International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 5, Sep-Oct 2024
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Machine Learning Analysis of Gut Microbiome Profiles as Predictors for Colorectal Cancer Risk: A Systematic Review and Meta-Analysis

Ernest E. Onuiri[1], Daniel Amorue[2], and Kelechi C. Umeaka[3]
[1]Department of Computer Science, School of Computing, Babcock University, Nigeria
[2]Department of Computer Science, School of Computing, Babcock University, Nigeria
[3]Department of Computer Science, School of Computing, Babcock University, Nigeria
Corresponding Author Email: onuirie@babcock.edu.ng

Abstract

Colorectal cancer (CRC) ranks as the third most deadly cancer worldwide, with high mortality rates in both developed and developing regions, largely due to late-stage diagnosis. This systematic review and meta-analysis, conducted in line with PRISMA guidelines, evaluates studies published between 2016 and 2024 that explore the potential of machine learning algorithms to predict colorectal cancer risk based on gut microbiome profiles. The inclusion criteria focused on studies employing machine learning methods to develop predictive models for CRC. A total of 12 studies were reviewed, each utilizing various machine learning approaches to enhance the accuracy of CRC diagnosis. The findings underscore the transformative role of machine learning in improving diagnostic precision for CRC, particularly through the analysis of gut microbiome profiles. These results demonstrate that machine learning-driven analyses hold great promise in advancing early detection and risk prediction of colorectal cancer, thus offering a valuable tool for improving patient outcomes.

Key Words — **Colorectal Cancer, Diagnosis, Gut Microbiome, Machine Learning, Prediction, Risk Assessment, Systematic Review, Meta-analysis.**

I. INTRODUCTION

Cancer is a predominant chronic illness on earth while also having a high mortality rate worldwide with over 19 million newly reported cases annually which has led to about 10 million deaths around the world [1]. Cancer risks like many other chronic diseases might be managed through taking steps to consume healthy diet, reduced consumption of substances like alcohol or cigarettes and exercising regularly [2].

Colorectal cancer (CRC) ranks third on the cancer prevalence list, and the fourth leading reason leading to cancer related mortalities [3]. It is said to be detected in about 1.2 million people each year worldwide and more than 600,000 die from the disease [4]. Predominantly detected in Western countries, its incidence continues to rise annually. The likelihood of encountering colorectal cancer hovers around 4%–5%, with the risk influenced by personal elements including age, history of chronic disease, and lifestyle choices. Presently, due to the absence of early indicators of CRC [5], many CRC patients are usually diagnosed when has advanced to the progressive phase [6], suggesting that the growth has metastasized, which can lead to cancer-related death [7][8]. Within this framework, the gut microbiota assumes a pivotal role in predicting a patient's predisposition to colorectal cancer [9]. The human gut microbiome holds a diverse array of bacteria, viruses, and fungi that, under optimal conditions, exist symbiotically in the gut with their human host and under specific conditions, individual species or the combined functions of bacteria can provide numerous benefits throughout the host's life. These advantages may include digesting dietary compounds, training the

human immune system, helping in the fight against pathogens, while also playing a role in the overall human health. [10], [11].

New developments in Machine learning and Artificial Intelligence have allowed scientist test and develop Machine Learning Algorithms in the Biomedical fields to predict and diagnose disease using available biomarkers. The human gut microbiome, has been explore by scientist to examine its use in predicting the risks of individuals been exposed to Colorectal cancers using Machine learning algorithms to learn from available datasets. This systematic review aims to explore how our human gut microbiome profiles can be employed to predict colorectal cancer risks using advanced machine learning analysis. Utilizing high-throughput sequencing technologies, our primary objective is to discern subtle microbial patterns that might elude traditional statistical methods. This Systematic Review and Meta-analysis, aims to pinpoint consistent microbial signatures, assess the reliability of machine learning models, and explore translational implications for the early detection of colorectal cancer (CRC). Through the synthesis of diverse literature, we endeavor to provide valuable enlightenment into the possibility of microbiome-based predictions for CRC risk.

A. Rationale

This study endeavours to examine a spectrum of existing literature that explores machine learning methods to predict colorectal cancer risk using biomarkers present in the human gut microbiome. The key objective is identifying the most effective models by assessing the model's performance accuracy, specificity, metrics, and sensitivity.

B. Objectives

The main objective of this study was to perform a standard review of existing literature on prediction of colorectal cancer in patients using gut microbiome profiles. The study aimed to explore the most efficient machine learning models for this predictive analysis. The PICOS framework was deployed on the study in order to address important questions for the research.

C. Population

What are the characteristics of the patient population involved in these studies which make use of machine learning models to predict colorectal cancer risk using gut microbiome profiles?

D. Intervention

What were the main machine learning models deployed in these studies and what type of data sets were used for the machine learning model.

E. Comparison

How do we examine the way machine learning compares with each other in terms of their prediction abilities.

F. Outcome

What machine learning model would be selected as the most efficient model among the models deployed in the research.

G. Study Design

How do these studies compare in terms of efficiency, and what factors from the study can influence the final result.

