International Journal of Medicine
pISSN 1468-3814 | eISSN 2667-7008
https://theinternationalmedicine.com/

Original Researcher Article *DOI: 10.61336/im/26-09-20*

From Anatomy to Outcome: Integrating CT Radiomics, Artificial Intelligence, and Clinical Features to Predict Metastatic Spread, Survival, and Treatment Response in Colorectal Cancer.

Dr. Salman Khan[1], Dr. Muhammad Rashed[2], Dr. Tariq Hayat Khan[3]*, Dr. Summaya zafar jalal[4], Dr. Sara Ishaq[5], Zainab Nabeel[6].

[1]*Senior Registrar Saidu Group of Teaching hospitals Swat*
[2]*Surgical Specialist type D hospital Garhi Habibullah Mansehra*
[3]*Assistant professor lady reading hospital, MTI Peshawar*
[4]*Assistant professor Jinnah Medical College Peshawar*
[5]*Assistant professor Jinnah Medical College Peshawar*
[6]*Jinnah Medical College Peshawar.*

Correspondence:
Dr. Tariq Hayat Khan.
Assistant professor lady reading hospital, MTI peshawar.

Email:
drtariqhayat2011@gmail.com

Received: 07-08-2026
Revised: 15-08-2026
Accepted: 25-08-2026
Published: 03-09-2026

Abstract

Background: Despite the continuous advances in CRC diagnosis and treatment, colorectal cancer (CRC) is still a leading cause of cancer deaths and illness, especially due to the unpredictable metastatic spread, recurrence and response to treatment when relying on anatomical staging alone. Although computed tomography (CT) is an imaging modality that is routinely used for staging and surveillance, conventional interpretation of the image only captures a fraction of the phenotypic information within the image. Radiomics and artificial intelligence (AI) provide a way to transform computed tomography (CT) data into quantitative biomarkers which could represent tumour heterogeneity, microenvironmental properties, and treatment-responsive phenotypes. ***Aim and Objectives:*** To discuss the role of CT radiomics, AI, and integrated clinical parameters in prediction of metastatic spread, survival, and therapeutic response in CRC with a critical perspective and evaluate their clinical applicability to meaningful risk stratification. ***Methods:*** Three outcome domains were chosen for a structured narrative synthesis of current literature: metastatic progression, survival/recurrence, and response to systemic or locoregional treatment. Emphasis was placed on studies that employed the use of contrast-enhanced CT, studies that involved machine-learning or deep-learning models, clinical-radiomic integration, internal or external validation, and studies that adhered to reproducibility standards. ***Results:*** Based on publicly published evidence, imaging phenotypes, such as using CT-derived radiomic signatures, can be identified as predicting metachronous liver metastasis, treatment response, pathological tumour regression, and survival. In various studies, models that included both the radiomic and clinical variables were as accurate as models that included either set of variables individually. But the performance was not universal and certain survival analyses demonstrated a marginal benefit of radiomics after controlling for already established and relevant clinical parameters. There is still significant variability between image acquisition and segmentation, feature extraction, sample size, model calibration, and external validation which continue to impede generalizability. ***Conclusion:*** While the potential of using CRC imaging for extending from anatomy to quantitative prediction of outcomes is promising, the most plausible application of CT radiomics and AI is as complementary decision support tools that can be combined with clinical and pathological data and molecular data. Routine clinical use is not feasible until standardized radiomic pipelines are developed, validated across multiple centres and reported in a transparent way and evaluated prospectively.

Keywords: colorectal cancer; computed tomography; radiomics; artificial intelligence; machine learning; liver metastasis; survival; treatment response.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).

INTRODUCTION

Colorectal cancer (CRC) is a significant human health problem, and one of the top causes of cancer mortality. Newer estimates from GLOBOCAN suggest that colorectal malignancies represent around 10% of new cancer cases globally, and a similar percentage of cancer deaths1. Despite advances in screening, surgery, systemic therapy, molecular profiling and multidisciplinary treatment, the prognosis is still highly dependent on the extent of metastasis and on significant interpatient treatment response variability. The liver is a highly significant site of metastasis and prediction of high-risk patients prior

to the appearance of metastases or to failure of treatment could have a significant impact on the intensity of surveillance and therapeutic planning. Cross sectional imaging plays a crucial role in the staging and follow up of CRC. In particular, computed tomography (CT), which is widely available, fast, has high spatial resolution and can evaluate the primary tumour, lymph nodes, liver, lungs, peritoneum and other disease locations. However, conventional CT evaluation tends to be purely anatomic in terms of tumour size, morphology, attenuation, nodal size and the presence and absence of visible metastatic deposits. These descriptors are clinically relevant and do not necessarily reflect intratumoral and peritumoral diversity that arises from differences in underlying cellularity, necrosis, vascularity, stromal organisation or clonal diversity2 High-dimensional quantitative features extracted from routinely acquired medical images have become a computational method known as radiomics. These features encompass first order measures of intensity, shape features, and texture features that are higher-order descriptors of spatial relationships among voxels3.

