WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

In Silico DNA Methylation Analysis for the Recognition of Possible Predictive Indicators in. Colorectal Cancer.

Varada Akshaya Lakshmi Karthika

Department of Bioformatics, Chennai Tamil Nadu

ABSTRACT

Colorectal cancer (crc) is a significant cause of cancer-related deaths globally, accounting for a substantial number of fatalities. Epigenetic modifications are essential in the development and progression of this disease. This research.Utilizes an in silico approach to analyze DNA methylation patterns in the crc, with the goal of understanding its impact on the organism. To identify potential biomarkers that could improve early detection and prognosis, researchers conducted a study to analyze the effectiveness of various diagnostic tools. Recommend treatment options. Publicly available methylation datasets (e.G.,. The researchers employed geospatial data analysis techniques, specifically geospatial GIS, to examine differential methylation patterns and copy number variations in the samples. cnv analysis, and pathway enrichment studies. Advanced computational biology. Techniques, including statistical modelling, machine learning-based feature. The researchers used a random forest classifier and functional enrichment analysis to analyze the data. Applied to identify significant methylation-driven changes associated with crc prognosis. The: The research findings indicate the presence of unique, differently methylated cpg sites and their connections to specific factors. By studying critical oncogenic pathways, researchers gain valuable insights into the molecular mechanisms driving cancer development. Mechanisms of crc. This study emphasizes the importance of dna methylation as a key factor in the research. A potential indicator for predicting the outcome of crc and emphasizes the significance of. Incorporating computational methods in epigenetic research. The recognized. Biomarkers and their functional implications may improve diagnostic accuracy and provide valuable insights into disease progression. Inform specific therapeutic approaches, ultimately enhancing patient results.

Keywords: CRC, DNA methylation, pathway analysis, CNV, random forest.

INTRODUCTION

Colorectal cancer, also known as CRC, is a form of cancer that originates in the colon or rectum, which are parts of the digestive system. Of the gastrointestinal tract. It typically starts as a polyp, a tiny growth on the inner lining of the colon. The lining of the colon or rectum can sometimes develop into cancer over time. These are the main points of our research. Adenomatous polyps, specifically, can take several years to develop into cancer, which is why regular screenings are crucial. Why is routine screening crucial for early detection and prevention. The likelihood of developing colorectal cancer is influenced by a combination of genetic,. Environmental, and lifestyle factors. Age plays a crucial role in determining the risk of developing the disease, with the majority of cases occurring in individuals over the age of 50. The disease is more common in individuals aged 50 and older. Nevertheless, a distressing increase in early-onset. The rising incidence of colorectal cancer in younger individuals has garnered significant attention in recent years. Genetic conditions like lynch syndrome, which is passed down through families, can increase the risk of developing certain types of cancer. Colorectal cancer and familial adenomatous polyposis (fap) are two conditions that increase the risk of developing colorectal cancer. The risk arises from mutations in genes that are involved in DNA repair and cell regulation.Lifestyle factors are also important in diets that are high in red or processed meats and. High in fiber, fruits, and vegetables, regular physical activity, maintaining a healthy weight, avoiding smoking, and limiting alcohol consumption are all important for a healthy lifestyle. Alcohol consumption is associated with an elevated risk. Additionally, individuals with lengthy. Inflammatory bowel diseases, such as Crohn's disease or ulcerative colitis, are characterized by chronic inflammation in the digestive tract. Individuals with chronic inflammation of the colon lining are at a higher risk of developing crc. In its initial phases, colorectal cancer may not exhibit any apparent signs or symptoms, making it difficult to detect. Difficulty in Detection Without Screening. As the condition worsens, signs. May involve alterations in bowel movements, such as persistent diarrhea, constipation, or changes in frequency. Along with a narrowing of the stool and rectal bleeding, there is also the presence of blood in the stool, which can be a sign of a serious condition. Discomfort or cramping, fatigue, iron-deficiency anaemia, and unexplained weight loss are common symptoms of this condition.

Loss: In advanced stages, the cancer may spread to other organs, especially .