II. METHODS

Working with PRISMA [12] guidelines for systematic review and meta-analysis, this systematic review explored databases like Scopus and PubMed in search of papers in relation to the area of study. The systematic review worked with papers ranging between the year 2016 and 2024. The main key words deployed in this search are 'machine learning', 'artificial intelligence', 'prediction', 'prognosis', 'gut microbiome' and 'colorectal cancer'.

A. Scope of the review

The scope of this systematic review covers papers which make use of machine learning methods to predict the risk of developing

colorectal cancers based on biomarkers present in the gut microbiome.

B. *Eligibility criteria*

The systematic review employed the use PICOS framework to select eligible research works.

1) **Participants:** Studies which involve colorectal cancer patients.
2) **Interventions:** Research studies involving the use of machine learning and artificial intelligence techniques to predict the risk of colorectal cancer using Gut microbiomes.
3) ***Comparisons:*** Studies which contrast machine learning models with regular prediction of colorectal cancer risks.
4) ***Outcomes:*** Outlined report of the prediction accuracy of models.
5) **Study *Design:*** The study considers observational studies and clinical trials.

C. *Inclusion criteria*

- Research papers employing machine learning techniques to study gut microbiome activities in relation to colorectal cancer.
- Articles published between 2014 and 2024.
- English Language based studies.
- Research outlining the various accuracy metrics for prediction models.
- Research where colorectal cancer (CRC) patients are the main study.
- Studies which are open and free to the public.

D. *Exclusion* criteria

- Research works which exclude the use of machine learning methods to perform analysis.
- Research not working with the gut microbiome as the primary item for analysis.
- Studies on Cancers not involving the colorectal region.
- Non-English based studies.
- Inaccessible Articles
- Conferences, abstracts, letters, editorials, case reports, reviews, and meta-analyses.

E. *Information Sources*

The source of information for this review were Scopus and PubMed databases. The use of advanced search and customized query was used to extract as much results as accurately as possible. The search query involved keywords pertaining to the subject area such as machine learning, colorectal cancer, gut microbiome and prediction. Although sources like Google scholar, were explored, the resources gotten from this source were duplicates of what had been gotten from earlier sources. The search queries were carved specifically to fit into each database's modus operandi.

F. *Search Strategy*

The search strategy was set to encompass the significant areas of studies related to the topic. The search did not set any date filters. But the results mostly spanned about 11 years i.e. Between 2013 to 2024 mostly due to the fact that the area of study is still at an infantile phase.

- Works which made use of machine learning models to predict colorectal cancer risks.
- Studies which used gut microbiome as the main predictor.
- Research studies which reported prediction models accuracy metrics.
- Language was restricted to English Language.

Scopus Database query returned 55 document results

TITLE-ABS-KEY (("Gut Microb*" OR "Intestin* Microb*" OR "Abdom* Microb*") AND (predict* OR prognosis OR forecast*) AND ("Colo* Cancer" OR "Colo* Carcinoma" OR "Colo* Neoplasm*" OR "rectal cancer") AND ("Machine Learning" OR "Deep Learning" OR "machine intelligence" OR "predictive learning" OR "artificial modeling" OR ai OR "Artificial Intelligence" OR "Neural Network")) AND (LIMIT-TO (LANGUAGE , "English")).

PubMed Database query returned 24 document results

(("Gut Microbiome" OR "Intestinal Microbe" OR "Abdominal Microbiome") AND (prediction OR prognosis OR forecasting OR Predicting) AND

("Colon Cancer" OR "Colon Carcinoma" OR "Colon Neoplasm" OR "Rectal cancer" OR "Colorectal Cancer" OR "Colorectal Carcinoma" OR "Colorectal Neoplasm") AND ("Machine Learning" OR "Deep Learning" OR "Machine Learning Analysis" OR "Deep Learning Analysis" OR "machine intelligence" OR "predictive learning" OR "artificial modeling" OR "ml" OR "AI" OR "artificial intelligence"))

G. Data Management

The search outcomes were retrieved from Scopus and PubMed as RIS and PubMed formats respectively. The PubMed was then converted to RIS format using Zorato software. After which both files were uploaded on Hubmeta for evaluation and screening. A total of 79 article references were uploaded on Hubmeta. The final search query was run on 29th February, 2024.

H. Study Selection

During the paper selection process, the inclusion criteria and exclusion criteria were strictly adhered to during the selection process and papers were scrutinized by an independent reviewer using the Hubmeta systematic review web application. The studies passed through titles and abstract screening before going into the full text screening process where only the most relevant articles were allowed to go through the screener. Multiple screenings were done to ensure qualified studies were involved and the unqualified were left out. Initially a total collection of 79 documents were recorded from both databases for examination, but after thorough review of the documents a total of 12 documents qualified for the final review. The inclusion and exclusion process are very well outlined in Figure 1, which employes the PRISMA flow chart 2000.

I. Data Extraction

The data extracted covers study features, gut microbiome profiles, patient demographics, sample size, machine learning techniques employed, prediction models, and final outcomes. The screening process has been documented using the PRISMA flow diagram shown on Figure 1 [13].