These features can be turned into predictive signatures for clinically relevant outcomes, when combined with machine-learning or deep-learning algorithms. This can be of particular interest in CRC since the CT images are routinely acquired as part of routine therapy, offering the potential to gain further prognostic information without the need for an additional invasive procedure or targeted imaging exam. There is growing evidence that the radiomic phenotypes can supply information at various points in the course of CRC disease. Multicentre studies have shown that baseline CT radiomics could be useful to identify patients at risk of developing the metachronous colorectal liver metastases, and a combined clinical-radiomic approach had a useful discrimination in validation cohorts4. Other studies have used CT radiomics to predict chemotherapy response and pathological regression in colorectal liver metastases, including models that performed better than conventional radiological response assessment in a subset of patients5.

Also promising is the results of survival modelling, but benchmark analysis indicates that radiomics' incremental value is not consistent and can be reduced if strong clinical prognostic variables are added6. This is important as the variability will be important if a statistically sound model is not necessarily a clinically valid biomarker. Methodological heterogenity adds to the complexity of translation. Radiomic signatures may be affected by the acquisition and reconstruction parameters of the computed tomography (CT), the timing of the contrast, the segmentation of the tumour, the image resampling, the calculation of features, the selection of the features, the complexity of the model, and the way it is validated7. The Image Biomarker Standardisation Initiative (IBSI) has therefore identified the need for defining features in a standardized fashion and processing workflows in a reproducible manner8, and current guidance for prediction models, including TRIPOD+AI, requires open reporting on model development, evaluation, calibration and validation9.

In this context, the current paper reviews the evolution from the traditional anatomical evaluation of CT towards comprehensive prediction of outcome in CRC. It aims to integrate evidence related to the capacity of CT radiomics and AI to predict metastatic spread, survival, recurrence, and treatment response and to specifically determine the added value of using combination with clinical features. The paper also addresses the methodological challenges currently hindering generalizability, as well as the steps needed to take the models developed in radiomic-AI research and make them available for use in clinical decision support systems.

MATERIALS AND METHODS

Study Design and Review Framework

This study was organized into a structured systematic-style review and narrative evidence synthesis, along with an illustrative dataset from local hospitals and a small survey of the clinicians. The literature review was dedicated to the application of computed tomography (CT) radiomics, artificial intelligence (AI), machine learning and integrated clinical variables in predicting the metastatic spread, survival, recurrence and treatment response in colorectal cancer (CRC). The review process followed PRISMA 2020 principles to enhance transparency of study identification, screening and reporting10. Due to the expected significant methodological and clinical differences between different radiomics studies, quantitative meta-analysis was not a pre-specified analysis, but rather effect estimates and performances were summarized in comparison.

Literature Search Strategy

A structured search strategy was created for PubMed/MEDLINE and further enhanced by screening of reference lists of the most relevant reviews and primary studies. The search was performed between 2016/01 and 2026/08 and older landmark radiomics papers were included if they were methodologically important.

The following search terms were used to combine: colorectal cancer, colon cancer, rectal cancer, colorectal liver metastasis, computed tomography, CT, radiomics, texture analysis, artificial intelligence, machine learning, deep learning, metastasis, survival, prognosis, recurrence, treatment response, and chemotherapy. Peer reviewed human studies published in English and studies from 2018 onward were given priority.

Eligibility Criteria

Studies were included when they included adult patients with histologically confirmed CRC, applied quantitative imaging features from computed tomography (CT), machine learning or deep learning analyses, and predicted a prespecified clinical outcome. Improved outcomes were defined as lymph-node or distant metastases, colorectal liver metastases, overall survival, disease-free survival, progression-free survival, recurrence, response to systemic chemotherapy, pathological response, and response to locoregional chemotherapy. Studies that included only image segmentation without outcome prediction, non-CT imaging alone, mixed tumour populations with the results of CRC reported separately, did not have outcome data that could be extracted or were limited to conference abstracts, editorials, case reports, and non-peer-reviewed articles were excluded.

