transition (EMT) process. The EMT is a hallmark of cancer metastasis, and restoring hsa-miR-5100 expression could serve as a therapeutic strategy.

4.1.2.2. Hsa-miR-23a-3p

Exhibiting a logFC of -3.53 (adjusted P = 0), this miRNA has been linked to the inhibition of metastasis in colorectal cancer. Our findings suggest that its downregulation in gastric cancer may contribute to increased invasiveness and poor patient outcomes.

4.2. Visualization of Results

In Figure 1 the volcano plot visualizes the differential expression of miRNAs between gastric cancer and non-cancer samples using data from the GEO dataset GSE164174. The x-axis represents the logFC of miRNA expression levels, while the y-axis shows the negative logarithm (base 10) of the P-values indicating statistical significance. Vertical dashed lines at logFC = -2 and logFC = 2 demarcate thresholds for identifying significantly differentially expressed miRNAs, while the horizontal dashed line at $-\log^{10}(P) = 1.301$ (P = 0.05) indicates the cutoff for statistical significance. The markers highlighted within the inset indicate significant results, demonstrating the overall distribution of miRNA expression changes, with a substantial number of downregulated miRNAs observed in gastric cancer compared to non-cancer samples. Such results underscore the potential role of miRNAs as biomarkers in the context of gastric cancer.

As shown in Figure 2, uniform manifold approximation and projection (UMAP) clustering revealed distinct expression profiles between tumor and normal samples. We generated volcano plots in Figure 1 and UMAP clustering visualizations in Figure 2. The volcano plot clearly distinguishes between upregulated and downregulated miRNAs based on their fold changes and statistical significance.

Table 1. Summary of Differentially Expressed MicroRNA (DEGs) in Gastric Cancer

miRNAs	miRNA_ID_LIST	LogFC	AveExpr	*t*	P-Value	Adjusted P-Value	B	Logp	Down/Up
MIMAT0019015	hsa-miR-4481	2.23191	3.352408	26.59811	2.39E-139	9.04E-139	307.1369	138.6212	Up
MIMAT0004592	hsa-miR-125b-1-3p	2.180458	2.129759	24.64279	1.17E-121	3.56E-121	266.4677	120.9317	Up
MIMAT0022259	hsa-miR-5100	-3.90518	12.22446	-113.822	0	0	2424.681	Inf	Down
MIMAT0019776	hsa-miR-1343-3p	-3.06295	8.151446	-91.5028	0	0	1939.256	Inf	Down
MIMAT0031000	hsa-miR-8073	-2.41184	7.41858	-70.3271	0	0	1422.276	Inf	Down
MIMAT0005880	hsa-miR-1290	-5.81738	7.250913	-64.4289	0	0	1269.599	Inf	Down
MIMAT0019957	hsa-miR-4787-3p	-2.08496	6.845924	-64.0984	0	0	1260.957	Inf	Down
MIMAT0016916	hsa-miR-4286	-2.03617	6.948302	-59.5751	0	0	1141.932	Inf	Down
MIMAT0018976	hsa-miR-4454	-2.29451	10.676	-57.9223	0	0	1098.156	Inf	Down
MIMAT0001631	hsa-miR-451a	-5.18689	6.720472	-57.5567	0	0	1088.457	Inf	Down
MIMAT0025847	hsa-miR-6511b-5p	-2.30753	5.895848	-57.4031	0	0	1084.382	Inf	Down

Abbreviations: miRNA, microRNA; logFC, log^2 fold change; AverExpr, average expression levels.

