Systematic Review / Meta-analysis

ANNALS OF MEDICINE & SURGERY

OPEN

Artificial intelligence models for survival prediction in colorectal cancer: a systematic review of time-to-event approaches

Abdulrahman Ali H. Alshamrani, MD[a,*], Khalid Abdullah Alghamdi, MD[b], Mohammed Faisal Alshahrani, MD[b], Tariq Mohammed S. Bin Ladnah, MD[b], Abdullah Zafer Alshihri, MD[b], Khalid Ali Bakri, MD[c], Mohammed Mesfer Alqarni, MD[b], Tasnim Mansour H. Abu Shaqah, MD[c], Zafer Ali Alshahrani, MD[b], Afnan Meshal H. Alotaibi, MD[b], Razan Sultan Al-Hufayyan, MD[c], Lamar Tariq Alassiry, MD[d], Feras Ibrahim A. Alshehri, MD[e], Mohamed Al-Azab, MD[f,*], Saad Dhafer Alshahrani, MD[g,*]

Background: Colorectal cancer (CRC) is a leading cause of cancer-related mortality worldwide, with marked heterogeneity in patient survival. Conventional clinicopathologic staging provides limited individualized prognostic information. Artificial intelligence (AI) and machine learning methods have enabled survival-aware modeling approaches that integrate complex clinical, imaging, and molecular data to improve time-to-event survival prediction. This systematic review aimed to synthesize current evidence on AI-based models for survival prediction in CRC.

Methods: A systematic search of PubMed, Scopus, Web of Science Core Collection, IEEE Xplore, and Google Scholar was conducted in accordance with the PRISMA 2020 guidelines. Studies were included if they applied AI methods to model time-to-event survival outcomes in CRC. Records were independently screened, data were extracted, and risk of bias and applicability were assessed using the PROBAST + AI tool.

Results: Eight retrospective cohort studies published between 2021 and 2025, encompassing 11 811 patients, met the inclusion criteria. The included studies evaluated diverse data modalities, including clinical variables, radiomics, histopathology, transcriptomics, and genomics. Model performance was moderate to high, with concordance index values ranging from approximately 0.70 to 0.85. All studies demonstrated effective risk stratification of patients into distinct survival groups. Models incorporating high-dimensional imaging or molecular data generally outperformed those based on clinical variables alone. The overall risk of bias was low to unclear, with no study rated as high risk.

Conclusion: AI-based time-to-event survival models demonstrate potential prognostic value for prognostic stratification in CRC. Further prospective studies, standardized reporting, and external validation are required to support clinical translation.

Keywords: artificial intelligence, colorectal cancer, machine learning, survival analysis, systematic review

Background

Colorectal cancer (CRC) is one of the most common malignancies worldwide and remains a leading cause of cancer-related mortality[1]. Globally, CRC accounts for approximately 1.9 million new cases each year and ranks among the top causes of cancer death[2]. Despite advances in screening, surgical techniques, systemic therapies, and targeted treatments, survival outcomes remain heterogeneous, particularly in patients with advanced disease[3]. Up to half of patients develop metastatic CRC during the disease course, with reported 5-year survival rates ranging from 15 to 40%[4]. Even among patients with non-metastatic disease who undergo curative resection, recurrence, and cancer-related mortality remain substantial[4].

Prognostic assessment in CRC has traditionally relied on clinicopathologic factors such as tumor size, histologic grade, and the American Joint Committee on Cancer tumor, node, and metastasis staging system[5]. Although widely used, these

[a]Department of General Surgery, King Abdulaziz Specialist Hospital, Taif, Saudi Arabia, [b]Department of General Surgery, King Abdullah Hospital, Bisha, Saudi Arabia, [c]Department of General Surgery, Aseer Central Hospital, Aseer Health Cluster, Abha, Saudi Arabia, [d]College of Medicine, King Khalid University, Abha, Saudi Arabia, [e]College of Medicine, Batterjee Medical College, Asir, Saudi Arabia, [f]Faculty of Medicine, Sana'a University, Sana'a, Yemen and [g]Department of Surgery, College of Medicine, University of Bisha, Bisha, Saudi Arabia

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

*Corresponding authors. Address: Department of General Surgery, King Abdulaziz Specialist Hospital, 26521 Qurwa, Qurwa, Taif 26521, Kingdom of Saudi Arabia. E-mail: Dr.a.alshamrani0@gmail.com (A.A.H. Alshamrani); Department of Surgery, College of Medicine, University of Bisha, 9054 King Saud Road, Al Matar District, Bisha 67614, Kingdom of Saudi Arabia. E-mail: Shohob.research@gmail.com (S.D. Alshahrani); Faculty of Medicine, Sana'a University, Sana'a, Madhbah Area, Next to the College of Medicine, Sana'a, Yemen. E-mail: dr.mohammed.alazab2@gmail.com (M. Al-Azab).

Annals of Medicine & Surgery (2026) 88:3556–3563

Received 12 January 2026; Accepted 15 April 2026

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.lww.com/annals-of-medicine-and-surgery.

