Table 1. Summary of Studies on AI-Driven Classification Models for CRLM

Reference	Application	DL method	Target	Training cohort	Validation mode	Validation cohort	Best performance			Description
							AUC	Accuracy	Other Available Metrics	
Tharmaseelan *et al.* (*57*)	CT	CNN (DenseNet-121), DT, Gaussian Naive Bayes, kNN, K-SVM, LR, RF, SVM, XGBoost	Identifying primary tumor source	43 patients	Internal	4 patients for all models except CNN; 5 patients in validation cohort and 8 patients in test cohort for CNN	DT= 0.43 DenseNet-121= 0.80 Gaussian Naive Bayes= 0.52 kNN= 0.87 K-SVM= 0.79 LR= 0.66 RF= 0.62 SVM= 0.79 XGBoost= 0.71	DT= 0.36 DenseNet-121= 0.83 Gaussian Naive Bayes= 0.52 kNN= 0.67 K-SVM= 0.72 LR= 0.60 RF= 0.84 SVM= 0.72 XGBoost= 0.84	--	To identify the primary tumor site using CT images
Jia *et al.* (*58*)	CT	DLM, DT, RF, kNN	Identifying primary tumor source	460 lesions	Internal, external	197 lesions in internal validation cohort and 112 lesions (61 patient) in external validation cohort	DLM= 0.811 DT= 0.762 kNN= 0.755 RF= 0.775	DLM= 0.714 DT= 0.613 kNN- 0.622 RF= 0.655	--	To identify the primary tumor site using CT images
Höppener *et al.* (*59*)	Whole-slide images	CNN (Neural Image Compression, NIC)	Tumor HGP classification	932 patients (3,641 images)	External	294 patients (870 images)	NIC= 0.95 (95% CI: 0.93-0.96)	NIC= 89%	Sensitivity: NIC= 87% Specificity: NIC= 91%	To classify tumor HGP mode
Li *et al.* (*63*)	CT	Radiomics Intelligent Analysis Toolkit, RIAT (compiling RF, GBDT, SVM, LR, MLP, stacking classifier)	Classifying CRLM from CRC	80 patients	Internal	20 patients	RF= 0.848 (95% CI: 0.640-1.000) GBDT= 0.848 (95% CI:0.640-1.000) LR= 0.899 (95% CI: 0.761-1.000), MLP= 0.768 (95% :CI: 0.542-0.994) Stacking classifier= 0.843 (95% CI: 0.674-1.000) SVM= 0.778 (95% CI: 0.515-1.000)	--	Sensitivity: RF= 0.78 GBDT= 0.78 LR= 0.78 MLP= 0.78 Stacking classifier= 0.78 SVM= 0.78 Specificity: RF= 0.91 GBDT= 0.91 LR= 0.82 MLP= 0.82 Stacking classifier= 0.91 SVM= 0.91	To predict liver metastases from CRC patient
Starmans *et al.* (*60*)	CT	CNN	Tumor HGP classification	61 patients (74 lesions)	Internal	15 patients (19 lesions)	CNN= 0.70 (95% CI:0.59-0.81)	CNN= 0.65 (95% CI:0.54-0.76)	Sensitivity: CNN= 0.71 (95% CI:0.57-0.86) Specificity: CNN= 0.67 (95% CI:0.51-0.83)	To classify tumor HGP mode
Kim *et al.* (*64*)	CT	DLM (Deep learning-based lesion detection algorithm, DLLD)	Classifying CRLM from CRC	502 patients (4,386 images)	Internal	85 patients (99 lesions)	Other indexes as primary variables: AUAFROC DLLD= 0.631 (95% CI: 0.520-0.737)	--	Sensitivity: DLLD= 81.82% (95% CI: 72.68%-88.39%) Specificity: DLLD= 22.22% (95% CI: 12.40-36.60)	To predict liver metastases from CRC patient
Wesdorp *et al.* (*61*)	CT	RF, GBM, GBM+LightGBM, Ensemble (RF + GBM + GBM+LightGBM)	Identifying tumor genetic mutation status	204 patients	Internal, external	51 patients in test cohort and 129 patients in validation cohort	Ensemble= 0.86 (95% CI: 0.76-0.95) RF= 0.77 (95% CI: 0.62-0.93) GBM= 0.77 (95% CI: 0.64-0.90) GBM+LightGBM= 0.72 (95% CI: 0.57-0.87)	Ensemble= 0.77 RF= 0.80 GBM= 0.73 GBM+LightGBM= 0.71	Sensitivity: Ensemble= 0.93 RF= 0.85 GBM= 0.67 GBM+LightGBM= 0.75 Specificity: Ensemble= 0.74 RF= 0.75 GBM= 0.79 GBM+LightGBM= 0.67	To identify tumor genetic mutation status