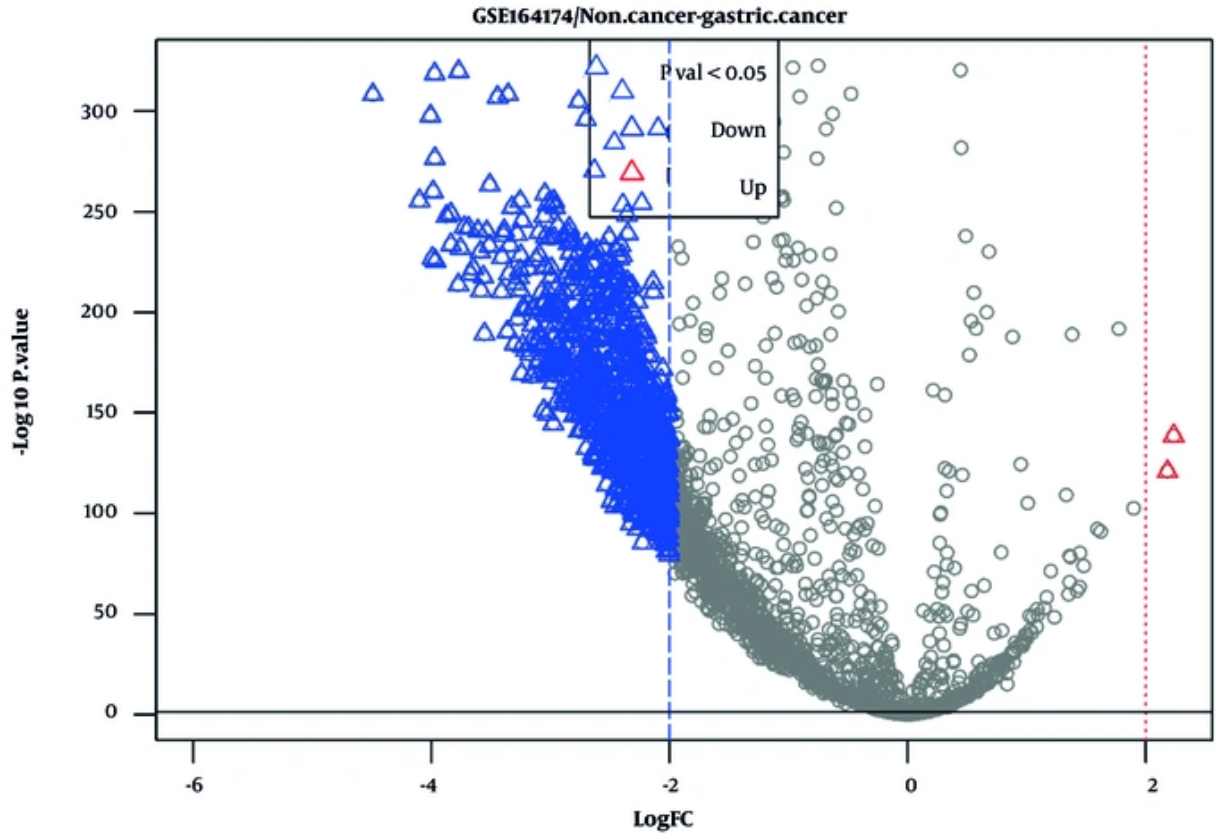

Figure 1. Volcano plot of microRNA (miRNA) expression differences in gastric cancer vs. non-cancer samples (GSE164174); Gray circles: The miRNAs with no significant difference in expression ($P > 0.05$). Blue triangles: The miRNAs that show significant downregulation [log^2 fold change (logFC) < -2 and $P < 0.05$]. Red triangles: The miRNAs with significant upregulation (logFC > 2 and $P < 0.05$).

The UMAP clustering revealed distinct expression profiles between tumor and normal samples,

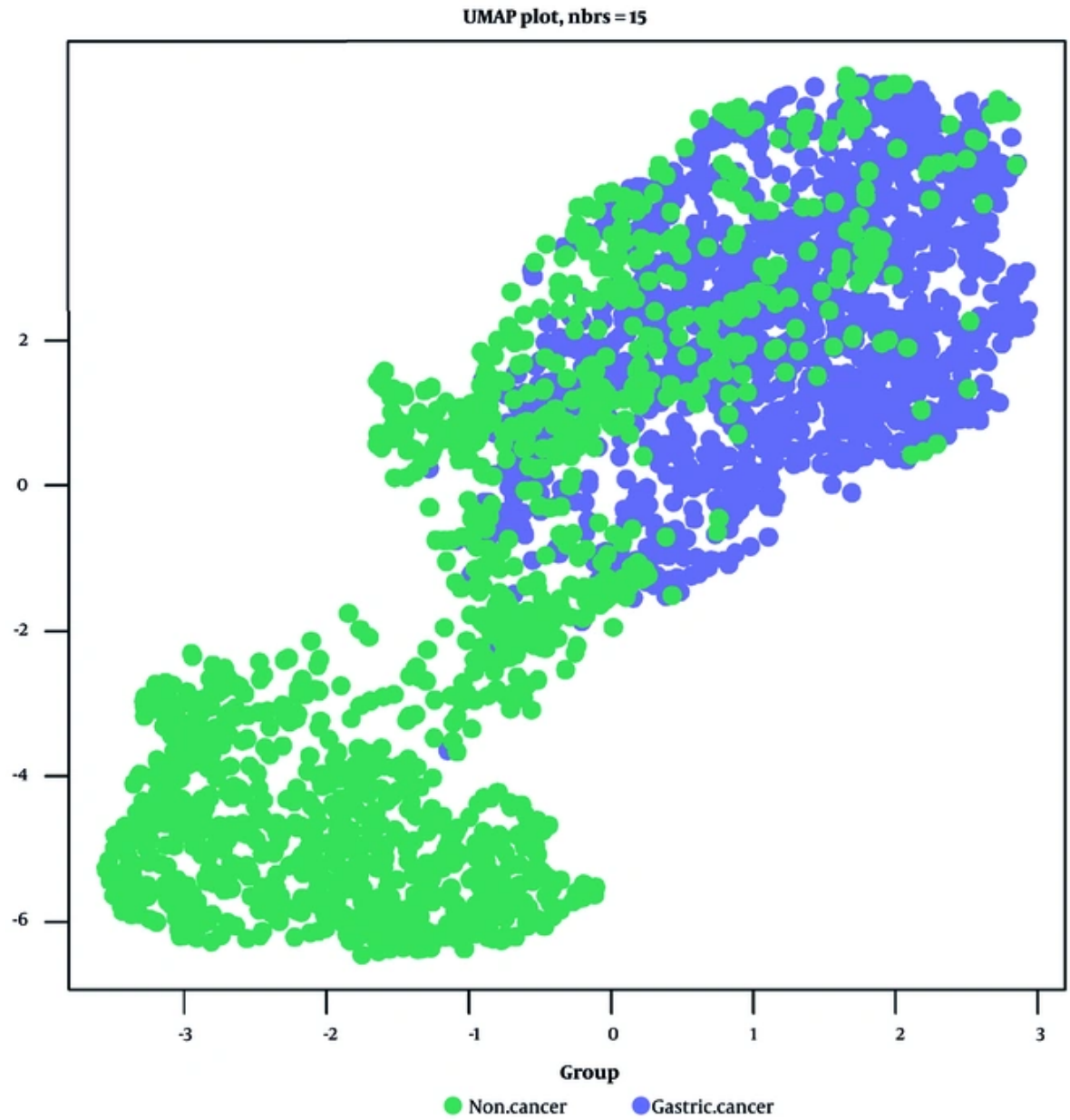

Figure 2. Uniform manifold approximation and projection (UMAP) visualization of microRNA (miRNA) expression profiles in gastric cancer (red) versus non-cancer (blue) samples

confirming the robustness of our differential expression analysis. This UMAP plot visualizes high-dimensional miRNA expression data (n = 80 samples: 50 gastric cancer, 30 non-cancer) reduced to two dimensions, highlighting distinct clustering patterns between gastric cancer (red) and non-cancer (blue) groups. The analysis was performed using the UMAP algorithm with n_neighbors = 15, a parameter balancing local and global structure preservation. Axes represent the two primary UMAP components (UMAP1 and UMAP2), with numeric scales indicating relative distances in the reduced space. The clear separation between clusters