Data Extraction and Evidence Synthesis

For all the studies included, information regarding publication year, study design, number of subjects, CRC population, imaging phase, tumour or metastatic lesion segmentation, class of radiomic features, AI or statistical algorithm, clinical variables, validation strategy, and target outcome was retrieved. Predictive performance was measured as area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, concordance index, hazard ratio or reported survival, depending on the specific study. A special focus was placed on the comparisons between the radiomics-only, clinical-only, and combined clinical-radiomic models. Random internal train-test splitting was not as good as external validation. The methodological quality was assessed based on the PRospective Biostatistical Assessment of Study and Translation of AI (PROBAST+AI) for prediction-model studies and the CLAIM for medical-imaging AI reporting11,12. Radiomic feature extraction and reproducibility were explained in the context of the recommendations from the Image Biomarker Standardisation Initiative8,9. The reporting of predictive models was assessed based on TRIPOD+AI guidance9.

Illustrative Local-Hospital Dataset and Clinician Survey

A synthetic hospital cohort was planned to illustrate how the published evidence could be applied to a local clinical situation, in the draft analysis. The illustrative dataset was designed to reflect clinically realistic distributions reported in CRC cohorts, but is not a true representation of clinically observed patients. Variables include the age, sex, primary tumour site, TNM stage, mass of carcinoembryonic antigen, liver metastasis, nodal metastasis, selected CT-radiomic risk category, type of treatment received, treatment response, recurrence, and survival status. Submissions of future patient institutional data would need to be replaced with de-identified patient data and ethical approval from the institution. A small exploratory survey was also recommended for radiologists, oncologists, surgeons and clinicians. The survey items include questions about radiomics and AI knowledge, perceived value of knowledge of radiomics and AI in metastatic-risk prediction and treatment planning, confidence in AI-assisted decision making, and perceived barriers such as validation, interpretability, workflow integration and data standardization.

Statistical Analysis

Descriptive synthesis of secondary published results. Means and standard deviations or medians and interquartile ranges were used to present continuous variables in the illustrative data set, and frequencies and percentages were used for categorical variables. Chi-square or Fisher's exact tests were used for categorical variables, and logistic regression was used where applicable, to assess exploratory associations. For survival-oriented illustration, Kaplan-Meier estimates were used and the Cox proportional-hazards modelling used only where sample size and event number were sufficient. A two-sided P value <0.05 was considered statistically significant and an AUC value was not considered if it was evaluated without a confidence interval.

RESULTS

Overview of the Evidence

The literature reviewed showed that radiomic features derived from computed tomography (CT) images and machine learning methods are being applied in the colorectal cancer pathway, and that their application is growing. The three major clinically relevant questions that brought together the main applications were: can quantitative imaging be used to identify patients likely to develop metastatic disease; can radiomic phenotypes be used to stratify survival or recurrence risk; and can pretreatment or post-treatment CT predict response to systemic therapy.

In each of these domains, predictive performance was widely different across the different cohorts, imaging protocols, segmentation strategies, feature selection methods, algorithms, and design for validation. However, a common characteristic was that radiomics may be able to provide information that was not easily discerned by the conventional visual interpretation of the CT. Prior to the individual outcome domains, representative studies discussed in the present synthesis are summarized in Table 1.

Table 1: Representative CT radiomics and AI studies in colorectal cancer outcome prediction

Study	Population	Outcome	Model/approach	Main reported finding
Li et al., 2022 [4]	323 CRC patients, two institutions	Metachronous liver metastasis	CT radiomics + clinical fusion	Fusion AUC 0.79 internally and 0.72 externally
Taghavi et al., 2021 [13]	91 CRC patients, multicentre	Metachronous liver metastasis	Random forest radiomics	Validation AUC 0.86 for radiomics; clinical model 0.71
Bülbül et al., 2024 [14]	73 colon cancer patients	Lymph-node involvement/T stage	Multiple ML algorithms	LN-prediction AUC range 0.557–0.800
Stüber et al., 2023 [6]	491 patients with hepatic CRC metastases	Overall survival	RSF, elastic net, XGBoost	Radiomics comparable with clinical information but little additional prognostic gain
Luo et al., 2024 [16]	180 CRLM patients	Disease-free survival	Elastic net and random survival forest	High radiomic-risk groups showed significantly shorter DFS
Mian et al., 2026 [17]	399 patients; 959 metastases	3-year mortality/overall survival	Clinical-radiomic ML	Combined model AUC 0.83; radiomics alone 0.77
Ammirabile et al., 2026 [18]	306 CRLM patients	Overall survival	Clinical + tumour/peritumoral radiomics	C-index increased from 0.629 clinically to 0.717 in selected combined modelling
Wei et al., 2021 [15]	192 CRLM patients	Chemotherapy response	Deep-learning radiomics	Validation AUC 0.82; combined with CEA 0.83
Karagkounis et al., 2024 [5]	85 patients; 95 CRLM lesions	Pathological response	ML CT radiomics	Validation AUC 0.87 vs 0.53 for RECIST
Miyamoto et al., 2025 [19]	150 CRLM patients	First-line chemotherapy response	Random forest/Boruta radiomics	Validation AUC 0.87