Granata *et al.* (*62*)	CT	DT, kNN, LRM, NLRM, SVM	Identifying tumor genetic mutation status	54 images	Internal	15 images in test cohort and 8 images in validation cohort	LRM= 0.953 NLRM= 0.955	LRM= 0.978 NLRM= 0.968	Sensitivity: LRM= 1.000 NLRM= 1.000 Specificity: LRM= 1.000 NLRM= 0.957	To identify tumor RAS mutation status
Yu *et al.* (*65*)	Blood	AdaBoost, ERT, MLP, RF, SGD, XGBoost	Classifying CRLM from CRC	400 patients	Internal	100 patients	Other indexes as primary variables: Accuracy	AdaBooost= 89.3% ± 3.5% ERT = 84.2% ± 3.5% MLP= 79.6% ± 1.9% RF=89.3% ± 5.0% SGD= 80.4% ± 4.3% XGBoost= 86.7% ± 4.2%	Precision: AdaBooost= 89.4% ± 3.5% ERT = 84.7% ± 3.1% MLP= 80.1% ± 1.7% RF=89.2% ± 5.0% SGD= 83.0% ± 2.0% XGBoost= 86.8% ± 4.3%	To predict liver metastases from CRC patient
Kiritani *et al.* (*66*)	Sequence data (MS)	LR	Classifying CRLM from normal tissue	183 samples	Internal	40 samples	LR= 0.999	LR= 99.5%	Sensitivity: LR= 90.0% Specificity: LR= 100&	To diagnose CRLM from non-cancerous tissue
Katipally *et al.* (*29*)	Sequence data (mRNA and miRNA)	ANN	Molecular subtyping	93 patients	External	147 patients	Other indexes as primary variables: Accuracy	ANN= 96%	--	To subtype tumor tissue
Moosavi *et al.* (*67*)	Sequence data (mRNA)	RF	Molecular subtyping	171 patients (283 samples)	External	308 samples	Other indexes as primary variables: Cohen's κ RF= 0.98 (95% CI: 0.95-1)	RF= 0.9763 (95% CI: 0.9405-0.9935)	--	To subtype tumor tissue
Nemlander *et al.* (*68*)	Text data (EHR)	SGB	Classifying non-CRLM from patient	2,013 patients	Internal	668 patients	SGB= 0.832 (95% CI: 0.790-0.874)	--	Sensitivity: SGB= 77.3% Specificity: SGB= 83.5%	To identify non-metastatic CRC patient in primary care
Krishnan *et al.* (*69*)	Multimodal (clinical data, laboratory data, sequence data)	ANN (BRANN)	Classifying CRLM from CRC	66 samples	Internal	59 samples	Other indexes as primary variables: :R^2 BRANN= 0.68	BRANN= 77%	--	To classify CRC and identify CRLM
Li *et al.* (*70*)	Multimodal (clinical data, laboratory data, pathological results)	DT, Extra tree, kNN, NLP, Nomogram, RF, SVM,	Predicting CRLM from CRC	1,463 patients in total	--	1,463 patients in total	DT= 0.579 (95% CI: 0.522-0.637) Extra tree= 0.582 (95% CI: 0.515-0.649) kNN= 0.551 (95% CI: 0.486-0.617) NLP= 0.887 (95% CI: 0.823-0.931) Nomogram= 0.782 RF= 0.620 (95% CI: 0.555-0.684) SVM= 0.640 (95% CI: 0.574-0.706)	DT= 0.589 Extra tree= 0.613 kNN= 0.558 NLP= 0.808 Nomogram= 0.760 RF= 0.613 SVM= 0.640	Precision: DT= 1.000 Extra tree= 0.776 kNN= 0.131 NLP= 0.803 Nomogram 0.763 RF= 0.447 SVM= 0.620	To predict liver metastases from postoperative CRC patient