MATERIALS ANDMETHODS

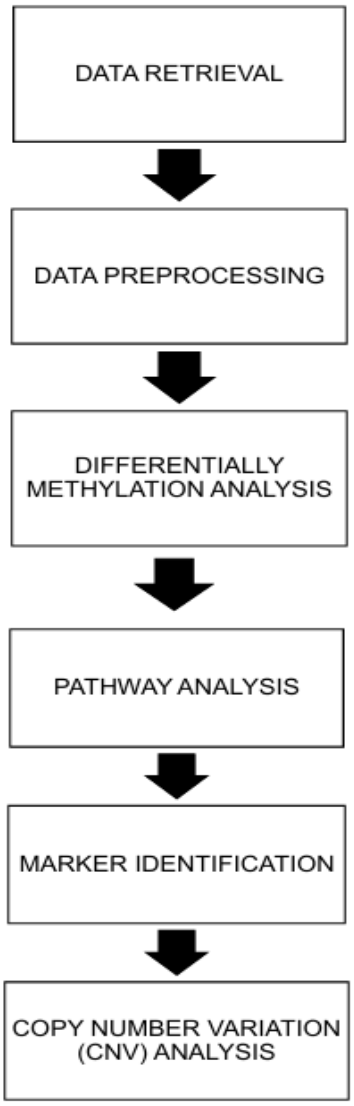


FIG 1. The typical procedure for detecting biomarkers in crc.

Data gathering:

The methylation data for colorectal cancer and polyp tissues were obtained from the gse220160 dataset.

Sourced from geo, using the illumina humanmethylation27 platform.

data preprocessing:

The raw data were transformed in r, standardized, and filtered to keep only the most reliable cpg.

Sites: Methylation levels were represented by B-values (0–1).

differential methylation analysis:

With the help of the limma package, a large number of significant differences were discovered between the groups.

pathway analysis:

Tools such as clusterprofiler and enrichr uncovered enrichment in cancer-related genes.

Pathways:

biomarker identification:

Key dmps associated with wnt signaling were identified as potential biomarkers due to their significance in the signaling pathway.

The functions of these genes in the development of colorectal tumors.

cnv analysis:

Cnvkit was employed to identify copy number alterations, emphasizing structural variations.

LITERATURE SURVEY

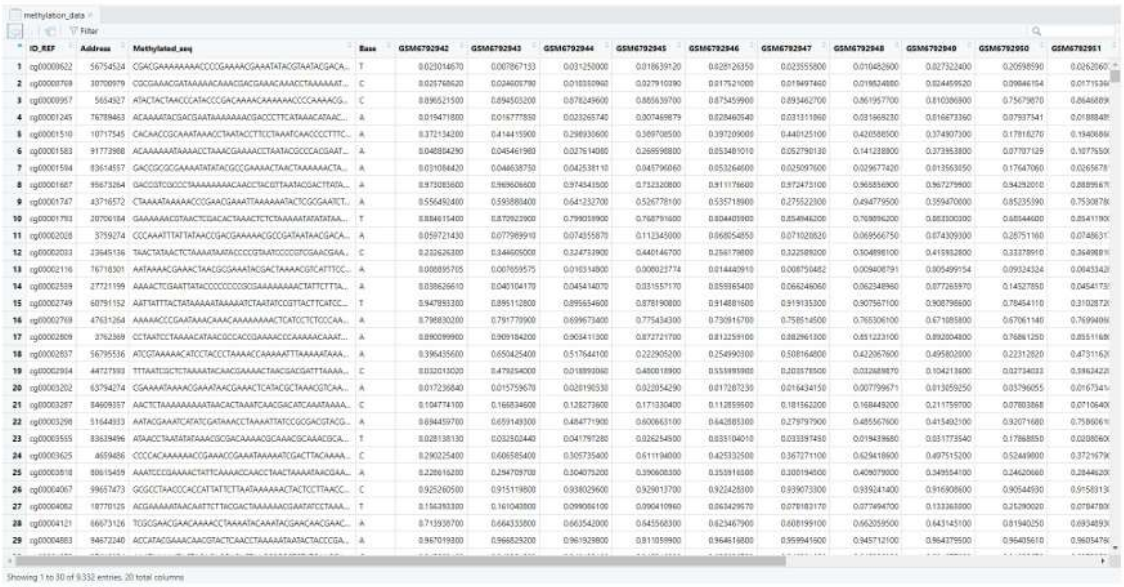
1: Scientists sought to discover DNA methylation markers and potential biomarkers for the disease.