J. Risk of Bais

A deep assessment of bias was conducted by applying the eligibility criteria to include the studies aligning with the search criteria and excluding studies that did not align with the search criteria or address the main topic [14].

III. RESULTS

The Systematic Review and Meta-Analysis included 12 studies which employed several Machine learning models like KNN, Support Vector Machine, Lasso, Gradient Boost, Elastic Network, Random Forest, Neural Network, Bayesian additive regression trees (BART) in order to predict the risk of CRC based on gut microbiomes. The models reported a high accuracy level from the final results gathered. The main metrics for evaluating were specificity, sensitivity, negative predictive value, positive predictive value, and the area under the curve (AUC).

After getting the initial pool of 79 studies, 12 studies were settled on for the final review. Following the step-by-step process outlined in the PRISMA flowchart in Figure 1, the study was able to ensure a quality and in-depth review of all selected studies. The summary of the reviewed studies is outlined on Table 1.

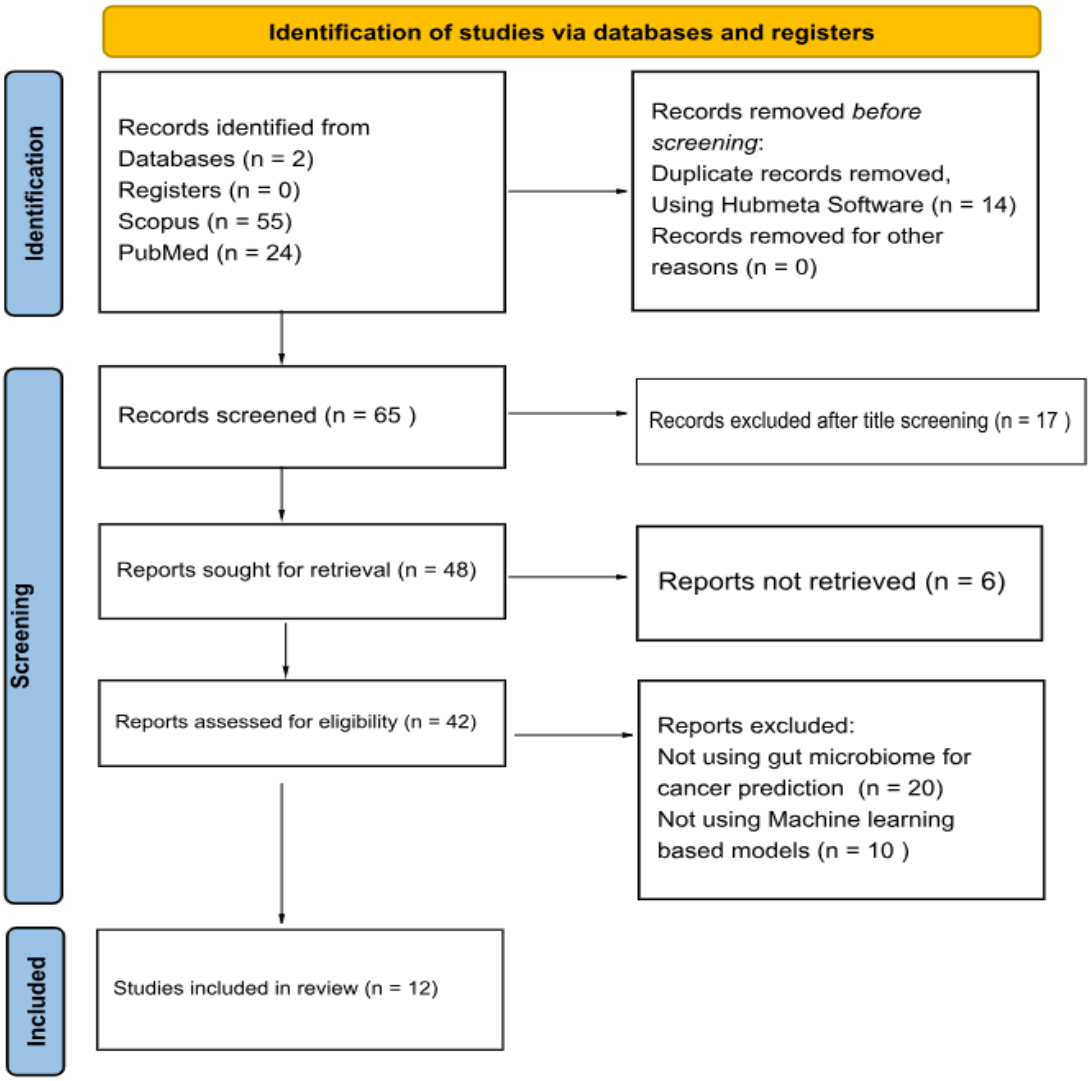

Fig .1 A PRISMA 2020 Flow diagram showing the studies screened

TABLE 1.

