Review

DOI: 10.5582/bst.2025.01045

# Artificial intelligence in colorectal cancer liver metastases: From classification to precision medicine

**Runze Huang[1,2,§], Xin Jin[1,2,§], Qinyu Liu[1,2], Xuanci Bai[3], Kenji Karako[4], Wei Tang[4,5], Lu Wang[1,2,*], Weiping Zhu[1,2,*]**

[1] Department of Hepatic Surgery, Fudan University Shanghai Cancer Center, Shanghai Medical College, Fudan University, Shanghai, China;
[2] Department of Oncology, Shanghai Medical College Fudan University, Shanghai, China;
[3] Department of Clinical Medicine, Shanghai Medical College, Fudan University, Shanghai, China;
[4] Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan;
[5] National Center for Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan.

**SUMMARY**: Colorectal cancer liver metastasis (CRLM) remains the leading cause of mortality among colorectal cancer (CRC) patients, with more than half eventually developing hepatic metastases. Achieving long-term survival in CRLM necessitates early detection, robust stratification, and precision treatment tailored to individual classifications. These processes encompass critical aspects such as tumor staging, predictive modeling of therapeutic responses, and risk stratification for survival outcomes. The rapid evolution of artificial intelligence (AI) has ushered in unprecedented opportunities to address these challenges, offering transformative potential for clinical oncology. This review summarizes the current methodologies for CRLM grading and classification, alongside a detailed discussion of the machine learning models commonly used in oncology and AI-driven applications. It also highlights recent advances in using AI to refine CRLM subtyping and precision medicine approaches, underscoring the indispensable role of interdisciplinary collaboration between clinical oncology and the computational sciences in driving innovation and improving patient outcomes in metastatic colorectal cancer.

***Keywords***: colorectal cancer liver metastasis, artificial intelligence, classification, precision medicine

## 1. Introduction

Colorectal cancer (CRC) is among the most prevalent malignancies globally, with over 50% of patients eventually developing colorectal liver metastasis (CRLM) (*1,2*). The high incidence of CRLM and its pivotal role in degrading patient survival underscores the importance of early and accurate classification within this population. Precise classification of CRLM serves as a cornerstone for optimizing therapeutic strategies. Moreover, it plays a pivotal role in predicting treatment responses and patient outcomes, thereby enabling more personalized and effective clinical management (*3*).

Conventional classification methods, such as histopathological analysis, imaging evaluation, and clinical risk scoring (CRS), while valuable, have notable limitations, including subjectivity, time spent, and dependency on expert interpretation (*4*). In contrast, artificial intelligence (AI) offers the potential for automated, efficient, and scalable classification, addressing the constraints of conventional approaches. AI excels in handling multimodal data, integrating information from imaging, genomics, and clinical parameters to enhance the accuracy of classification models (*5,6*).

This review provides a comprehensive overview of the role of machine learning (ML) in CRLM classification, focusing on current methodologies, data applications, and future directions. Specifically, the discussion covers established classification frameworks for CRLM, including intrapatient stratification (*e.g.* sensitivity to treatment) and interpatient subgrouping (*e.g.* distinguishing CRLM from liver metastases of non-CRC origins). By integrating AI advances in clinical use, this review aims to highlight the transformative potential of AI in CRLM management, promoting the advancement of precision medicine in oncology.

## 2. The concept of integrating AI into CRLM classification

In clinical practice, metastatic liver cancer staging primarily relies on the TNM system established by the AJCC in 2017, which classifies cancer based on

tumor invasion, lymph node involvement, and distant metastases. However, growing evidence suggests that this pathological classification often fails to fully reflect patient heterogeneity at specific stages. For example, with current advances in medical care, many patients classified as having advanced-stage disease under the TNM system still demonstrate the potential for long-term survival (*7*). Consequently, clinical guidelines are increasingly reevaluating its role in determining surgical indications.

A more comprehensive and precise staging system is urgently needed to guide personalized cancer treatment. Advances in AI, multi-omics sequencing, and clinical data integration are enabling more accurate and efficient classification models. ML-based systems outperform conventional methods in predicting treatment responses and prognoses, while AI-driven clinical decision support systems (CDSS) are transforming oncology care. This review aims to explore and validate these emerging possibilities (Figure 1)

## 2.1 Current methods of clinical classification of CRLM

### 2.1.1. Histopathological classification (HGP)

As early as 2001, Vermeulen *et al*. identified three histopathological growth patterns (HGPs) in HE-stained sections of CRLM: desmoplastic (dHGP), pushing (pHGP), and replacement (rHGP) (*8*). In dHGP, the metastatic lesion is separated from the liver parenchyma by a stromal layer, with tumor cells infiltrating the matrix but not directly contacting hepatocytes. In pHGP, only a thin reticulin fiber layer separates tumor cells from hepatocytes, with metastatic lesions compressing and displacing hepatic plates. Unlike these patterns, rHGP preserves liver architecture, as tumor cells replace hepatocytes within hepatic plates while maintaining direct contact with normal hepatocytes. Notably, pHGP exhibits a higher ratio of proliferative endothelial cells compared to the other two (*9*).

Studies have demonstrated that different HGP patterns have a significant impact on patient prognosis. For instance, pHGP is associated with poorer survival outcomes (*10*), whereas dHGP correlates with better survival (*11,12*). Moreover, HGP classification also aids in developing various clinical strategies. As an example, Lazaris *et al*. demonstrated that bevacizumab is more effective in treating dHGP-type CRLM with abundant angiogenesis compared to rHGP-type CRLM (*13*). However, despite being a valuable classification system, HGPs has several limitations. For example, the growth patterns of tumor tissues may be altered by chemotherapy, and the classification still relies on postoperative histological analysis of tissue sections.

### 2.1.2. MMR/MSI classification

The DNA mismatch repair (MMR) system plays a critical role in correcting base mismatches or insertion/deletion errors that occur during DNA replication, and it was first identified as being associated with the progression of CRC (*14*). Defects in the MMR system led to the microsatellite instability (MSI) phenotype, also known as deficient MMR (dMMR). Based on the MMR/

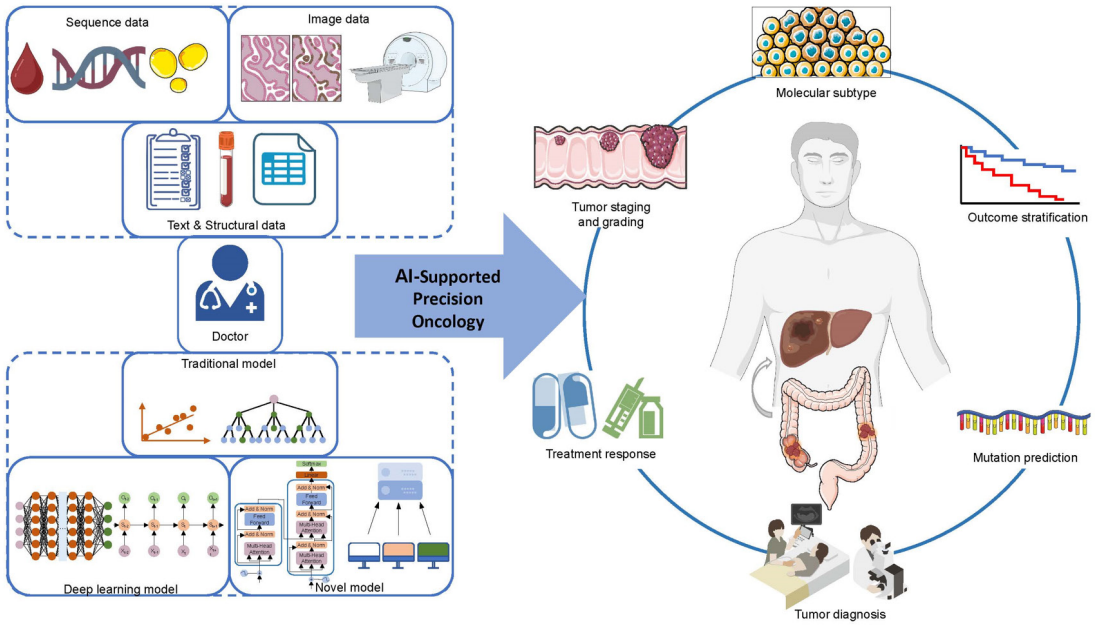


**Figure 1. AI-driven Framework for Precision Oncology in Colorectal Cancer Liver Metastases.** This figure depicts the integration of sequence data, imaging, and structured data with AI models to enhance tasks such as tumor classification, molecular subtyping, mutation prediction, and evaluation of treatment response. Central to this workflow is the use of advanced machine learning and deep learning techniques to facilitate personalized clinical decision-making.

microsatellite status, metastatic CRC patients can be classified into microsatellite stable (MSS, also referred to as proficient MMR, pMMR) and MSI (dMMR). Studies have shown that mCRC patients with dMMR generally have poorer survival outcomes compared to those with pMMR (*15,16*). However, these dMMR patients represent a small subset, accounting for only 3–5% of cases. Recent research has highlighted that monoclonal antibodies targeting immune checkpoints such as programmed cell death protein-1 and cytotoxic T-lymphocyte-associated protein 4 exhibit remarkable and durable benefits in this minority of MSI patients (*17-20*). Additionally, MSI status serves as a predictive biomarker for sensitivity to immune checkpoint blockade (ICB) therapy (*21*). In contrast, ICB therapies have not demonstrated superior efficacy over standard treatments in pMMR mCRC patients, underscoring the need for further exploration in this area (*22*).

