

A retrospective analysis of key predictors and patient outcomes: Using artificial intelligence for precision survival prediction in colorectal cancer

Abstract

Objective: The study aims to identify critical predictors of mortality and evaluate the performance of different artificial intelligence (AI) models among patients with colorectal cancer (CRC). Furthermore, the study also sought to enhance our comprehension of survival outcomes by identifying key predictors and evaluating the accuracy of AI-driven prediction methods for patients with CRC.

Methods: The study employed a retrospective-predictive design, using data from the electronic health records of patients with colorectal cancer (CRC) admitted between 2016 and 2023. Among the eight AI models created by the SPSS Modeler version 18.0, the Bayesian network model was the most effective of the eight models in this study.

Results: The researchers identified the most relevant variables associated with mortality among patients with CRC through data visualization. The study analysed 1,159 colorectal cancer patients, with 45.7% living up to six years and 54.3% living between seven and 16 years post-diagnosis. The Bayesian network AI model identified stage, age, recurrence, sex, marital status, and smoking status as key predictors.

Conclusion: This study model's structure emphasizes these predictors' interconnectedness because parent nodes directly connect to child nodes. The model shows how age, smoking status, marital status, cancer stage, and recurrence affect patient survival. The model clarifies these variables' interactions.

Keywords: artificial intelligence, colorectal cancer, survival prediction, mortality, Bayesian network

Muayyad M. Ahmad¹
Islam Bani
Mohammad²

1 Clinical Nursing Department,
School of Nursing, The
University of Jordan, Amman,
Jordan

2 Department of Nursing,
Faculty of Nursing, Al-Balqa
Applied University, Al-Salt,
Jordan

Introduction

Colorectal cancer (CRC) is one of the leading causes of cancer deaths globally, and recurrence risk affects patient survival [1]. According to studies, recurrent CRC patients' median survival duration depends on the location and timing of the CRC recurrence [2], [3].

Artificial intelligence (AI) is crucial for health prediction due to its capacity to rapidly and accurately evaluate and process vast quantities of data [4], [5]. Traditional approaches may fail to detect subtle patterns and correlations in complex datasets that AI algorithms can find [6], [7]. Furthermore, traditional health prediction uses statistical techniques that are sometimes limited by human biases and the inability to consider multiple variables [8]. AI excels in managing multidimensional and complex data, resulting in more accurate forecasts and deeper insights. Prediction driven by AI can enhance patient outcomes and save lives by enabling early disease detection and treatment [8].

Few studies have shown that AI can improve diagnostic accuracy, treatment planning, and patient care in CRC prediction [9]. For early detection of CRC, Schiele et al.

[10] used deep learning algorithms to assess histopathological pictures and predict recurrence. Gonzalez et al. [11] used machine learning algorithms to predict CRC survival and recurrence using Electronic Health Records (EHR) data. Tamang and Kim [12] created an AI-based approach, which has improved polyp identification rates over traditional approaches, minimizing misses and enhancing early diagnosis.

The survival rates of CRC for men and women are significantly different [13]. According to research, women have better CRC survival rates than men [14]. Due to genetic differences in tumor biology and hormonal variables, females often have less severe cancer types [14], [15]. Behavioral and lifestyle factors, including screening and treatment adherence, can affect survival results [16]. Other research shows that while women have better survival rates, they often have more advanced CRC upon diagnosis than men [13]. Male patients are diagnosed earlier and may proceed more aggressively [14].

Age, which reflects the patient's health and disease stage upon diagnosis, affects CRC survival. Younger CRC patients have a more aggressive disease and a worse prognosis, according to research [17]. Despite being discov-

ered later due to lower screening rates and comorbidities, elder people usually have a less aggressive disease [18], [19]. However, due to cancer stage and therapy access, senior people may have better survival rates [20].

According to research, married people have better five-year survival rates than single people [21]. This outcome is possibly due to better support systems, treatment adherence, and psychological well-being [22]. Social support from a spouse or partner can help a patient manage the disease, follow treatment, and preventive measures, improving survival [23].

Non-smokers have a decreased chance of CRC, less severe illness, and better survival [24]. Recent Cancer Research findings link smoking to a higher rate of colon cancer and more aggressive tumor behavior, which lowers survival rates [25]. The carcinogenic effects of tobacco, which accelerate tumor progression and metastasis, raise the chance of smokers developing advanced-stage disease and having greater treatment response problems [26], [27]. Smoking can also reduce cancer treatment efficacy and exacerbate treatment-related problems, lowering survival rates [26].