AUAFROC, area under the alternative free-response receiver operating characteristic curve; AUC, area under the curve; ANN, artificial neurol network; CNN, convolutional neural networks; DLM, deep learning model; DT, decision tree; EHR, electronic health record; ERT, extremely randomized tree; GBDT, gradient boosting decision tree; GBM, gradient boosting machine; HGP, histopathological growth patterns; LR, logistic regression; RF, random forest; kNN, k-nearest neighbor; LRM, linear regression model; MLP, multi-layer perceptron; MS, mass spectrometry; NLP, natural language processing; NLRM, non-logistic regression model; SGB, stochastic gradient boosting; SGD, stochastic gradient descent; SVM, support vector machine; XGBoost, extreme gradient boosting.

*All metrics recorded reflect performance on validation and/or test datasets, excluding training set results.

Table 2. Summary of Studies on AI-Assisted Prognostic Stratification and Treatment Optimization for CRLM

Reference	**Application**	**DL method**	**Target**	**Training cohort**	**Validation mode**	**Validation cohort**	**Best performance**			**Description**
							AUC	**Accuracy**	**Other Available Metrics**	
Davis *et al.* (*73*)	CT	MIL	Predicting response to neoadjuvant chemotherapy	71 patients	Internal	24 patients	MIL= 0.77	--	Sensitivity: MIL= 81% Specificity: MIL= 58%	To predict patient pathologic response
Maaref *et al.* (*72*)	CT	ANN, CNN, DT, SVM	Classifying treatment history and predicting response to neoadjuvant chemotherapy	311 samples	Internal	44 samples in validation cohort and 89 samples in test cohort	Classification of treatment status ANN= 0.62 CNN= 0.97 DT= 0.66 SVM= 0.60 Prediction of chemotherapy response CNN= 0.88	Classification of treatment status ANN= 0.60 CNN= 0.91 (95% CI: 0.88-0.93) DT= 0.67 SVM= 0.59 Prediction of chemotherapy response CNN= 0.766 (95% CI: 0.71-0.82)	Sensitivity: Classification of treatment status CNN= 90% (95% CI: 86%-93%) Prediction of chemotherapy response CNN= 98% (95% CI: 96%-99%) Specificity: Classification of treatment status CNN= 91% (95% CI: 85%-94%) Prediction of chemotherapy response CNN= 54% (95% CI: 50%-60%)	To predict patient response to FOLFOX-based chemotherapy
Lu *et al.* (*75*)	Multimodal (clinical data, CT images)	RNN	Predicting response to anti-VEGF treatment	301 patients	Internal	201 patients in tuning cohort and 526 patients in test cohort	RNN= 0.76 (95% CI: 0.72-0.80) Other major index used: Harrell C-Index RNN= 0.694 (95% CI: 0.661-0.720)	--	Sensitivity: RNN= 60.6% Specificity: RNN= 78.6%	To predict patient response to anti-VEGF therapy
Endo *et al.* (*76*)	Clinical data	DT	Predicting response to adjuvant chemotherapy	679 patients	External	679 patients	Other indexes as primary variables: Harrell C-Index DT= 0.69	--	--	To predict patient response to adjuvant chemotherapy
Qi *et al.* (*74*)	Multimodal (clinical data, CT images)	AdaBoost, ANN, GBDT, LR, RF, RBF-SVM, XGBoost	Predicting response to chemotherapy	81 patients	Internal	35 patients	AdaBoost= 0.591 (95% CI: 0.338-0.760) ANN= 0.752 (95% CI: 0.581-0.904) GBDT= 0.734 (95% CI:0.520-0.927) LR= 0.692 (95% CI:0.531-0.883) RF= 0.682 (95% CI:0.493-0.936) RBF-SVM= 0.657 (95% CI: 0.476-0.866) XGBoost= 0.710 (95% CI:0.484-0.889) ANN subtype: Model$_{4channel}$= 0.752 (95% CI: 0.581-0.904) Model$_{3channel}$= 0.684 (95% CI: 0.529-0.890) Model$_{Clinical}$= 0.545 (95% CI: 0.360-0.785)	--	--	To predict patient response to first-line irinotecan-based chemotherapy
Zhu *et al.* (*77*)	MRI	CNN (DC3CNN)	Predicting response to chemotherapy	101 patients	Internal, external	54 patients in test cohort and 25 patients in external cohort	DC3CNN= 0.833 (95% CI: 0.695-1.000)	DC3CNN= 0.885 (95% CI: 0.805-0.965)	Sensitivity: DC3CNN= 0.918 (95% CI: 0.842-0.995) Specificity: DC3CNN= 0.75 (95% CI: 0.505-0.995)	To predict patient response to chemotherapy
Giannini *et al.* (*78*)	CT	GNB	Predicting response to targeted therapy	28 patients (108 lesions)	Internal	10 patients (33 lesions)	Other indexes as primary variables: NPV/PPV GNB= 73% (95% CI: 62%-82%) (NPV) GNB= 71% (95% CI: 36%92%) (PPV)	--	Sensitivity: GNB= 90% (95% CI: 70%-99%) Specificity: GNB= 42% (95% CI: 15%-72%)	To predict patient response to HER-2 targeted therapy
Wang *et al.* (*79*)	CT	Unsupervised clustering	Prognosis stratification	158 patients	Internal	39 patients	Unsupervised clustering= 0.66 (95% CI: 0.59-0.73)	--	--	To stratify patient prognosis from preoperative CT images