CRC: colorectal cancer; CRLM: colorectal liver metastases; AUC: area under the receiver-operating-characteristic curve; CEA: carcinoembryonic antigen; DFS: disease-free survival; RSF: random survival forest; ML: machine learning.

Prediction of Metastatic Spread

One of the most promising applications of CT radiomics was prediction of metastatic disease. Li et al. analyzed 323 patients from 2 institutions, and derived over 1200 quantitative values from baseline contrast-enhanced computed tomography (CT). The internal performance of the radiomics-only model and clinical model were comparable, while the fusion of imaging and clinical information yielded the best overall model. Importantly, the combined approach had an AUC of about 0.72 when evaluated in an external cohort, reflecting not only the potential reduction in performance when the algorithm is transferred outside of its development cohort[4], but the anticipated decrease as well.

In apparently normal liver parenchyma, in the absence of apparent liver metastases, an earlier multicentre study by Taghavi et al. assessed liver parenchyma on primary staging CT. With 24 months of follow-up, the validation AUC for radiomics was ~0.86, while the AUC of clinical variables was 0.71[13]. The observation is of particular conceptual interest, as it implies that the heterogeneity of the images beyond the primary tumour limits of the visible tumour bears information concerning a pre-metastatic or host tissue phenotype. There has also been some research using radiomics to predict lymph-node involvement. Bülbül et al. found the AUCs for machine learning algorithms ranged from ~0.56 to 0.80 in predicting nodal involvement, and slightly better for advanced T stage[14].

The results indicate that the use of CT texture may be useful for staging, although it is not as powerful as for some liver metastasis applications when it comes to predicting metastasis. The findings of the metastatic spread studies suggest that radiomics could be most helpful for risk enrichment rather than definitive diagnosis. An AI score alone is not considered an indication of occult metastatic disease, and may be a justification for intensifying surveillance, thorough multidisciplinary evaluation, or combining with circulating, molecular, pathological or clinical biomarkers.

Survival, Recurrence, and Prognostic Stratification

Encouraging, but less uniform results were seen in the survival prediction. Luo et al. tested elastic-net and random-survival-forest signatures they created for disease-free survival using contrast-enhanced CT in 180 patients with colorectal liver

metastases. Both resulted in the separation of higher and lower risk patients; in the case of the random-survival-forest model, hazard ratios of 2.54 in the training set and 1.84 in the testing set were obtained for the identified risk groups[15]. More recent multicentre evidence has confirmed the prognostic value, and has put the size of its effects into context. Mian et al. reported the AUC of the radiomics was 0.77 and the AUC of clinical variables was 0.82 for predicting the three-year survival in 399 patients with colorectal liver metastases (CRLM).

The random-forest performance was slightly boosted by the combination of the two to 0.83. The radiomic risk score was, however, still useful for stratification of survival, as high-risk patients had significantly poor outcomes in an independent validation cohort[16]. In the same way, a study of 306 patients in 2026 had a clinical-model C-index of 0.629. When a patient's CT exam was performed within 30 days prior to surgery, the C-index for patients who had tumour radiomics added on to the model was 0.691, and the C-index for those who had peritumoral features added to the model was 0.717[17].

This finding reinforces the increasing trend of studying the tumour margin and the surrounding microenvironment instead of the tumour core only when using radiomic analysis. Not every analysis showed clinically significant improvement, though. On 60 machine-learning pipeline configurations, Stüber et al. evaluated radiomic and clinical data from 491 patients. Radiomics also have contented prognostic information similar to clinical variables; however, when added to the clinical dataset, there was no relevant additional survival benefit[6]. This negative finding is relevant as it shows that predictive information contained within CT at least partly overlaps with existing markers of tumour burden, liver function, molecular status and disease severity.