SUMMARY OF STUDIES

S/N	Author/Year/Publisher	Title	Machine Learning Methodology	Findings
1.	Yi-Hui Zhou et al. 2022[15] Frontiers in Molecular Biosciences	Improve the Colorectal Cancer Diagnosis Using Gut Microbiome Data[15]	KNN, Neural Network, Elastic Network, Support Vector Machine, Lasso, Gradient Boost, Random Forest, Bayesian additive regression trees (BART)	The machine learning pipeline demonstrated an 8.7% higher prediction accuracy compared to other published works when predicting colorectal cancer status using gut microbiome data. The pipeline also showed a 13% higher area under the receiver operator characteristic curve (AUC-ROC) than existing methods, indicating improved performance in distinguishing between colorectal cancer cases and controls.
2.	Abazar Arabameri et al. 2018[16] Frontiers in IEEE	Detection of Colorectal Carcinoma Based on Microbiota Analysis using Generalized Regression Neural Networks and Nonlinear Feature Selection[16]	Various machine learning classifiers were tested, including Support Vector Machines (SVM), Random Forest (RF), naïve Bayesian classifier, Multi-Layer Perceptron (MLP), General Regression Neural Networks (GRNN), and k-Nearest Neighbors (k-NN). GRNN was selected as the optimal classifier for identifying CRC from healthy cases due to its high performance and simplicity	The document also emphasized the importance of microbial biomarkers in the initiation, progression, and metastasis of CRC, underlining the need for effective diagnostic panels. Various microbial markers were categorized based on their statistical significance and validation technologies used. The research study focuses on the set up of knowledge base of CRC related microbes to help facilitate, the prediction accuracy in CRC diagnosis.
3.	Luoyan Ai et al 2017[17] Oncotarget	Systematic evaluation of supervised classifiers for fecal microbiota-based prediction of colorectal cancer[17]	The machine learning methods which were deployed in the study involved algorithms Bayes Net, Random Forest, Simple Logistic, JRip, J48, and SMO. Bayes Net and Random Forest demonstrated higher accuracies in predicting CRC.	The research study finding focuses on the importance of early detection of CRC using machine learning algorithms to analyze the existing human gut microbiomes. It also emphasizes this as a non-invasive method.
4.	Zhongkun Zhou et al. 2020[18] Frontiers in Microbiology	Human Gut Microbiome-Based Knowledgebase as a Biomarker Screening Tool to Improve the Predicted Probability for Colorectal Cancer[18]	The ML algorithms used for diagnosis included random forest (RF), support vector machine (SVM), logistic regression (LR), and leave-one-dataset-out (LODO) analyses. Random forest was the most deployed algorithm in the study for its predictive capabilities and ability to determine accurate models using a minimal microbial signature. SVM was also noted for its effectiveness in classifying small data volumes.	The research study identified some specific microbial markers which are associated with CRC and then grouping them into 3 classes based on their significance. A database was constructed to enable inquiry and comparison of important information, on microbiome data for CRC diagnosis. The challenges in integrating microbiome data for CRC diagnosis, including sample types, taxonomic resolutions, robustness among different countries or regions, variability in microbial markers, and specificity.
5.	Daniel K Chan et al. 2016[19] World Journal of Gastroenterology	Diagnosing gastrointestinal illnesses using fecal headspace volatile organic compounds[19]	The study deployed the use of Artificial Neural Networks as a machine learning method to analyze and classify the collective patterns of VOCs in fecal samples to facilitate the diagnosis of gastrointestinal illnesses. It aimed to diagnose gastrointestinal conditions using fecal headspace VOC analysis.	The study explains that fecal headspace VOC analysis, together with advanced analytical methods and machine learning techniques have potentials of diagnosing multiple types of gastrointestinal diseases.
6.	Miodrag Cekikj et al. 2022 [20] Multidisciplinary Digital Publishing Institute (MDPI)	Understanding the Role of the Microbiome in Cancer Diagnostics and Therapeutics by Creating and Utilizing ML Models[20]	The study combined supervised, deep learning and ensemble methods, together with advanced feature selection techniques enabling a detailed model to help understand the gut microbiome-cancer relationship.	The findings in the study discuss the deep relationship between the gut microbiome and cancer. It showcases the possibility of machine learning models to unravel the complex biological interactions to bring about improved diagnosis and therapy.