Published online 12 May 2026

http://dx.doi.org/10.1097/MS9.0000000000005157

approaches provide population-level risk estimates and often fail to capture interpatient variability in tumor biology, spatial heterogeneity, and treatment response[6]. Patients within the same stage frequently experience markedly different survival outcomes, underscoring the limitations of conventional staging systems and the need for more individualized prognostic tools[7].

Advances in high-throughput technologies, including next-generation sequencing, transcriptomics, and medical imaging, have enabled comprehensive characterization of tumor genomic, molecular, and phenotypic features[8]. Gene expression signatures, radiomic features, and histopathologic patterns have been shown to reflect underlying tumor aggressiveness and the tumor microenvironment[9]. These data-rich modalities provide an opportunity to move beyond single-factor predictors toward integrative models that better represent disease complexity and biological heterogeneity[9].

Artificial intelligence (AI) and machine learning methods are increasingly applied in oncology to analyze high-dimensional data and model complex, nonlinear relationships between predictors and clinical outcomes[10]. More broadly, recent advances in AI have substantially transformed biomedical research and medical diagnostics. For example, deep learning models such as AlphaFold have enabled highly accurate prediction of protein three-dimensional structures, accelerating discoveries in structural biology and drug development[11]. In parallel, large language models have demonstrated growing capabilities in biomedical knowledge synthesis, clinical decision support, and medical text analysis[12].

These advances illustrate the expanding role of AI across biomedical disciplines and highlight its potential to enhance predictive modeling and precision medicine. In the context of survival analysis, machine learning approaches offer advantages over traditional Cox proportional hazards models by accommodating nonlinearity, interactions, and time-dependent effects without strict parametric assumptions[13]. Survival-specific machine learning models have demonstrated encouraging performance in predicting overall survival, cancer-specific survival, and recurrence-free survival across multiple cancer types, including CRC[14]. However, the rapid growth of this field has resulted in substantial methodological heterogeneity, variable reporting quality, and uncertainty regarding model validity, interpretability, and clinical applicability[14].

Given these challenges, a systematic synthesis of current evidence is needed to evaluate how AI-based survival models have been developed, validated, and applied in CRC. This review aims to summarize study characteristics, modeling approaches, performance, and risk of bias to inform future research and guide the responsible integration of AI into prognostic assessment for CRC.

HIGHLIGHTS

- This systematic review summarizes artificial intelligence–based time-to-event survival models in colorectal cancer published between 2021 and 2025.
- Eight retrospective studies involving 11 811 patients demonstrated moderate-to-high prognostic performance using survival-aware machine learning methods.
- Models integrating radiomics, histopathology, transcriptomics, or genomics generally outperformed those based solely on clinical variables.
- Explainable AI approaches enhance model transparency and the identification of key prognostic features.
- Despite promising results, limited external validation and heterogeneous methodologies highlight the need for prospective studies and standardized reporting.

Methods

Study design

This study was conducted as a systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines[15]. This review was not registered in the PROSPERO database. However, a predefined review protocol specifying the research question, eligibility criteria, search strategy, screening process, and data extraction approach was developed by the authors prior to study initiation. Although prospective registration can enhance transparency and reduce the risk of reporting bias, the review was conducted in accordance with established methodological standards and PRISMA reporting recommendations.

The review question and study design were structured according to the Population, Intervention, Comparator, and Outcome (PICO) framework. The population comprised human patients diagnosed with CRC. The intervention of interest was the development or evaluation of AI or machine learning–based models for survival prediction. Comparators included traditional statistical survival models, such as Cox proportional hazards regression, or alternative AI approaches when available. The outcomes of interest were time-to-event survival endpoints, including overall survival, cancer-specific survival, disease-specific survival, and recurrence-free survival. This framework was used to guide the search strategy, eligibility criteria, and data synthesis.

Search strategy

A comprehensive and systematic literature search was conducted to identify relevant studies from database inception to 30 December 2025, which served as the final search date. Five electronic databases were searched, including PubMed (MEDLINE), Scopus, Web of Science Core Collection, IEEE Xplore, and Google Scholar, to identify eligible studies. The search strategy was designed to capture studies investigating the use of AI or machine learning techniques for time-to-event survival prediction in CRC.

Controlled vocabulary terms and free-text keywords related to CRC, AI or machine learning, and survival or time-to-event outcomes were combined using Boolean operators. The search strategy was adapted for each database to optimize sensitivity and specificity. No restrictions on publication year or language were applied during the search phase. In addition to the electronic database search, the reference lists of included studies were manually screened to identify potentially relevant articles that may not have been captured in the initial search. The complete search strings for all databases are provided in Supplemental Digital Content File 1, available at: http://links.

lww.com/MS9/B224 in accordance with PRISMA-S reporting recommendations.