Prediction of Treatment Response

Some of the best discriminations were obtained for the prediction of treatment response. In 192 patients who received first-line chemotherapy for CLM, Wei et al. created a deep learning radiomics model. The model has an AUC of 0.903 when it is trained and 0.820 when it is tested in time. While modest, addition of CEA yielded a modest improvement: the AUC for validation was 0.830[18]. Miyamoto et al. then analysed 150 patients receiving first-line doublet chemotherapy. Individual CT texture features exhibited moderate discriminatory ability, while machine learning integration resulted in significant improvement, yielding a training AUC of 0.94, and a validation AUC of 0.87[19]. This difference highlights one key point: while each radiomic variable might not be a clinically meaningful feature individually, combinations of variables might be more informative and encode imaging phenotypes. The machine-learning radiomics compared with conventional radiological assessment after chemotherapy, of particular clinical interest, revealed that radiomics is superior. Their model had a validation AUC of 0.87 for pathological response while the one for subjective RECIST assessment was 0.53, and the one for CT morphological assessment was 0.56[5]. The findings indicate that dimensional shrinkage alone does not sufficiently reflect treatment-related biological changes, which may be detectable by quantitative imaging.

Illustrative Local-Hospital Dataset

In order to demonstrate how these findings can be tested locally, an explicitly simulated data set of CRC patients (n = 120) was created. It should not be interpreted as real institutional evidence and must be substituted by real patient data extracted from institutions, which are acceptable from an ethical standpoint, before being used as original clinical study for submission. The synthetic cohort was carefully selected to simulate plausible CRC distributions, as indicated in Table 2, and to be similar to no published set.

Table 2: Simulated local-hospital colorectal cancer cohort (n = 120)

Variable	Simulated result
Median age	61 years (IQR 52–70)
Male sex	68 (56.7%)
Colon primary	74 (61.7%)
Rectal primary	46 (38.3%)
Stage II	24 (20.0%)
Stage III	56 (46.7%)
Stage IV	40 (33.3%)
Elevated CEA	65 (54.2%)
Liver metastasis	34 (28.3%)
High CT-radiomic risk category	48 (40.0%)
Recurrence during follow-up	38 (31.7%)
Death during follow-up	22 (18.3%)

In the simulated dataset, there were 72 low-risk patients and 48 high-radiomic-risk patients, with liver metastasis occurring in 20 of the 48 high-risk patients (41.7%) and 14 of the 72 low-risk patients (19.4%). The odds ratio for this was 2.96 (P≈

0.013) illustrative. In addition, the high-risk group had a higher number of cases of recurrence, albeit this simulated association was not statistically significant at the usual level (P=0.072). The results are meant to be illustrative of a realistic scenario in which an imaging biomarker may show association with outcome without reaching the threshold of significance in all endpoints.

Illustrative Clinician Survey

A small simulated survey dataset was also created to illustrate the reporting of clinician acceptance if a real local survey is conducted.

Table 3: Illustrative clinician survey concerning CT radiomics and AI (n = 20)

Survey item	Positive responses
Familiar with radiomics/AI concepts	13/20 (65%)
Believe AI could improve metastatic-risk assessment	16/20 (80%)
Believe AI could assist treatment-response prediction	15/20 (75%)
Currently comfortable using AI output in clinical decisions	8/20 (40%)
Identify insufficient external validation as a major barrier	17/20 (85%)
Identify model interpretability as a major barrier	14/20 (70%)
Identify workflow integration as a major barrier	12/20 (60%)

One of the simulated surveys provides an example of an important gap that might exist in the transition to the clinic: clinicians might realize the potential of radiomics, but be wary of taking action on basis of algorithmic predictions. The difference between technical performance and clinical trust is of the essence in evaluating the published evidence and will be expanded upon in the Discussion.

DISCUSSION

The conclusions that can be drawn from the evidence contained in this review suggest that the ability of artificial intelligence and CT radiomics to go beyond anatomic assessment of colorectal cancer is enormous. Most quantitative CT features showed some relationship with clinically important outcomes across the various studies examined for metastatic progression, survival, recurrence, and treatment response. The general message is, however, not that radiomics should be used in lieu of the established staging or clinical assessment, but that imaging-derived data may be most useful when combined with clinical and pathological and molecular information. One of the prominent features of CT radiomics is its non-invasive capability to query tumour heterogeneity. Traditional CT interpretation includes the size of the tumour, the contour of the tumour, the attenuation of the tumour, enlargement of the nodes, and presence of metastatic disease. These images can also be transformed into numerical data detailing the distribution of intensity, the morphology of the tumour, and the spatial texture of the image, which can indirectly yield information on cellular heterogeneity, necrosis, angiogenesis, stromal composition and tumour–host interactions2,3. This is especially applicable to CRC where tumours of similar anatomical stage can have very different biological progression. Radiomic analysis is thus an attempt to exploit phenotypic information that is present in the routine imaging but which is not consistently noticed by the human eye.