7.	Zhang Qi et al 2022[21] BMC Microbiology	Prediction model of poorly differentiated colorectal cancer (CRC) based on gut bacteria[21]	The study deployed Machine Models such as Logistic regression, Random Forest, Neural network, Support vector machine, Gradient boosted decision tree, Cat Boost. The ML models deployed analyzed the differences in stool sample flora, and a prediction model was also built to differentiate colorectal cancer based on gut bacteria. In this study Random Forest Model was reported to have the highest accuracy.	The study findings highlight that the degree of pathological differentiation in CRC has a relation to gut flora. Poorly differentiated CRC exhibits distinct bacterial flora, which suggest that the intestinal bacteria can serve as biomarkers for predicting poorly differentiated CRC.
8.	Beatriz Alessandra et al. 2024[22] Frontiers in Microbiology	Identification of taxonomic changes in the fecal bacteriome associated with colorectal polyps and cancer: potential biomarkers for early diagnosis.[22]	The machine learning model employed in this study is Random Forest analysis. This method was used to identify predictive features (biomarkers) through a machine learning algorithm for classification. The study utilized 5,000 trees and seven predictors for the classification of the three groups (control, polyps, and CRC)	The main finding of the study is that individuals with polyps and colorectal cancer (CRC) exhibit distinct differences in the composition of gut bacteria compared to the control group, indicating potential microbial biomarkers for early detection and prognosis.
9.	Filippo Grazioli et al. 2022[23] PLOS Computational Biology	Microbiome-based disease prediction with multimodal variational information bottlenecks[23]	This study deployed machine learning models which include Random Forest, Support Vector Machine (SVM), Variational Autoencoder (VAE) used in the Deep Micro method, PopPhy-CNN, Multimodal Variational Information Bottleneck (MVIB). These methods were compared using their performance on various datasets.	The study highlights that the Multimodal Variational Information Bottle neck model performed better than other machine learning models in predicting multiple disease bases on data extracted from the gut microbiome.
10.	Shaobo Mo et al. 2022[24] Frontiers in Oncology	Fecal Multidimensional Assay for Non-Invasive Detection of Colorectal Cancer: Fecal Immunochemical Test, Stool DNA Mutation, Methylation, and Intestinal Bacteria Analysis[24]	The study deployed a linear support vector classification model to enable analysis of the data from multidimensional stool samples. Integrated features such as FIT, age, DNA mutation, DNA methylation, and bacteria levels were used to develop a diagnostic tool for the Colorectal Cancer detection.	The study highlights that a multidimensional stool analysis combining FIT and stool DNA tests improved the detection of colorectal cancer (CRC), with a specificity of 94.4% and a sensitivity of 81.5%. The test was more sensitive for stage II and III CRC, and showed a trend of higher sensitivity in smoking individuals.
11.	Hui Chong 2022[25] Oxford	EXPERT: transfer learning-enabled context-aware microbial community classification[25]	The study made use of an ontology-aware neural network framework for microbial community classification. Transfer learning techniques were used to adapt the knowledge of fundamental models to different classification contexts. This study primarily focused on the use of neural network-based models for microbial community classification.	The study highlights the ability of expert for high fidelity microbial community classification in a wide array of uses. These include disease classification, infant gut microbiome analysis and more.
12.	Hai Thanh Nguyen 2020 [26] Advances in Sciences, Technology and Engineering	Efficient Discretization Approaches for Machine Learning Techniques to Improve Disease Classification on Gut Microbiome Composition Data[26]	The stud deployed multiple machine learning algorithms including Random Forest, Linear Regression, as well as deep learning algorithms, such as Convolutional Neural Network (CNN). These models were used for prediction based on metagenomic data.	The study established that the proposed binning methods improved disease prediction tasks in metagenomics. The combination of scaler algorithms and binning methods showed promising results and outperformed existing methods in most cases. Data discretization using 5 bins achieved the best results. Overall, the study concluded that the proposed framework, Metagenomic-To-Bins (Met2Bin), has the potential to enhance the performance of machine learning algorithms in metagenomics.

IV. DISCUSSION

A complete overview of studies which used machine learning techniques to analyze gut microbiome profiles data in diagnosing and predicting colorectal cancer risks, was presented in the systematic review. The studies selected and reviewed focused on using machine learning methodologies like k-Nearest Neighbors (KNN), Neural Networks, Support Vector Machine (SVM), Random Forest, etc.

The findings from the studies selected for the systematic review are enlightening and remarkable. The main finding points to the fact that the use of machine learning techniques demonstrates a much higher prediction accuracy and enhanced performance when trying to distinguish between Colorectal Cancer Patients cases. This is an indication that machine learning models will potentially improve cancer and other disease diagnosis in the future. Great emphasis was laid on some machine learning algorithms like the General Regression Neural Networks (GRNN) and feature selection in accurate CRC detection based on gut microbiota analysis. Moving further the review pointed the review pointed to the potential of machine learning serving as a non-invasive method to detect colorectal cancer at an early stage, which can further help in the diagnosis and successful treatment of the disease before it reaches a fatal stage.

Secondly, the studies in the systematic review highlighted the possibility of machine learning in improving the diagnosis while employing the use of machine learning algorithms like, Random Forest and Linear Regression, as well as deep learning algorithms, including Convolutional Neural Network (CNN) to perform disease prediction tasks on metagenomic data. The outcome of these studies shows the potential impacts machine learning can have in the area of colorectal cancer prediction and most especially employing the use of gut microbiome profiles to diagnose the disease accurately. Improving the early detection window, by the use of the potential biomarkers available.

V. REPORT ON META-ANALYSIS

The Meta-analysis report on the study was conducted using GraphPad Prism. The human gut microbiome plays an important role in our overall human health, and understanding its composition and trends across different human populations offers valuable insights into the risks of colorectal cancer.