Eligibility criteria

Studies were included if they met the following criteria: (1) involved human patients with colorectal, colon, or rectal cancer; (2) developed or evaluated AI or machine learning–based prediction models; (3) reported time-to-event survival outcomes, such as overall survival, disease-specific survival, cancer-specific survival, or recurrence-free survival; and (4) applied survival analysis methods that appropriately accounted for censoring. Studies were excluded if they: (1) modeled survival as a binary classification outcome without time-to-event analysis; (2) relied solely on traditional statistical methods without machine learning or AI components; (3) were reviews, editorials, conference abstracts, or non-original research; or (4) did not report sufficient methodological details to assess model development or performance.

Study selection and screening

All identified records were imported into Rayyan, a web-based platform for systematic review screening[16]. Duplicate records were automatically identified and removed. Title and abstract screening was conducted independently by eight reviewers, working in pairs, according to the predefined eligibility criteria. Inter-rater agreement between reviewer pairs was assessed using Cohen's kappa statistic and demonstrated substantial agreement ($\kappa > 0.75$)[17]. Discrepancies were resolved through discussion and consensus. Full-text screening was subsequently performed independently by the same reviewers to confirm eligibility. Reference management and full-text organization were conducted using EndNote.

Quality and risk of bias assessment

Risk of bias and applicability of the included studies were assessed using the Prediction Model Risk of Bias Assessment Tool with AI extensions (PROBAST + AI)[18]. PROBAST + AI is an updated version of the original PROBAST tool, developed to address advances in prediction modeling methodology and the increasing use of AI and machine learning techniques[18]. The tool evaluates both model development and model evaluation through structured signaling questions across four domains: participants and data sources, predictors, outcome, and analysis. Applicability was assessed for the participants and data sources, predictors, and outcome domains in relation to the review question. Each domain was judged as having low, high, or unclear risk of bias, and the overall risk of bias for each study was determined by the highest risk rating across domains. Two reviewers independently performed the assessment, with disagreements resolved through discussion and consensus.

Data extraction and synthesis

Data were extracted independently by four reviewers working in pairs, using a standardized data extraction form. Extracted variables included study characteristics (author, year, country, and study design), patient population, sample size, data modality, AI or machine learning models, survival outcomes, validation strategies, and model performance metrics. Given the substantial heterogeneity across studies with respect to model architectures, predictor types, survival outcomes, performance measures, and validation approaches, quantitative meta-analysis was not feasible. Findings were therefore synthesized narratively and summarized in descriptive tables to facilitate structured comparison of study characteristics, methodological approaches, and key results.

Results

Study selection

The database search identified a total of 1990 records, including 680 from PubMed, 801 from Scopus, 441 from the Web of Science Core Collection, 18 from IEEE Xplore, and 50 from Google Scholar. After the removal of 780 duplicate records, 1210 records were screened based on titles and abstracts. Of these, 1196 records were excluded because they did not address CRC, did not use AI or machine learning methods, did not report survival or time-to-event outcomes, modeled survival as a binary classification outcome, or represented non-original publications. Fourteen full-text articles were retrieved and assessed for eligibility. Eight studies met the inclusion criteria and were included in the final qualitative synthesis. The study selection process is summarized in the PRISMA flow diagram (Fig. 1).

Study characteristics

The main characteristics of the included studies are summarized in Table 1. The eight studies were published between 2021 and 2025 and comprised a total of 11 811 patients with CRC. Most studies were conducted in China ($n = 5$), followed by the United States ($n = 2$) and Austria ($n = 1$). All included studies employed retrospective cohort designs, with one study using a multicenter cohort and one employing a computational cohort based on publicly available datasets.

The included patient populations covered a broad spectrum of disease stages, including stage II to III colorectal cancer, stage IV metastatic disease, T1 colorectal cancer, rectal cancer, and mixed-stage cohorts. Sample sizes ranged from 206 to 5629 patients. Data modalities varied substantially across studies and included structured clinical variables, radiomics derived from magnetic resonance imaging and computed tomography, whole-slide histopathology images, transcriptomic data, single-cell RNA sequencing data, microRNA targeting information, and genomic alterations derived from next-generation sequencing.

A wide range of AI and machine learning models were used, including random survival forests, gradient boosting methods, deep learning architectures, XGBoost survival models, and Cox-based nomograms enhanced with machine learning feature selection. Survival outcomes assessed included overall survival, cancer-specific survival, disease-specific survival, and recurrence-free survival. Validation strategies varied across studies and included internal cross-validation, train and validation splits, and external validation using independent cohorts.

Model performance and key findings

Model performance and key findings are summarized in Table 2. All included studies reported discrimination metrics appropriate for time-to-event survival analysis, most commonly the