Earlier TNM (Tumor, Node, and Metastasis) stages at diagnosis greatly improve CRC survival [28]. Gastroenterology studies show that patients with stage I and stage II CRC have a higher five-year survival rate [29]. Due to early treatment and lower illness severity at diagnosis, these patients often have a better quality of life and longer survival rate [29].

Objectives

This study aimed to examine the AI models in predicting the duration of life from the time of CRC diagnosis until mortality. The study also sought to enhance our comprehension of survival outcomes by identifying key predictors and evaluating the accuracy of AI-driven prediction methods for patients with CRC.

Methods

Study design

The study employed a retrospective-predictive and design, using data from the EHR of patients admitted between 2016 and 2023.

Eligibility criteria

The necessary variables were implemented to retrieve the records of patients with CRC from the chosen institution. The inclusion criteria consist of data regarding patients who died with CRC.

Data processing

The data acquisition procedure was completed in four months after consent was obtained to extract the necessary data. To identify missing values, outliers, and inconsistencies, the data were subjected to frequency analysis and inspection. The retrieved data was sorted, cleaned, and structured, resulting in the elimination of redundant material. The Statistical Package for Social Sciences, version 29.1, was employed to conduct descriptive statistics [30].

The researchers identified the most relevant variables associated with mortality among patients with CRC through data visualization. Furthermore, SPSS Modeler version 18.0 was employed to visualize, analyze, and administer the data [31]. This software has the capacity to manage data for descriptive and predictive modeling, as well as to display data with high statistical power.

As shown in Table 1, eight AI models were generated to predict the length of time lived after diagnosis with CRC in relation to the study variables. Based on the area under the curve (AUC) (0.845) and the overall accuracy of 81.828%, the Bayesian network model was selected to examine the predictors and the probabilities for age lived until death among patients with CRC.

Table 1: The eight models created from the study data

Model	Area under curve	Overall accuracy (%)
Bayesian network	0.845	81.828
C&R tree	0.812	80.864
C5	0.784	78.173
CHAID	0.778	74.244
Logistic regression	0.744	72.653
Neural net	0.706	69.980
Decision list	0.681	66.762
Quest	0.663	64.880

Ethical consideration

The study was authorized by the Scientific Research Committees at The University of Jordan. Additionally, the study facility's ethical committee granted permission for data acquisition under the reference number (IRB-JUH 10/24/1503; dated on 16/1/2024). Patient information was managed with assurance by employing an ID as a unique identifier for each record.

Results

A total of 1,159 patients with CRC who died within one to 16 years of their diagnosis comprised the study sample. The duration of the study sample's life after being diagnosed with CRC was divided into two groups: one group, which lived up to six years ($n=530$, 45.7%), and the other group, which lived between seven and 16 years ($n=629$, 54.3%). The study sample consisted of slightly more

males (55.9%, n=648). The largest cohort consisted of patients aged 65 to 79 years (37.3%). The number of nonsmokers (n=674, 58.2%) was greater than that of smokers. The sample consisted of approximately 83% married individuals. Most of the patients (62.3%) were in stage 4, while 26% were in stage 3, as per the TNM staging system. The disease recurred in only 25% of the patients following treatment (Table 2).

Table 2: Descriptive statistics for the study variables (N=1,159)

Variables	n	%
Age lived until death		
1–6 years	530	45.7
7–16 years	629	54.3
Sex		
Male	648	55.9
Female	511	44.1
Age		
20–49	167	14.4
50–64	343	29.6
65–79	432	37.3
80–104	217	18.7
Smoking		
Nonsmoker	674	58.2
Smoker	485	41.8
Marital status		
Married	967	83.4
Single, divorces, or widow	192	16.6
TNM stage		
0	62	5.3
1	10	0.9
2	64	5.5
3	301	26.0
4	722	62.3
Recurrence		
No	293	74.7
Yes	866	25.3

The Bayesian network model (Figure 1) illustrates the relationships between the outcome variable (length of time lived before death, 1–6 years versus 7–16 years) and the six predictors. The graphics-based Bayesian network model shows probabilistic links between variables. A Bayesian network parent node has direct edges to one or more child nodes. The child node's status depends conditionally on its parents. Predecessor nodes affect offspring probability distribution. The link between parent-child pairs, caused by the parent nodes' different conditions, suggests the child node's probability evolution [32].