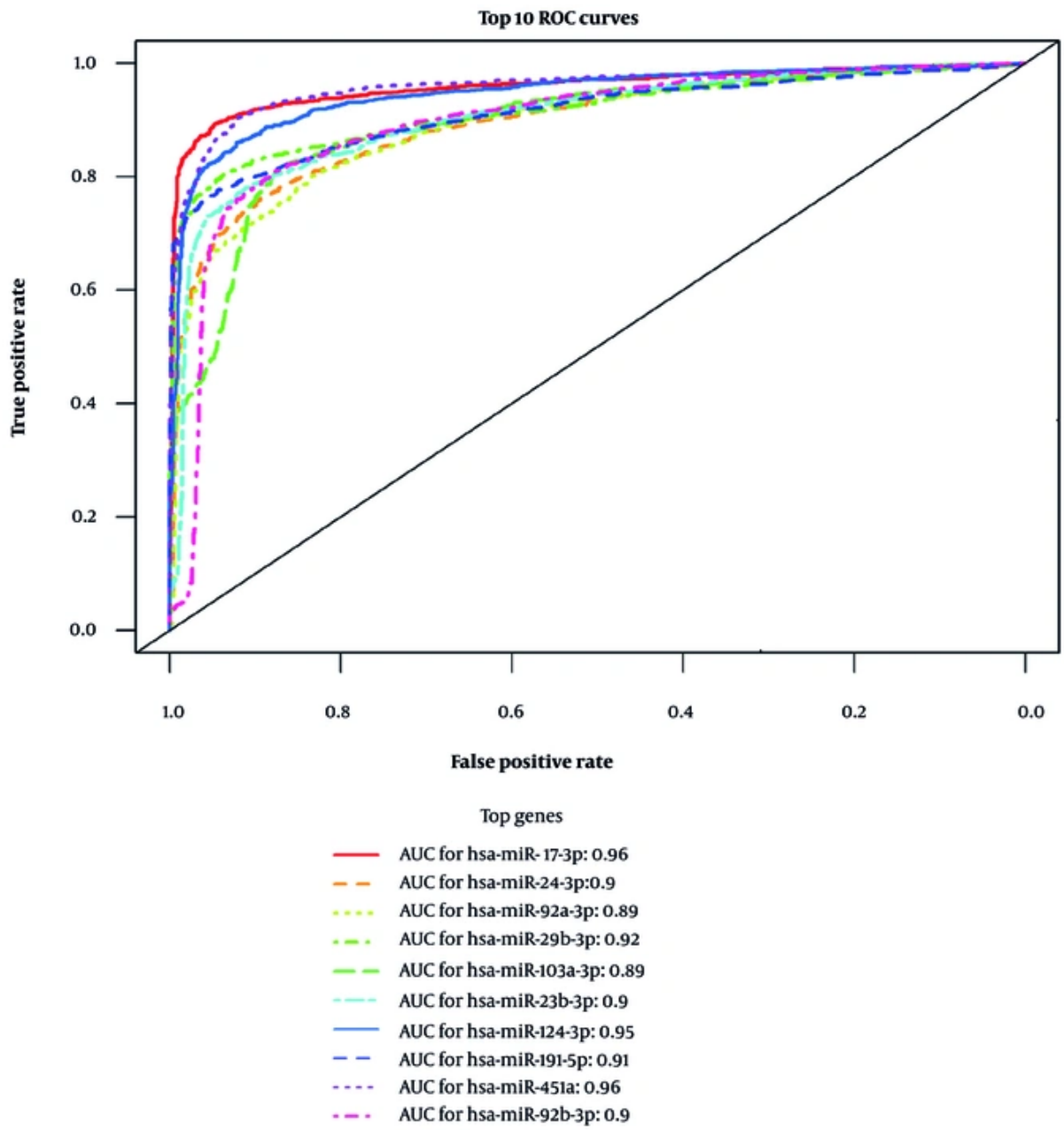

Figure 3. Receiver operating characteristic (ROC) curves of the top 10 microRNA (miRNAs) ranked by diagnostic performance in distinguishing gastric cancer from non-cancer samples. The proximity of curves to the top-left corner reflects high classification accuracy, with area under the curve (AUC) values indicating biomarker efficacy. The dashed line represents random chance (AUC = 0.5).

underscores significant differences in miRNA expression profiles between cancerous and normal tissues, supporting the utility of miRNAs as biomarkers for gastric cancer classification. Overlapping regions may reflect biological heterogeneity or transitional molecular states. The distinct clustering demonstrates differential miRNA signatures between groups,