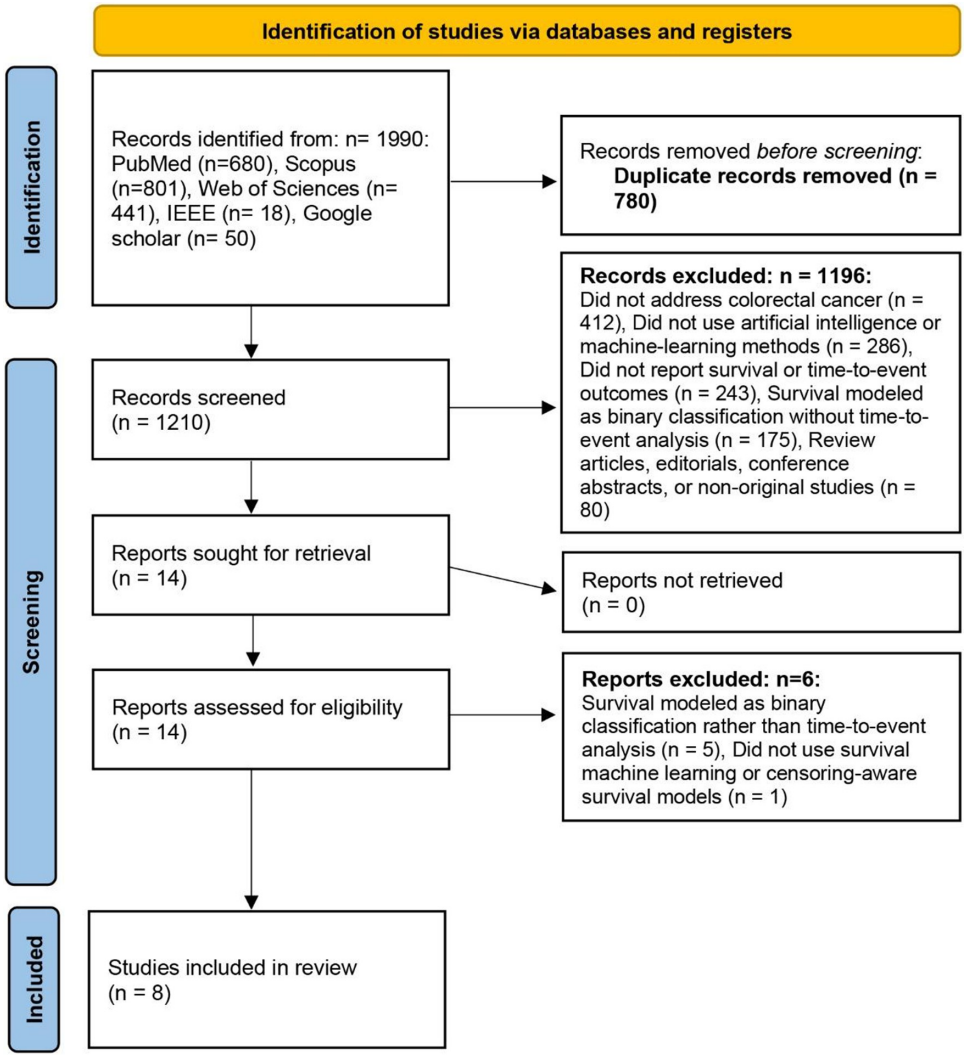

Figure 1. PRISMA 2020 flow diagram illustrating the study selection process.

concordance index and time-dependent area under the curve (AUC). Several studies additionally reported calibration measures, integrated Brier scores, and decision curve analyses.

Reported model performance varied across studies. Concordance index values ranged from approximately 0.70 to 0.85. Studies incorporating high-dimensional data, such as radiomics, transcriptomics, genomics, or histopathology, generally demonstrated higher predictive performance compared with models based solely on clinical variables. Time-dependent AUC values exceeded 0.80 in several studies, particularly those employing radiomics or gene expression signatures.

All studies demonstrated the ability to stratify patients into distinct risk groups with statistically significant differences in survival outcomes using Kaplan–Meier analysis. Several studies employed explainable AI approaches, most commonly SHapley Additive exPlanations, to identify key predictors and enhance interpretability. Across studies, important prognostic features included tumor stage, lymph node involvement, radiologic texture features, histopathologic patterns, and molecular signatures.

Risk of bias and applicability assessment

The risk of bias and applicability assessment using PROBAST + AI is presented in Table 3. Overall, the risk of bias was judged as low in three studies and unclear in five studies. No study was judged to be at high risk of bias. The most frequent source of uncertainty related to the analysis domain, primarily due to limited reporting on hyperparameter tuning, handling of missing data, or reliance on internal validation without external validation. All studies were judged to have low concerns regarding applicability in relation to the review question, as the included populations, predictors, outcomes, and modeling approaches were appropriate for artificial intelligence–based time-to-event survival prediction in colorectal cancer.

Discussion

This systematic review provides a comprehensive synthesis of contemporary evidence on AI-based models for time-to-event survival prediction in CRC. Eight studies published

Table 1

Characteristics of included studies evaluating artificial intelligence models for survival prediction in colorectal cancer.