The AI model identified six predictors regarding the studied outcome (length of time lived before death), as illustrated in Figure 2. The TNM stages are the most significant predictor, followed by the age categories, recurrence of CRC, sex of individuals, marital status, and smoking status.

The Bayesian network models with predictor variables and the probabilities of the length of time lived before death are illustrated in Table 3. The conditional probability was 0.46 for individuals who lived for 1 to 6 years and 0.54 for those who lived for 7 to 16 years. The conditional probability of death within 1 to 6 years was 83% for females who did not smoke, compared to 42% for males.

Regarding the duration of time until mortality, males had conditional probabilities of 0.93 and 0.92, respectively, while married females had identical probabilities of 0.72. The highest conditional probabilities that resulted from the parent nodes' sex of patients and the length of time before death were 0.42 for males aged 65 to 79 and 0.38 for females in the same age group. In general, the probability of males living for 7 to 16 years was higher than that of females (0.58). The age group with the earliest age group had a 0.75 probability of living 1 to 6 years and was in TNM stage 4. The age group 65–79 and TNM stage 4 had a 0.71 probability of living 1 to 6 years. The youngest age group, which lived from 1 to 6 years and 7 to 16 years, had the highest conditional probabilities of recurrence (0.31 and 0.32).

Discussion

The study developed eight AI models to predict the duration of life following a CRC diagnosis and represents a substantial advancement in the application of machine learning to oncology. The Bayesian network model, which was chosen for its overall accuracy of 81.828% and area under the curve (AUC) of 0.845, demonstrates a robust predictive quality. The incorporation of a diverse array of predictors, such as age, smoking status, marital status, cancer stage, and recurrence, facilitates a broad assessment of factors that affect survival outcomes.

The conditional probabilities of survival for patients with CRC are comprehensively understood by the Bayesian network model introduced in this study. Similar to a study conducted by Ghebrial et al. [33], our model illustrates that there are significant disparities in survival probabilities, with nonsmoking females having a substantially higher likelihood of mortality within 1 to 6 years than their male counterparts. Additionally, the model suggests that males have a higher probability of surviving for 7 to 16 years, particularly in the 65 to 79 age group. These findings emphasize the complexity of predicting patient outcomes, offering an understanding of the interaction between demographic and clinical factors in the context of CRC survival [34].

Traditional studies frequently emphasize the significant impact of early detection and treatment advancements on the enhancement of CRC survival rates, particularly for stage 3 and 4 patients [35], [36], [37]. However, the high conditional probabilities of mortality within 1 to 6 years for both younger and older patients in stage 4 in this study suggest that late-stage CRC continues to present significant survival challenges, despite the progress that has been made. Furthermore, and consistent with a study by Gheybi et al. [38], the finding that the risk of recurrence is highest in the youngest age cohort within both survival brackets (1–6 years and 7–16 years) may indicate that the model fails to adequately account for the underlying biological factors or disparities in treatment efficacy. This emphasizes the importance of conducting further research to investigate the potential benefits of personal-