Radiomics and Metastatic Risk

One of the more interesting possibilities for application is the prediction of the spreading of metastasis. The multicentre study by Li et al. showed that pretreatment primary tumour CT features could be useful in predicting the presence of metachronous liver metastases. While radiomics and clinical information each achieved sufficient discrimination, there was a significant improvement in overall prediction using both together4. In a similar study, Taghavi et al. demonstrated that apparently non-metastatic liver parenchyma-derived radiomic features could separate those patients who later developed colorectal liver metastases, achieving a validation AUC of 0.8613. This is an intriguing biological question. The visible primary tumour will not be the only factor that indicates metastatic risk; subtle features of the target organ and of the surrounding tissue may also play a role. This aligns with the general idea of a tumour microenvironment and premetastatic niche, but does not provide any information regarding the underlying biological mechanisms, which can only be inferred from radiomic associations. Currently, therefore, radiomic risk scores can be considered predictive imaging biomarkers and not as direct measures of metastatic biology. The ability to interpret the clinical situation is also important. Even if the AUC is just 0.80–0.86, it is not enough to consider an individual patient as having occult metastases. Rather, these models could be used in the future to predict risks. The findings may lead to closer monitoring, a multi-disciplinary review or a combination with carcinoembryonic antigen, circulating tumour DNA, histopathology and molecular markers in a high-risk result. This complimentary approach is safer, and more realistic, than relying on a radiomic score as a standalone diagnostic decision.

Survival and Recurrence Prediction

Evidence of survival prediction is encouraging but is more varied. Luo et al. showed that by employing a CT radiomic signature, patients with colorectal liver metastases (cRLMs) could be stratified into subsets with significantly different disease free survival15. Later multicentre studies have reinforced the findings of the prognostic value of radiomic features. However, Mian et al. reported moderate discrimination of 3-year mortality based on radiomics, with slightly better discrimination when clinical variables were added (AUC of about 0.83)16. In the same way, Ammirabile et al. found that the inclusion of tumour and peritumoral radiomic features in clinical models resulted in a better survival discrimination, especially if the time between the CT and surgery was short17. These studies help to demonstrate two key points. Firstly, the performance of the model might be affected by the timing of the image acquisition as the tumour biology is dynamic. Images acquired several weeks prior to surgery may reflect a different state of biology than is present at surgery. Secondly, the surrounding tissue of the tumour may provide useful information for prognosis. Peritumoral radiomics may reflect tumour–host interactions, vascular changes, inflammatory changes and infiltrative growth patterns, which are not represented by just segmenting the tumour area. However, some studies have failed to show any incremental value. Stüber et al. showed that radiomic features could provide a survival prediction similar to clinical variables; however, when combined with clinical variables, adding the radiomic features was not of statistically significant benefit6. The discovery is especially significant given that positive predictive studies can lead to an inflated view of AI capabilities when published without any negative outcomes reported. This should be a central issue between statistical prediction and clinical utility. The improvement in AUC or concordance-index of a model does not necessarily translate into a change in treatment decisions or patient outcomes. Future studies should be directed towards assessing discrimination as well as calibration, decision-curve analysis, net clinical benefit, and the impact of predictions on multidisciplinary management.

Prediction of Treatment Response

The prediction of treatment response seems to be a promising area with respect to size being the main criterion for conventional treatment response. Although, RECIST is still clinically applicable, the changes in tumour diameter do not always correspond directly to the histological tumour regression. The process of therapy can induce necrosis, fibrosis, vascular alteration and internal textural change before significant dimensional reduction. This constraint is a good argument for the use of radiomic assessment. In CRC-LM, deep-learning radiomics was found to be capable of predicting chemotherapy response, with a validation AUC of 0.82, even with the addition of CEA, which led to an AUC of 0.8316. Miyamoto et al. also found a validating AUC of 0.87 with a machine-learning model that used several CT texture parameters19. Perhaps more impressively, Karagkounis et al. report a study that found a machine-learning model based on CT radiomics outperformed RECIST (AUC of 0.87 for pathological response vs 0.53) and conventional morphological assessment of CT (AUC of 0.56)5. This does not mean that radiologists or RECIST is no longer needed. Instead, it shows that there are changes associated with treatment that do not get captured by diameter measurements and may be detected by quantitative analysis. Potential future use of this approach may allow for earlier identification of likely non-responders, prevent unnecessary and prolonged chemotherapy, identify patients for conversion surgery, and improve the timing of liver-directed treatment. The aforementioned applications, however, need prospective studies demonstrating the added value of radiomics-guided management, and that needs to happen.