Study	Country	Study design	Cancer type/stage	Sample size	Data modality	AI/ML model	Survival outcome	Validation strategy
Wuliczyn et al., 2021[19]	Austria	Retrospective multicenter cohort	Stage II–III colorectal cancer	5629	Whole-slide histopathology images	Weakly supervised deep learning	Disease-specific survival	Two independent external cohorts
Chuanji et al., 2022[20]	China	Retrospective cohort	Rectal cancer	206	MRI T2-weighted radiomics + clinical	Radiomics + LASSO-Cox nomogram	3-year overall survival	Internal training/ validation split
Willems et al., 2023[21]	USA	Retrospective computational cohort	Colon and rectal adenocarcinoma	633	scRNA-seq + bulk RNA-seq + miRNA targeting	Gene prioritization + LASSO-Cox	Overall survival	10-fold cross-validation
Yang et al., 2023[22]	China	Retrospective cohort	Colorectal cancer (all stages)	2157	Structured clinical variables	RSF, GBM, DeepSurv, DeepHit, Cox-Time, N-MTLR	CRC-specific survival	Repeated stratified 5-fold CV
Altaf et al., 2025[23]	USA	Retrospective cohort	Stage IV colorectal cancer	688	Genomic alterations (NGS)	XGBoost survival model	Overall survival	5-fold cross-validation
Hong et al., 2025[24]	China	Retrospective cohort	Non-metastatic CRC (curatively resected)	794	CT radiomics + clinicopathologic	AutoML radiomics + Cox-based nomogram	Cancer-specific survival	Internal training/ validation split
Jun Li et al., 2025[25]	China	Retrospective cohort	Colorectal cancer (all stages)	624	Transcriptomics (TCGA + GEO)	Random survival forest	Overall survival	External validation (GEO cohort)
Zhihong Li et al., 2025[26]	China	Retrospective multicenter cohort	T1 colorectal cancer	580	Clinicopathologic and histopathologic	RSF, GB, ST, CoxPH, Coxnet	Recurrence-free survival	Train–test split with internal CV

AI, artificial intelligence; CRC, colorectal cancer; CV, cross-validation; GBM, gradient boosting machine; GEO, Gene Expression Omnibus; ML, machine learning; NGS, next-generation sequencing; RSF, random survival forest; scRNA-seq, single-cell RNA sequencing; ST, survival tree; TCGA, The Cancer Genome Atlas.

between 2021 and 2025 were identified, encompassing diverse patient populations, data modalities, and modeling strategies. Collectively, the findings demonstrate that AI and machine learning approaches can meaningfully improve prognostic stratification in CRC, particularly when high-dimensional data sources are integrated with traditional clinical variables.

A key finding of this review is the consistent improvement in predictive performance observed in models incorporating complex data modalities such as radiomics, transcriptomics, genomics, and histopathology. Across studies, concordance index values generally ranged from 0.70 to 0.85, with several models achieving time-dependent AUC values exceeding 0.80[20,21]. These results suggest that AI models are capable of capturing nonlinear relationships and complex interactions that are not adequately modeled by conventional statistical approaches alone[27]. In particular, studies using deep learning on whole-slide histopathology images, radiomics features extracted from computed tomography or magnetic resonance imaging, and gene expression signatures demonstrated robust survival discrimination and clinically meaningful risk stratification[24,26].

The ability of these models to stratify patients into distinct risk groups with significantly different survival outcomes is of particular clinical relevance. All included studies reported statistically significant separation of survival curves using Kaplan–Meier analysis, indicating potential utility for identifying high-risk patients who may benefit from intensified surveillance or adjuvant therapy. Notably, several studies demonstrated superior prognostic performance compared with established staging systems, including the American Joint Committee on Cancer tumor, node, and metastasis classification[24]. This finding aligns with growing evidence that categorical staging systems provide limited individualized prognostic information and fail to fully capture tumor biological heterogeneity[28,29].

Another important observation is the increasing emphasis on explainability and model transparency. More than half of the included studies employed interpretability techniques such as SHapley Additive exPlanations, feature importance analysis, or biologically informed pathway enrichment[22,23]. These approaches enabled the identification of clinically and biologically meaningful predictors, including tumor stage, lymph node involvement, radiologic texture features, histopathologic patterns, and molecular signatures. The integration of explainability methods is essential for promoting clinician trust, facilitating clinical translation, and addressing concerns related to the "black box" nature of AI models[30,31].

From a methodological perspective, this review highlights both the strengths and limitations of current evidence. Most studies used appropriate survival-specific machine learning methods that accounted for censoring and time-to-event outcomes, addressing a common limitation in earlier AI studies that modeled survival as a binary classification problem[32]. In addition, the use of performance metrics such as the concordance index, time-dependent AUC, integrated Brier score, and decision curve analysis reflects increasing methodological rigor[32]. However, substantial methodological heterogeneity across model types, predictor sets, outcomes, and validation strategies precluded quantitative meta-analysis.

The risk of bias assessment using PROBAST + AI indicated that no included study was judged to be at high risk of bias, which is reassuring. However, several studies were rated as

Table 2

Performance and key findings of artificial intelligence models for survival prediction in colorectal cancer.