2.1.3. Clinical staging (TNM)

The TNM staging system is based on the tumor, lymph node, and metastasis (TNM) concept first proposed by Pierre Denoix in the 1940s and 1950s. It remains the most commonly used staging system in the clinical management of CRLM (*23,24*). This system classifies cancer based on three key parameters: T refers to the size and depth of tumor invasion; N describes the involvement of regional lymph nodes; M indicates the presence of distant metastases (*25*). The TNM system provides a comprehensive framework for evaluating the severity and extent of tumor spread. However, due to its relatively narrow evaluation criteria, recent studies have suggested incorporating additional factors, such as tumor burden and the number of metastatic lesions, to improve the TNM system's prognostic accuracy for CRC patients and to better guide treatment strategies (*26,27*).

2.1.4. Clinical risk score

The clinical risk score has emerged in recent years as one of the most prominent tools for evaluation of colorectal cancer liver metastases (CRLM). A clinical risk score was initially proposed by Fong *et al*. in 1999. This landmark study analyzed clinical and pathological data from 1,001 consecutive patients and identified five key clinical indicators for the scoring system: the nodal status of the primary tumor, a disease-free interval of less than 12 months between the primary tumor and the detection of liver metastases, the presence of more than one tumor, a preoperative CEA level exceeding 200 ng/ml, and a maximum tumor diameter greater than 5 cm. Each criterion is assigned 1 point, and the total score stratifies patients by risk. The aforementioned study demonstrated that patients with lower CRS scores had significantly better 5-year survival rates compared to those with higher scores, and the CRS outperformed other scoring combinations available at the time. However, despite its pivotal role in clinical practice over the years, an increasing number of studies have sought to enhance the predictive power of the CRS system. This is being explored through the incorporation of additional clinical or molecular indicators, as well as by integrating CRS with other classification systems to address its limitations in specific contexts (*28,29*).

2.2. AI models commonly used in tumor classification

The standard workflow in ML research typically involves key steps such as data preprocessing, model construction, model training, parameter optimization, and external validation. Model selection plays a pivotal role in both data analysis and the advancement of research. An appropriate model not only significantly enhances research efficiency but also improves the accuracy and reliability of analytical results. This section systematically describes and discusses ML models, which are categorized into three types: conventional models, deep learning models, and emerging models.

2.2.1. Conventional ML models

*Support vector machines*: Support vector machines (SVM) were first developed by Cortes *et al*. in 1995 as a method for binary classification (*30*). Before the resurgence of deep learning, SVM was one of the most widely used ML models in various domains. The fundamental principle of SVM is to identify an optimal separating hyperplane in the feature space that divides data into distinct classes while maximizing the margin between them. To achieve this, SVM uses kernel methods, utilizing a mapping function to transform input data from its original feature space into a higher-dimensional space, where a separating hyperplane can be more easily identified. The strengths of SVM lie in its ability to handle high-dimensional data, its suitability for small sample sizes, and its excellent generalization capabilities. SVM has been widely applied to tumor subtyping and classification, including applications that incorporate imaging data (*31*) or transcriptomic data (*32*). However, SVM has certain limitations, such as longer training times for large datasets or datasets with high feature dimensions, and its sensitivity to noise and outliers, which can lead to overfitting.

*Random forest*: Random forest (RF), first developed by Breiman in 2001, is an ensemble learning method that solves classification and regression problems by constructing multiple independent decision trees (*33*). In the structure of RF, each internal node of a decision tree represents a feature, while each leaf node corresponds to a classification (category) or regression (numerical) outcome. The final prediction is generated by aggregating the outputs of all trees using a voting mechanism (for classification problems) or an averaging mechanism

(for regression problems). RF is highly robust against overfitting and can be parallelized to efficiently handle large-scale datasets. As a non-parametric approach, it effectively models complex nonlinear relationships and high-dimensional feature data, and it has been extensively applied to tumor subtyping and classification, including studies on breast cancer and pancreatic cancer (*34,35*). However, the ensemble nature of RF reduces its interpretability, and processing large datasets with many decision trees can require substantial time and computational resources.

*Regression models*: Regression models are widely used in statistics and ML to analyze relationships between a dependent variable and one or more independent variables. While regression methods can predict outcomes and explain variable influences, they are primarily statistical tools rather than standalone ML models. The core of regression models lies in finding a function that optimally maps the input values of independent variables to the output values of the dependent variable, typically by minimizing the error between predicted and observed values. Regression models encompass various types, including linear regression, logistic regression, lasso regression, and Cox regression. Cox regression, developed by Cox in 1972 (*36*), and lasso regression, developed by Tibshirani in 1996 (*37*), are the most commonly utilized in oncology research. For instance, Liu *et al*. utilized Cox regression to investigate the relationship between metabolic-associated fatty liver disease (MAFLD) and multiple cancers (*38*). Li *et al*. used a Bayesian lasso model to integrate multi-omics data for lung cancer classification (*39*). Each regression model has its own strengths and limitations. In general, regression models are simple, interpretable, and computationally efficient, with various regularization methods available to enhance their generalizability. However, they are also sensitive to outliers, heavily dependent on data characteristics, and often less effective when dealing with complex high-dimensional datasets.

*Gradient boosting algorithm*: Gradient boosting machine (GBM) is an ensemble learning method developed by Friedman in 2001 (*40*). It iteratively optimizes a target function to achieve the best possible solution by sequentially combining multiple weak learners (typically decision trees) into a strong learner. Each weak learner focuses on correcting the prediction errors of the previous model. GBM demonstrates exceptional performance in handling nonlinear, high-dimensional, and large-scale datasets, effectively capturing complex data patterns while maintaining robustness against noise and outliers. Popular implementations of gradient boosting include XGBoost, LightGBM, and CatBoost. For example, Rodriguez *et al*. used XGBoost combined with imaging and clinical parameters for risk stratification of hepatocellular carcinoma (HCC) patients (*41*), while Qi *et al*. used LightGBM as the optimal algorithm for predicting cardiovascular disease (CVD)-cancer comorbidity (*42*). However, GBM has some limitations, such as a lengthy training time for large datasets and a high dependence on hyperparameter tuning, which often requires extensive optimization to achieve peak performance.

*k-nearest neighbors*: The k-nearest neighbors (kNN) algorithm is an instance-based, non-parametric learning method known for its simplicity and intuitive nature. It predicts outcomes by measuring the similarity between samples. Specifically, kNN calculates the distance between an input sample and all training samples, selects the k nearest neighbors, and infers the target category (for classification problems) or value (for regression problems) based on the labels or values of these neighbors. A defining feature of kNN is its lack of an explicit training phase, as it relies primarily on the stored training data and distance computations. This simplicity makes it easy to implement and adapt. The algorithm has been used in various medical research fields. For instance, Wang *et al*. used kNN for lung cancer subtype classification (*43*), and modified kNN methods have been used to classify CRC tissues (*44*).

#### 2.2.2. Deep learning models

Deep learning models (DLMs) are an extension of artificial neural networks (ANNs) and represent a more advanced and sophisticated branch of ML. Broadly, neural networks can be categorized into shallow neural networks (typically consisting of one or two hidden layers) and deep neural networks (DNNs, generally comprising three or more hidden layers). The latter forms the foundation and most prevalent framework for DLMs.

DLMs excel in non-linear modeling, making them effective for pattern recognition and predictive tasks. These networks act as multi-layer feature extractors, transforming input data (*e.g.* images or text) into abstract representations. Using these features, models interpret and process data for various applications, such as analyzing histopathological slides to distinguish tumor from non-tumor regions.

The following sections will delve into the foundational concepts of DLMs, focusing on deep neural networks as a paradigm of DLMs and their commonly implemented architectures, such as convolutional neural networks and recurrent neural networks.

*Deep neural network*: Deep neural networks (DNNs) are fundamental in deep learning, enabling hierarchical feature extraction and complex pattern recognition. A typical DNN consists of an input layer, multiple hidden layers, and an output layer, with each neuron in a layer connected to all neurons in the previous layer. DNNs operate through two key processes: forward propagation and backpropagation. In forward propagation, input data is passed layer by layer, with each neuron computing a weighted sum of its inputs, followed by a non-linear

activation function. This allows the network to model complex patterns. Backpropagation is the cornerstone of training DNNs. It calculates the error between predicted and true values, propagates it backward, and updates weights using optimization algorithms. This iterative process minimizes errors and refines model performance.

DNNs have displayed exceptional performance in handling highly complex datasets. For instance, Khan *et al*. integrated over 23,000 CT and pathology images to develop a multimodal DNN for predicting metastasis and variant classification of liver tumors, achieving an accuracy of 96.06% and an AUC of 0.832 (*45*). Nevertheless, DNNs have notable limitations. In addition to their high demands for large-scale data and computational resources, they are inherently "black box" models, making their learning processes and decision logic difficult to interpret—an issue that constrains their broader adoption in medicine. Moreover, conventional DNNs may suffer from shallow feature loss when dealing with high-dimensional, non-linearly distributed complex data, often requiring compensation or refinement through the introduction of attention mechanisms or skip connections.