Why Integrated Models Are Likely to Be Stronger

Across the reviewed evidence, integrated models provide the most convincing direction for future development. Imaging phenotype represents only one dimension of CRC biology. Age, performance status, TNM stage, CEA, tumour location, number and distribution of metastases, treatment exposure, RAS/BRAF status, microsatellite instability, histopathological characteristics, and other clinical factors carry independent information. Expecting an image-only algorithm to consistently outperform all these variables is therefore unrealistic.

A more appropriate model is one in which AI integrates complementary information. Radiomics can provide quantitative phenotypic information, clinical variables describe the patient and disease state, pathology provides microscopic biology, and molecular biomarkers characterize genetic and signalling pathways. In this framework, AI functions as an integrative analytical tool rather than as an autonomous replacement for clinical expertise.

Reproducibility and Generalizability

The issue of reproducibility is one of the main challenges to translation, although results are promising. Radiomic features may vary depending on the type of CT scanner, tube voltage, slice thickness, reconstruction kernel, injection of contrast medium, enhancement phase, resampling of voxels, the segmentation method, and image-processing software7. Signatures that are acquired through one procedure might not function the same way in different hospitals. This problem is demonstrated in some studies by the drop in predictive performance from the development to the independent validation set. For instance, Li et al. found that their clinical-radiomic fusion model achieved an AUC of 0.79 on their internal dataset and ~0.72 on their test set4. This loss of performance should not be attributed to model failure, but instead it will be measured through an external test to find out how much of the initial performance can be attributed to transferable signal

and how much to signal specific to the development test. This standardization effort is thus essential. The Image Biomarker Standardisation Initiative is providing standardisation of radiomic feature calculation definitions and reference values8. Meanwhile, the TRIPOD+AI recommendations emphasize the need to report the development and evaluation of prediction models9 and the PROBAST+AI checklist examines quality, risk of bias, applicability, and the potential for algorithmic bias11. In medical imaging AI studies, as CLAIM does, transparency in reporting is a key focus. It should be normal, not optional, to follow these standards.

Artificial Intelligence, Interpretability, and Clinical Trust

There is another challenge with complexity: the model. While deep neural network approaches can learn relationships that are not recognizable through the traditional regression approaches, it can sometimes be hard to understand the meaning of highly complex algorithms. For instance, a prediction of '73% probability of non-response' may be of little clinical value without knowledge of the model's training population, the definition of the outcome of interest, calibration, the model's uncertainty, and pertinent patient features. Not all mathematical expressions are necessarily easy to interpret. However, it is important for clinicians to know why the model is being adopted, what the model is predicting, the accuracy of the model and when the prediction is not to be trusted. Algorithmic outputs must be used in conjunction with the other disciplines and not in place of.

Interpretation of the Illustrative Local Data

For the purpose of demonstrating the structure for institutional validation, the simulated local-hospital cohort presented herein has been added. Liver metastasis was more common in the high-radiomic-risk group in this illustrative finding, which is directionally concordant with literature but does not represent clinical evidence, as these were synthetic findings. The same model can be used in a real study in the same hospital using archived, de-identified CRC patients with baseline CT scans. Radiomic scores could be then validated with local outcomes such as nodal involvement, liver metastasis, recurrence, treatment response and survival. Important: A locally developed model should be tested against an existing signature built by another group, not build a very complex algorithm from a small sample of data, thereby risking an over-fitted model. Likewise, the illustrative clinician survey reveals a potential disconnect between enthusiasm for AI and the potential use of its output. This could offer helpful implementation information on validation, interpretability, training needs, integration into workflow, and clinician confidence in a true survey. These factors are important when implementing the prediction model, since it is not much good if an accurate prediction model cannot be easily and safely added to clinical practice.

Limitations and Future Directions

There are a number of limitations to the existing evidence base. The majority are retrospective studies, many are single-centre, and sample sizes are still relatively small relative to the number of imaging features of interest in candidate studies. Manual segmentation is observer dependent and imaging protocols vary between centres, limiting reproducibility between centres. Definitions of outcomes are also inconsistent, especially when it comes to survival and treatment response, making it difficult to easily pool performance estimates. Prospective multicentre validation, harmonized acquisition and reconstruction protocols, automated but quality-controlled segmentation, predefined feature extraction pipelines and transparent registration of analysis plans should be emphasized in future research. Studies should include calibration and clinical utility in addition to AUC and clinical testing should be performed in clinically relevant sub-groups. The incorporation of genomic, pathological, and laboratory and circulating biomarkers may further enhance the ability to predict individualization. So, the real question isn't if an AI model can achieve a high retrospective AUC, but if it can provide reproducible information that can enhance decisions and outcomes when applied in real clinical pathways.

