Journal of Medical and Health Studies (JMHS)
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

AI-Assisted Literature Mining Framework for miRNA–Target Mapping in Colorectal Cancer

Swati Kulshrestha
Associate Scientist, Celltheon Cooperation, Union City, CA, USA
Corresponding Author: Swati Kulshrestha, **E-mail**: swati@celltheon.com

| ABSTRACT

Colorectal cancer is an important cause of mortality worldwide, and the complex molecular mechanisms that govern the development, progression, and spread of the disease are complex. Among these complex mechanisms, the role of miRNAs in the regulation of gene expression at the post-transcriptional level through the modulation of the expression of multiple target genes that are integral parts of oncogenic pathways is crucial. The discovery of the interactions between miRNAs and their targets is therefore essential in the understanding of colorectal cancer and the discovery of potential biomarkers and therapeutic targets. The exponential growth of scientific literature in the field of biomedicine has made the discovery of such interactions from the scientific literature an inefficient and difficult task. To overcome the inefficiencies associated with the traditional approaches for the discovery of miRNA–target interactions, the present study proposes a conceptual framework for AI-assisted literature mining for the discovery of miRNA–target interactions that are crucial in the understanding of colorectal cancer. The framework for the discovery of miRNA–target interactions is structured and aims at synthesizing the scientific evidence that is disseminated in the scientific community. The framework is therefore crucial in the development of the field of bioinformatics.

| KEYWORDS

MicroRNA (miRNA), Colorectal Cancer, AI-assisted Literature Mining, miRNA–Target Interaction, Biomedical Text Mining

| ARTICLE INFORMATION

ACCEPTED: 01 March 2026 **PUBLISHED:** 18 March 2026 **DOI:** 10.32996/jmhs.2026.7.5.9

1. Introduction

1.1 Colorectal Cancer and Molecular Regulation

Colorectal cancer (CRC) is an extremely prevalent type of malignancy that is responsible for contributing substantially to the overall burden of cancer-related mortality. Global statistics on cancer reveal that CRC is among the top three types of malignancies in terms of incidence rates. It significantly adds to the overall burden of cancer-related mortality. Colorectal cancer is known to result from the action of intricate molecular and genetic mechanisms that affect cellular proliferation, differentiation, apoptosis, and metastasis. Recently, the scientific community has witnessed an increasing focus on the involvement of non-coding RNAs in the regulation of gene expression during tumorigenesis. Among various classes of non-coding RNAs, microRNAs (miRNAs) play an important role in the regulation of genes during tumorigenesis [1].

These are small non-coding RNA molecules measuring about 20-24 nucleotides in length and are known to play an important role in gene expression by binding to complementary sequences in the mRNA transcript. The binding action degrades the mRNA transcript or represses translation, effectively regulating protein synthesis at the cellular level. In the context of colorectal cancer development and progression, there is considerable evidence to suggest aberrant levels of specific miRNA species that play an important role in the regulation of key oncogenic pathways. While some miRNA species are known to act as oncogenes by repressing tumor suppressor genes, others are known to act as tumor suppressors by repressing oncogenes. Thus, miRNA and

their target interaction is an emerging area of research with considerable potential in the context of tumor development and progression and the development of biomarkers for early diagnosis and intervention [2].

1.2 Importance of miRNA–Target Mapping

It is important to map the interactions that exist between the miRNAs and their corresponding gene targets in the development of colorectal cancer. A single miRNA has the capacity to regulate multiple genes, and similarly, a single gene can be regulated by multiple miRNAs, thus leading to complex interactions that form the regulatory network that influences various pathways in the development of cancer. The interactions include those that influence cell proliferation, invasion, angiogenesis, and resistance to chemotherapy, thus providing crucial insights into the development of colorectal cancer [3].