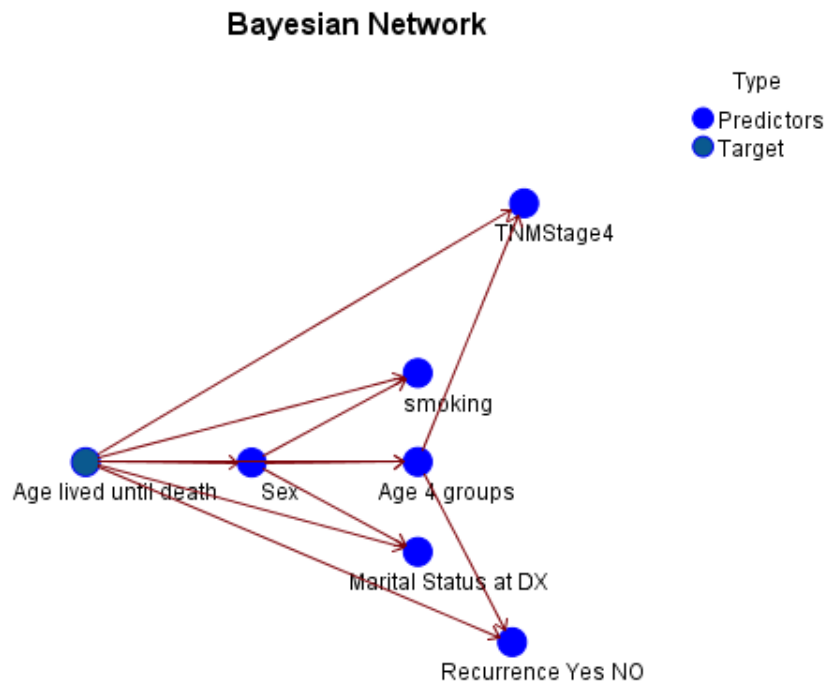


Figure 1: The Bayesian model with the study variables' interrelationship

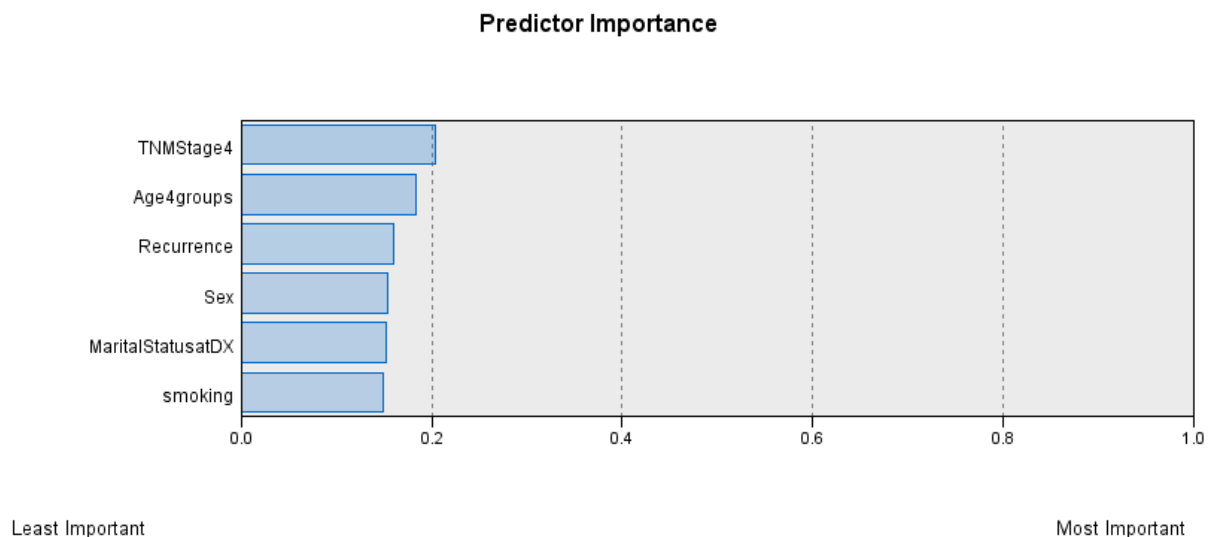


Figure 2: The study predictors' importance in the Bayesian model

ized treatment strategies and age-specific recurrence mechanisms.

The study's findings indicate that nearly half of the patients diagnosed with CRC survived for a period of up to six years, while a slightly larger proportion survived for a period of seven to sixteen years. This division implies that the survival outcomes of CRC patients are subject to significant variation, which may be influenced by a variety of factors, such as the stage of the disease at diagnosis, the efficacy of treatment, genetic predispositions, and overall health [15]. The predominance of males and the largest cohort in the age group of 65 to 79 years emphasize the demographic trends that are consistent with the existing literature, which recognizes the significant role of gender and older age in the prognosis of CRC [3], [33].

However, there are certain arguments that can be made when comparing these results to the broader literature. This study does not provide information regarding the distribution of cancer stages at the time of diagnosis or the specific treatments that were implemented, even though numerous studies have emphasized the importance of early detection and sophisticated treatment options in improving survival rates [39]. Furthermore, the results of the study suggest that the sample population is predominantly composed of married individuals and nonsmokers, with most of the stage 4 CRC cases. The increased proportion of nonsmokers is in accordance with specific literature that implies that nonsmokers may experience superior health outcomes and longevity because of the absence of smoking-related complications following a diagnosis [24]. In contrast to other studies,

Table 3: The Bayesian network model's predictors and the probabilities for age lived until death among patients with CRC