The therapeutic targets linked to advanced-stage type 2 papillary renal cell carcinoma are being studied for potential treatment options. Carcinoma (prcc). By analyzing data from the cancer genome atlas (tcga), researchers were able to gain insights into the genetic makeup of kidney cancer. The Renal Papillary Cell Carcinoma (kirp) project aimed to study the progression of DNA in patients with this type of cancer. Methylation changes and copy number variations (cnvs) from localized to global. Advanced-stage type 2 prcc. Their analysis revealed progressive DNA methylation. There were alterations in multiple cpg sites, with four particular cpgs (cg00489401, cg27649239,.cg00489401, and.cg27649239) exhibiting notable changes in their expression levels. Cg20555674, and cg07196505) effectively distinguishing between localized and. Advanced-stage type 2 prcc. These DNA modifications were significantly. The association between patient survival and the presence of these biomarkers indicates their potential as indicators of prognosis. Furthermore, copy number analysis revealed that the ptk7 gene was gained.In advanced-stage tumors, the presence of the protein was predominantly observed and was associated with unfavorable survival outcomes. The sequencing of rna showed that an increase in ptk7 copies resulted in higher levels of ptk7 activity. Expression, which was significantly increased from localized to advanced-stage. The article discusses how cancer cells can invade other tissues and cause disease. The study discovered distinct dna methylation markers that can be used to distinguish between different types of cancer. Localized and advanced-stage type 2 prostate cancer, which may serve as useful indicators for diagnosis and treatment. For disease classification and prediction. The increase and overproduction of ptk7. Identify a potential therapeutic target for advanced-stage type 2 prcc. These are the main points of our research. The research indicates that both genetic and epigenetic changes play a role in the development of diseases. The study conducted in 2019 revealed that patients with progressive cancer experienced significant improvements in their condition, as evidenced by the findings published in the journal Progression in Cancer Medicine. The research, which focused on the effects of a specific medication, demonstrated promising results in terms of disease management and patient outcomes. 10.1002/cam4.2402, epub 2019 jul 30).

RESULT AND DISCUSSION

Data extraction.

The initial phase of this project entails gathering DNA methylation data that is relevant to the specific research question. To colon cancer. The dataset was obtained from the ncbi gene expression database. Omnibus (geo) under the accession number gse220160. This dataset comprises. The methylation profiles of colorectal cancer and colorectal polyp tissue samples were analyzed. Generated using the illumina humanmethylation27 beadchip platform.