The discovery of relationships between miRNA-targets is a challenge due to the distribution of relevant data in thousands of biomedical publications and studies. Conventional methods for delineating relationships between miRNA-targets involve manual literature surveys and database searches. This is a time-consuming task and is likely to be incomplete. As the biomedical literature is growing at a tremendous rate, there is a need for efficient computational methods to systematically harvest this growing body of literature [4].

1.3 Growth of Biomedical Literature and the Need for AI-Based Mining

The ever-increasing rate of expansion of biomedical literature poses not only opportunities but also challenges for scholars. Although the ever-increasing literature provides tremendous biological knowledge, it is increasingly difficult to identify valuable relationships within this ever-expanding literature. Artificial intelligence (AI) and natural language processing (NLP) techniques have shown tremendous potential in addressing this issue by allowing the automatic extraction of biological entities and relationships within literature. These techniques have shown tremendous potential in processing huge amounts of literature and extracting relevant biological terms and relationships within it [5].

1.4 Research Gap and Study Contributions

In spite of the growing trend of using artificial intelligence in biomedical informatics, there is a lack of well-structured frameworks that are particularly focused on identifying and mapping disease-oriented miRNA-target interactions using vast literature resources. The existing literature resources are mostly based on manual curation and limited automated methods, which may not effectively identify newly published literature related to colorectal cancer. In order to address the above-identified gap in literature resources, the present study aims to suggest a conceptual AI-assisted literature mining framework for identifying and mapping miRNA-target interactions related to colorectal cancer. The suggested framework is particularly focused on efficient literature-based knowledge discovery in cancer bioinformatics studies.

2. Background and Related Work

2.1 miRNA Regulatory Mechanisms in Cancer

MicroRNAs (miRNAs), being the major regulators of gene expression in various biological processes, have emerged as key regulators in cancer biology as well. The importance of miRNAs in cancer biology lies in their ability to regulate signaling pathways, which are significant in cancer development and progression. In colorectal cancer, it has been observed that there is significant dysregulation of miRNA expression, which is responsible for the abnormal regulation of genes controlling cell proliferation, migration, and invasion. Some miRNAs have been reported to function as oncogenic miRNAs, which are responsible for the suppression of tumor suppressor genes, thereby promoting cancer development and progression. On the other hand, there are miRNAs that have been reported to function as tumor suppressor miRNAs, which are responsible for the suppression of oncogenic pathways. For example, miR-21 and miR-155 are reported to be overexpressed in colorectal cancer and are responsible for promoting cancer development and progression through the regulation of genes controlling apoptosis and cell cycle control. On the other hand, miR-34a has been reported to function as a tumor suppressor miRNA and is responsible for regulating pathways controlling differentiation and growth inhibition. The detailed understanding of this regulation is possible only through the systematic identification of miRNA-target gene interaction, which collectively forms a complex system [6][7].

2.2 Existing miRNA Databases and Knowledge Sources

In order to further improve the knowledge on miRNA biology, several databases specializing in miRNA-target interactions have been established and are used for the storage of experimentally validated and predicted miRNA-target interactions. Some of the prominent databases include miRTarBase, TargetScan, and miRDB. These databases are significant tools for scientists working on the biology of genes and disease. Despite the significance and potential for these databases, several limitations are associated with these databases. One such constraint is that most of these databases are based on manual curation and, therefore, may not

readily incorporate new research articles published on miRNA-target interactions. In addition, several databases are based on predicted interactions and not experimentally validated interactions, which may cause confusion for scientists working on miRNA biology. Another constraint is that most databases are not disease-specific, and therefore, interactions relevant to colorectal cancer may not be easily identified. Consequently, knowledge on miRNA-target interactions published in research articles may not be fully harnessed [8].