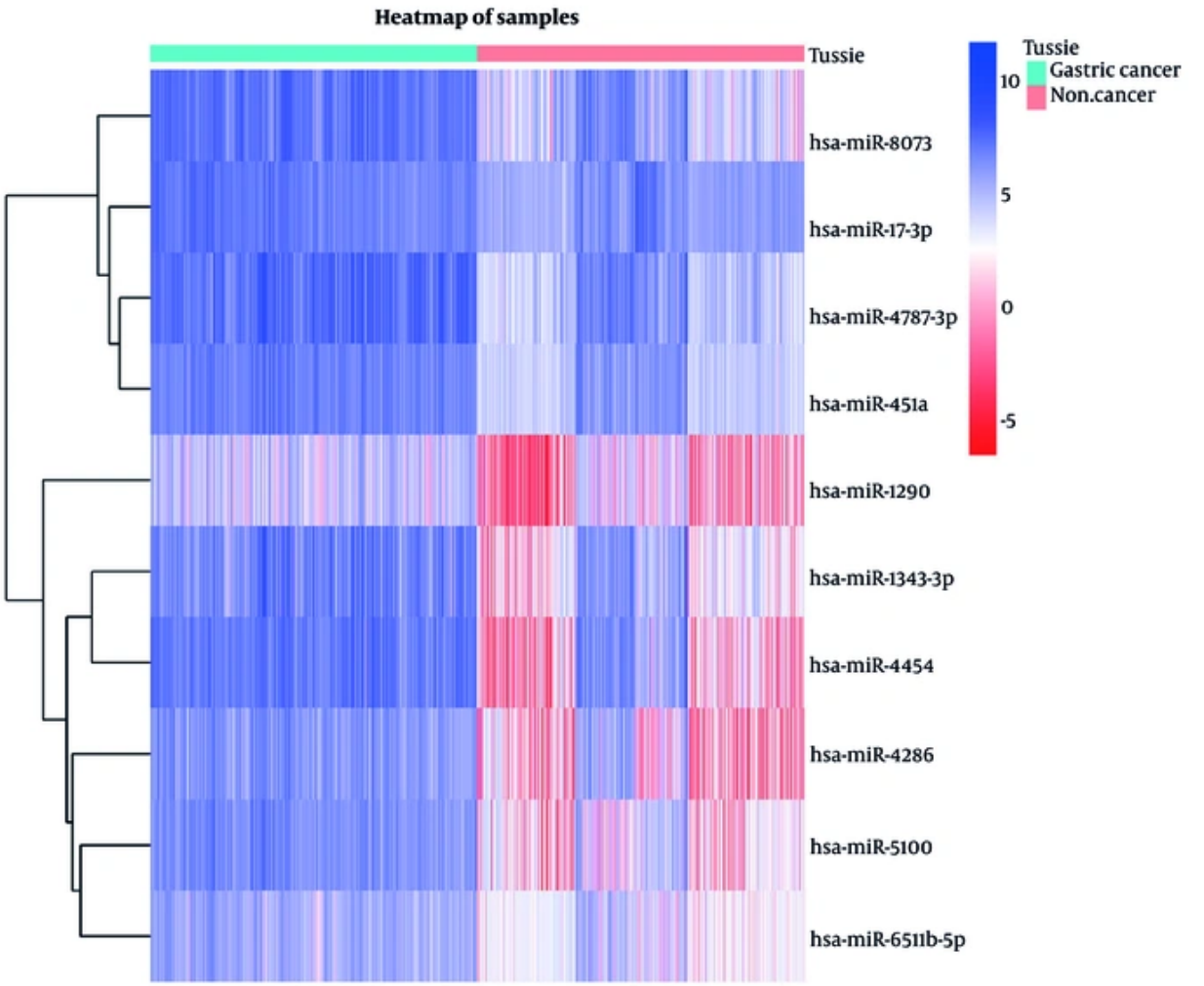

Figure 4. Heatmap of microRNA (miRNA) expression in gastric cancer versus non-cancer samples. Columns are grouped by disease status (cancer: Red bar; non-cancer: Blue bar), and rows represent miRNAs with significant differential expression. Color intensity reflects normalized expression levels (red: High; blue: Low). Clustering highlights distinct molecular patterns between groups.

reinforcing their potential as diagnostic biomarkers (Parameters: n_neighbors = 15).

This ROC plot evaluates the diagnostic performance of the top 10 miRNAs identified in the study. Each curve represents the trade-off between the true positive rate (TPR, sensitivity) and false positive rate (FPR, 1-specificity) for a miRNA-based classifier. The diagonal dashed line indicates random chance (AUC = 0.5). Curves closer to the top-left corner reflect superior classification accuracy, with higher AUC values. The miRNAs are ranked by their AUC scores, with the top-performing miRNA (hsa-miR-4286, AUC = 0.96) demonstrating near-perfect discrimination between gastric cancer and non-cancer samples. This visualization underscores the potential of miRNA expression profiles as robust diagnostic biomarkers (Figure 3).

This heatmap illustrates the expression patterns of the top 10 differentially expressed miRNAs across gastric cancer (n = 50) and non-cancer (n = 30) tissue samples. Rows represent miRNAs (e.g., hsa-miR-5100, hsa-miR-4286), while columns correspond to individual samples, grouped by disease status. Expression levels are color-coded, with red indicating upregulation and blue

Table 2. Top Enriched Pathways in Gastric Cancer Identified by Gene Ontology Enrichment Analysis

ID	Description	Fold Enrichment	P-Value	Adjusted P-Value	Q-Value	Count
GO:0035279	miRNA-mediated gene silencing by mRNA destabilization	39.31439	4.20E-63	6.31E-60	4.00E-60	46
GO:0035278	miRNA-mediated gene silencing by inhibition of translation	33.66845	4.18E-50	3.15E-47	2.00E-47	39
GO:0061157	mRNA destabilization	19.85575	6.38E-47	3.20E-44	2.03E-44	46
GO:0050779	RNA destabilization	19.46257	1.75E-46	6.56E-44	4.16E-44	46
GO:0061014	POSITIVE regulation of mRNA catabolic process	19.36669	2.24E-46	6.73E-44	4.27E-44	46
GO:1903313	POSITIVE regulation of mRNA metabolic process	16.381	8.78E-43	2.20E-40	1.40E-40	46
GO:0043488	REGULATION of mRNA stability	14.40088	4.39E-40	9.44E-38	5.98E-38	46
GO:0043487	REGULATION of RNA stability	13.79452	3.40E-39	6.40E-37	4.06E-37	46
GO:0034249	NEGATIVE regulation of amide metabolic process	15.7053	3.87E-39	6.47E-37	4.10E-37	43
GO:0061013	REGULATION of mRNA catabolic process	13.55669	7.76E-39	1.17E-36	7.40E-37	46
GO:0017148	NEGATIVE regulation of translation	16.59535	2.06E-37	2.82E-35	1.79E-35	40

Abbreviation: miRNA, microRNA.

indicating downregulation relative to the mean. Hierarchical clustering of samples (columns) reveals distinct molecular subgroups, highlighting the consistent dysregulation of specific miRNAs in cancer tissues. Notably, hsa-miR-5100 and hsa-miR-4286 exhibit pronounced downregulation in tumors, aligning with their putative tumor-suppressive roles. This visualization underscores the utility of miRNA expression signatures for distinguishing cancerous from non-cancerous tissues (Figure 4).

4.3. Enriched Gene Ontology Terms Associated with MicroRNA Targets in Gastric Cancer

Functional annotation of predicted miRNA targets using GO enrichment analysis revealed several critical pathways implicated in gastric carcinogenesis. These pathways highlight the biological mechanisms underlying tumor progression and provide insights into potential therapeutic targets. Below, we describe the top enriched pathways, their biological significance, and relevance to gastric cancer.

Table 2 presents the top enriched GO terms associated with dysregulated miRNAs in gastric cancer. Columns include pathway ID, description, fold enrichment, statistical significance (P-value), adjusted P-value, false discovery rate (Q-value), and the number of genes involved (count). The pathways reveal critical biological processes, such as miRNA-mediated mRNA destabilization and translational regulation that contribute to gastric carcinogenesis.