Paro *et al.* (*80*)	Multimodal (clinical data, laboratory data, surgery records)	SHC (ML-TB)	Prognosis stratification	672 patients	External	672 patients	Other indexes as primary variables: Cohen's d ML-TB= 2.73	--	--	To stratify patient postoperative prognosis based on tumor burden and other data
Lam *et al.* (*81*)	Multimodal (clinical data, laboratory data, pathological results, surgery records, PET-CT images)	CR	Prognosis stratification	400 patients	External	172 patients	Other indexes as primary variables: Harrell C-Index CR= 0.651 (95% CI: 0.537-0.765)	--	--	To stratify patient postoperative prognosis based on Multimodal data
Elforaici *et al.* (*82*)	Whole-slide images	DNN + GAN + Transformer (SSL)	Prognosis prediction, Tissue classification and TRG prediction	180 patients	Internal, external	39 patients for both validation cohort and test cohort	Prognosis prediction: Other indexes as primary variables: Harrell C-Index SSL= 0.804 (OS); 0.735 (TTR) Tissue classification: Other indexes as primary variables: loU SSL= 0.885 TRG prediction-other index used: Accuracy	Prognosis prediction: NA Tissue classification: SSL= 0.912 (cancer) TRG prediction: SSL= 90.3%	--	To predict patient prognosis and TRG status ,and classify tissue
Moro *et al.* (*83*)	Multiomics (clinical data, laboratory data, surgery records)	SRT (CART)	Prognosis prediction	1,123 patients in total	Internal	1,123 patients in total	Other indexes as primary variables: AIC SRT= 3334 (wtKRAS) SRT= 1356 (mtKRAS)	Brier Score: SRT= 0.1741 (wtKRAS) SRT= 0.1793 (mtKRAS)	--	To predict patient postoperative prognosis
Saber *et al.* (*84*)	Multimodal (pathological data, CT images)	LR, RF, SVM, Transformer (TabNet, rad-CD73), XGBoost	CD73 and prognosis prediction	125 lesions	Internal	35 lesions	CD73 prediction: LR= 0.51 (95% CI: 0.34-0.68) RF= 0.59 (95% CI: 0.43-0.75) SVM= 0.60 (95% CI: 0.44-0.76) TabNet= 0.79 (95% CI: 0.65-0.92) XGBoost= 0.61 (95% CI: 0.45-0.77) Prognosis prediction-Other indexes as primary variables: HR rad-CD73= 1.88 (95% CI: 1.11-3.18) (DFS) rad-CD73= 2.11 (95% CI: 1.30-3.45) (TTR)	CD73 prediction: LR= 0.46 RF= 0.63 SVM= 0.49 TabNet= 0.71 XGBoost= 0.60 prognosis prediction-NA	Sensitivity: CD73 prediction: LR= 0.63 RF= 0.75 SVM= 0.94 TabNet= 0.63 XGBoost= 0.69 prognosis prediction-NA Specificity: CD73 prediction: LR= 0.32 RF= 0.53 SVM= 0.11 TabNet= 0.79 XGBoost= 0.53 prognosis prediction-NA	To predict CD73 expression status and prognosis
Zhou *et al.* (*85*)	Multimodal (clinical data, pathological results, PET-CT images)	RF(DERBY, DERBY+)	Bevacizumab response prediction and prognosis stratification	103 patients	Internal, external	65 patients in internal validation cohort, 82 patients in histology cohort, 102 patients in external validation cohort and 37 patients in negative validation cohort	Bevacizumab response prediction: DERBY= 0.84 (95% CI: 0.74-0.93) DERBY+= 0.95 (95% CI: 0.91-1.00) prognosis stratification-Other indexes as primary variables: HR DERBY+= 0.50 (95% CI: 0.29-0.86) (OS) DERBY+= 0.52 (95% CI: 0.35-0.79) (PFS)	--	Sensitivity: Bevacizumab response prediction: DERBY= 0.68 (95% CI: 0.54-0.80) DERBY+= 0.96 (95% CI: 0.85-0.99) prognosis stratification-NA Specificity: Bevacizumab response prediction: DERBY= 0.84 (95% CI: 0.66-0.94) DERBY+= 0.85 (95% CI: 0.68-0.94) prognosis stratification-NA	To predict patient response to bevacizumab treatment and stratify prognosis

Zhao *et al.* (*86*)	Multimodal (clinical data, CEUS images)	2D-CNN + AM + Bi-LSTM + MLP (DL model, DL-C model)	Recurrence prediction	153 patients	Internal, external	32 patients in internal test cohort and 22 patients in external test cohort	DL-model= 0.76 (95% CI: 0.71-0.81) DL-C model= 0.82 (95% CI: 0.77-0.86)	DL-model= 0.79 (95% CI: 0.75-0.83) DL-C model= 0.83 (95% CI: 0.77-0.86)	Precision: DL-model= 0.8 (95% CI: 0.76-0.85) DL-C model= 0.86 (95% CI: 0.83-0.92) Sensitivity: DL-model= 0.91 (95% CI: 0.85-0.94) DL-C model= 0.88 (95% CI: 0.74-0.92) Specificity: DL-model= 0.71 (95% CI: 0.55-0.85) DL-C model= 0.74 (95% CI: 0.68-0.80)	To predict CRLM early recurrence after thermal ablation
Luo *et al.* (*87*)	CT	EN, RSF	Prognosis prediction	129 patients	Internal	51 patients	Other indexes as primary variables: Harrell C-Index EN= 0.78 RSF= 0.77	--	--	To predict patient postoperative prognosis
Amygdalos *et al.* (*88*)	Multimodal (clinical data, laboratory data, surgery records)	GBDT, GBDT-Top6	Prognosis stratification	389 patients	Internal	98 patients	Other indexes as primary variables: Harrell C-Index GBDT= 0.65 GBDT-Top6= 0.70	--	--	To predict patient postoperative prognosis

AIC, akaike's information criterion; AM, attention model; ANN, artificial neural network; CEUS, contrast-enhanced ultrasound; CNN, convolutional neurol network; CR, cox regression; DFS, disease-specific survival; DNN, deep neural network; DL, deep learning; DT, decision tree; EN, elastic net; GAN, generative adversarial network; GBDT, gradient-boosted decision tree; GNB, Gaussian Naïve Bayes; RF, random forest; RFS, recurrence free survival; RSF, random survival forest; LR, logistic regression; MIL, multiple-instance learning; SHC, stochastic hill-climbing; SRG, survival regression tree; SRT, survival regression tree; SSL, semi-supervised learning; SVM, support vector machine; TRG, tumor regression grade; TTR, time to recurrence.

*All metrics recorded reflect performance on validation and/or test datasets, excluding training set results.