2.3 Biomedical Literature Mining Techniques

In response to the increasing volume of literature published in the biomedical domain, computational approaches for literature mining have received increasing research attention. NLP is extensively used for literature mining to derive structured knowledge from unstructured scientific literature. The most commonly used approaches are named entity recognition (NER), which is used to recognize biological entities such as genes, proteins, and miRNAs from literature, and relation extraction approaches that are used to determine relationships between these entities. In recent times, contemporary machine learning and deep learning technologies, such as transformer models, have greatly improved the accuracy of literature mining approaches for the biomedical domain. These approaches are used to analyze large volumes of research literature and determine patterns for gene-disease relationships, regulatory interactions, and other molecular pathways [9].

Despite all these advancements, existing literature mining literature is more focused on general knowledge extraction in the field of biomedicine. Therefore, there is a need for a framework that is capable of systematically extracting and organizing miRNA–target relationship knowledge specifically related to colorectal cancer. The conceptual framework proposed in this study aims at filling this gap by integrating literature mining with regulatory networks in order to boost knowledge discovery in cancer bioinformatics with the help of AI [10].

3. Conceptual AI-Assisted Literature Mining Framework

The sheer volume of biomedical literature published in recent times has created numerous challenges in the identification and organization of knowledge related to molecular interactions in cancer research. In order to address these issues and challenges in cancer research related to miRNA–target interactions in colorectal cancer, this study proposes a conceptual framework of AI-assisted literature mining for extracting and organizing knowledge related to miRNA–target interactions in cancer research. Unlike traditional literature reviews that rely heavily on human expertise in conducting literature reviews, the proposed framework utilizes artificial intelligence and natural language processing techniques for processing numerous scientific literature and identifying key regulatory interactions between miRNA and gene targets. The framework is composed of four layers that work in conjunction with each other to efficiently extract knowledge related to miRNA–target interactions in cancer research.

The first layer of the framework involves the literature acquisition layer, which primarily focuses on the collection of appropriate scientific literature on the subject of colorectal cancer and microRNA (miRNA) research. The main sources of literature would include biomedical literature databases such as PubMed, PubMed Central, and other open-access literature databases. This process could be automated with the use of appropriate search terms such as “microRNA,” “miRNA,” “colorectal cancer,” “gene regulation,” and “target interaction.” The main idea behind the literature acquisition layer is to obtain an appropriate set of literature articles on the molecular mechanisms of miRNAs and their role in colorectal cancer. This process would be automated, thereby ensuring the collection of a wide spectrum of literature on the subject.

Once relevant literature is acquired in the next step, biomedical entity recognition is employed to identify primary entities in the collected literature. The prominent entities in biomedical literature are generally miRNAs, targets of genes, proteins, and disease-related terms. The Named Entity Recognition (NER) models are used to recognize entities in biomedical literature. Advanced natural language processing tools, like transformer-based language models, can be employed to improve the accuracy of biomedical literature recognition. This is a critical step in biomedical literature analysis since it converts unstructured data into a structured form for further analysis.

The third part of this framework is the relationship extraction layer. The function of this layer is to identify the presence of any interaction between the identified entities. After identifying the presence of miRNA and gene targets, it is necessary to analyze the sentence structure and semantic content to determine whether there is an interaction between them. Various methods can be adopted to determine this interaction, including analyzing whether there is a statement regarding the regulation of gene targets by miRNA. For example, statements regarding the suppression of specific genes related to colorectal cancer can be identified and classified as part of this interaction between miRNA and gene targets.

The final step of this framework involves the construction of a miRNA-target interaction network. In this step, it is possible to create a structured network of interactions that are visually represented and that describe interactions between miRNA and gene targets. Graph-based models are also possible in this step, in which nodes are used to describe miRNA and gene targets, and edges are utilized to describe interactions that are identified in the literature. Network analysis techniques can also be utilized to

identify nodes that are most interconnected, which may describe key miRNA and significant gene targets in colorectal cancer pathways. Network models provide a precise description of the intricate interactions that are involved in molecular interactions in cancer.