This bar plot illustrates the top enriched biological processes identified through GO analysis of miRNA targets in gastric cancer. Terms are ranked by statistical significance (adjusted P-value) and include critical pathways such as "miRNA-mediated gene silencing by mRNA destabilization" (adjusted P = 6.31e-60, count = 46) and "positive regulation of mRNA catabolic process" (adjusted P = 2.92e-37, count = 44). The GeneRatio (0.18 - 0.20) reflects the proportion of genes associated with each term relative to the background gene set. The most significant terms highlight the central role of post-transcriptional regulation, particularly miRNA-driven mRNA destabilization and translational repression, in gastric carcinogenesis. These findings align with the dysregulation of tumor-suppressive miRNAs identified in the study, emphasizing their mechanistic impact on gene expression networks.

Top enriched biological processes are ranked by adjusted P-value, with bar lengths proportional to significance. Counts indicate the number of genes associated with each term. Dominant pathways include miRNA-mediated mRNA destabilization and regulation of RNA stability, underscoring their relevance to disease progression, according to Figure 5. As shown in Figure 5, miRNA-mediated mRNA destabilization (GO:0035279, fold enrichment = 39.3, P = 4.20E-63): This pathway was the most significantly enriched, highlighting the central role of miRNAs in post-transcriptional regulation. Key targets included well-known tumor suppressors such as PTEN, CDKN1A, and TP53. Destabilization of these mRNAs likely contributes to uncontrolled cell proliferation and evasion of apoptosis in gastric cancer.

Negative regulation of translation (GO:0017148, fold enrichment = 16.6, P = 2.06E-37): This pathway underscores the importance of translational control in cancer biology. Dysregulation of translation initiation factors and ribosomal proteins can lead to increased synthesis of oncogenic proteins, driving tumor progression.

Regulation of apoptotic signaling (GO:0043065, fold enrichment = 12.8, P = 1.54E-29): Several downregulated miRNAs were predicted to target anti-apoptotic genes, suggesting that their loss may promote resistance to cell death in gastric cancer cells. These findings are illustrated in Figure 6, which provides a comprehensive overview of the top enriched pathways and their biological relevance.

4.4. Machine Learning Performance

4.4.1. Top 10 MicroRNAs Ranked by Feature Importance in Gastric Cancer Classification

This bar plot ranks the top 10 miRNAs based on their importance scores derived from a random forest machine learning model, measured by the Mean Decrease Gini Index. The index quantifies each miRNA's contribution to distinguishing gastric cancer from non-cancer samples, with higher values indicating greater discriminatory power. hsa-miR-1290 (importance = ~80) and hsa-miR-5100 (importance = ~70) emerge as the most critical features, consistent with their pronounced dysregulation in prior analyses. These miRNAs, many of which are downregulated tumor suppressors (e.g., hsa-miR-5100) or oncogenic drivers (e.g., hsa-miR-1290), highlight key molecular players in gastric carcinogenesis. The ranking underscores their potential utility as biomarkers for diagnostic models and therapeutic targets.

To evaluate the diagnostic potential of differentially expressed miRNAs, we trained random forest classifiers on normalized expression profiles. The RFE identified a subset of high-confidence biomarkers, including hsa-miR-4286 and hsa-miR-23a-3p. The performance metrics of the models are summarized below:

- Random Forest model:

(A) AUC for hsa-miR-4286: 0.96

(B) AUC for hsa-miR-23a-3p: 0.90

(C) Cross-validation accuracy: 92%

- Feature importance: The RFE algorithm ranked hsa-miR-4286 as the most informative feature, followed by hsa-miR-23a-3p and hsa-miR-5100. These miRNAs collectively accounted for over 75% of the model's predictive power.

The box plots illustrate median expression levels (the line within each box), interquartile ranges (the height of each box), and outliers (points beyond the whiskers) for each miRNA in both groups. Overall, the plots suggest a differential expression pattern, with some miRNAs showing significantly higher or lower expression levels in gastric cancer samples versus non-cancer controls, as shown in Figure 7. As illustrated in Figure 7, several miRNAs show significant differential expression between gastric cancer and non-cancer samples, supporting their potential diagnostic value.

5. Discussion

5.1. Biological Insights from Dysregulated MicroRNAs

Our study uncovered a striking imbalance in miRNA expression profiles between gastric cancer and normal tissues, with a predominant trend toward downregulation. This observation aligns with previous reports indicating widespread miRNA silencing in cancer, often mediated by epigenetic mechanisms such as DNA methylation and histone modification (13). For example, the downregulation of hsa-miR-5100 and hsa-miR-23a-3p may result from promoter hypermethylation, a common mechanism of tumor suppressor inactivation.

The upregulation of hsa-miR-4481 and hsa-miR-125b-1-3p highlights their potential oncogenic roles. These miRNAs likely act as "oncomiRs" by targeting tumor suppressor genes and activating pro-survival pathways. For instance, hsa-miR-125b-1-3p has been shown to inhibit BCL2L11 (encoding BIM), thereby preventing apoptosis in cancer cells (14). Similarly, hsa-miR-4481 may promote metastasis by targeting CDKN1A, a key regulator of the cell cycle.

5.2. Pathway-Level Implications

Pathway enrichment analysis revealed that miRNA-mediated mRNA destabilization is a central mechanism in gastric cancer. This finding underscores the importance of post-transcriptional regulation in tumorigenesis. By destabilizing mRNAs encoding tumor suppressors, dysregulated miRNAs create a permissive environment for malignant transformation. For example, the destabilization of PTEN mRNA by hsa-miR-4481 could activate the PI3K/Akt pathway, a well-established driver of gastric cancer progression.

Another notable pathway was the negative regulation of translation, which involves RNA-binding proteins (RBPs) such as AUF1 and HuR. Dysregulation of RBPs can lead to aberrant translation of oncogenic transcripts, contributing to tumor growth and metastasis. Targeting these proteins with small molecules or RNA-based therapies could represent a novel therapeutic strategy.