Parents nodes	Conditional probabilities of smoking		Conditional probabilities of marital status		Conditional probabilities of age				Conditional probabilities of sex	
1–6 vs. 7–16 years/sex	Nonsmoker	Smoker	Single, widow, divorced	Married	21–49	50–64	65–79	80–104	Female	Male
1–6/female	0.83	0.17	0.28	0.72	0.25	0.36	0.27	0.12	0.47	
7–16/female	0.81	0.19	0.28	0.72	0.15	0.29	0.38	0.18	0.42	
1–6/male	0.42	0.58	0.07	0.93	0.11	0.31	0.42	0.16		0.53
7–16/male	0.37	0.63	0.08	0.92	0.09	0.25	0.40	0.26		0.58
	Conditional probabilities of TNM stages					Conditional probabilities of recurrence				
1–6 vs. 7–16 years/age 4 groups	0	1	2	3	4	No	Yes			
7–16/21–49 years	0.07	0.00	0.00	0.22	0.71	0.69	0.31			
1–6/21–49 years	0.05	0.00	0.02	0.18	0.75	0.68	0.32			
7–16/50–64 years	0.05	0.00	0.04	0.34	0.57	0.69	0.31			
1–6/50–64 years	0.07	0.00	0.03	0.20	0.69	0.73	0.27			
7–17/65–79 years	0.03	0.02	0.07	0.30	0.57	0.73	0.27			
1–6/65–79 years	0.04	0.01	0.03	0.21	0.71	0.79	0.21			
7–16/80–104 years	0.04	0.03	0.15	0.27	0.50	0.78	0.22			
1–6/80–104 years	0.14	0.00	0.05	0.30	0.51	0.92	0.08			

which frequently report higher recurrence rates for late-stage CRC [1], the recurrence rate of 25% is relatively modest. This discrepancy may suggest that the efficacy of the treatment, the practices of follow-up, or the characteristics of the sample have varied.

The Bayesian network model's graphical representation and accuracy are compelling. To ensure that the predictions are consistent with real-world outcomes and can be trusted in clinical practice, they should be validated against broader datasets and integrated with clinical expertise. Furthermore, the Bayesian network model demonstrates potential; consequently, it is essential to execute a critical evaluation of its outcomes in comparison to the current body of literature. Traditional models frequently underscore the importance of early detection and the role of specific treatment modalities in improving survival rates. To ensure that the models are not only statistically valid but also effectively applicable in a diverse array of healthcare environments, the existing literature frequently suggests the integration of clinical expertise with AI predictions. The implementation of AI models, such as the Bayesian network, should be meticulously evaluated and compared to the broader clinical experiences and outcomes documented in the literature, even though they provide substantial advancements.

Even though the Bayesian network model offers valuable probabilistic insights, it is imperative that its predictions are tested through comprehensive longitudinal studies and interpreted in conjunction with clinical expertise to ensure that they are consistent with real-world patient outcomes and effectively contribute to clinical decision-making.

Conclusion

The Bayesian network model employed in this study illustrates the probabilistic relationships between the duration of time lived after a CRC diagnosis and six critical predictors. The model visually represents the relationships between various variables, such as age, smoking status, marital status, cancer stage, and recurrence, to influence patient survival. This is achieved using graphics-based nodes and edges. The model provides a clear comprehension of the interactions between these variables. The model's structure emphasizes the interconnectedness of these predictors, as parent nodes have direct interconnections to child nodes and the status of the child node is conditionally dependent on its parents' node. By facilitating analysis of how changes in one variable can probabilistically alter the outcomes in another, this method facilitates a more comprehensive and dynamic comprehension of survival probabilities.

Notes

Authors' ORCIDs

- Muayyad M. Ahmad: 0000-0002-4388-8332
- Eslam Bani Mohammad: 0000-0003-3569-7875

Competing interests

The authors declare that they have no competing interests.