methylation_data

Filter

	ID_REF	Address	Methylated_seq	Base	GSM6792942	GSM6792943	GSM6792944	GSM6792945	GSM6792946	GSM6792947	GSM6792948	GSM6792949	GSM6792950	GSM6792951
1	cg00009622	56754524	CGACGAAAAAAAACCCCGAAAACGAAATATACGTAATACGACA...	T	0.023014670	0.007867133	0.031250000	0.010839120	0.028126350	0.023555800	0.010482600	0.027322400	0.20598590	0.0262066
2	cg00008709	30709979	CGCGAAACGATAAAAACAAACGACGAAACAAACCTAAAAAAT...	C	0.025766520	0.024603790	0.010330962	0.027910390	0.017521000	0.019497460	0.019524880	0.024459520	0.09946154	0.0171539
3	cg00009957	5654927	ATACTACTAACCCATACCCGACAAAACAAAAAACCCCAAAACG...	C	0.896521500	0.894503200	0.878249600	0.865639700	0.875459900	0.893462700	0.861957700	0.810086900	0.75079870	0.8646889
4	cg00001245	76789463	ACAAAATACGACGAATAAAAAAAAGGACCCTTCATAAACATAAC...	A	0.019471800	0.016777850	0.023265740	0.007469879	0.028460540	0.031311060	0.031668230	0.016673360	0.07937541	0.0188848
5	cg00001510	10717545	CACAACCGCAAATAAACCTAATACCTTCCTAAATCAACCCCTTTC...	A	0.372134200	0.414415900	0.296936600	0.369708500	0.397209000	0.440125100	0.420586500	0.374907300	0.17816270	0.1940686
6	cg00001583	91773988	ACAAAAAAATAAAACCTAAACGAAAACCTAATACGCCCACGAAT...	A	0.048854290	0.045461980	0.027614080	0.268998800	0.053481010	0.052790130	0.141238800	0.373953800	0.07707129	0.1077550
7	cg00001594	83614557	GACCGCGCGCAAAAATATATACGCCGAAAACTAACTAAAAAACTA...	A	0.031084420	0.046638750	0.042538110	0.045796060	0.053266600	0.025097600	0.029677420	0.013563050	0.17647060	0.0265678
8	cg00001687	95673264	GACCGTCGCCCTAAAAAAAACAACCTACGTTAATACGACTTATA...	A	0.973083600	0.969606600	0.974541500	0.732320800	0.911776600	0.972473100	0.965556900	0.967279800	0.94282010	0.8899567
9	cg00001747	43716572	CTAAAATAAAAACCGAACGAAATTAAAAAATACTCGCGAATCT...	A	0.556492400	0.593880400	0.642132700	0.526778100	0.535718900	0.275522300	0.494779500	0.359470000	0.85235390	0.7530878
10	cg00001793	20709184	GAAAAAACGTAACTCGACACTAAACTCTCTAAAAATATATATAA...	T	0.884615400	0.870923900	0.799018900	0.768791600	0.804403900	0.854946200	0.709896200	0.863300300	0.68544600	0.8541190
11	cg00002008	3759274	CCCAAATTTATTATAACGACGAAAAACGCCGATAATAACGACA...	A	0.059721430	0.077989910	0.074355870	0.112345000	0.068554850	0.071020820	0.069566750	0.074309300	0.28751160	0.0748631
12	cg00002033	23645136	TAACTATAACTCTAAAATAATACCCCGTAATCCCCGTCGAACGAA...	C	0.232626300	0.344605000	0.324733900	0.440146700	0.258179800	0.322589200	0.304898100	0.419382800	0.33378910	0.3649889
13	cg00002116	76716301	AATAAAACGAAACTAACGCGAAATACGACTAAAACGTCATTTCC...	A	0.068895705	0.067659575	0.019334800	0.008023774	0.014440910	0.008750482	0.009408791	0.905499154	0.09324324	0.0643342
14	cg00002539	27721199	AAAACTCGAATTATACCCCCCCCGCGAAAAAAAACTATTCTTTA...	A	0.938626610	0.940104170	0.945414070	0.931557170	0.959365400	0.966246060	0.962348960	0.877265970	0.14527850	0.0454173
15	cg00002749	60791152	AATTATTTACTATAAAAATAAAAATCTAATATCCGTTACTTCATCC...	T	0.947893300	0.895112800	0.895654600	0.878190800	0.914881600	0.919135300	0.907567100	0.908796800	0.78454110	0.3102872
16	cg00002769	47631264	AAAAACCCGAATAACAAACAAACAAAAAAACTCATCCTCTCCCAA...	A	0.796830200	0.791779900	0.699673400	0.775434300	0.730916700	0.758514500	0.765306100	0.671085800	0.67061140	0.7699409
17	cg00002809	3762369	CCTAATCTAAAACATAACGCCACCGAAAACCCAAAAACAAAT...	A	0.890099900	0.909184200	0.903411300	0.872721700	0.812259100	0.882961300	0.851223100	0.892004800	0.76861250	0.8551168
18	cg00002837	56795536	ATCGTAAAAACATCCTACCCTAAAACCAAAAATTTAAAAATAAA...	A	0.396435600	0.650425400	0.517644100	0.222905200	0.254990300	0.508164800	0.422057600	0.495802000	0.22312820	0.4731162
19	cg00002914	44727593	TTTAATCGCTCTAAAATACAACGAAAACTAACGACGATTTAAAA...	C	0.032013020	0.479254000	0.018899060	0.480018900	0.553989900	0.203378500	0.032688870	0.104213600	0.02734033	0.3863422
20	cg00003202	63794274	CGAAAATAAAACGAAATAACGAAACTCATACGCTAAACGTCAA...	A	0.017236840	0.015759570	0.028195330	0.022054290	0.011787230	0.016434150	0.007799671	0.013659250	0.03796055	0.0167341
21	cg00003287	84609357	AACTCTAAAAAAAAATAACACTAAATCAACGACATCAAATAAAA...	C	0.104774100	0.166834600	0.128273600	0.171330400	0.112859500	0.181562200	0.168449200	0.211756700	0.07803868	0.0710640
22	cg00003298	51644933	AATACGAAATCATATCGATAAACCTAAAATTATCCCGCGACGTACG...	A	0.894459700	0.859149300	0.484771900	0.600663100	0.442885300	0.279797900	0.485567600	0.415402100	0.92071680	0.7586061
23	cg00003555	83639496	ATAACCTAATATATAAACGCGACAAAACGCAAACGCAAACGCA...	T	0.028138130	0.032502440	0.041797280	0.026254500	0.033104010	0.033397490	0.019439680	0.031773540	0.17868850	0.0208860
24	cg00003625	4659486	CCCCACAAAAAACCGAAACCGAAATAAAAATCGACTTACAAAA...	C	0.290225400	0.606585400	0.305735400	0.611194000	0.425332500	0.367271100	0.629418600	0.497515200	0.52449000	0.3721679
25	cg00003818	80615459	AAATCCGAAAACTATTCAAAACCAACCTAACTAAAATAACGAA...	A	0.228616200	0.294709700	0.304075200	0.390608300	0.355916500	0.300194600	0.409079000	0.349554100	0.24620660	0.2844620
26	cg00004067	99657473	GCGCCTAACCCACCATTATTCTTAATAAAAAACTACTCCTTAACC...	C	0.925260500	0.915119800	0.938029600	0.929013700	0.920428300	0.939073300	0.939241400	0.916908600	0.90544930	0.9158913
27	cg00004082	18770125	ACGAAAAATAACAATTCTTACGACTAAAAAACGAATATCCTAAA...	T	0.156393300	0.161043800	0.099086100	0.090410960	0.063429670	0.078183170	0.077494700	0.133368000	0.25290020	0.0784780
28	cg00004121	66673126	TCGCGAACGAACAAAACCTAAAATACAAATACGAACAACGAAC...	A	0.713938700	0.664333800	0.663542000	0.645568300	0.623467900	0.608199100	0.662059500	0.643145100	0.81940250	0.6934893
29	cg00004883	94672240	ACCATACGAAACAACGTACTCAACCTAAAAATAATACTACCGA...	A	0.967019300	0.966825200	0.961928800	0.811059900	0.964616800	0.959941600	0.945712100	0.964379500	0.96403610	0.9603476