The Meta-analysis explores the various machine learning methods, which focus on the application of advanced machine learning algorithms such as convolutional neural networks (CNNs), to improve the accuracy of colorectal cancer predictions. This analysis shows the innovative methods which aim to improve the sensitivity and specificity of predictive models based on the gut microbiome data profiles.

The study highlights the utilization of CNN-based models for classifying gut microbiome profiles linked with the risks of colorectal cancer. By using deep learning techniques, the models demonstrate the ability to predict the risk of CRC based on the biomarkers present. The higher accuracy rates point towards the effectiveness of the models in helping healthcare practitioners assess patients and diagnose the disease early on with any invasive process.

In summary, the Meta-analysis points out the substantial strides in the prediction of colorectal cancer by using machine learning methods using gut microbiome profiles. The integration of new methods and an array of methodologies, and then coupled with a focus on data augmentation, contributes to enhancing the accuracy and efficiency of colorectal cancer risk prediction models.

VI. METHOD

A. Data Collection and Preparation

The Human Gut Microbiome Profile from various studies were collected and organized for this meta-analysis. All the data was thoroughly reviewed for to ensure the met the quality assurance and completeness benchmark before proceeding for further analysis.

B. Statistical Analysis

The use of the advanced statistical software GraphPad Prism was employed for the data analysis. The descriptive statistics were all computed to summarize the central tendency and variability of gut microbiome profiles across the different selected studies. The key statistical metrics like mean, minimum and maximum values were also calculated for each of the study's data, in order to provide the valuable insights into the distribution and characteristics of gut microbiome profiles associated with the risks of colorectal cancers.

TABLE 2

DATA EXTRACTED: FROM SYSTEMATIC REVIEWED PUBLICATIONS

Year/Sample Size	**116**	**156**	**162**	**390**
2017		1		
2018				
2020				4
2022	2		3	

TABLE 3

DESCRIPTIVE STATISTIC TABLE

SAMPLE SIZE	**116**	**156**	**162**	**390**
Number of values	1	1	1	1
Minimum	2.000	1.000	3.000	4.000
25% Percentile	2.000	1.000	3.000	4.000
Median	2.000	1.000	3.000	4.000
75% Percentile	2.000	1.000	3.000	4.000
Maximum	2.000	1.000	3.000	4.000
Range	0.000	0.000	0.000	0.000
Mean	2.000	1.000	3.000	4.000
Std. Deviation	0.000	0.000	0.000	0.000
Std. Error of Mean	0.000	0.000	0.000	0.000
Coefficient of variation	0.000 %	0.000 %	0.000 %	0.000 %
Skewness				
Kurtosis				
Sum	2.000	1.000	3.000	4.000

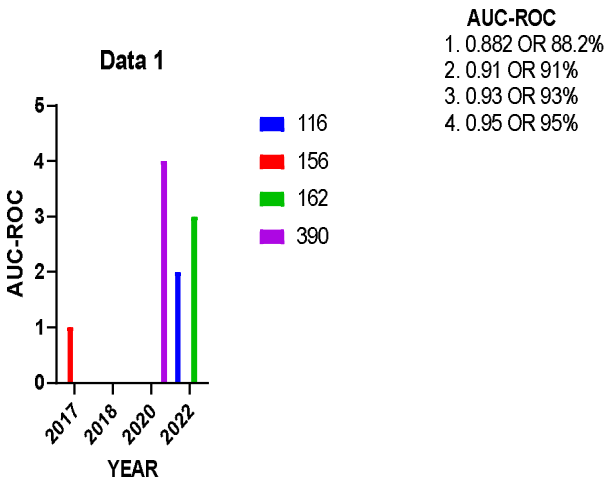

Figure 2. A graph showing statistical representation of the data

1) ***DATA 1 Analysis:*** A graph demonstrating a moderate predictive performance compared to higher AUC-ROC values.

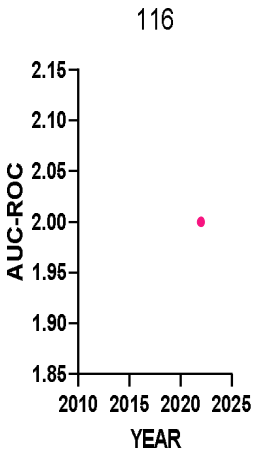

Figure 3. A graph showing data 1 analysis

2) ***DATA 2 ANALYSIS:*** The graph here maintains a relatively lower AUC-ROC value, suggesting comparatively weaker predictive performance supported by a moderate sample size.

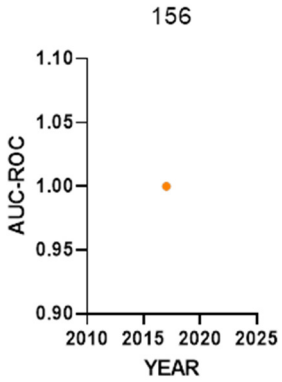

Figure 4. A graph showing data 2 analysis

3) ***DATA 3 ANALYSIS:*** Graph showing that despite the slightly lower AUC-ROC value compared to Data 4, Data 3 demonstrates strong predictive performance.