References

1. Tsai KY, Huang PS, Chu PY, Nguyen TNA, Hung HY, Hsieh CH, Wu MH. Current Applications and Future Directions of Circulating Tumor Cells in Colorectal Cancer Recurrence. *Cancers (Basel)*. 2024 Jun 24;16(13):2316. DOI: 10.3390/cancers16132316
2. Shahabi F, Mehri A, Abdollahi A, Hoshyar SHH, Ghahramani A, Noei MG, Orafaie A, Ansari M. Post recurrence survival in early versus late period and its prognostic factors in rectal cancer patients. *Sci Rep*. 2024 Jul;14(1):17661. DOI: 10.1038/s41598-024-67852-7
3. Rasola C, Laurent-Puig P, André T, Falcoz A, Lepage C, Aparicio T, Bouché O, Lievre A, Mineur L, Bennouna J, Louvet C, Bachet JB, Borg C, Vernerey D, Lonardi S, Taieb J. Time to recurrence and its relation to survival after recurrence in patients resected for stage III colon cancer. *Eur J Cancer*. 2023 Nov;194:113321. DOI: 10.1016/j.ejca.2023.113321
4. Ghaffar Nia N, Kaplanoglu E, Nasab A. Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discov Artif Intell*. 2023;3(1):5. DOI: 10.1007/s44163-023-00049-5
5. Ahmad M, Bani Hani SH, Abu-Abeeleh M, Aljarah I. Mortality in young Jordanians with cardiovascular disease: Intelligent algorithms predicting model. *Electron J Gen Med*. 2024;24(2).
6. Devi KJ, Alghamdi W, Divya N, Alkhayyat A, Sayyora A, Sathish T. Artificial intelligence in healthcare: diagnosis, treatment, and prediction. *E3S Web of Conferences*. 2023;399:04043. DOI: 10.1051/e3sconf/202339904043
7. Ahmad M, Bani Hani SH, Sabra MA, Almahmoud O. Big data can help prepare nurses and improve patient outcomes by improving quality, safety, and outcomes. *Frontiers of Nursing*. 2023;10(2):241-8. DOI: 10.2478/fo-n-2023-0026
8. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput*. 2023;14(7):8459-86. DOI: 10.1007/s12652-021-03612-z
9. Mohammad EB, Ahmad M. A systematic evaluation of big data-driven colorectal cancer studies. *Med Glas (Zenica)*. 2024 Feb 1;21(1):63-77. DOI: 10.17392/1684-23
10. Schiele S, Arndt TT, Martin B, Miller S, Bauer S, Banner BM, Brendel EM, Schenkirsch G, Anthuber M, Huss R, Märkl B, Müller G. Deep Learning Prediction of Metastasis in Locally Advanced Colon Cancer Using Binary Histologic Tumor Images. *Cancers (Basel)*. 2021 Apr;13(9):2074. DOI: 10.3390/cancers13092074
11. Gonzalez R, Nejat P, Saha A, Campbell CJV, Norgan AP, Lokker C. Performance of externally validated machine learning models based on histopathology images for the diagnosis, classification, prognosis, or treatment outcome prediction in female breast cancer: A systematic review. *J Pathol Inform*. 2024 Dec;15:100348. DOI: 10.1016/j.jpi.2023.100348
12. Tamang LD, Kim BW. Deep learning approaches to colorectal cancer diagnosis: a review. *Applied Sciences*. 2021;11(22):10982. DOI: 10.3390/app112210982
13. van Erning FN, Greidanus NEM, Verhoeven RHA, Buijsen J, de Wilt HW, Wagner D, Creemers GJ. Gender differences in tumor characteristics, treatment and survival of colorectal cancer: A population-based study. *Cancer Epidemiol*. 2023 Oct;86:102441. DOI: 10.1016/j.canep.2023.102441
14. Baraibar I, Ros J, Saoudi N, Salvà F, García A, Castells MR, Tabernero J, Élez E. Sex and gender perspectives in colorectal cancer. *ESMO Open*. 2023 Apr;8(2):101204. DOI: 10.1016/j.esmoop.2023.101204