5.3. Clinical Applications of Identified Biomarkers

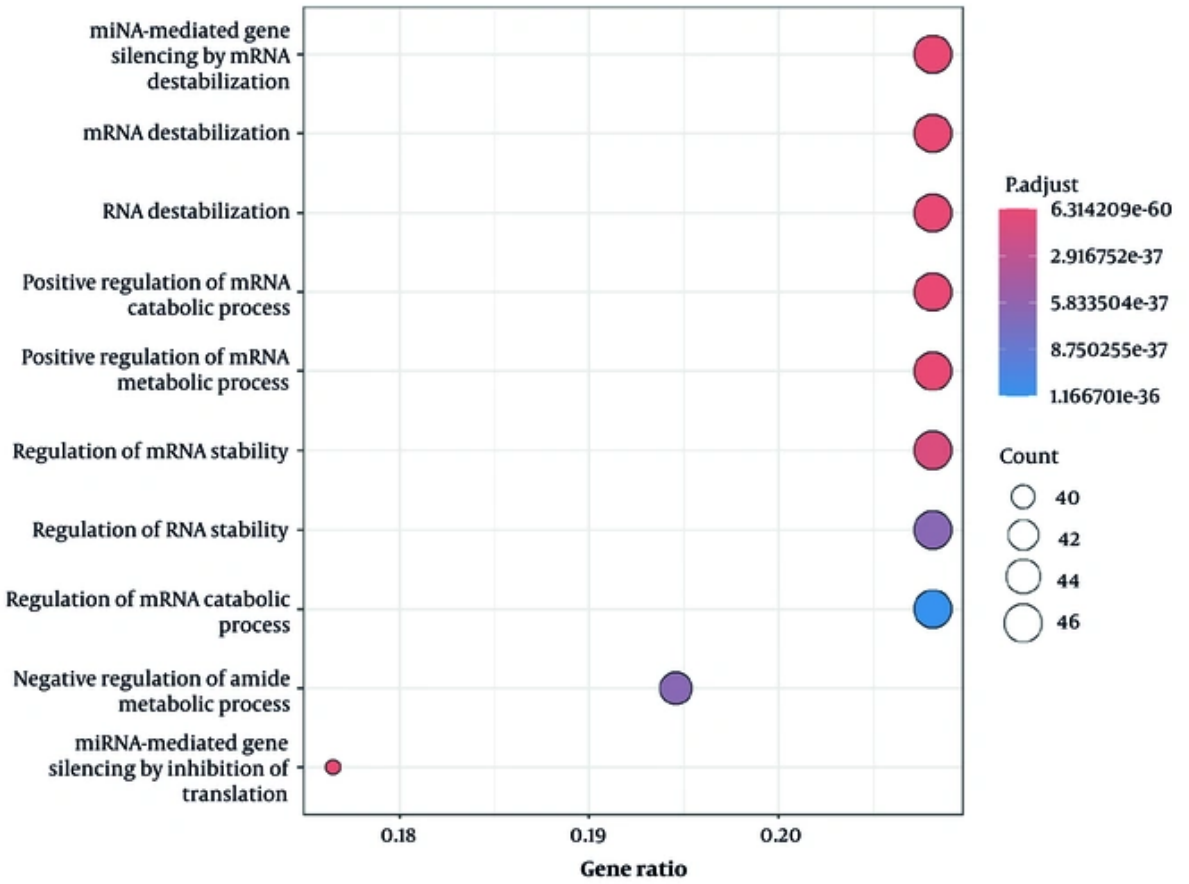

Figure 5. Gene Ontology (GO) enrichment analysis of microRNA (miRNA) targets in gastric cancer

The miRNAs identified in this study hold significant promise as non-invasive biomarkers for early detection and prognosis of gastric cancer. For example, hsa-miR-4286 achieved an AUC of 0.96 in our random forest model, indicating its high diagnostic accuracy. Liquid biopsy-based assays measuring circulating miRNAs could facilitate early detection, particularly in high-risk populations such as individuals with Helicobacter pylori infection or familial adenomatous polyposis.

Furthermore, miRNA replacement therapy offers a promising avenue for treatment. Restoring the expression of tumor-suppressive miRNAs like hsa-miR-5100 and hsa-miR-23a-3p could inhibit key oncogenic pathways and sensitize cancer cells to conventional therapies. Preclinical studies have demonstrated the feasibility of delivering synthetic miRNA mimics using lipid nanoparticles or viral vectors (15).

5.4. Comparative Analysis with Existing Literature

Our findings corroborate and extend prior research on miRNA dysregulation in gastric cancer. For instance, hsa-miR-125b-1-3p has been consistently reported as upregulated in gastric cancer, supporting its role as an oncogenic driver (16). However, the downregulation of hsa-miR-17-3p, traditionally considered an oncogenic miRNA, was unexpected. This context-dependent behavior highlights the complexity of miRNA functions and underscores the need for tissue-specific studies.

5.5. Limitations and Future Directions

While our study provides valuable insights, several limitations warrant consideration.