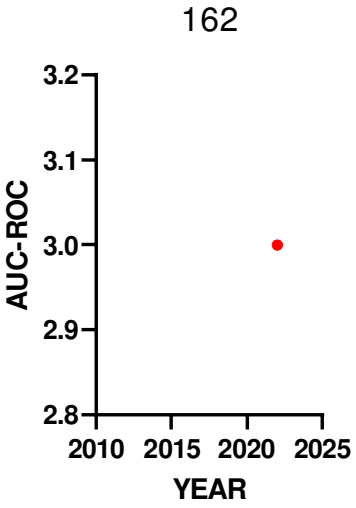

Figure 5. A graph showing data 4 analysis.

4) ***DATA 4 ANALYSIS:*** Data Analysis 4 suggestes a stable and high predictive performance for the given AUC-ROC value.

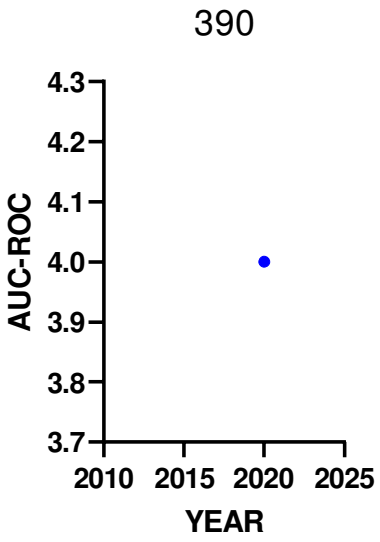

Figure 6. A graph showing data analysis 4

.

VII. RESULTS

This Meta-analysis, assesses the gut microbiome profiles as predictors for the risk of colorectal cancer by using AUC-ROC scores which were derived from multiple studies analyzed with GraphPad Prism software. AUC-ROC scores ranged from 0.882 to 0.95, indicating the predictive power of gut microbiome profiles for colorectal cancer risk. Variability in predictive performance was observed across cohorts, emphasizing the diverse association between gut microbiome profiles and colorectal cancer risk. These findings of the metanalysis underscore the potential of gut microbiome profiles as predictors for colorectal cancer risk, warranting further research and validation for clinical utility.

VIII. CONCLUSION

The systematic review and meta-analysis explored the existing studies which explore the potentials of machine learning methodologies in predicting colorectal cancer risks based on human gut microbiome profiles. The findings from the research reveal the various machine learning methodologies used to sort gut microbiome profiles and their links with the occurrence of colorectal cancer in various patients. The study

used metrics such as sensitivity, specificity, and AUC-ROC, to assess and provide a detailed understanding of the predictive capabilities of the machine learning models deployed. The study also emphasizes the necessity for standardized methodologies and larger sample sizes in future researches, in order to improve the reliability and accuracy of the models. Finally, the future of diseases and specifically cancer diagnoses using machine learning holds a lot of potentials, and will continue to improve as more data and refinement is made available.

REFERENCES

[1] J. Ferlay et al., "Cancer statistics for the year 2020: An overview," Int J Cancer, vol. 149, no. 4, pp. 778–789, 2021, doi: 10.1002/ijc.33588.

[2] E. E. Onuiri, O. Akande, O. B. Kalesanwo, T. Adigun, K. Rosanwo, and K. C. Umeaka, "A Systematic Review of Machine Learning Prediction Models for Colorectal Cancer Patient Survival Using Clinical Data and Gene Expression Profiles," Revue d'Intelligence Artificielle, vol. 37, no. 5, pp. 1273–1280, 2023, doi: 10.18280/ria.370520.

[3] R. Siegel, C. Desantis, and A. Jemal, "Colorectal Cancer Statistics , 2014," 2014, doi: 10.3322/caac.21220.

[4] H. Brenner, M. Kloor, and C. P. Pox, "Colorectal cancer," The Lancet, vol. 383, no. 9927, pp. 1490–1502, Apr. 2014, doi: 10.1016/S0140-6736(13)61649-9.

[5] M. Tepus and T. O. Yau, "Non-Invasive Colorectal Cancer Screening: An Overview," Gastrointest Tumors, vol. 7, no. 3, pp. 62–73, May 2020, doi: 10.1159/000507701.

[6] Y.-Q. Yu et al., "Health-related quality of life in advanced colorectal cancer patients in China: a nationwide hospital-based survey," Annals of Translational Medicine; Vol 10, No 6 (March 31, 2022): Annals of Translational Medicine, 2022, [Online]. Available: https://atm.amegroups.org/article/view/92240

[7] C. Chen et al., "Identification of intestinal microbiome associated with lymph-vascular invasion in colorectal cancer patients and predictive label construction," Front Cell Infect Microbiol, vol. 13, no. May, pp. 1–15, 2023, doi: 10.3389/fcimb.2023.1098310.

[8] S. Vatandoust, T. J. Price, and C. S. Karapetis, "Colorectal cancer: Metastases to a single organ," World J Gastroenterol, vol. 21, no. 41, p. 11767, 2015.