15. Dominguez-Valentin M, Haupt S, Seppälä TT, Sampson JR, Sunde L, Bernstein I, Jenkins MA, Engel C, Aretz S, Nielsen M, Capella G, Balaguer F, Evans DG, Burn J, Holinski-Feder E, Bertario L, Bonanni B, Lindblom A, Levi Z, Macrae F, Winship I, Plazzer JP, Sijmons R, Laghi L, Della Valle A, Heinimann K, Dębniak T, Fruscio R, Lopez-Koestner F, Alvarez-Valenzuela K, Katz LH, Laish I, Vainer E, Vaccaro C, Carraro DM, Monahan K, Half E, Stakelum A, Winter D, Kennelly R, Gluck N, Sheth H, Abu-Freha N, Greenblatt M, Rossi BM, Bohorquez M, Cavestro GM, Lino-Silva LS, Horisberger K, Tibiletti MG, Nascimento ID, Thomas H, Rossi NT, Apolinário da Silva L, Zaránd A, Ruiz-Bañobre J, Heuveline V, Mecklin JP, Pylvänäinen K, Renkonen-Sinisalo L, Lepistö A, Peltomäki P, Therkildsen C, Madsen MG, Burgdorf SK, Hopper JL, Win AK, Haile RW, Lindor N, Gallinger S, Le Marchand L, Newcomb PA, Figueiredo J, Buchanan DD, Thibodeau SN, von Knebel Doeberitz M, Loeffler M, Rahner N, Schröck E, Steinke-Lange V, Schmiegeler W, Vangala D, Perne C, Hüneburg R, Redler R, Büttner R, Weitz J, Pineda M, Duenas N, Vidal JB, Moreira L, Sánchez A, Hovig E, Nakken S, Green K, Lalloo F, Hill J, Crosbie E, Mints M, Goldberg Y, Tjandra D, Ten Broeke SW, Kariv R, Rosner G, Advani SH, Thomas L, Shah P, Shah M, Neffa F, Esperon P, Pavicic W, Torrezan GT, Bassaneze T, Martin CA, Moslein G, Moller P. Mortality by age, gene and gender in carriers of pathogenic mismatch repair gene variants receiving surveillance for early cancer diagnosis and treatment: a report from the prospective Lynch syndrome database. *EClinicalMedicine*. 2023 Mar 20;58:101909. DOI: 10.1016/j.eclinm.2023.101909
16. Ulanja MB, Ntafam C, Beutler BD, Antwi-Amoabeng D, Rahman GA, Ulanja RN, Mabrouk T, Governor SB, Djankpa FT, Alese OB. Race, age, and sex differences on the influence of obesity on colorectal cancer sidedness and mortality: A national cross-sectional study. *J Surg Oncol*. 2023 Jan;127(1):109-18. DOI: 10.1002/jso.27096
17. Schellerer VS, Merkel S, Schumann SC, Schlabrakowski A, Förtsch T, Schildberg C, Hohenberger W, Croner RS. Despite aggressive histopathology survival is not impaired in young patients with colorectal cancer: CRC in patients under 50 years of age. *Int J Colorectal Dis*. 2012 Jan;27(1):71-9. DOI: 10.1007/s00384-011-1291-8
18. Zhao L, Bao F, Yan J, Liu H, Li T, Chen H, Li G. Poor prognosis of young patients with colorectal cancer: a retrospective study. *Int J Colorectal Dis*. 2017 Aug;32(8):1147-56. DOI: 10.1007/s00384-017-2809-5
19. Ahmad M, Bani Hani SH, Abu-Abeeleh M, Aljarah I. Prediction of mortality in young adults with cardiovascular disease using artificial intelligence. *Electron J Gen Med*. 2024;21(3):em584. DOI: 10.29333/ejgm/14575
20. Papamichael D, Audisio R, Horiot JC, Glimelius B, Sastre J, Mitry E, Van Cutsem E, Gosney M, Köhne CH, Aapro M; SIOG. Treatment of the elderly colorectal cancer patient: SIOG expert recommendations. *Ann Oncol*. 2009 Jan;20(1):5-16. DOI: 10.1093/annonc/mdn532
21. Li Q, Gan L, Liang L, Li X, Cai S. The influence of marital status on stage at diagnosis and survival of patients with colorectal cancer. *Oncotarget*. 2015 Mar;6(9):7339-47. DOI: 10.18632/oncotarget.3129
22. Wang X, Cao W, Zheng C, Hu W, Liu C. Marital status and survival in patients with rectal cancer: An analysis of the Surveillance, Epidemiology and End Results (SEER) database. *Cancer Epidemiol*. 2018 Jun;54:119-24. DOI: 10.1016/j.canep.2018.04.007
23. Yang CC, Cheng LC, Lin YW, Wang SC, Ke TM, Huang CI, Su YC, Tai MH. The impact of marital status on survival in patients with surgically treated colon cancer. *Medicine (Baltimore)*. 2019 Mar;98(11):e14856. DOI: 10.1097/MD.00000000000014856