CONCLUSION

CT radiomics and artificial intelligence represent an important step forward in the imaging of colorectal cancer, transforming the role of CT from purely anatomical tool into a means of quantitatively characterising tumour phenotype and outcome risk outcome. First evidence shows that radiomic and machine-learning models can be used to predict metachronous metastatic spread, recurrence, survival, response to chemotherapy, and pathological treatment response. Colorectal liver metastases have shown specific encouraging results with quantitative imaging that may provide information on prognosis and therapy that is not fully captured by traditional size-based evaluation. But the evidence suggests that routine clinical evaluations should not be replaced by independent AI prediction.

There is population and institutional variability in performance, and radiomics does not always deliver significant incremental value beyond powerful clinical models. The most plausible route is hence that of integrated prediction, which integrates various clinical, pathological, molecular and laboratory features with imaging biomarkers derived from the CT. Future research is needed to highlight the importance of external and prospective validation, model calibration, uniformity of image acquisition and feature extraction, interpretability, fairness, and clinical utility of radiomics in order to move them from research methodology to clinical routine decision support. Thus, radiomics using CT should be viewed as a

complementary quantitative information source, rather than as a substitute for the radiologist or oncologist. Workflow standardization and validation with the right tools, could make a significant contribution to more individualized treatment of colorectal cancer from anatomy to outcome prediction.

REFERENCES

1. Bray F, Laversanne M, Sung H et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians 2024;74(3):229–63.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278(2):563–77.
3. Lambin P, Leijenaar RT, Deist TM et al. Radiomics: the bridge between medical imaging and personalized medicine. Nature Reviews Clinical Oncology 2017;14(12):749–62.
4. Li Y, Gong J, Shen X et al. Assessment of primary colorectal cancer CT radiomics to predict metachronous liver metastasis. Frontiers in Oncology 2022;12:861892.
5. Karagkounis G, Horvat N, Danilova S et al. Computed tomography-based radiomics with machine learning outperforms radiologist assessment in estimating colorectal liver metastases pathologic response after chemotherapy. Annals of Surgical Oncology 2024;31(13):9196–204.
6. Stüber AT, Coors S, Schachtner B et al. A comprehensive machine learning benchmark study for radiomics-based survival analysis of CT imaging data in patients with hepatic metastases of CRC. Investigative Radiology 2023;58(12):874.
7. Traverso A, Wee L, Dekker A et al. Repeatability and reproducibility of radiomic features: a systematic review. International Journal of Radiation Oncology* Biology* Physics 2018;102(4):1143–58.
8. Zwanenburg A, Vallières M, Abdalah MA et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295(2):328–38.
9. Collins GS, Moons KG, Dhiman P et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. Bmj 2024;385.
10. Page MJ, McKenzie JE, Bossuyt PM et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj 2021;372.
11. Moons KG, Damen JA, 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.
12. Tejani AS, Klontzas ME, Gatti AA et al. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. Radiology: Artificial Intelligence 2024;6(4):e240300.
13. Taghavi M, Trebeschi S, Simões R et al. Machine learning-based analysis of CT radiomics model for prediction of colorectal metachronous liver metastases. Abdominal Radiology 2021;46(1):249–56.
14. Bülbül HM, Burakgazi G, Kesimal U. Preoperative assessment of grade, T stage, and lymph node involvement: machine learning-based CT texture analysis in colon cancer. Japanese Journal of Radiology 2024;42(3):300–7.
15. Luo X, Deng H, Xie F et al. Prognostication of colorectal cancer liver metastasis by CE-based radiomics and machine learning. Translational Oncology 2024;47:101997.
16. Mian A, Young R, Lakha AS et al. CT radiomics for survival risk stratification in resectable colorectal liver metastases: a multi-centre study. Scientific Reports 2026;16(1):19580.
17. Ammirabile A, Matteucci G, Fiz F et al. CT-based radiomics improves survival prediction in colorectal liver metastases: beyond clinical scores. Updates in Surgery 2026:1–12.
18. Wei J, Cheng J, Gu D et al. Deep learning-based radiomics predicts response to chemotherapy in colorectal liver metastases. Medical Physics 2021;48(1):513–22.
19. Miyamoto Y, Nakaura T, Ohuchi M et al. Radiomics-based machine learning approach to predict chemotherapy responses in colorectal liver metastases. Journal of the Anus, Rectum and Colon 2025;9(1):117–26.