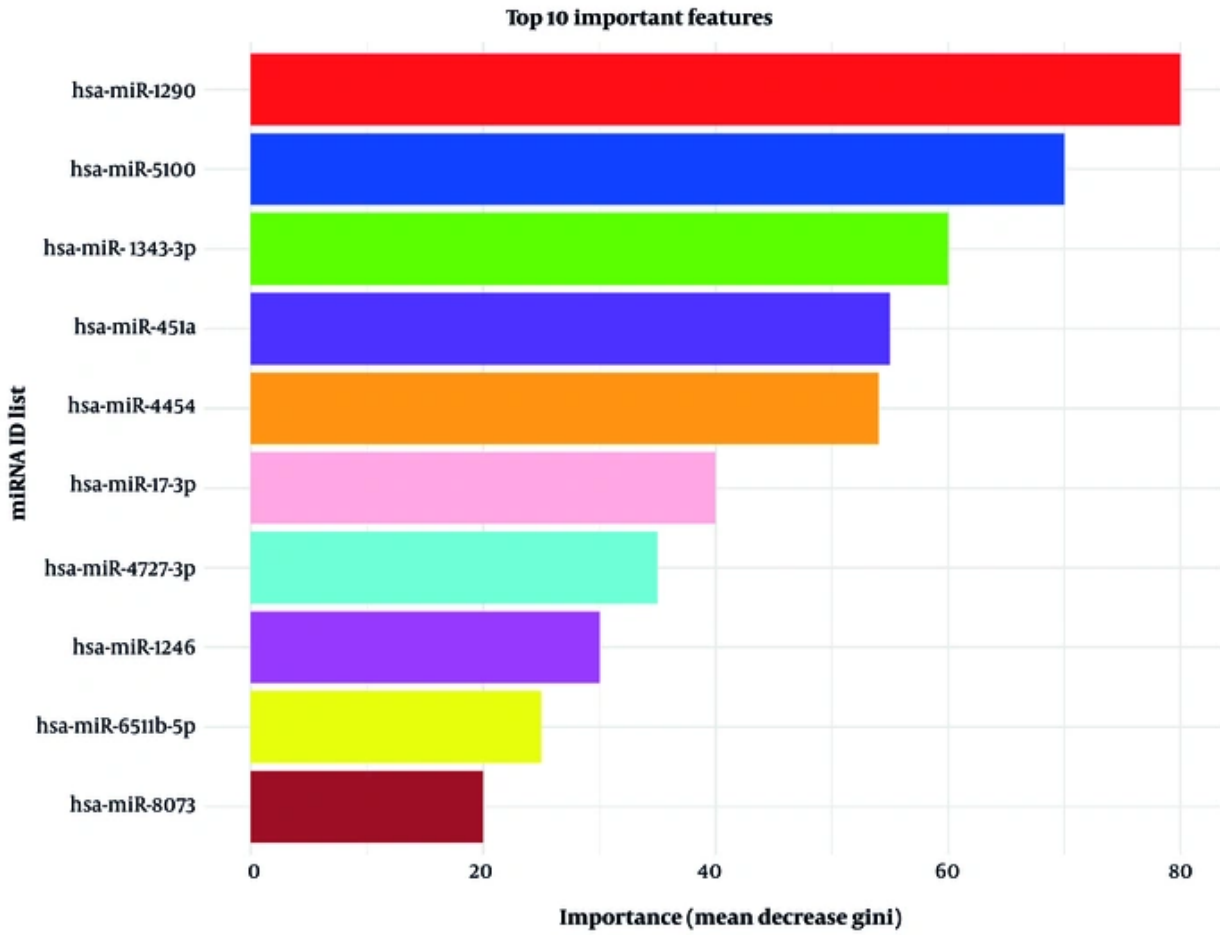

Figure 6. Feature importance of top 10 microRNAs (miRNAs) in gastric cancer classification, ranked by Mean Decrease Gini scores from a random forest model. Longer bars indicate greater contributions to distinguishing cancer from non-cancer samples, emphasizing their diagnostic relevance

5.5.1. Data Heterogeneity

Batch effects and inter-patient variability in GEO datasets may influence the reproducibility of findings. Future studies should incorporate larger, multi-center cohorts to validate these results.

5.5.2. Experimental Validation

Functional studies, such as CRISPR-based knockout experiments and qPCR validation, are essential to confirm the mechanistic roles of identified miRNAs.

5.5.3. Multi-Omics Integration

Combining miRNA data with proteomics, metabolomics, and methylation profiles could unravel complex regulatory networks and identify novel therapeutic targets.

5.5.4. Clinical Trials

Prospective trials are needed to evaluate the clinical utility of miRNA-based biomarkers and therapies.

5.6. Proposed Experimental Designs

To address these gaps, we propose the following experimental designs:

- In vitro studies: Use CRISPR/Cas9 to knock out candidate miRNAs in gastric cancer cell lines and assess