24. Japuntich SJ, Kumar P, Pendergast JF, Juarez Caballero GY, Malin JL, Wallace RB, Chrischilles EA, Keating NL, Park ER. Smoking Status and Survival Among a National Cohort of Lung and Colorectal Cancer Patients. *Nicotine Tob Res.* 2019 Mar;21(4):497-504. DOI: 10.1093/ntr/nty012
25. Budianto A, Andarini S, Hariyanti T, Muslihah N. The prediction of colorectal cancer: Perspective of smoking and socioeconomic influence of culture in East Java with SEM analysis. *Nurture.* 2024;18(2):315-27. DOI: 10.55951/nurture.v18i2.614
26. Keivanlou MH, Amini-Salehi E, Hassanipour S, Mahapatro A, Raghuma N, Joukar F, Letafatkar N, Habibi A, Norouzi N, Aleali MS, Javid M, Mirdamadi A, Mansour-Ghanaei F. Association between smoking and colorectal cancer in Eastern Mediterranean Regional Office (EMRO): A systematic review and meta-analysis. *Saudi J Gastroenterol.* 2023;29(4):204-11. DOI: 10.4103/sjg.sjg_163_23
27. Shamoun S, Ahmad M. Literature review on breast cancer-related lymphedema and related factors. *Arch Oncol.* 2023;29(2):22-9. DOI: 10.2298/A00230313003S
28. Lu X, Wang G, Feng L, Kan Y, Wang W, Yang J, Zhang M. Preoperative metabolic parameters of F-FDG PET/CT are associated with TNM stage and prognosis of colorectal cancer patients. *Quant Imaging Med Surg.* 2024 Jan;14(1):462-75. DOI: 10.21037/qims-23-966
29. Zhang H, Wang R, Yu T, Yu D, Song C, Ma B, Li J. A prognostic nomogram integrating carcinoembryonic antigen levels for predicting overall survival in elderly patients with stage II-III colorectal cancer. *J Gastrointest Oncol.* 2024 Feb;15(1):164-78. DOI: 10.21037/jgo-23-863
30. IBM. IBM SPSS Statistics for Windows, Version 29.1. Armonk, NY: IBM Corp. 2023.
31. IBM. IBM SPSS Modeler 18.0 Batch User's Guide. 2016. Available from: <https://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerBatch.pdf>
32. Constantinou AC. Learning Bayesian networks that enable full propagation of evidence. *IEEE Access.* 2020;8:124845-56. DOI: 10.1109/ACCESS.2020.3006472
33. Ghebrial M, Aktary ML, Wang Q, Spinelli JJ, Shack L, Robson PJ, Kopciuk KA. Predictors of CRC Stage at Diagnosis among Male and Female Adults Participating in a Prospective Cohort Study: Findings from Alberta's Tomorrow Project. *Curr Oncol.* 2021 Nov;28(6):4938-52. DOI: 10.3390/curroncol28060414
34. Rogers CR, Moore JX, Qeadan F, Gu LY, Huntington MS, Holowatyj AN. Examining factors underlying geographic disparities in early-onset colorectal cancer survival among men in the United States. *Am J Cancer Res.* 2020 May 1;10(5):1592-1607.
35. Ciombor KK, Wu C, Goldberg RM. Recent therapeutic advances in the treatment of colorectal cancer. *Annu Rev Med.* 2015;66:83-95. DOI: 10.1146/annurev-med-051513-102539
36. Maida M, Dahiya DS, Shah YR, Tiwari A, Gopakumar H, Vohra I, Khan A, Jaber F, Ramai D, Facciorusso A. Screening and Surveillance of Colorectal Cancer: A Review of the Literature. *Cancers (Basel).* 2024 Aug;16(15):2746. DOI: 10.3390/cancers16152746
37. Bani Hani S, Ahmad M. Effective Prediction of Mortality by Heart Disease Among Women in Jordan Using the Chi-Squared Automatic Interaction Detection Model: Retrospective Validation Study. *JMIR Cardio.* 2023 Jul;7:e48795. DOI: 10.2196/48795
38. Gheybi K, Buckley E, Vitry A, Roder D. Associations of advanced age with comorbidity, stage and primary subsite as contributors to mortality from colorectal cancer. *Front Public Health.* 2023;11:1101771. DOI: 10.3389/fpubh.2023.1101771
39. Abedizadeh R, Majidi F, Khorasani HR, Abedi H, Sabour D. Colorectal cancer: a comprehensive review of carcinogenesis, diagnosis, and novel strategies for classified treatments. *Cancer Metastasis Rev.* 2024 Jun;43(2):729-53. DOI: 10.1007/s10555-023-10158-3

Corresponding author:

Prof. Muayyad M. Ahmad, PhD, RN
Clinical Nursing Department, School of Nursing, The University of Jordan, Queen Rania St, Amman – 11942, Jordan, Phone: +962 799313745, Fax: +962 65300244
mma4@ju.edu.jo

Please cite as

Ahmad MM, Bani Mohammad I. A retrospective analysis of key predictors and patient outcomes: Using artificial intelligence for precision survival prediction in colorectal cancer. *GMS Ger Med Sci.* 2025;23:Doc09.
DOI: 10.3205/000345, URN: urn:nbn:de:0183-0003453

This article is freely available from

<https://doi.org/10.3205/000345>

Received: 2024-09-11

Revised: 2025-05-21

Published: 2025-08-29

Copyright

©2025 Ahmad et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.