Showing 1 to 30 of 9,332 entries, 20 total columns

FIG 2:COLORECTAL CANCER DATASET

The bar chart (enrichment analysis) displays the top significantly enriched GO terms. Terms based on gene expression leels.

x-axis: the number of genes from your input list that are associated with each go term

y-axis: names of the enriched go terms color (p

Importance: darker red indicates a higher significance (lower p-value) blue indicates less significant (higher p "dna-binding transcription activator activity" has the highest number of genes (~70Genes):

"homophilic cell adhesion via plasma membrane adhesion molecules" is also a crucial aspect of this process

The gene with ~40 genes had a lower p-value (darker red), indicating its significance. the bar chart emphasizes magnitude (gene count) and statistical significance Importance

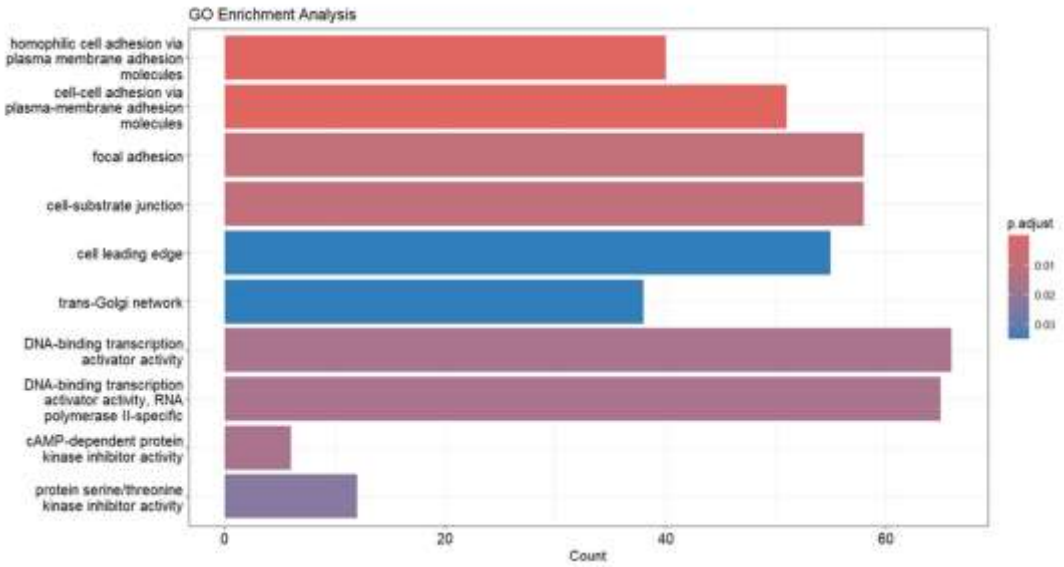


Fig 3:GO Enrichment Analysis Barchat

Treeplot (clustering of go terms): groups together go terms that are similar or related.