*Convolutional neural networks*: Convolutional neural networks (CNN) are a class of DNNs that incorporate convolutional layers and that are particularly suited for processing data with grid-like topology, such as images and time-series data. Derived from multi-layer perceptrons, CNNs are designed to efficiently capture spatial locality in data by introducing specialized structures like convolutional and pooling layers.

A typical CNN architecture consists of five main components: an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer. The convolutional layer is the backbone of CNNs, utilizing convolutional kernels (or filters, typically small matrices) to slide over the input data and extract local features such as edges, textures, and shapes. Each kernel learns specific feature patterns, with its parameters optimized through backpropagation. The pooling layer reduces the dimensions of the feature maps using down-sampling techniques, thereby decreasing computational complexity and enhancing translational invariance. Finally, the fully connected layer maps the extracted high-dimensional features into specific outputs, such as classification or regression predictions.

Thanks to their powerful feature extraction capabilities, CNNs excel in various computer vision tasks. For instance, Cho *et al*. used deep convolutional neural networks (DCNNs) in conjunction with image data to distinguish between benign and malignant lip skin lesions (*46*). Similarly, Chang *et al*. utilized CNNs combined with self-attention mechanisms to analyze histopathological slides in order to predict MSI status in CRC patients (*47*). However, CNNs also have notable limitations. They require large-scale training datasets and significant computational resources to achieve optimal performance. Moreover, CNNs are sensitive to hyperparameter settings, often necessitating extensive tuning to refine the model for specific applications.

*Recurrent neural networks*: Recurrent neural networks (RNNs) are specifically designed to process sequential data and can trace their origins back to Hopfield Networks of Associative Memory, developed by Hopfield in 1982 (*48*). Unlike conventional feedforward neural networks (*e.g.* DNNs), RNNs possess recurrent connections and memory capabilities, allowing them to retain information across time steps and respond to current inputs in the context of past information. This unique structure makes RNNs particularly suitable for tasks involving temporal dependencies, such as speech recognition, natural language processing, and time-series forecasting.

Despite their advantages, RNNs face challenges such as vanishing and exploding gradients, particularly when processing long sequences. To address these issues, several variants of RNNs have been developed, with the most prominent being the long short-term memory (LSTM) network (*49*). LSTM introduces a gating mechanism that regulates the retention and forgetting of information, overcoming the limitations of conventional RNNs in learning long-term dependencies. LSTM features three core gates—input, forget, and output gates—that collectively govern the flow and storage of information within the hidden states, enabling it to effectively capture long-range dependencies.

RNNs have accelerated advances in oncology research. For example, Yun *et al*. developed a transfer recurrent feature learning framework for intraoperative imaging and diagnosis of epithelial cancers (*50*). Similarly, a study combined CNNs with RNNs to differentiate benign from malignant fibroepithelial breast lesions, achieving promising results (*51*).

#### 2.2.3. Emerging models and learning strategies

With the rapid advancement of AI technologies, an increasing number of emerging models and learning strategies are being applied to tumor classification and other related medical tasks. These approaches place greater emphasis on multimodal data integration, few-shot learning, and model interpretability, addressing the limitations of conventional models while driving innovation in the use of ML in medicine.

*Transformer*: The Transformer is a deep learning architecture based on attention mechanisms, initially designed for natural language processing tasks. With its typically deeper and more sophisticated layer design, the Transformer is categorized as a type of DNN. It processes input text or data sequences by dividing them into segments and using attention scores to determine the weight of each segment in the output module.

Compared to previous models, the Transformer uses multi-head self-attention mechanisms to process

input sequences in parallel, significantly improving computational efficiency and enabling effective modeling of long-range dependencies. The Transformer architecture consists of two main components: the encoder and the decoder. The encoder transforms input data into abstract contextual representations, while the decoder generates target sequences based on these contextual representations.

In tasks such as image segmentation and tumor classification, Transformers demonstrate exceptional performance. For instance, Xin *et al*. developed an improved Transformer model for skin cancer classification, achieving an accuracy exceeding 94% (*52*). Similarly, Xu *et al*. proposed a Transformer-based model, Prov-GigaPath, which not only classified subtypes across multiple cancer types but also identified molecular expression patterns and predicted gene mutations from histopathological slides, outperforming conventional models in various aspects (*53*). Nevertheless, the Transformer architecture faces certain challenges. Its complex and opaque internal mechanisms make the decision-making process difficult to interpret, and its high demand for computational resources remains a significant obstacle to its widespread use in oncology.

*Multimodal ML*: Multimodal ML is an approach that integrates information from different data sources to extract complementary features from multiple types of data simultaneously. Given the diverse data involved in oncology research—such as imaging, genomic, and transcriptomic data—multimodal ML has high compatibility and significant potential for advancing cancer research.

Just like DNNs, multimodal learning does not refer to a specific model but rather demonstrates the variety of data. For instance, Qian *et al*. reported the development of a multimodal model named BMU-Net, which integrates clinical data, mammographic images, and trimodal ultrasound data to diagnose benign and malignant breast tumors, achieving an overall diagnostic accuracy exceeding 90% (*54*). Multimodal ML models are evolving toward more efficient data fusion, improved interpretability, and enhanced clinical applicability. By integrating data from multiple modalities, these models can capture deeper insights that are often unavailable from a single data source, thereby offering more accurate and comprehensive support for tumor classification and diagnosis.

*Self-supervised learning*: In AI, ML approaches can be categorized into supervised learning, unsupervised learning, and reinforcement learning, based on whether the analyzed data includes specific labels or annotations. Self-supervised learning (SSL) is considered an extension of unsupervised learning. Unlike supervised methods, SSL does not require extensive labeled datasets. Instead, its core principle is to construct pretext tasks that enable models to extract meaningful feature representations from unlabeled data for downstream tasks. SSL methods are broadly divided into two main categories: generative methods and contrastive methods. Generative methods train models by reconstructing data, such as completing images or predicting missing words, making them ideal for reconstruction tasks. Contrastive methods, in contrast, use positive and negative sample pairs to help models distinguish similarities and differences.

These methods are particularly effective in image classification and data representation tasks. For example, Schirris *et al*. developed DeepSMILE, a contrastive SSL framework for classifying whole slide images of HE-stained tissue sections (*55*). Similarly, Zhang *et al*. developed SANDI, a model capable of spatial cellular classification, by first learning pairwise similarities among unlabeled data and subsequently incorporating reference data (*56*). While SSL eliminates the dependency on large-scale labeled datasets, developing high-performance SSL models requires carefully designed pretext tasks and significant computational resources.

In conclusion, while ML models face challenges in data dependency and interpretability, AI integration into medicine is a key direction for the future. By combining AI with conventional methodologies, particularly in cancer detection, diagnosis, subtyping, and personalized treatment, AI-driven research is advancing precision medicine and overcoming technical barriers in healthcare.

## 3. Application of ML to CRLM classification

Based on task requirements and technical methodologies, the application of ML to CRC liver metastasis classification can be divided into two major categories: single-modality task classification and multi-modality task classification. The former can be further subdivided into three subcategories: classification based on imaging data, classification based on omics data, and classification based on structured data.

### 3.1. Single-modality task classification

#### 3.1.1. Based on imaging data

ML has been extensively applied to imaging data for CRLM classification, with researchers exploring both conventional ML models and advanced deep learning frameworks. The following studies demonstrate the diversity of approaches and highlight their respective strengths and limitations (onlin data: Table 1, *https://www.biosciencetrends.com/supplementaldata/252*).

Tharmaseelan *et al*. conducted a study using CT imaging data from 78 patients, encompassing 1,296 metastatic liver lesions, to evaluate the performance of various ML models (*57*). These models included conventional ML classifiers such as XGBoost and kNN, as well as a DLM based on CNN. The CNN model was

derived from the DenseNet-121 architecture and trained using the PyTorch platform. The objective was to identify the primary tumor site in gastrointestinal cancer patients with liver metastases. Interestingly, the kNN model achieved the highest discriminative ability (AUC: 0.87), outperforming the CNN model (AUC: 0.80). However, the CNN model demonstrated superior accuracy (0.83 *vs.* 0.67). These findings suggest that conventional ML models may, in certain classification tasks, perform comparably or even better than advanced DLMs, and especially with limited datasets.

Building upon this, Jia *et al*. proposed a DLM based on CT imaging to identify the primary tumor sites in patients with liver metastases (*58*). Their study included imaging data from 489 patients and a total of 769 metastatic liver lesions. To provide a comparative analysis, the researchers also developed conventional ML models, including Decision Tree, RF, and kNN. With a larger sample size, the DLM outperformed conventional ML models in all metrics. Specifically, the DLM achieved an accuracy of 0.714 and an AUC of 0.811 on the validation set, and external validation yielded an accuracy of 0.667 and an AUC of 0.784. In comparison, the best-performing conventional model, RF, achieved a maximum AUC of 0.775 and an accuracy of 0.655. These results highlight the advantages of DLMs in using larger datasets for superior performance.

Moving into histopathological growth pattern analysis, Höppener *et al*. developed a CNN-based model for binary classification of HGPs in liver metastases (desmoplastic *vs.* non-desmoplastic) using digitalized whole-slide images (*59*). Their algorithm, neural image compression (NIC), is a multi-task learning framework that compresses high-dimensional image patches into low-dimensional embeddings while preserving spatial information and suppressing noise. The study used 3,641 slides from 932 patients for training and 870 slides for external validation. The model achieved outstanding results, with an AUC of 0.93 on the training set and 0.95 on the validation set. By using supervised training across multiple histopathological tasks, NIC demonstrated the potential of multi-task learning in extracting transferable features for robust classification.

Similarly, Starmans *et al*. explored the use of CNNs to classify the HGPs of CRLM using CT data (*60*). The study used multi-observer segmentation, combining data from three human observers to train the model. Each lesion was segmented three times, effectively tripling the training sample size. Interestingly, the performance of the multi-observer model (AUC: 0.69) was comparable to the single-observer models (maximum AUC: 0.72). Despite exploring ICC-based feature selection and ComBat for further analysis, these methods did not significantly improve performance. The aforementioned study highlights the challenges of utilizing multi-observer data and suggests the importance of optimizing segmentation techniques for better performance.

Turning to genetic mutation prediction, Wesdorp *et al*. developed models based on RF and gradient boosting algorithms to identify KRAS mutation status using CT imaging data (*61*). The study included 255 CRLM patients, split into training (n = 204) and test (n = 51) sets. While the ensemble classifier performed well on the test set (AUC: 0.86), it underperformed in external validation (AUC: 0.47). In contrast, RF demonstrated relatively better external performance (AUC: 0.54). These results reflect ongoing challenges in linking imaging features to genetic mutations, exacerbated by small sample sizes and insufficient preprocessing.

Similarly, Granata *et al*. utilized CT data to predict RAS mutations (*62*). They extracted 851 radiomic features from 77 liver metastases in 28 patients and constructed multiple ML models, including logistic regression, decision trees, kNN, and SVM. Multivariable analysis using logistic regression achieved superior performance (AUC: 0.953, accuracy: 98%), especially after z-score normalization. However, the authors noted no significant improvements when applying normalization techniques, raising questions about their utility in radiomic analysis.

Finally, Li *et al*. developed a comprehensive platform, the Radiomics Intelligent Analysis Toolkit (RIAT), for predicting liver metastasis risk (*63*). By integrating multiple ML methods and clinical data, RIAT demonstrated the value of combining advanced statistical and ML techniques for robust diagnostic tool development. Similarly, Kim *et al*. applied YOLO-based deep learning to large-scale CT imaging (*64*), achieving sensitivity comparable to radiologists but emphasizing its role as an assistive, rather than standalone, diagnostic tool.

Together, these studies highlight the diversity of imaging-based ML applications in CRLM classification, emphasizing the importance of task-specific adaptations, model optimization, and data integration.

#### 3.1.2. Based on omics data

In the context of CRLM classification, omics data provide a rich source of biological insights, enabling ML models to predict risk, classify subtypes, and identify molecular features with significant diagnostic and prognostic implications. The integration of multi-omics datasets with ML not only offers enhanced classification accuracy but also deepens our understanding of the underlying molecular mechanisms driving CRLM.

Yu *et al*. used the AdaBoost algorithm to predict the risk of liver metastases in CRC patients using blood test markers (*65*). The study compared AdaBoost to five other algorithms, including Extremely Randomized Trees (ERT), Multilayer Perceptron, Stochastic Gradient Descent (SGD), RF, and XGBoost. AdaBoost, which dynamically adjusts sample weights to optimize weak

learners, achieved the highest diagnostic accuracy (89.3%) and precision (89.4%). Interestingly, MLP demonstrated the weakest performance, with an accuracy of 79.6% and a precision of 80.1%. The superior performance of AdaBoost was attributed to its robustness with small datasets, whereas MLP's reliance on larger, high-dimensional data likely limited its effectiveness in that study. This underscores the potential of simpler, adaptive algorithms in data-limited clinical settings.

Extending the analysis to tissue-level investigations, Kiritani *et al*. developed a logistic regression model using mass spectrometry data from 103 CRLM samples and 80 normal tissue samples to distinguish metastatic from non-malignant tissues (*66*). The model underwent validation using leave-one-out cross-validation (LOOCV), 10-fold cross-validation, and an independent cohort of 40 samples (20 CRLM and 20 non-cancerous tissues). Phosphatidylcholine, phosphatidylethanolamine, and monounsaturated fatty acids were found to be enriched in CRLM tissues, with the model achieving an exceptional accuracy of 99.5% and an AUC of 0.9999. These findings highlight the synergy of mass spectrometry and ML in identifying molecular markers for CRLM diagnosis.

Taking a step further into molecular subtyping, Katipally *et al*. utilized data from the Phase 3 new EPOC randomized clinical trial to construct a neural network model for CRLM molecular subtyping (*29*). Sequencing data from 93 patients revealed 31 optimal features, including 24 mRNAs and 7 miRNAs, which were used for subtyping. In a validation cohort of 147 patients, the model classified CRLM into canonical, immune, and stromal subtypes, with immune subtype patients having the best 5-year OS (63%) and canonical subtype patients having the worst prognosis (43%). Incorporating molecular subtypes into clinical risk scores improved predictive performance (OS AUC increased from 0.59 to 0.63). The aforementioned study demonstrates how molecular subtyping can enhance both prognostic stratification and personalized therapeutic strategies.

Finally, Moosavi *et al*. developed an RF-based CRLM classification model using transcriptomic data from 171 patients (*67*). The study compared the new LMS subtyping framework to the CMS and CRIS classification systems, using 829 CRC samples, including CRLM, primary CRC tumors, non-malignant liver tissues, organoids, and cell lines. Unlike CMS, which struggled to classify CRLM and which was influenced by prior systemic treatments, LMS effectively stratified samples into five subtypes (LMS1-5). LMS1 was associated with the poorest prognosis (5-year OS of 15%, HR = 2.2, $p = 9 \times 10^{-4}$), while LMS5 exhibited stromal-like characteristics. LMS demonstrated superior prognostic stratification and independence from treatment-related biases, outperforming CMS and CRIS in this regard.

Together, these studies illustrate the potential of integrating omics data with ML for CRLM classification. From simple blood markers to comprehensive transcriptomic analyses, omics-driven ML approaches offer unparalleled opportunities to provide precision oncology, unravel molecular complexities, and provide robust frameworks for diagnosis and prognosis.

#### 3.1.3. Based on structured data

Building on the success of imaging-based approaches, the application of ML to omics data has opened up new avenues for CRLM classification. By using molecular and biological datasets, these studies aim to glean deeper insights into tumor biology while improving diagnostic accuracy and prognostic predictions.

Nemlander *et al*. developed a gradient boosting model to identify non-metastatic colorectal cancer (NMCRC) patients during their first clinical visit (*68*). The study included 2,681 participants, consisting of 542 NMCRC patients and 2,139 matched controls. Clinical data used for model construction included age, sex, primary healthcare (PHC) unit, NMCRC stage (I-III), the number of general practitioner consultations in the previous year, and all diagnoses reported in VEGA within the preceding year. The dataset contained 360 different ICD-10 or KSH97-P diagnostic codes. Of the participants, 75% were used for training, while 25% were used for validation.

The model was constructed using the GBM package in R, with class-stratified 10-fold cross-validation. The final model correctly identified 99 out of 135 NMCRC cases, achieving a sensitivity of 73.3%, a specificity of 83.5%, and an AUC of 0.832. Among the 361 predictors, 184 variables were found to have predictive value, with 16 factors showing a normalized relative influence (NRI) >1%. Notable predictors included changes in bowel habits, other diseases of the anus and rectum, iron deficiency anemia, and other and unspecified non-infective gastroenteritis and colitis. These findings suggest that such symptoms may indicate an elevated risk of NMCRC.

Although studies utilizing structured data for CRLM classification are relatively scarce, this research highlights the potential of such data to contribute to early cancer detection. Structured data offers a non-invasive, cost-effective means of identifying diagnostic patterns that can complement other ML approaches in precision oncology.

### 3.2. Multimodal task classification

Moving beyond single-modality approaches, multimodal task classification integrates diverse datasets to improve predictive accuracy and uncover complex patterns in CRLM. By combining biological, clinical, and electronic health record (EHR) data, these models provide a comprehensive framework for understanding and

predicting disease progression.

Krishnan *et al*. developed a model using a Bayesian regularized neural network (BRANN) and sparse multilinear regression to classify CRC patients (*69*). The study integrated multiple biological datasets, including plasma lipid and protein levels, chemokines, gene mutation status, and clinical information. Initially, a regression model, MLR-EM, was constructed to extract key feature data, identifying 9 lipids as significant predictors for distinguishing CRLM patients. Using these features, the BRANN model, a variant of ANN with Bayesian regularization, successfully classified cancer-free individuals, CRC patients, and CRLM patients. The model had an $R^2$ of 0.77 and an accuracy of 87% on the training set and an $R^2$ of 0.68 and an accuracy of 77% on the test set. The aforementioned study demonstrates the potential of integrating biochemical and clinical data for accurate classification of disease stages in CRC patients.

Li *et al*. combined EHR information and laboratory data to construct NLP and ML models in order to predict the likelihood of postoperative liver metastases in CRC patients (*70*). The study included 1,463 CRC patients, 609 with CRLM and 854 without. A total of 18 features were analyzed using five conventional models and a bidirectional encoder representations from Transformer (BERT)-based NLP model. Among the conventional models, SVM demonstrated the best performance (AUC: 0.64, accuracy: 0.64), comparable to the NLP model (AUC: 0.676, accuracy: 0.636). When these two approaches were fused into a single model, the combined framework exhibited significantly enhanced performance, achieving an accuracy of 80.8% and precision of 80.3%. Moreover, the combined model outperformed physicians in an external validation cohort in both accuracy (0.760 *vs.* 0.658 and 0.640) and precision (0.763 *vs.* 0.697 and 0.670). These results highlight the potential of combining EHR data with advanced NLP and ML techniques to improve predictive accuracy for CRLM.

These studies underscore the advantages of multimodal approaches in CRLM classification, using complementary datasets to refine predictions and improve patient stratification. By integrating diverse data sources, multimodal models address the limitations of single-modality methods and pave the way for more robust and clinically actionable insights.

## 4. Clinical decision-making and treatment optimization based on cancer classification

### 4.1. Classification-guided personalized treatment

ML models are increasingly being used to guide clinical decision-making and optimize treatment strategies for CRLM patients. These models provide valuable tools for predicting therapeutic responses, stratifying patients, and personalizing treatment approaches. The following studies illustrate how classification results can inform clinical decisions and improve patient outcomes (onlin data: Table 2, *https://www.biosciencetrends.com/supplementaldata/252*).

To begin with, Karagkounis *et al*. developed an RF model to evaluate the pathological response of CRLM patients to chemotherapy (*71*). The study included 85 patients and 95 liver metastases, with 63 lesions classified as responders and 32 as non-responders based on histopathological assessments. To address a data imbalance, the authors implemented cost-sensitive learning by assigning higher penalties for misclassifying non-responders. The model outperformed conventional methods, including RECIST and morphological evaluation, achieving an AUC of 0.87 compared to 0.53 and 0.56, respectively. This demonstrates the potential of ML models to provide more accurate and nuanced assessments of chemotherapy responses.

Building on the use of CT data to predict chemotherapy response, Maaref *et al*. utilized CNNs to predict treatment responses in CRLM patients (*72*). The study included 202 patients with 444 lesions, where 230 had previously undergone FOLFOX-based chemotherapy. The CNN model achieved outstanding performance in distinguishing treated from untreated lesions (AUC: 0.97) and predicting chemotherapy responses (AUC: 0.88, sensitivity: 98.1%). These findings highlight the ability of CNNs to handle large imaging datasets and assist in managing metastatic lesions.

Expanding on the prediction of chemotherapy response, Davis *et al*. used an attention-based deep learning framework to analyze CT images and predict responses to neoadjuvant chemotherapy in CRLM patients (*73*). Using a dataset of 33,619 CT images from 95 patients, the model assigned attention weights to different image patches and achieved an AUC of 0.77, far surpassing the logistic regression model based on Fong scores (AUC: 0.41). These results emphasize the utility of attention mechanisms and multi-instance learning frameworks when analyzing complex imaging data with weak annotations.

Taking the next step toward multi-modal modeling, Qi *et al*. developed an artificial neural network (ANN) model to predict the sensitivity of unresectable CRLM patients to irinotecan-based chemotherapy (*74*). The study included 116 patients, randomly divided into training (n = 81) and validation (n = 35) sets. Feature selection using Pearson correlation and the MRMR algorithm identified key imaging and clinical variables for model construction. The primary ANN model (p-model) integrated multi-scale resampling of imaging features with clinical data, while three variant ANN models used only single-scale inputs. The p-model achieved an AUC of 0.754 (training) and 0.752 (validation), surpassing the best conventional model, XGBoost (AUC: 0.718 and 0.704). Further intra-ANN comparisons confirmed the superiority of multi-modal

integration, with the p-model outperforming single-scale ANN variants. The aforementioned study underscores the value of combining multi-scale imaging and clinical data to enhance chemotherapy response prediction, offering a promising tool for optimizing CRLM treatment strategies.

Focusing on precision medicine, Lu *et al*. developed a hybrid CNN-RNN model to predict VEGF therapy sensitivity in mCRC patients based on the VELOUR trial (*75*). By combining CNN-based feature extraction with RNN-based temporal sequence analysis, the model demonstrated superior performance in predicting early treatment responses (AUC: 0.76) compared to conventional RECIST (AUC: 0.66) and ETS (AUC: 0.60) standards. Moreover, responders identified by the model had a significantly longer median OS (18.0 months *vs.* 10.4 months for non-responders, HR = 0.49, $p = 1\times10^{-6}$). The aforementioned study highlights the potential of combining dynamic imaging data and ML for real-time therapeutic decision-making.

In terms of survival stratification, Endo *et al*. developed a decision-tree-based model to predict postoperative chemotherapy responses in CRLM patients (*76*). The study analyzed data from 1,358 patients, incorporating 18 demographic and clinicopathologic variables, including T stage, primary tumor location, and tumor burden score (TBS). Patients with lymph node metastasis, specific tumor locations, and certain KRAS statuses displayed significant benefits from adjuvant chemotherapy. Subgroup analyses revealed that patients with lymph node metastasis, left-sided or rectal primary tumors with low/high TBS, and right-sided tumors with KRAS mutations benefited significantly from adjuvant chemotherapy. The model demonstrated good predictive performance, with a C-index of 0.68 for OS and 0.69 for RFS in both training and test sets. These findings highlight the utility of incorporating clinicopathologic data into predictive models to guide adjuvant chemotherapy decisions and improve patient outcomes

Shifting focus to imaging data and biological response, Zhu *et al*. developed a ML model using pre- and post-chemotherapy MRI images to predict pathological tumor regression grade (TRG) in CRLM patients (*77*). The study included 180 patients (389 lesions) divided into training, test, and external validation sets. Implemented with TensorFlow and Keras, the model utilized multi-stream inputs and center cropping to enhance CNN performance. Three models with varying input streams were compared: Model A (four input streams), Model B (pre-treatment images), and Model C (post-treatment images). Model A achieved the highest AUC (0.849) with the training set, significantly outperforming Models B, C ($p = 0.04$), and RECIST ($p = 0.03$). In external validation, Model A maintained superior performance (AUC: 0.833, accuracy: 0.885) compared to RECIST (AUC: 0.558, accuracy: 0.533). Additionally, Model A effectively stratified survival outcomes, while RECIST-defined groups displayed no significant differences (DFS and OS, $p = 0.12$, $p = 0.99$). The aforementioned study underscores the potential of CNN-based models in improving chemotherapy response prediction and survival stratification over conventional RECIST assessments.

Giannini *et al*. utilized imaging data to predict treatment responses in HER2-amplified CRLM patients receiving HER2 dual-targeted therapy (*78*). The study included CT data from 38 patients and 141 metastatic lesions, with 28 patients (108 lesions) in the training set and 10 patients (33 lesions) in the validation set. The authors extracted 24 radiomic features from CT images and applied a Gaussian Naïve Bayes (GNB) classifier for feature selection, ultimately retaining 12 significant features. The GNB model performed better on the training set compared to the validation set, particularly in sensitivity (training: a sensitivity of 0.89, a specificity of 0.85; validation: a sensitivity of 0.90, a specificity of 0.42). The model correctly classified 24 of the 38 patients, partially misclassified 12, and completely misclassified 2. The authors noted that while the model effectively predicted responsive lesions (R+), it struggled to accurately identify non-responsive lesions (R-). The aforementioned study underscores the potential of radiomic feature-based models to predict treatment response in HER2-targeted therapies, while highlighting challenges in generalizability and specificity.

Together, these studies demonstrate the pivotal role of ML models in guiding clinical decisions and optimizing treatment strategies for CRLM. By improving the accuracy of therapeutic response predictions, stratifying patients based on clinical and molecular characteristics, and integrating multi-modal data, these models are driving precision oncology forward.

### 4.2. Patient prognostic stratification

ML has significantly enhanced prognostic stratification for CRLM patients, utilizing diverse data types to improve survival predictions and patient management. From imaging-based models to multi-modal approaches, these studies illustrate the versatility of ML in addressing clinical challenges.

Wang *et al*. developed an unsupervised ML model based on preoperative CT imaging and clinical data to stratify survival risks in 197 CRLM patients (*79*). Using hierarchical clustering, the study filtered imaging features from 851 to 56 through Cox regression and divided patients into favorable and poor prognosis groups, with the latter exhibiting an OS HR of 1.78 (95% CI: 1.12–2.83). The model outperformed CRS and TBS scores in predicting long-term survival, with a time-dependent AUC of 0.66 compared to 0.58 and 0.55, respectively.

Building on this, Paro *et al*. used a tumor burden-focused ML model, ML-TB, to optimize thresholds for tumor size and number, maximizing five-year survival

stratification (*80*). The study analyzed 1,344 patients from five centers and noted superior OS stratification compared to conventional Fong scores, with Cohen's d values of 1.61, 0.84, and 2.73, highlighting the model's ability to redefine tumor burden parameters for better clinical outcomes.

In a similar vein, Lam *et al*. incorporated lasso regression and Cox models to identify key predictors from 36 clinical variables in 572 patients (*81*). Variables such as CEA levels, tumor size, and KRAS mutation status were critical for OS and RFS predictions, achieving a concordance index of 0.651 and significantly outperforming Fong CRS in one- and five-year OS predictions. This comprehensive analysis underscores the importance of integrating clinicopathologic and molecular data into ML models for precise risk stratification.

Adding a histopathological dimension, Elforaici *et al*. used deep learning frameworks with GANs and Vision Transformers to analyze 1,620 pathology slides from 258 patients (*82*). The model extracted tumor and peritumoral features, achieving a c-index of 0.804 for OS and 0.735 for time-to-recurrence. By using multi-task deep learning, this approach demonstrated the potential to enhance prognostic precision through advanced histological insights.

Moro *et al*. utilized a classification and regression tree (CART) model to identify risk factors for CRLM prognosis in 1,123 patients (*83*). Based on demographic and clinicopathologic data, the model revealed distinct survival profiles for wtKRAS and mtKRAS patients. For instance, wtKRAS patients with small (<4.3 cm) solitary metastases and no nodal involvement exhibited the highest five-year OS (68.5%). The CART model also outperformed conventional Fong scores, and particularly for wtKRAS patients (AIC 3334 *vs.* 3341).

Incorporating imaging and molecular characteristics, Saber *et al*. utilized an attention-based TabNet model to predict levels of CD73 expression in 122 patients (*84*). By integrating immunofluorescence and CT data, the model achieved an AUC of 0.95 and yielded significant prognostic implications, with high levels of CD73 expression linked to shorter recurrence (13.0 *vs.* 23.6 months, p: 0.0098) and disease-specific survival (53.4 *vs.* 126.0 months, p: 0.0222). The aforementioned study emphasizes the role of molecular markers in stratifying treatment responses and outcomes.

Expanding the focus to targeted therapy, Zhou *et al*. developed the DERBY+ model to predict bevacizumab response using PET-CT and clinical data (*85*). Trained on multi-center cohorts, the model achieved an AUC of 0.95 with independent datasets, outperforming individual predictors such as clinical (AUC: 0.66) and imaging features (AUC: 0.72). The identified responders exhibited prolonged OS (27.6 *vs.* 18.5 months, p = 0.010), underscoring the utility of integrated ML frameworks for precision oncology.

Turning to recurrence prediction, Zhao *et al*. designed a hybrid DLM combining 2D-CNN, Bi-LSTM, and attention modules to predict early recurrence after thermal ablation (*86*). Analyzing 13,248 ultrasound images and clinical data from 207 patients, the combined model achieved an AUC of 0.78 and demonstrated significant prognostic stratification. Notably, the DL model consistently outperformed clinical models in all datasets, with significantly lower false-positive rates and better high-risk group identification ($p < 0.001$).

In the realm of disease-free survival prediction, Luo *et al*. compared elastic net (EN) and random survival forest (RSF) models using contrast-enhanced CT imaging data from 180 patients (*87*). The EN model outperformed RSF in the test set (C-index = 0.78), while RSF excelled in the training set (C-index = 0.74). Both models effectively stratified DFS outcomes, illustrating the complementarity of regression- and forest-based approaches in survival analysis.

Finally, Amygdalos *et al*. developed a gradient-boosted decision tree model to predict OS in 487 CRLM patients (*88*). By focusing on six top-ranked predictors, such as CEA levels and metastatic lesion size, the GBDT-Top6 model achieved a superior C-index of 0.70, outperforming the original GBDT (C-index: 0.65). This highlights the potential of feature selection in enhancing ML model performance and clinical interpretability.

Together, these studies underscore the transformative potential of ML in CRLM prognostic stratification. By integrating diverse data sources and using cutting-edge algorithms, these models will pave the way for more personalized and effective patient care.

In conclusion, ML has advanced clinical decision-making and prognostic stratification for CRLM patients by integrating clinical, imaging, and molecular data. Techniques such as RFs, regression trees, and deep learning have demonstrated effectiveness in predicting chemotherapy responses, stratifying survival risks, and enhancing prognostic accuracy. These advances highlight AI's potential to optimize personalized treatment and improve patient outcomes in CRLM management.

## 4. Conclusion

AI has shown great promise in classifying and managing CRLM, yet challenges remain in its clinical integration. The complexity of multimodal data, limited access to large annotated datasets, and ethical concerns such as data privacy and model transparency hinder their widespread use. Additionally, CRLM's biological heterogeneity requires AI models that are both adaptive and interpretable.

To overcome these barriers, future research should focus on federated learning to enable secure multi-institution collaboration, self-supervised and transfer learning to reduce dependence on labeled data, and improved model interpretability to enhance clinical

trust. Longitudinal studies integrating AI into real-world workflows will be essential for validation.

Despite these challenges, AI is transforming CRLM management by integrating clinical, imaging, and omics data for personalized treatment strategies. Advancing AI-driven solutions through interdisciplinary collaboration will further enhance precision medicine, optimizing outcomes for CRLM patients.

*Funding*: This work was supported by the Shanghai Natural Science Foundation Project (22ZR1413300), the National Natural Science Foundation of China (81874056, 81874182), the National Key Research and Development Plan of the Ministry of Science and Technology (2022YFE0125300), and the Public Health Bureau Foundation of Shanghai (202240240, 201940043).

*Conflict of Interest*: The authors have no conflicts of interest to disclose.

**References**

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71:209-249.
2. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, Diaz LA Jr. Immunotherapy in colorectal cancer: Rationale, challenges and potential. Nat Rev Gastroenterol Hepatol. 2019; 16:361-375.
3. Chen Q, Chen J, Deng Y, Bi X, Zhao J, Zhou J, Huang Z, Cai J, Xing B, Li Y, Li K, Zhao H. Personalized prediction of postoperative complication and survival among colorectal liver metastases patients receiving simultaneous resection using machine learning approaches: A multi-center study. Cancer Lett. 2024; 593:216967.
4. Rompianesi G, Pegoraro F, Ceresa CD, Montalti R, Troisi RI. Artificial intelligence in the diagnosis and management of colorectal cancer liver metastases. World J Gastroenterol. 2022; 28:108-122.
5. Wu X, Li W, Tu H. Big data and artificial intelligence in cancer research. Trends Cancer. 2024; 10:147-160.
6. Joshi RM, Telang B, Soni G, Khalife A. Overview of perspectives on cancer, newer therapies, and future directions. Oncol Transl Med. 2024; 10:105-109.
7. Yu C, Shi Z, Zhou G, Chang X. Revisiting the survival paradox between stage IIB/C and IIIA colon cancer. Sci Rep. 2024; 14:22133.
8. Vermeulen PB, Colpaert C, Salgado R, Royers R, Hellemans H, Van Den Heuvel E, Goovaerts G, Dirix LY, Van Marck E. Liver metastases from colorectal adenocarcinomas grow in three patterns with different angiogenesis and desmoplasia. J Pathol. 2001; 195:336-342.
9. van Dam PJ, van der Stok EP, Teuwen LA, *et al*. International consensus guidelines for scoring the histopathological growth patterns of liver metastasis. Br J Cancer. 2017; 117:1427-1441.
10. Falcão D, Alexandrino H, Caetano Oliveira R, Martins J, Ferreira L, Martins R, Serôdio M, Martins M, Tralhão JG, Cipriano MA, Castro E Sousa F. Histopathologic patterns as markers of prognosis in patients undergoing hepatectomy for colorectal cancer liver metastases - Pushing growth as an independent risk factor for decreased survival. Eur J Surg Oncol. 2018; 44:1212-1219.
11. Galjart B, Nierop PMH, van der Stok EP, van den Braak RRJC, Höppener DJ, Daelemans S, Dirix LY, Verhoef C, Vermeulen PB, Grünhagen DJ. Angiogenic desmoplastic histopathological growth pattern as a prognostic marker of good outcome in patients with colorectal liver metastases. Angiogenesis. 2019; 22:355-368.
12. Takasu C, Morine Y, Yoshikawa K, Tokunaga T, Nishi M, Kashihara H, Wada Y, Yoshimoto T, Shimada M. Impact of pure desmoplastic histological growth patterns in colorectal liver metastasis. BMC Cancer. 2024; 24:1528.
13. Lazaris A, Amri A, Petrillo SK, Zoroquiain P, Ibrahim N, Salman A, Gao ZH, Vermeulen PB, Metrakos P. Vascularization of colorectal carcinoma liver metastasis: Insight into stratification of patients for anti-angiogenic therapies. J Pathol Clin Res. 2018; 4:184-192.
14. Cunningham JM, Kim CY, Christensen ER, Tester DJ, Parc Y, Burgart LJ, Halling KC, McDonnell SK, Schaid DJ, Walsh Vockley C, Kubly V, Nelson H, Michels VV, Thibodeau SN. The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas. Am J Hum Genet. 2001; 69:780-790.
15. Venderbosch S, Nagtegaal ID, Maughan TS, *et al*. Mismatch repair status and BRAF mutation status in metastatic colorectal cancer patients: A pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies. Clin Cancer Res. 2014; 20:5322-5330.
16. Tran B, Kopetz S, Tie J, Gibbs P, Jiang ZQ, Lieu CH, Agarwal A, Maru DM, Sieber O, Desai J. Impact of BRAF mutation and microsatellite instability on the pattern of metastatic spread and prognosis in metastatic colorectal cancer. Cancer. 2011; 117:4623-4632.
17. Overman MJ, McDermott R, Leach JL, *et al*. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): An open-label, multicentre, phase 2 study. Lancet Oncol. 2017; 18:1182-1191.
18. Overman MJ, Lonardi S, Wong KYM, *et al*. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/microsatellite instability-high metastatic colorectal cancer. J Clin Oncol. 2018; 36:773-779.
19. André T, Shiu KK, Kim TW, *et al*. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. N Engl J Med. 2020; 383:2207-2218.
20. Bi F, Dong J, Jin C, *et al*. Iparomlimab (QL1604) in patients with microsatellite instability-high (MSI-H) or mismatch repair-deficient (dMMR) unresectable or metastatic solid tumors: A pivotal, single-arm, multicenter, phase II trial. J Hematol Oncol. 2024; 17:109.
21. Le DT, Uram JN, Wang H, *et al*. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med. 2015; 372:2509-2520.
22. Kawazoe A, Xu RH, García-Alfonso P, *et al*. Lenvatinib plus pembrolizumab versus standard of care for previously treated metastatic colorectal cancer: Final analysis of the randomized, open-label, Phase III LEAP-017 study. J Clin Oncol. 2024; 42:2918-2927.
23. AJCC Cancer Staging Handbook. *https://link.springer.*

*com/book/9780387884424* (accessed February 1, 2025)
24. Sobin LH, Gospodarowicz MK, Wittekind C, eds. TNM Classification of Malignant Tumours. 8th ed. Wiley-Blackwell, Oxford, UK, 2017.
25. Weiser MR. AJCC 8th Edition: Colorectal Cancer. Ann Surg Oncol. 2018; 25:1454-1455.
26. Jin X, Wu Y, Feng Y, Lin Z, Zhang N, Yu B, Mao A, Zhang T, Zhu W, Wang L. A population-based predictive model identifying optimal candidates for primary and metastasis resection in patients with colorectal cancer with liver metastatic. Front Oncol. 2022; 12:899659.
27. Takeda Y, Mise Y, Matsumura M, Hasegawa K, Yoshimoto J, Imamura H, Noro T, Yamamoto J, Ishizuka N, Inoue Y, Ito H, Takahashi Y, Saiura A. Accuracy of modern clinical risk score including RAS status changes based on whether patients received perioperative chemotherapy for colorectal liver metastases. World J Surg. 2021; 45:2176-2184.
28. Brudvik KW, Jones RP, Giuliante F, *et al*. RAS mutation clinical risk score to predict survival after resection of colorectal liver metastases. Ann Surg. 2019; 269:120-126.
29. Katipally RR, Martinez CA, Pugh SA, Bridgewater JA, Primrose JN, Domingo E, Maughan TS, Talamonti MS, Posner MC, Weichselbaum RR, Pitroda SP; with the S:CORT Consortium. Integrated cinical-molecular classification of colorectal liver metastases: A biomarker analysis of the Phase 3 new EPOC randomized clinical trial. JAMA Oncol. 2023; 9:1245-1254.
30. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995; 20:273-297.
31. Sun T, Wang J, Li X, Lv P, Liu F, Luo Y, Gao Q, Zhu H, Guo X. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. Comput Methods Programs Biomed. 2013; 111:519-524.
32. Chu F, Wang L. Applications of support vector machines to cancer classification with microarray data. Int J Neural Syst. 2005; 15:475-484.
33. Breiman L. Random forests. Mach Learn. 2001; 45:5-32.
34. Zare A, Postovit LM, Githaka JM. Robust inflammatory breast cancer gene signature using nonparametric random forest analysis. Breast Cancer Res. 2021; 23:92.
35. Verschuur AVD, Hackeng WM, Westerbeke F, *et al*. DNA methylation profiling enables accurate classification of nonductal primary pancreatic neoplasms. Clin Gastroenterol Hepatol. 2024; 22:1245-1254.e10.
36. Cox D. Regression models and life-tables. J R Stat Soc Ser B-Stat Methodol. 1972; 34:187-220.
37. Tibshirani R. Regression shrinkage and selection *via* the lasso. J R Stat Soc Ser B (Methodol). 1996; 58:267-288.
38. Liu Z, Lin C, Suo C, Zhao R, Jin L, Zhang T, Chen X. Metabolic dysfunction-associated fatty liver disease and the risk of 24 specific cancers. Metabolism. 2022; 127:154955.
39. Li J, Wang Y, Song X, Xiao H. Adaptive multinomial regression with overlapping groups for multi-class classification of lung cancer. Comput Biol Med. 2018; 100:1-9.
40. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat. 2001; 29:1189-1232.
41. Rodriguez LA, Schmittdiel JA, Liu L, Macdonald BA, Balasubramanian S, Chai KP, Seo SI, Mukhtar N, Levin TR, Saxena V. Hepatocellular carcinoma in metabolic dysfunction-associated steatotic liver disease. JAMA Netw Open. 2024; 7:e2421019.
42. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. Redox Biol. 2024; 79:103470.
43. Wang C, Long Y, Li W, Dai W, Xie S, Liu Y, Zhang Y, Liu M, Tian Y, Li Q, Duan Y. Exploratory study on classification of lung cancer subtypes through a combined K-nearest neighbor classifier in breathomics. Sci Rep. 2020; 10:5880.
44. Li Q, Hao C, Kang X, Zhang J, Sun X, Wang W, Zeng H. Colorectal cancer and colitis diagnosis using Fourier transform infrared spectroscopy and an improved k-nearest-neighbour classifier. Sensors (Basel). 2017; 17:2739.
45. Khan RA, Fu M, Burbridge B, Luo Y, Wu FX. A multi-modal deep neural network for multi-class liver cancer diagnosis. Neural Netw. 2023; 165:553-561.
46. Cho SI, Sun S, Mun JH, Kim C, Kim SY, Cho S, Youn SW, Kim HC, Chung JH. Dermatologist-level classification of malignant lip diseases using a deep convolutional neural network. Br J Dermatol. 2020; 182:1388-1394.
47. Chang X, Wang J, Zhang G, Yang M, Xi Y, Xi C, Chen G, Nie X, Meng B, Quan X. Predicting colorectal cancer microsatellite instability with a self-attention-enabled convolutional neural network. Cell Rep Med. 2023; 4:100914.
48. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A. 1982; 79:2554-2558.
49. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997; 9:1735-1780.
50. Yun G, Vyas K, Yang J, Yang GZ. Transfer recurrent feature learning for endomicroscopy image recognition. IEEE Trans Med Imaging. 2019; 38:791-801.
51. Cheng CL, Md Nasir ND, Ng GJZ, *et al*. Artificial intelligence modelling in differentiating core biopsies of fibroadenoma from phyllodes tumor. Lab Invest. 2022; 102:245-252.
52. Xin C, Liu Z, Zhao K, Miao L, Ma Y, Zhu X, Zhou Q, Wang S, Li L, Yang F, Xu S, Chen H. An improved transformer network for skin cancer classification. Comput Biol Med. 2022; 149:105939.
53. Xu H, Usuyama N, Bagga J, *et al*. A whole-slide foundation model for digital pathology from real-world data. Nature. 2024; 630:181-188.
54. Qian X, Pei J, Han C, *et al*. A multimodal machine learning model for the stratification of breast cancer risk. Nat Biomed Eng. 2025; 9:356-370.
55. Schirris Y, Gavves E, Nederlof I, Horlings HM, Teuwen J. DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. Med Image Anal. 2022; 79:102464.
56. Zhang H, AbdulJabbar K, Grunewald T, *et al*. Self-supervised deep learning for highly efficient spatial immunophenotyping. EBioMedicine. 2023; 95:104769.
57. Tharmaseelan H, Vellala AK, Hertel A, Tollens F, Rotkopf LT, Rink J, Woźnicki P, Ayx I, Bartling S, Nörenberg D, Schoenberg SO, Froelich MF. Tumor classification of gastrointestinal liver metastases using CT-based radiomics and deep learning. Cancer Imaging. 2023; 23:95.
58. Jia W, Li F, Cui Y, Wang Y, Dai Z, Yan Q, Liu X, Li Y, Chang H, Zeng Q. Deep learning radiomics model of contrast-enhanced CT for differentiating the primary source of liver metastases. Acad Radiol. 2024; 31:4057-

4067.
59. Höppener DJ, Aswolinskiy W, Qian Z, Tellez D, Nierop PMH, Starmans M, Nagtegaal ID, Doukas M, de Wilt JHW, Grünhagen DJ, van der Laak JAWM, Vermeulen P, Ciompi F, Verhoef C. Classifying histopathological growth patterns for resected colorectal liver metastasis with a deep learning analysis. BJS Open. 2024; 8:zrae127.
60. Starmans MPA, Buisman FE, Renckens M, Willemssen FEJA, van der Voort SR, Groot Koerkamp B, Grünhagen DJ, Niessen WJ, Vermeulen PB, Verhoef C, Visser JJ, Klein S. Distinguishing pure histopathological growth patterns of colorectal liver metastases on CT using deep learning and radiomics: A pilot study. Clin Exp Metastasis. 2021; 38:483-494.
61. Wesdorp N, Zeeuw M, van der Meulen D, *et al*. Identifying genetic mutation status in patients with colorectal cancer liver metastases using radiomics-based machine-learning models. Cancers (Basel). 2023; 15:5648.
62. Granata V, Fusco R, Setola SV, Brunese MC, Di Mauro A, Avallone A, Ottaiano A, Normanno N, Petrillo A, Izzo F. Machine learning and radiomics analysis by computed tomography in colorectal liver metastases patients for RAS mutational status prediction. Radiol Med. 2024; 129:957-966.
63. Li M, Li X, Guo Y, Miao Z, Liu X, Guo S, Zhang H. Development and assessment of an individualized nomogram to predict colorectal cancer liver metastases. Quant Imaging Med Surg. 2020; 10:397-414.
64. Kim K, Kim S, Han K, Bae H, Shin J, Lim JS. Diagnostic performance of deep learning-based lesion detection algorithm in CT for detecting hepatic metastasis from colorectal cancer. Korean J Radiol. 2021; 22:912-921.
65. Yu Z, Li G, Xu W. Rapid detection of liver metastasis risk in colorectal cancer patients through blood test indicators. Front Oncol. 2024; 14:1460136.
66. Kiritani S, Yoshimura K, Arita J, Kokudo T, Hakoda H, Tanimoto M, Ishizawa T, Akamatsu N, Kaneko J, Takeda S, Hasegawa K. A new rapid diagnostic system with ambient mass spectrometry and machine learning for colorectal liver metastasis. BMC Cancer. 2021; 21:262.
67. Moosavi SH, Eide PW, Eilertsen IA, Brunsell TH, Berg KCG, Røsok BI, Brudvik KW, Bjørnbeth BA, Guren MG, Nesbakken A, Lothe RA, Sveen A. De novo transcriptomic subtyping of colorectal cancer liver metastases in the context of tumor heterogeneity. Genome Med. 2021; 13:143.
68. Nemlander E, Ewing M, Abedi E, Hasselström J, Sjövall A, Carlsson AC, Rosenblad A. A machine learning tool for identifying non-metastatic colorectal cancer in primary care. Eur J Cancer. 2023; 182:100-106.
69. Krishnan ST, Winkler D, Creek D, Anderson D, Kirana C, Maddern GJ, Fenix K, Hauben E, Rudd D, Voelcker NH. Staging of colorectal cancer using lipid biomarkers and machine learning. Metabolomics. 2023; 19:84.
70. Li J, Wang X, Cai L, Sun J, Yang Z, Liu W, Wang Z, Lv H. An interpretable deep learning framework for predicting liver metastases in postoperative colorectal cancer patients using natural language processing and clinical data integration. Cancer Med. 2023; 12:19337-19351.
71. 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. Ann Surg Oncol. 2024; 31:9196-9204.
72. Maaref A, Romero FP, Montagnon E, Cerny M, Nguyen B, Vandenbroucke F, Soucy G, Turcotte S, Tang A, Kadoury S. Predicting the response to FOLFOX-based chemotherapy regimen from untreated liver metastases on baseline CT: A deep neural network approach. J Digit Imaging. 2020; 33:937-945.
73. Davis JMK, Niazi MKK, Ricker AB, Tavolara TE, Robinson JN, Annanurov B, Smith K, Mantha R, Hwang J, Shrestha R, Iannitti DA, Martinie JB, Baker EH, Gurcan MN, Vrochides D. Predicting response to neoadjuvant chemotherapy for colorectal liver metastasis using deep learning on prechemotherapy cross-sectional imaging. J Surg Oncol. 2024; 130:93-101.
74. Qi W, Yang J, Zheng L, Lu Y, Liu R, Ju Y, Niu T, Wang D. CT-based radiomics for the identification of colorectal cancer liver metastases sensitive to first-line irinotecan-based chemotherapy. Med Phys. 2023; 50:2705-2714.
75. Lu L, Dercle L, Zhao B, Schwartz LH. Deep learning for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging. Nat Commun. 2021; 12:6654.
76. Endo Y, Alaimo L, Moazzam Z, *et al*. Optimal policy tree to assist in adjuvant therapy decision-making after resection of colorectal liver metastases. Surgery. 2024; 175:645-653.
77. Zhu HB, Xu D, Ye M, Sun L, Zhang XY, Li XT, Nie P, Xing BC, Sun YS. Deep learning-assisted magnetic resonance imaging prediction of tumor response to chemotherapy in patients with colorectal liver metastases. Int J Cancer. 2021; 148:1717-1730.
78. Giannini V, Rosati S, Defeudis A, *et al*. Radiomics predicts response of individual HER2-amplified colorectal cancer liver metastases in patients treated with HER2-targeted therapy. Int J Cancer. 2020; 147:3215-3223.
79. Wang Q, Nilsson H, Xu K, Wei X, Chen D, Zhao D, Hu X, Wang A, Bai G. Exploring tumor heterogeneity in colorectal liver metastases by imaging: Unsupervised machine learning of preoperative CT radiomics features for prognostic stratification. Eur J Radiol. 2024; 175:111459.
80. Paro A, Hyer MJ, Tsilimigras DI, Guglielmi A, Ruzzenente A, Alexandrescu S, Poultsides G, Aucejo F, Cloyd JM, Pawlik TM. Machine learning approach to stratifying prognosis relative to tumor burden after resection of colorectal liver metastases: An international cohort analysis. J Am Coll Surg. 2022; 234:504-513.
81. Lam CSN, Bharwani AA, Chan EHY, Chan VHY, Au HLH, Ho MK, Rashed S, Kwong BMH, Fang W, Ma KW, Lo CM, Cheung TT. A machine learning model for colorectal liver metastasis post-hepatectomy prognostications. Hepatobiliary Surg Nutr. 2023; 12:495-506.
82. Elforaici MEA, Montagnon E, Romero FP, Le WT, Azzi F, Trudel D, Nguyen B, Turcotte S, Tang A, Kadoury S. Semi-supervised ViT knowledge distillation network with style transfer normalization for colorectal liver metastases survival prediction. Med Image Anal. 2025; 99:103346.
83. Moro A, Mehta R, Tsilimigras DI, Sahara K, Paredes AZ, Bagante F, Guglielmi A, Alexandrescu S, Poultsides GA, Sasaki K, Aucejo FN, Pawlik TM. Prognostic factors differ according to KRAS mutational status: A classification and regression tree model to define prognostic groups after hepatectomy for colorectal liver metastasis. Surgery. 2020; 168:497-503.
84. Saber R, Henault D, Messaoudi N, Rebolledo R,

Montagnon E, Soucy G, Stagg J, Tang A, Turcotte S, Kadoury S. Radiomics using computed tomography to predict CD73 expression and prognosis of colorectal cancer liver metastases. J Transl Med. 2023; 21:507.
85. Zhou S, Sun D, Mao W, Liu Y, Cen W, Ye L, Liang F, Xu J, Shi H, Ji Y, Wang L, Chang W. Deep radiomics-based fusion model for prediction of bevacizumab treatment response and outcome in patients with colorectal cancer liver metastases: A multicentre cohort study. EClinicalMedicine. 2023; 65:102271.
86. Zhao QX, He XL, Wang K, Cheng ZG, Han ZY, Liu FY, Yu XL, Hui Z, Yu J, Chao A, Liang P. Deep learning model based on contrast-enhanced ultrasound for predicting early recurrence after thermal ablation of colorectal cancer liver metastasis. Eur Radiol. 2023; 33:1895-1905.
87. Luo X, Deng H, Xie F, Wang L, Liang J, Zhu X, Li T, Tang X, Liang W, Xiang Z, He J. Prognostication of colorectal cancer liver metastasis by CE-based radiomics and machine learning. Transl Oncol. 2024; 47:101997.
88. Amygdalos I, Müller-Franzes G, Bednarsch J, Czigany Z, Ulmer TF, Bruners P, Kuhl C, Neumann UP, Truhn D, Lang SA. Novel machine learning algorithm can identify patients at risk of poor overall survival following curative resection for colorectal liver metastases. J Hepatobiliary Pancreat Sci. 2023; 30:602-614.

Received February 2, 2025; Revised March 30, 2025; Accepted April 12, 2025.

[§]These authors contributed equally to this work.
*Address correspondence to:
Lu Wang and Weiping Zhu, Department of Hepatic Surgery, Fudan University Shanghai Cancer Center, Shanghai Medical College, Fudan University, Shanghai 200032, China.
E-mail: wangluzl@fudan.edu.cn (LW), wpzhush@hotmail.com (WZ)

Released online in J-STAGE as advance publication April 15, 2025.