Study	Performance metrics	Best model performance	Risk stratification	Explainability	Key findings
Wulczyn et al., 2021[19]	5-year AUC, Cox HR, Kaplan–Meier analysis	5-year AUC 0.69–0.70 in external cohorts; HR ≈ 1.4–1.6	Clear separation of disease-specific survival across risk groups	Interpretable histologic feature clustering	Deep learning on routine histopathology provided independent and externally validated survival prediction in stage II–III CRC
Chuanji et al., 2022[20]	C-index, AUC, calibration, DCA	C-index 0.872 (training), 0.944 (validation)	Significant overall survival separation between high- and low-risk groups	Radiomics feature weights and nomogram	MRI-based radiomics combined with clinical factors improved postoperative survival prediction in rectal cancer
Willems et al., 2023[21]	C-index, Kaplan–Meier analysis	Mean C-index ≈ 0.75 (COAD), ≈ 0.84 (READ)	Significant overall survival stratification	Gene weights and pathway enrichment	AI-assisted integration of single-cell transcriptomics and miRNA targeting enhanced CRC survival prediction
Yang et al., 2023[22]	Time-dependent C-index, IBS, calibration, DCA	DeepHit C-index 0.789; RSF IBS 0.096	Clinically meaningful CRC-specific survival stratification	SHAP	Time-to-event ML models outperformed traditional Cox regression and provided transparent CRC-specific survival prediction
Altaf et al., 2025[23]	C-index, time-dependent AUC, KM analysis	C-index 0.701; AUC 0.742–0.793 (12–36 months)	Significant overall survival separation using a 20-gene signature	SHAP	Genomic ML models accurately stratified survival risk in metastatic colorectal cancer
Hong et al., 2025[24]	C-index, time-dependent AUC, calibration, DCA	C-index 0.803 (training), 0.772 (validation); 5-year AUC up to 0.86	Improved cancer-specific survival stratification vs AJCC staging	Radiomics signature and nomogram	CT-based AutoML radiomics nomogram outperformed clinicopathologic models in non-metastatic CRC
Jun Li et al., 2025[25]	Time-dependent AUC, C-index, Cox HR	AUC 0.910–0.935 at 1–5 years	Robust overall survival stratification in training and external cohorts	Feature importance and biological validation	Random survival forest-based gene signature enabled accurate and externally validated CRC survival prediction
Zhihong Li et al., 2025[26]	C-index, IBS, time-dependent AUC	RSF C-index 0.848; AUC 0.918	Clear recurrence-free survival separation	SHAP	Survival ML models effectively predicted postoperative recurrence in T1 colorectal cancer

AJCC, American Joint Committee on Cancer; AUC, area under the curve; COAD, colon adenocarcinoma; CRC, colorectal cancer; DCA, decision curve analysis; HR, hazard ratio; IBS, integrated Brier score; READ, rectal adenocarcinoma; RSF, random survival forest; SHAP, SHapley Additive exPlanations.

having an unclear risk of bias in the analysis domain. This uncertainty was primarily related to incomplete reporting of key methodological details, including hyperparameter tuning procedures, handling of missing data, and model validation strategies. Limited reporting of these aspects may hinder reproducibility and obscure potential methodological weaknesses. In addition, many studies relied primarily on internal validation approaches, such as cross-validation or training–validation splits, without independent external validation. While internal validation is useful during model development, it may produce optimistic estimates of predictive performance if models capture dataset-specific patterns rather than generalizable prognostic signals. Only a minority of studies performed external validation using independent cohorts, which remains essential for assessing the generalizability and real-world applicability of AI-based prediction models[19,25]. Future studies should also adhere to emerging reporting guidelines, such as TRIPOD-AI, to improve transparency, reproducibility, and methodological rigor in AI-based prediction model research[33].

Several limitations of the current evidence base should be considered. All included studies employed retrospective designs, which are susceptible to selection bias and unmeasured confounding[34]. In addition, most studies relied on single-center or region-specific datasets, with a notable geographic concentration in China and the United States. Differences in population demographics, genetic backgrounds, healthcare systems, screening practices, and treatment protocols across regions may influence model performance and limit the generalizability of findings to other populations. Broader geographic representation and multinational validation efforts will therefore be important for ensuring the robustness and global applicability of AI-based survival prediction models in CRC. Further limitations include variability in data quality, preprocessing pipelines, feature extraction methods, and feature selection strategies across studies, all of which may affect model performance and reproducibility. The absence of standardized benchmarks, shared datasets, and consistent reporting frameworks further complicates cross-study comparison and external validation.

Despite these limitations, the findings of this review highlight the potential of AI to advance prognostic assessment in CRC. By integrating multidimensional data sources and leveraging survival-specific modeling techniques, AI models offer opportunities

Table 3

Risk of bias and applicability assessment of included studies using PROBAST + AI.