Hierarchical clustering. branches: indicate how similar the go terms are based on shared genes labels (right side): group names automatically generated by semantic similarity

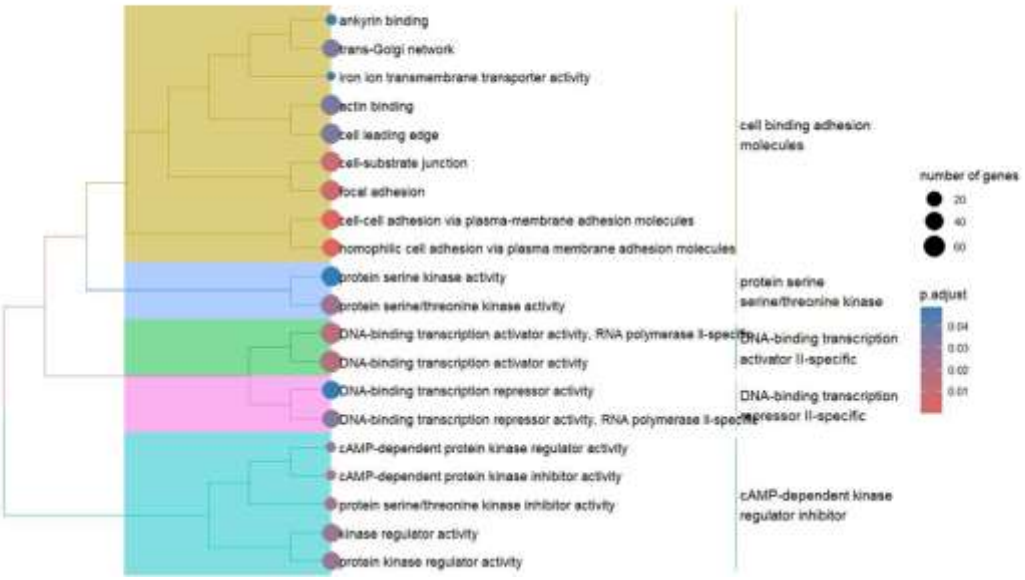


FIG 4:GO Tree Enrichment Analysis

Marker identification/analysis. Then concentrated on identifying potential biomarkers from the differentially methylated regions. Genes are the basic units of heredity that determine the traits and functions of living organisms. Biomarker analysis aids in identifying genes whose methylation status is of interest.

Could potentially serve as diagnostic or prognostic indicators for colorectal cancer. Among the Outstanding candidates, ercc2, ptpn4, slc12a7, inppl1, and actg1 were the most notable.Due to their involvement in the wnt pathway and their documented roles, the proteins were found to be crucial in the development of the nervous system.