The framework proposed in the paper presents a systematic approach for combining artificial intelligence with literature mining in biomedicine for knowledge extraction. By combining literature retrieval, entity recognition, relationship extraction, and network modeling, it aims to improve the efficiency of identifying miRNA-target interactions related to colorectal cancer. This framework is a conceptual framework for establishing the basis for future computational systems that can incorporate new literature and help scientists identify key molecular regulatory mechanisms in cancer biology.

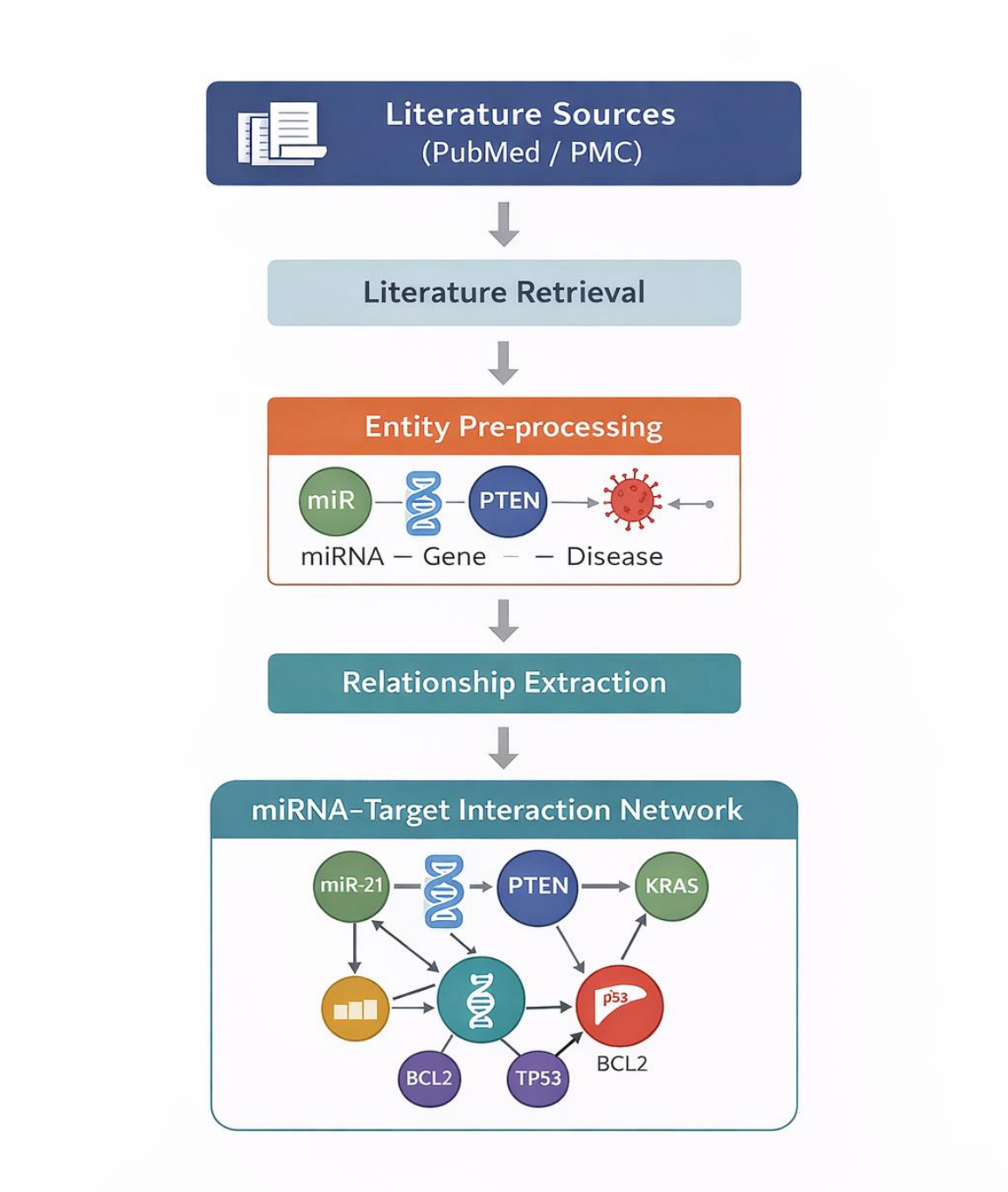

Fig 1. Comprehensive architecture

Figure 1 presents the comprehensive architecture of the proposed AI-assisted literature mining framework for the identification of miRNA–target interactions related to colorectal cancer. As presented in the figure, the proposed framework begins with

literature sources that include various biomedical literature databases such as PubMed and PubMed Central. The literature retrieval module employs various automated literature search strategies to retrieve literature related to microRNAs, genes, and colorectal cancer. Subsequently, the retrieved literature is processed using the entity preprocessing module that employs various natural language processing techniques to identify the presence of various biological entities related to microRNAs, genes, and disease within the literature. Finally, the literature mining framework employs the relationship extraction module that identifies the relationships between microRNAs and genes based on the literature. The proposed framework presents the extracted relationships between microRNAs and genes in the form of the miRNA–target interaction network. The proposed literature mining framework provides an architecture that presents the relationships between microRNAs and genes in the context of colorectal cancer bioinformatics research.

4. Conceptual Framework Demonstration

To exemplify the prospective capabilities of the proposed framework for AI-assisted literature mining, this section conceptually demonstrates how the framework can be used for extracting and organizing miRNA–target relationships in literature related to colorectal cancer. Due to its hypothetical nature, it is essential to highlight that the framework is presented with emphasis on its workflow and potential biological insights that may be generated when it is used for processing large-scale scientific literature.

The first phase of the conceptual demonstration process involves the knowledge extraction process, whereby scientific articles on the subject of colorectal cancer and the regulation of microRNA are extracted from the databases. This information is then processed utilizing natural language processing, which identifies the relevant information on biological entities such as microRNAs, gene targets, and specific keywords related to the disease. For example, if the framework identifies statements in the articles on the regulation of microRNAs, which regulate the expression of genes responsible for the development of colorectal cancer, the framework will identify the microRNAs and the gene targets as significant biological entities. The framework will then utilize the relationship extraction process to analyze the sentence structure and identify the interaction between the two entities, thereby identifying the specific relationships that describe the role of specific microRNAs in the development of colorectal cancer.

The obtained relationships are then assembled into a conceptual network of miRNA-target interaction, with each miRNA and target gene represented as a node and the interaction between the two represented as an edge connecting the two nodes. Nodes with high interconnectivity in the network may represent miRNAs with many target genes or target genes with many modulating miRNAs, which could be important regulatory elements in the signaling pathways of colorectal cancer. By displaying the relationships extracted as a network, researchers can more easily identify important molecular regulators which could be important in the development of the tumor or in the response to therapy.

Another important aspect that this conceptual framework is able to demonstrate is scalability and automation. As a result of its reliance on AI-based literature mining techniques, it is possible to analyze thousands of research publications and incorporate emerging literature into its database. This is important since it enables the framework to provide a current state of knowledge regarding miRNA and target regulatory interactions in colorectal cancer. As it is known that there is a rapid increase in literature regarding this topic, it is believed that this kind of framework can contribute to a wider level of bioinformatics research.