Study	Participants & data sources	Predictors	Outcome	Analysis	Overall risk of bias	Applicability concern
Wulczyn et al., 2021[19]	Low	Low	Low	Low	Low	Low
Chuanji et al., 2022[20]	Low	Low	Low	Unclear	Unclear	Low
Willems et al., 2023[21]	Low	Unclear	Low	Unclear	Unclear	Low
Yang et al., 2023[22]	Low	Low	Low	Low	Low	Low
Altaf et al., 2025[23]	Low	Low	Low	Unclear	Unclear	Low
Hong et al., 2025[24]	Low	Low	Low	Unclear	Unclear	Low
Jun Li et al., 2025[25]	Low	Low	Low	Low	Low	Low
Zhihong Li et al., 2025[26]	Low	Low	Low	Unclear	Unclear	Low

Risk of bias and applicability were assessed using the Prediction model Risk Of Bias Assessment Tool with artificial intelligence extensions (PROBAST + AI). Each domain was rated as low, high, or unclear risk of bias according to signaling questions. Overall risk of bias was determined by the highest risk rating across domains. Applicability concerns were judged in relation to the review question of artificial intelligence–based time-to-event survival prediction in colorectal cancer.

for more refined and individualized risk prediction than is currently achievable using traditional clinicopathologic factors alone. With improved methodological transparency, external validation across diverse populations, and prospective evaluation in clinical settings, these models may contribute to more personalized treatment planning, optimized surveillance strategies, and improved patient counseling.

Future research should focus on prospective validation, multicenter collaboration, and integration of AI models into clinical workflows[34]. Emphasis should also be placed on evaluating clinical utility, cost-effectiveness, and impact on patient outcomes. Importantly, transparent reporting, explainability, and ethical considerations must remain central as AI continues to evolve in oncology[35].

Conclusion

This systematic review synthesizes current evidence on the use of AI and machine learning models for time-to-event survival prediction in CRC. Overall, the included studies demonstrate that AI-based approaches can achieve moderate to high predictive performance and effectively stratify patients into clinically meaningful risk groups. Models integrating high-dimensional data sources such as radiomics, transcriptomics, genomics, and histopathology generally outperformed approaches relying solely on conventional clinical variables, highlighting the potential value of multidimensional data in prognostic modeling. Across studies, the increasing adoption of survival-specific machine learning methods that appropriately account for censoring and time dependency reflects methodological progress in this field. The use of explainable AI techniques further enhanced model interpretability by identifying clinically relevant prognostic factors and supporting transparent decision-making. However, several limitations remain. Most studies were retrospective and relied primarily on internal validation, which may lead to optimistic estimates of model performance. External validation was performed in only a minority of studies, and the geographic concentration of datasets in China and the United States may limit generalizability to more diverse populations. Future research should therefore prioritize prospective study designs, rigorous external validation, standardized reporting, and evaluation across diverse populations to support reliable clinical translation.

Ethical approval

Ethical approval was not required for this study because it is a systematic review and meta-analysis based exclusively on previously published data. No new data were collected, and no direct involvement of human participants occurred.

Consent

Informed consent was not required for this study as it involved the analysis of data extracted from published studies in which informed consent had already been obtained by the original investigators.

Sources of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

A.A.H.A. conceived and designed the study, supervised the systematic review process, coordinated the research team, and served as the guarantor of the study. A.A.H.A., K.A.A., M.F.A., and T.M.S.B.L. conducted title and abstract screening. A.Z.A., K.A.B., M.M.A., and T.M.H.A.S. performed full-text screening and eligibility assessment. Z.A.A., A.M.H.A., R.S.A.H., and L.T.A. carried out data extraction and data verification. F.I.A.A., M.A.-A., and S.D.A. conducted quality and risk of bias assessment using the PROBAST+AI tool. A.A.H.A. and S.D.A. contributed to data synthesis and interpretation. A.A.H.A., K.A.A., and M.F.A. drafted the initial manuscript. A.A.H.A., M.A.-A., and S.D.A. critically revised the manuscript for important intellectual content and finalized the manuscript. All authors reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Conflicts of interest disclosure

The authors declare no conflicts of interest related to this study.

Research registration unique identifying number (UIN)

Not applicable.

Guarantor

Saad Dhafer Alshahrani.

Provenance and peer review

Not commissioned, externally peer-reviewed.

Data availability statement

All data generated or analyzed during this study are included in this published article and its supplementary materials. Additional data are available from the corresponding author upon reasonable request.

Acknowledgements

We would like to express our gratitude to Shohob Research Services Center for their efforts in training the research team and supervising the research process until the completion of this study.