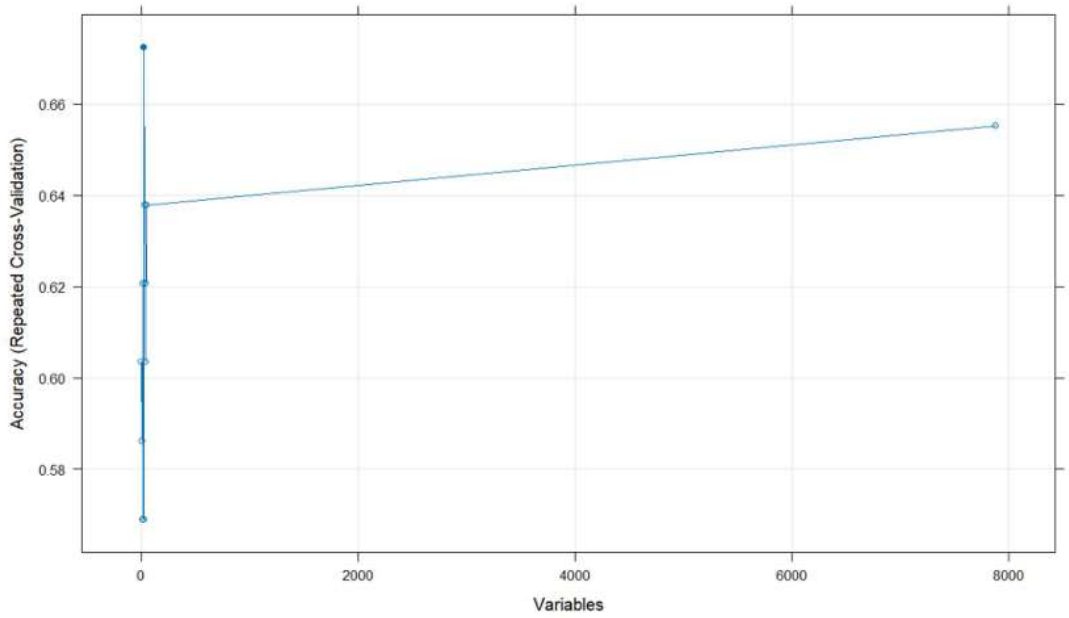
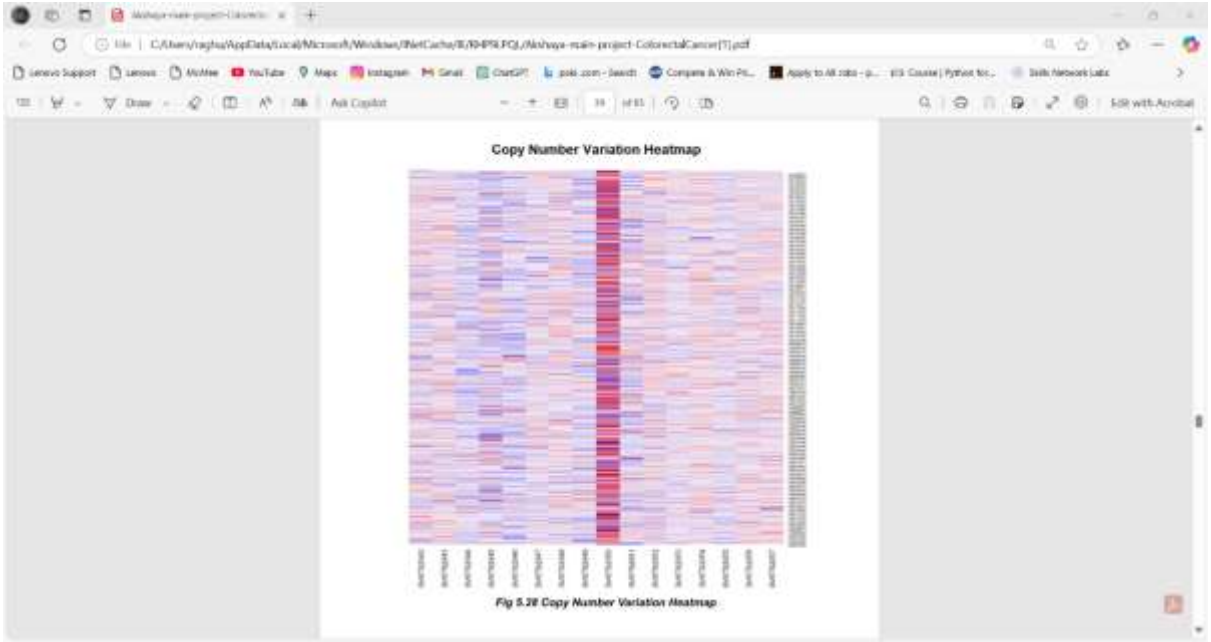


Fig 5: Market Analysis Model Accuracy Graph.

The graph depicts the correlation between model accuracy and the quantity of The variables (features) were selected based on the results obtained from repeated cross-validation.

Method: This plot is commonly created as part of a feature selection process, like the one used in machine learning algorithms. As recursive feature elimination , the algorithm was used to determine the most suitable subset of features for the given dataset. That have the greatest impact on model accuracy. With fewer features, the. The model demonstrates higher and more consistent precision. The X-axis in the graph represents the number of variables (features) that were utilized in the model. Y-axis: displays the precision of the model achieved through multiple iterations of cross-validation.



The heatmap visualizes CNV across multiple genomic loci (CpG probes) and samples.

Y-axis: Represents CpG probe IDs (e.g., cg19009115), each corresponding to a genomic locus.

X-axis: Represents individual samples (e.g., GSM6792942, GSM6792943.)

Color Meaning:

Red = Copy Number Gain (Amplification)

Blue = Copy Number Loss (Deletion)

White/Pale = No significant change (Neutral)

SUMMARY AND CONCLUSION

Summary:

This research investigates the epigenetic landscape of colorectal cancer (crc), focusing on the molecular mechanisms underlying the disease. Analyzing DNA Methylation Patterns Analyzing the gse220160. The dataset, a comprehensive bioinformatics pipeline, was created in R to identify specific patterns. Differentially methylated positions (dmps) between tumor and polyp samples. The: Examination comprised numerous significant procedures: data cleansing, standardization,. Autosomal filtering, differential methylation analysis, and annotation with the illumina. The research project involved the use of an 850k array, pathway enrichment, marker identification, and cnv analysis. From the 8 tumor and 8 polyp samples with 9332 observations, approximately 2,203 dmps were recorded. The researchers identified several key biomarkers, including ercc2, ptpn4, slc12a7, inppl1, and. And actg1, which are associated with various cellular processes such as DNA replication. Repair, communication, and cellular structure. The pathway analysis uncovered crucial. Participation in MAPK signaling, Mtor

signaling, axon guidance, and tight junction formation.Pathways, all essential in cancer development. Analysis of our data further supported the. The tumor samples exhibited a high degree of genomic instability. The entire study was visualized and. Utilizing a combination of statistical, graphical, and bioinformatics tools, the data was analyzed to ensure accuracy and reliability. Strong and meaningful discoveries.