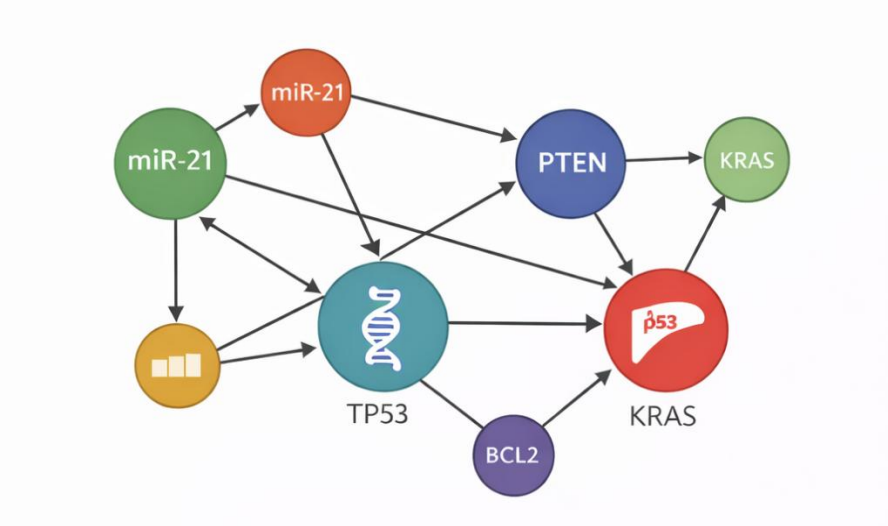

Fig 2. Conceptual network of miRNA-target regulatory interactions

A conceptual network of miRNA-target regulatory interactions based on the proposed AI-assisted literature mining framework in the context of colorectal cancer research is shown in Figure 2. As shown in the network diagram, each node in the network represents a biological entity, including microRNAs and gene targets. On the other hand, the directed edges represent the regulatory interactions between the biological entities. For example, microRNAs such as miR-21, miR-34a, and miR-155 have regulatory interactions with various crucial genes associated with colorectal cancer pathways, including PTEN, TP53, KRAS, and BCL2 genes. The directed edge in the network indicates the regulatory interactions between microRNAs and gene targets. For instance, a single microRNA may have regulatory interactions with multiple genes, and a single gene may have regulatory interactions with multiple microRNAs. Moreover, nodes with multiple interactions may represent crucial regulatory interactions in the network, which may play a crucial role in various biological processes, including cancer development and progression. Visualizing these regulatory interactions in a network format helps in understanding the complex gene regulatory mechanisms and various potential biomarkers and targets of interest in the context of colorectal cancer.

5. Discussion

This is due to the fact that the rapid advancement of genomic science has significantly increased our level of comprehension regarding molecular processes related to cancer development. In this regard, microRNAs have emerged as key gene expression regulators, controlling various biological pathways related to cancer development and treatment. The theoretical model presented in this paper has demonstrated the importance of artificial intelligence-based literature mining in developing a more systematic comprehension of microRNA-target interaction in colorectal cancer. By using this approach, it is possible to identify molecular interaction patterns that might have gone unnoticed within a multitude of scientific literature.

One of the major contributions of this framework is its potential to speed up research on cancer bioinformatics and molecular oncology. Generally, methods of identifying interactions involving miRNAs and targets often rely on manual literature reviews or experimental databases that may not be entirely comprehensive given the constantly expanding research literature. The use of artificial intelligence and natural language processing technologies may enable automated extraction of biological entities and relationships from unstructured scientific literature. Such an ability may enable researchers to quickly identify emerging trends in gene regulation and discover novel associations of miRNAs with cancer-related genes. It is therefore conceivable that this framework may lead to the discovery of new biomarkers and targets of relevance to colorectal cancer.

Another important advantage of this proposed framework is that it can help to overcome the limitations of a literature review process. It is a known fact that a manual literature review process is usually limited in its ability to analyze a large number of research papers in a given amount of time. Moreover, there is a possibility of inconsistencies in a literature review process when it comes to identifying relevant regulatory interactions. With a literature mining tool that uses AI technology, it is possible to analyze thousands of research papers in a short amount of time in a consistent manner. Such a process would help to provide a complete picture of all the knowledge that is involved in colorectal cancer.

Although this proposed framework provides certain benefits, it also provides a number of limitations that need to be addressed. The accuracy of literature mining techniques is largely dependent on the accuracy and clarity of the source literature. Confusion in biological terminologies and lack of clear descriptions of experimental results can lead to a number of inaccuracies in the results. Moreover, it is also possible that computer models may pick up co-occurring entities but fail to capture the biological significance of interactions. Hence, it is recommended that future implementations of this framework include a number of validation techniques and experimental results to increase accuracy.

Looking forward, it is possible to identify various ways in which future research can be conducted. For instance, it is possible to integrate various pieces of knowledge that are derived from literature and various multiomics datasets, such as transcriptomics and proteomics data. Such a framework is also likely to allow for the identification of various potential therapeutic targets and the formulation of a more personalized therapeutic strategy. As artificial intelligence continues to grow and develop, it is possible to identify that such a framework would be a vital tool in advancing knowledge discovery in the field of cancer genomics.

6. Conclusion

This study proposes a framework for literature mining with AI, which is conceptually driven and aims to assist in the systematic mining and mapping of the interactions between miRNAs and their targets, which are related to colorectal cancer. MicroRNAs play an important role in the control of gene expression and are also closely related to the molecular mechanisms of the development and progression of cancer, as well as the response to therapy. However, the rapid increase in the number of publications in the biomedical field has rendered the identification and organization of the substantial information related to these regulatory interactions impracticable. The framework proposed in the current work helps to mitigate such a challenge by integrating literature mining, natural language processing, and interaction mapping in one process.

By proposing a framework that describes a process to extract knowledge from scientific publications, this framework provides a scalable solution to identify any possible regulatory relationship that may exist between miRNA and gene targets in colorectal cancer research. Although this framework is still in a conceptual state, it highlights the possibility of artificial intelligence playing a vital role in synthesizing knowledge in cancer bioinformatics. It is possible that this framework could be expanded to include experimental data and multi-omics information to further elucidate the molecular regulatory network in colorectal cancer.

References

[1] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA Cancer J. Clin.*, vol. 74, no. 1, 2024, doi: 10.3322/caac.21820.

[2] P. Naeli, T. Winter, A. P. Hackett, L. Alboushi, and S. M. Jafarnejad, "The intricate balance between microRNA-induced mRNA decay and translational repression," 2023. doi: 10.1111/febs.16422.

[3] C. Liu *et al.*, "Comprehensive analysis of miRNA-mRNA regulatory pairs associated with colorectal cancer and the role in tumor immunity," *BMC Genomics*, vol. 24, no. 1, 2023, doi: 10.1186/s12864-023-09635-4.

[4] L. Cui, H. Lu, and Y. H. Lee, "Challenges and emergent solutions for LC-MS/MS based untargeted metabolomics in diseases," 2018. doi: 10.1002/mas.21562.

[5] S. E. Doneva *et al.*, "Large language models to process, analyze, and synthesize biomedical texts: a scoping review," 2024. doi: 10.1007/s44163-024-00197-2.

[6] Y. Chi and D. Zhou, "MicroRNAs in colorectal carcinoma - From pathogenesis to therapy," 2016. doi: 10.1186/s13046-016-0320-4.

[7] Z. Wang, H. Wang, S. Zhou, J. Mao, Z. Zhan, and S. Duan, "miRNA interplay: Mechanisms and therapeutic interventions in cancer," 2024. doi: 10.1002/mog2.93.

[8] M. Dey, P. Remiszewski, J. Piątkowski, P. Golik, P. Teterycz, and A. M. Czarnecka, "MicroRNA bioinformatics in precision oncology: an integrated pipeline from NGS to AI-based target discovery," 2025. doi: 10.1007/s13353-025-01024-9.

[9] P. Ernst, A. Siu, and G. Weikum, "KnowLife: A versatile approach for constructing a large knowledge graph for biomedical sciences," *BMC Bioinformatics*, vol. 16, no. 1, 2015, doi: 10.1186/s12859-015-0549-5.

[10] S. Ramadhani *et al.*, "Feature Selection Optimisation for Cancer Classification Based on Evolutionary Algorithms: An Extensive Review," 2025. doi: 10.32604/cmes.2025.062709.