References

[1] Trahearn N, Sakr C, Banerjee A, *et al*. Computational pathology applied to clinical colorectal cancer cohorts identifies immune and endothelial cell spatial patterns predictive of outcome. J Pathol 2025;265:198–210.
[2] Colorectal cancer. Accessed 10 January 2026. https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer .
[3] Fadlallah H, Masri JE, Fakhereddine H, *et al*. Colorectal cancer: recent advances in management and treatment. World J Clin Oncol 2024;15:1136.
[4] Zeineddine FA, Zeineddine MA, Yousef A, *et al*. Survival improvement for patients with metastatic colorectal cancer over twenty years. NPJ Precis Oncol 2023;7:16–.
[5] Kim MJ, Jeong SY, Choi SJ, *et al*. Survival paradox between stage IIB/C (T4N0) and stage IIIA (T1-2N1) colon cancer. Ann Surg Oncol 2015;22:505–12.
[6] Savu E, Vasile L, Serbanescu MS, *et al*. Clinicopathological analysis of complicated colorectal cancer: a five-year retrospective study from a single surgery unit. Diagnostics 2023;13:2016.
[7] Schneider NI, Langner C. Prognostic stratification of colorectal cancer patients: current perspectives. Cancer Manag Res 2014;6:291.
[8] Kolmar L, Autour A, Ma X, *et al*. Technological and computational advances driving high-throughput oncology. Trends Cell Biol 2022;32:947–61.
[9] Isaic A, Motofelea N, Hoinoiu T, *et al*. Next-generation sequencing: a review of its transformative impact on cancer diagnosis, treatment, and resistance management. Diagnostics 2025;15:2425.
[10] Huhulea EN, Huang L, Eng S, *et al*. Artificial Intelligence advancements in oncology: a review of current trends and future directions. Biomedicines 2025;13:951.
[11] Guo SB, Meng Y, Lin L, *et al*. Artificial intelligence alphafold model for molecular biology and drug discovery: a machine-learning-driven informatics investigation. Mol Cancer 2024;23. doi:10.1186/s12943-024-02140-6
[12] Guo S-B, Liu D-Y, Fang X-J, *et al*. Current concerns and future directions of large language model ChatGPT in medicine: a machine-learning-driven global-scale bibliometric analysis. Int J Surg 2026;112:8961–63.
[13] de Swart Wk, Loog M, Krijthe JH. A comparative study of methods for dynamic survival analysis. Front Neurol 2025;16:1504535.
[14] Vieira LM, Jorge NAN, Sousa JB, *et al*. Machine learning-based survival prediction in colorectal cancer combining clinical and biological features. Oncotarget 2025;16:834.
[15] Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372. doi:10.1136/BMJ.N71
[16] Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5:1–10.
[17] McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22:276.
[18] Moons KGM, Damen JAA, Kaul T, *et al*. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ 2025;388. doi:10.1136/BMJ-2024-082505
[19] Wulczyn E, Steiner DF, Moran M, *et al*. Interpretable survival prediction for colorectal cancer using deep learning. NPJ Digit Med 2021;4. doi:10.1038/s41746-021-00427-2
[20] Chuanji Z, Zheng W, Shaolv L, *et al*. Comparative study of radiomics, tumor morphology, and clinicopathological factors in predicting overall survival of patients with rectal cancer before surgery. Transl Oncol 2022;18. doi:10.1016/j.tranon.2022.101352
[21] Willems A, Panchy N, Hong T. Using single-Cell RNA sequencing and MicroRNA targeting data to improve colorectal cancer survival prediction. Cells 2023;12:228.
[22] Yang X, Qiu H, Wang L, *et al*. Predicting colorectal cancer survival using time-to-event machine learning: retrospective cohort study. J Med Internet Res 2023;25:e44417.
[23] Altaf A, Kawashima J, Khalil M, *et al*. Identification of a gene signature and prediction of overall survival of patients with stage IV colorectal cancer using a novel machine learning approach. Eur J Surg Oncol 2025;51:109718.
[24] Hong T, Zhang H, Zhao Q, *et al*. A hybrid machine learning CT-Based radiomics nomogram for predicting cancer-specific survival in curatively resected colorectal cancer. Acad Radiol 2025;32:2630–41.
[25] Li J, Yu K, Chen Z, *et al*. A machine-learning prognostic model for colorectal cancer using a complement-related risk signature. Oncol Res 2025;33:3469–92.
[26] Li Z, Aihemaiti Y, Yang Q, *et al*. Survival machine learning model of T1 colorectal postoperative recurrence after endoscopic resection and surgical operation: a retrospective cohort study. BMC Cancer 2025;25. doi:10.1186/s12885-025-13663-6
[27] Ibrahim KSMH, Huang YF, Ahmed AN, *et al*. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. Alexandria Eng J 2022;61:279–303.
[28] Wu H, Chen Y. Integrative prognostic modeling and mediation analysis of recurrence risk in extremely early-stage oral squamous cell carcinoma. J Stomatol Oral Maxillofac Surg 2026;127:102708.
[29] Piwocka O, Suchorska WM, Kulcenty K. Empowering personalized medicine: unleashing the potential of patient-derived explants in clinical practice. EXCLI J 2024;23:81.
[30] Muhammad D, Bendechache M. Unveiling the black box: a systematic review of explainable Artificial Intelligence in medical image analysis. Comput Struct Biotechnol J 2024;24:542.
[31] Sadeghi Z, Alizadehsani R, Cifci MA, *et al*. A review of explainable Artificial Intelligence in healthcare. Comp Electr Eng 2024;118:109370.
[32] O'Donnell A, Cronin M, Moghaddam S, *et al*. A systematic review on machine learning techniques for survival analysis in cancer. Cancer Med 2025;14:e71375.
[33] Collins GS, Moons KGM, Dhiman P, *et al*. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024;385. doi:10.1136/BMJ-2023-078378
[34] Haneuse S. Distinguishing selection bias and confounding bias in comparative effectiveness research. Med Care 2016;54:e23.
[35] Kolla L, Parikh RB. Uses and limitations of artificial intelligence for oncology. Cancer 2024;130:2101.