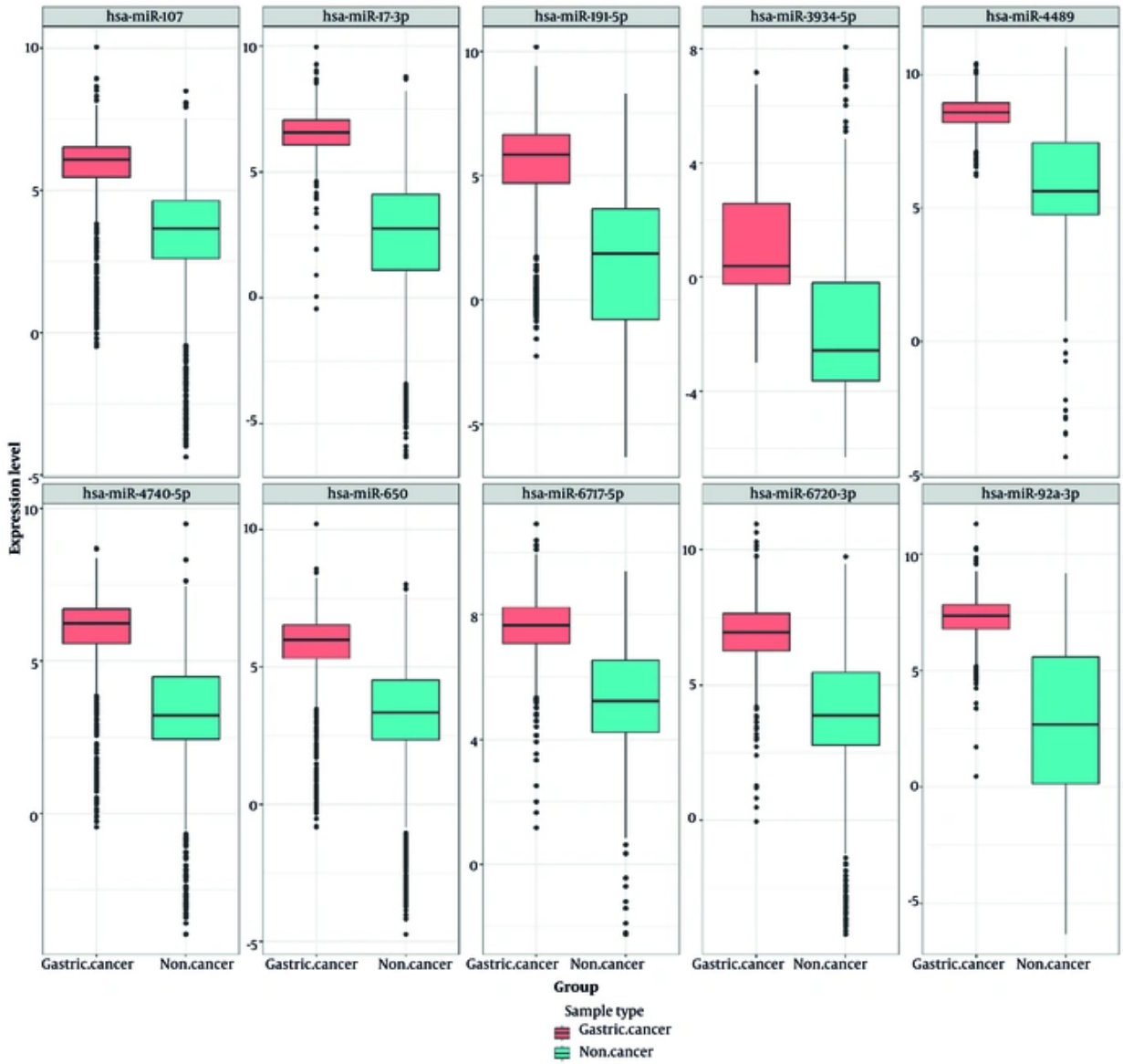

Figure 7. Box plots of microRNA (miRNA) expression levels of selected miRNAs in gastric cancer compared to non-cancer individuals. Each panel presents a different miRNA (hsa-miR-107, hsa-miR-17-3p, hsa-miR-191-5p, hsa-miR-3934-5p, hsa-miR-4489, hsa-miR-4740-5p, hsa-miR-650, hsa-miR-6717-5p, hsa-miR-6720-3p, and hsa-miR-92a-3p), with the x-axis representing the two groups: Gastric Cancer (red) and Non-Cancer (cyan).

their effects on proliferation, apoptosis, and migration.

- Animal models: Test the efficacy of miRNA mimics or inhibitors in xenograft models of gastric cancer.

- Liquid biopsy assays: Develop and validate assays for detecting circulating miRNAs in plasma or serum samples from gastric cancer patients.

Footnotes

AI Use Disclosure: The authors declare that no generative AI tools were used in the creation of this article.

Authors' Contribution: Study concept and design: H. N.; Analysis and interpretation of data: H. N.; Drafting of the manuscript: H. N.; Critical revision of the manuscript for important intellectual content: L. S.; Statistical analysis: H. N.

Conflict of Interests Statement: The authors declare no conflict of interest.

Data Availability: The data presented in this study are openly available in the Gene Expression Omnibus (GEO) repository under accession number GSE164174.

Funding/Support: The present study received no funding/support.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;**68**(6):394-424. [PubMed ID: 30207593]. https://doi.org/10.3322/caac.21492.
2. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet*. 2020;**396**(10251):635-48. [PubMed ID: 32861308]. https://doi.org/10.1016/S0140-6736(20)31288-5.
3. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;**6**(11):857-66. [PubMed ID: 17060945]. https://doi.org/10.1038/nrc1997.
4. Bartel DP. Metazoan MicroRNAs. *Cell*. 2018;**173**(1):20-51. [PubMed ID: 29570994]. [PubMed Central ID: PMC6091663]. https://doi.org/10.1016/j.cell.2018.03.006.
5. Zhang X, Zhu W, Zhang J, Huo S, Zhou L, Gu Z, et al. MicroRNA-125b promotes invasion and metastasis of gastric cancer by targeting STARD13 and NEAT1. *Cell Death Dis*. 2019;**10**(12):914.
6. Li X, Zhang Y, Zhang H, Liu X. MicroRNA-34a inhibits gastric cancer progression by targeting NOTCH1 signaling. *Oncol Rep*. 2020;**43**(4):1153-62.
7. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;**513**(7517):202-9. [PubMed ID: 25079317]. [PubMed Central ID: PMC4170219]. https://doi.org/10.1038/nature13480.
8. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;**26**(1):139-40. [PubMed ID: 19910308]. [PubMed Central ID: PMC2796818]. https://doi.org/10.1093/bioinformatics/btp616.
9. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;**15**(2):R29. [PubMed ID: 24485249]. [PubMed Central ID: PMC4053721]. https://doi.org/10.1186/gb-2014-15-2-r29.
10. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;**28**(6):882-3. [PubMed ID: 22257669]. [PubMed Central ID: PMC3307112]. https://doi.org/10.1093/bioinformatics/bts034.
11. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015;**4**. [PubMed ID: 26267216]. [PubMed Central ID: PMC4532895]. https://doi.org/10.7554/eLife.05005.
12. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;**16**(5):284-7. [PubMed ID: 22455463]. [PubMed Central ID: PMC3339379]. https://doi.org/10.1089/omi.2011.0118.
13. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;**435**(7043):834-8. [PubMed ID: 15944708]. https://doi.org/10.1038/nature03702.
14. Song G, Yu X, Shi H, Sun B, Amateau S. miRNAs in HCC, pathogenesis, and targets. *Hepatology*. 2024. [PubMed ID: 39626210]. [PubMed Central ID: PMC12119976]. https://doi.org/10.1097/hep.0000000000001177.
15. Trang P, Wiggins JF, Daige CL, Cho C, Omotola M, Brown D, et al. Systemic delivery of tumor suppressor microRNA mimics using a neutral lipid emulsion inhibits lung tumors in mice. *Mol Ther*. 2011;**19**(6):1116-22. [PubMed ID: 21427705]. [PubMed Central ID: PMC3129804]. https://doi.org/10.1038/mt.2011.48.
16. Burz C, Pop V, Silaghi C, Lupan I, Samasca G. Prognosis and Treatment of Gastric Cancer: A 2024 Update. *Cancers (Basel)*. 2024;**16**(9). [PubMed ID: 38730659]. [PubMed Central ID: PMC11083929]. https://doi.org/10.3390/cancers16091708.