[9] I. Mármol, C. Sánchez-de-Diego, A. Pradilla Dieste, E. Cerrada, and M. Rodriguez Yoldi, "Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer," Int J Mol Sci, vol. 18, no. 1, p. 197, Jan. 2017, doi: 10.3390/ijms18010197.

[10] A. L. Kau, P. P. Ahern, N. W. Griffin, A. L. Goodman, and J. I. Gordon, "Human nutrition, the gut microbiome and the immune system," Nature, vol. 474, no. 7351, pp. 327–336, 2011, doi: 10.1038/nature10213.

[11] S. V. Lynch and O. Pedersen, "The Human Intestinal Microbiome in Health and Disease," New England Journal of Medicine, vol. 375, no. 24, pp. 2369–2379, 2016, doi: 10.1056/nejmra1600266.

[12] S. Pahlevan Sharif, P. Mura, and S. N. R. Wijesinghe, "Systematic Reviews in Asia: Introducing the 'PRISMA' Protocol to Tourism and Hospitality Scholars," Perspectives on Asian Tourism, vol. Part F184, pp. 13–33, 2019, doi: 10.1007/978-981-13-2463-5_2.

[13] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness, "PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis," Campbell Systematic Reviews, vol. 18, no. 2, pp. 1–12, 2022, doi: 10.1002/cl2.1230.

[14] J. A. C. Sterne et al., "RoB 2: a revised tool for assessing risk of bias in randomised trials," BMJ, vol. 366, p. l4898, Aug. 2019, doi: 10.1136/bmj.l4898.

[15] Y.-H. Zhou and G. Sun, "Improve the Colorectal Cancer Diagnosis Using Gut Microbiome Data," Front Mol Biosci, vol. 9, 2022, doi: 10.3389/fmolb.2022.921945.

[16] A. Arabameri, D. Asemani, and P. Teymourpour, "Detection of Colorectal Carcinoma Based on Microbiota Analysis Using Generalized Regression Neural Networks and Nonlinear Feature Selection," IEEE/ACM Trans Comput Biol Bioinform, vol. 17, no. 2, pp. 547–557, 2020, doi: 10.1109/TCBB.2018.2870124.

[17] L. Ai, H. Tian, Z. Chen, H. Chen, J. Xu, and J.-Y. Fang, "Systematic evaluation of supervised classifiers for fecal microbiota-based prediction of colorectal cancer," Oncotarget, vol. 8, no. 6, pp. 9546–9556, 2017, doi: 10.18632/oncotarget.14488.

[18] Z. Zhou et al., "Human Gut Microbiome-Based Knowledgebase as a Biomarker Screening Tool to Improve the Predicted Probability for Colorectal Cancer," Front Microbiol, vol. 11, 2020, doi: 10.3389/fmicb.2020.596027.

[19] D. K. Chan, C. L. Leggett, and K. K. Wang, "Diagnosing gastrointestinal illnesses using fecal headspace volatile organic compounds," World J Gastroenterol, vol. 22, no. 4, pp. 1639–1649, 2016, doi: 10.3748/WJG.V22.I4.1639.

[20] M. Cekikj, M. Jakimovska Özdemir, S. Kalajdzhiski, O. Özcan, and O. U. Sezerman, "Understanding the Role of the Microbiome in Cancer Diagnostics and Therapeutics by Creating and Utilizing ML Models," Applied Sciences (Switzerland), vol. 12, no. 9, 2022, doi: 10.3390/app12094094.

[21] Z. Qi et al., "Prediction model of poorly differentiated colorectal cancer (CRC) based on gut bacteria," BMC Microbiol, vol. 22, no. 1, 2022, doi: 10.1186/s12866-022-02712-w.

[22] B. A. R. Grion et al., "Identification of taxonomic changes in the fecal bacteriome associated with colorectal polyps and cancer: potential biomarkers for early diagnosis," Front Microbiol, vol. 14, no. January, pp. 1–17, 2023, doi: 10.3389/fmicb.2023.1292490.

[23] F. Grazioli, R. Siarheyeu, I. Alqassem, A. Henschel, G. Pileggi, and A. Meiser, "Microbiome-based disease prediction with multimodal variational information bottlenecks," PLoS Comput Biol, vol. 18, no. 4, pp. 1–27, 2022, doi: 10.1371/journal.pcbi.1010050.

[24] S. Mo et al., "Fecal Multidimensional Assay for Non-Invasive Detection of Colorectal Cancer: Fecal Immunochemical Test, Stool DNA Mutation, Methylation, and Intestinal Bacteria Analysis," Front Oncol, vol. 11, 2021, doi: 10.3389/fonc.2021.643136.

[25] H. Chong et al., "EXPERT: transfer learning-enabled context-aware microbial community classification," Brief Bioinform, vol. 23, no. 6, 2022, doi: 10.1093/bib/bbac396.

[26] H. T. Nguyen, N. Y. K. Phan, H. H. Luong, T. P. Le, and N. C. Tran, "Efficient discretization approaches for machine learning techniques to improve disease classification on gut microbiome composition data," Advances in Science, Technology and Engineering Systems, vol. 5, no. 3, pp. 547–556, 2020, doi: 10.25046/aj050368.