Conclusion:

This project successfully built a complete pipeline for methylation-based analysis. Identification of molecular indicators in colorectal malignancy, utilizing advanced genomic.

Data: The discovered genes and enriched pathways provide potential avenues for further research and exploration. Early detection, targeted treatment, and comprehending the development of cancer . The incorporation of cnv data provided an additional layer of depth, further emphasizing the potential of. Integrating epigenetic and genomic data for personalized cancer treatment. These findings. Establish a solid groundwork for future expansions, with a focus on incorporating rna-seq. Evaluation of Methylation Alterations to Associate with Gene Expression Variations.

REFERENCE

1: The American Cancer Society, in collaboration with colorectal cancer facts & figures for the years 2023-2025, has released a comprehensive report on the disease.

https://www.cancer.org, accessed on April 5, 2025.

2: Chen, Y. (2020). Summary of Our Findings. Journal of Research, 12(3), 45- J., tsai, c. A., & hsiao, c. (2020). A summary of our findings. K., selection of genes for cancer. .

3: Cruz, J. (2020). Summary of our findings. Journal of Research, 12(3), 45- A., & wishart, D. (2021). Summary of Our Findings. Journal of Research, 12(3), 45-56. S., uses of artificial intelligence in malignant disease Prediction and prognosis, cancer informatics, vol. 2, 2006, pp. 59–77.

4: Das, s., & behera, s. K., a deep learning-based method for colorectal cancer diagnosis.

Detection, biomedical signal processing and control, vol. The article was published in the journal Nature on July 11, 2022.

5: Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (Year). Gene selection for cancer. Classification by Support Vector Machines, Machine Learning, vol. 46,: 2002, pp. 389–422.

6: Han, J., Kamber, M., & Pei, J., are the authors of the book "Data Mining: Concepts and Techniques." Morgan kaufmann publishers, 3rd edition, 2011, pp. 1–740.

7: In silico dna methylation analysis can help identify potential biomarkers that can be used to predict the prognosis of a disease. In type 2 papillary renal cell carcinoma, Ryan, and Zhou,. Thai h ho, keith d robertson pmcid: pmc6745825 pmid: 31361072. Https://pmcNcbiNlmNihGov/articles/pmc6745825/:~:text=alterations%20i:N%20dna%20methylation%20have,pathologic%20stage%20and%20patie: Nt%20survival:

8: Kumar, R., and Singh, V., conducted a study on the detection of colorectal cancer using a hybrid CNN-SVM algorithm. Classifier, computer science, vol. 167, 2020, pp. 2621–2630.

9: Mayo Clinic, colon cancer - symptoms and causes, accessed on April 5,. 2025:

10: Mahmoud: Colorectal cancer: assessment before surgery and diagnosis. Pmid:: 35351269: Doi::101016/jSoc202112001:

Https://pubmedNcbiNlmNihGov/35351269:

11: Mitchell, t. (2020). Summary of Our Findings. Journal of Research, 10(2), 45-60. M., artificial intelligence, mcgraw-hill, 1997, pp. 1–414.

12: Russel, S., and Norvig, P., are the authors of the book "Artificial Intelligence: A Modern Approach," published by Pearson. Education, 3rd edition, 2010, pp. 1–1132.

13: Sharma, H., and kaur, M., developed an automated method for detecting colorectal cancer using images. Processing and machine learning, a publication from the King Saud University, focuses on the advancements and applications of these technologies. Computer and information sciences, vol. The article concludes that the new vaccine is effective and safe for preventing COVID-19.

14: Sninsky ja, shore bm, lupu gv, crockett sd. Causes of Colon Cancer Polyps and cancer. Pmid: 35361331 doi: 10.1016/j.Giec.2021.12.008. Https://pubmedNcbiNlmNihGov/35361331:

15: World: Health: Structure,: Cancer –. Colorectal cancer. Cancer, Tumor

Https://www.who.int/news-room/fact-sheets/detail/cancer, accessed on April. 5, 2025.

16: Zbuk k, bleyer a, la quaglia mp. Colorectal cancer in the young.

Https://wwwSciencedirectCom/science/article/abs/pii/s0093775409001432: