Supplementary Table 1 Studies examining the upper gastrointestinal tract

Reference	Author (year)	Disease	Diagnostic Modality	Objective	Analytic model	Validation	Accuracy/AUC	Sensitivity/Specificity
13	de Groof *et al.* (2019)	BE	EGD, WLI	Detect early BE neoplasia	Supervised learning techniques	Internal and external	92.0%	95.0% / 85.0%
12	de Groof *et al.* (2020)	BE	EGD	Detect early BE neoplasia	Hybrid ResNet-UNet model	Internal and external	88.0%-89.0% (2 test datasets)	90-93% / 83-88% (2 test datasets)
10	Hashimoto *et al.* (2020)	BE	EGD, WL, NBI	Detect early BE neoplasia	CNN (Inception-ResNet-v2) + data augmentation	Internal	95.4%	96.4/94.2 (Sn WLI 98.6, NBI 92.4, Sn standard focus 96.6, near focus 96.2) NBI Sp WLI 99.2, WLI 88.8, near focus 98.4, standard focus 89.9.
11	van der Sommen *et al.* (2016)	BE	EGD, WLI	Detect early BE neoplasia	SVM. specific texture, color filters, and machine learning.	Internal and external	N/A	Per lesion: 83.0% / 83.0% Per patient: 86.0% / 87.0%
16	Swager *et al.* (2017)	BE	EGD, VLE	Detect early BE neoplasia on ex-vivo VLE	n/a	Internal	0.950	90.0% / 93.0%
17	Trindade *et al.*. (2019)	BE	EGD, VLE	Detect dysplasia in BE	IRIS	Internal	N/A	N/A
14	Ebigbo *et al.* (2020)	BE, EAC	EGD WL, NBI	Differentiation between BE and early EAC	ResNet based DCNN (based on DeepLab V.3+) with 101 layers	Internal	89.9%	83.7% / 100
15	Riaz *et al.* (2013)	BE, EC	EGD (NBI)	Detect and classify BE into normal, pre-cancer and cancer	SVM	Internal and external	91.8%	91.8% / 92.1%
29	Liu *et al.* (2020)	EC	EGD WLI	Classify and distinguish EAC from premalignant lesions	CNN Inception-ResNet (O-stream and P-stream) with data augmentation	Internal	85.8%	94.2% / 94.7%
30	Liu *et al.* (2016)	EC, GC	EGD	1. Detect early esophageal cancer. 2. Detect early gastric cancer	Joint diagonalization principal component analysis (JDPCA)	Internal	1. 90.8% 2. 90.8%	1. 93.3% / 89.2% 2. 90.8% / 90.7%
31	Nakagawa *et al.* (2019)	ESCC	EGD, non-ME, ME	Differentiate between SM1 and SM2/3 ESCC lesions	BP-CNN with 16 layers and Caffe framework.	Internal	91.0%	90.1% / 95.8%
32	Shimamoto *et al.* (2020)	ESCC	EGD, WLI, BLI/NBI, non-ME and ME	Detect invasion depth in ESCC	BP-CNN with 16 layers and PyTorch framework.	Internal	non-ME: 87.3%. ME: 89.2%	non-ME: 50.0% / 98.7% ME: 70.8% / 94.9%
20	Cai *et al.* (2019)	ESCC	EGD WL	Detect ESCC under WL imaging	CNN with 8-layers trained with	Internal and external	91.4%	97.8% / 85.4%

					augmented training data			
21	Fukuda *et al.* (2020)	ESCC	EGD, non-ME and ME NBI	Detect suspicious lesions and characterize cancer (ESCC) vs non-cancer under NBI	BP-CNN with 16 layers	Internal	Detection: 63.0%. Characterization: 88.0%	Detection: 91.0% / 51.0% Characterization: 86.0% / 89.0%
24	Kumagai *et al.* (2019)	ESCC	EGD, ECS	Investigate whether biopsy based ESCC histology can be replaced by ECS.	GoogLeNet	Internal	90.9% (AUC 85)	92.6% /89.3%
25	Li *et al.* (2021)	ESCC	EGD, non-ME NBI	Detect early ESCC, and compare diagnosis with 20 endoscopists (validation	CAD NBI	Internal and external	95.3% (AUC 97.61)	91.0% / 96.7%
26	Ohmori *et al.* (2020)	ESCC	1. EGD (non-ME with WLI) 2. EGD (non-ME with NBI/BLI) 3. EGD (ME with NBI/BLI)	Detect ESCC	CNN	Internal and external	1. 81.0% 2. 77.0% 3. 77.0%	1. 90.0% / 76.0% 2. 100% / 63.0% 3. 98.0% / 56.0%
27	Tan *et al.* (2021)	ESCC	EGD, HRME	Detection of ESCC by algorithm and endoscopists, and improvement of endoscopists with algorithm	Fully automated algorithm	Internal and external	79.4%	76.3% / 85.3%
28	Zhao *et al.* (2019)	ESCC	EGD, NBI-ME	Classification of IPCLs to improve detection of ESCC.	double-labeling FCN image segmentation and multitask learning. VGG16 net	Internal and external	89.2%	87.0% / 84.1%
23	Horie *et al.* (2019)	ESCC, EAC	EGD, WL, NBI, non-ME	Detect ESCC and EAC	CNN, 16 or more layers, Caffe framework.	Internal	99.0% for superficial cancer and 92.0% for advanced cancer	98.0% / 79.0%
22	Guo *et al.* (2020)	ESCC, precancerous lesions	EGD, non-ME and ME NBI	Develop real-time automated diagnosis of precancerous lesions and early ESCC in non-ME and ME setting	CAD DNN. SegNet lesion segmentation.	Internal and external	0.989	Image: (dataset A) 98.04/ (dataset B)95.03. Dataset C: Video per frame non-ME: Sn 60.8. per-lesion non-ME: Sn 100. Video per frame ME: 96.1. per lesion ME 100. Dataset D: Full range video per-frame Sp 99.9. per case Sp 90.9.

55	Chen *et al.* (2019)	Gastric cancer	Genetics	identify long non-coding RNA signatures able to classify microsatellite instability and create a predictive model for MSI	SVM	10-fold cross-validation	0.95	N/A
53	Gao *et al.* (2019)	Gastric cancer	CT	Diagnosis of metastatic LN in gastric cancer	FR-CNN	N/A	AUC: 0.8995	N/A
37	Guimaraes *et al.* (2020)	Atrophic gastritis	EGD	Detect atrophic gastritis	CNN	10-fold cross-validation	93	100.0% / 87.5%
38	Hirasawa *et al.* (2018)	Gastric cancer	EGD (WL, CE, NBI) white light. Chromoendoscopy Narrow-band imaging	Detect gastric cancer (early or advanced)	CNN	N/A	N/A	92.2% / N/A
39	Ishioka *et al.* (2019)	Gastric cancer	EGD (ESD)	Detect gastric cancer (early or advanced)	CNN	N/A	94.10%	N/A
54	Jagric *et al.* (2010)	Gastric cancer	Biomarkers, imaging, tumor size, histology, TNM, Lymph nodes	Prediction of liver metastasis	QNN	N/A	N/A	71.0% / 96.1%
57	Jiang *et al.* (2018)	Gastric cancer	Immunohistochemistry	Predict survival, predict treatment benefit	SVM	Unspecified internal/external	AUC for OS, DFS in training 0.796, 0.805, internal validation 0.809, 0.813 and external validation 0.834, 0.828 cohorts. This compared to the TNM's training (0.649, 0.659), internal (0.746, 0.678) and external (0.745, 0.737)	
50	Kanesaka *et al.* (2018)	Gastric cancer	EGD NBI + ME	Detect gastric cancer (early or advanced)	SVM	N/A	96.30%	96.7% / 95.0%
40	Korhani Kangi *et al.* (2018)	Gastric cancer	Medical record	Predict survival	ANN, Bayesian NN	N/A	ANN 89.1% (0.944), BNN: 93.5% (0.961)	ANN: 88.2% / 90.3%, BNN: 95.4% / 90.9%
41	Li *et al.* (2020)	Gastric cancer	EGD (ME-NBI) magnified narrow band imaging	Detect EGC	CNN	Internal	90.9%	91.2% / 90.6%
30	Liu *et al.* (2016)	Esophageal	EGD	1. Detect early	Joint diagonalization	10-fold	1. 90.8%	1. 93.3% / 89.2%

		and gastric cancer		esophageal cancer. 2. Detect early gastric cancer	principal component analysis (JDPCA)	cross-validation	2. 90.8%	2. 90.8% / 90.7%
62	Martin *et al.* (2020)	H Pylori	Gastric biopsies	Identification of HP on histopathology	CNN	Internal	99.1% (AUC: 1.000)	95.7% / 100%
42	Miyaki *et al.* (2015)	Gastric cancer	EGD (BLI)	Detect gastric cancer (early or advanced)	SVM	Internal	SVM output for cancerous lesions: 1.453e-17)	N/A
63	Nakashima *et al.* (2020)	H Pylori	EGD	Detect infection and prior infection		Internal	82.5% for current infection, 79.2% for prior infection	N/A
58	Que *et al.* (2019)	Gastric cancer	Biomarkers, medical record	Predict survival	ANN	5-fold cross validation	75.2%	86.5% / 43.8%
61	Shichijo *et al*. (2019)	H Pylori	EGD images	compare diagnostic ability of CNN vs endoscopists	CNN (GoogLeNet)	Internal	87.7%	88.9% / 87.4%
43	Togo *et al.* (2019)	Gastritis	Barium XR	Detect Gastritis	CNN	5-fold cross validation	N/A	96.2% / 98.3%
44	Wang *et al.* (2019)	Gastric cancer	Pathology slides	Detect gastric cancer (early or advanced)	Recalibrated multi-instance DL (RMDL)	N/A	86.5%	N/A
52	Wu *et al.* (2019)	Gastric cancer	EGD	Detect early gastric cancer (EGC)	CNN	5-fold cross-validation and early stopping	92.5%	94.0% / 91.0%
49	Zhang *et al.* (2020)	Atrophic gastritis	EGD	Detect and classify chronic atrophic gastritis	CNN	5-fold cross-validation	94.2%	94.5% / 94.0%
65	Zheng *et al.* (2019)	H Pylori	EGD	H pylori detection	CNN	Internal	84.5% (AUC: 0.93)	81.4% / 90.1%
59	Zhu *et al.* (2019)	Gastric cancer	EGD	Predict gastric cancer depth	CNN-CAD system	Internal	AUC: 0.94	76.5% / 95.6%
56	Nakahira *et al.* (2020)	Gastric cancer	EGD	Gastric cancer risk stratification	CNN	Internal	Kappa 0.27 (fair interobserver agreement among endoscopists)	N/A
45	Luo *et al.* (2019)	Gastric cancer	EGD	Detection of gastric cancer (automatic)	GRAIDS	Internal and external	97.7%	94.2% / 92.3%
51	Sakai *et al.* (2019)	Gastric cancer	EGD	Detection of gastric cancer (automatic)	CNN (GoogLeNet)	Internal	87.6%	80.0% / 94.8%
46	Namikawa *et al.* (2020)	Gastric cancer	EGD	Gastric cancer classification	CNN	Internal	95.9-100%	99.0% / 93.3%
47	Ueyama *et al.* (2020)	Gastric cancer	EGD (NBI)	Diagnosis of GC	CNN (ResNet50)	Internal	98.7%	98.0% / 100%

48	Zhou *et al.* (2021)	Gastric cancer	Medical record	Gastric cancer recurrence	Five models	5-fold cross validation	Highest algorithm: Logistic (80.1%)	N/A
69	Shung *et al.* (2020)	GIB	Health records	Develop prognostic score and compare to GBS, Rockall and AIMS65	Gradient-boosting model	Internal/external	AUC: 0.88 for GBS, 0.73 for Rockall and 0.78 for AIMS65	100% / 26.0%
68	Seo *et al.* (2020)	GIB	Health records	Algorithm that predicts adverse events in non-variceal UGIB	Random forest classifier	Internal	Mortality AUC of 0.917 vs 0.710 of GBS, VC model was best for hypotension (AUC:0.757 vs GBS: 0.668) and rebleeding (AUC 0.733 vs GBS: 0.694)	N/A
67	Wong *et al.* (2019)	GIB	Health records	Identify patients at high risk for recurrent ulcer bleeding at 1 year in patients with idiopathic PUD bleed	IPU-ML	Internal	Accuracy: 84.3%, AUC: 0.775	41.4% / 74.6%
71	Das *et al.* (2008)	GIB	Health records	Non-endoscopic triage of UGIB compared to Rockall	ANN	Internal/external	77-89% for stigmata, 61-81% for need of endoscopic therapy	SRH: 89.0%-96.0%, Endoscopic therapy 81.0-94.0%; Specificity: 63-89%, 48-82%
70	Das *et al.* (2003)	GIB	Health records	Compare predictive score to BLEED score	ANN	Internal/external	97.0%, 93.0%, 94.0% for mortality, recurrent bleed and endoscopic reintervention	For mortality, recurrent bleed and endoscopic reintervention: Sensitivity: 87.5%, 80.0%, 89.0% Specificity: 97.0%, 95.0%, 95.0%, respectively
72	Loftus *et al.* (2017)	GIB	Health records	Compare if ANN can outperform Strate rule to predict severe GIB and predict surgical intervention	ANN	Internal	AUC: 0.954	N/A
73	Ayaru *et al.* (2015)	GIB	Health records	Prediction of LGIB outcomes	Gradient-boosting model	Internal/external	88.0%, 91.0% and 83.0% for recurrent bleeding, therapeutic	N/A

							reintervention and severe bleeding, respectively	

BE: Barrett's esophagus; EGD: Esophagogastroduodenoscopy; WLI: White light imaging; NBI: Narrow band imaging; SVM: Support vector machine; VLE: Volumetric laser endomicroscopy; EAC: Esophageal adenocarcinoma; EC: Esophageal cancer; CNN: Convoluted neural network; GC: Gastric cancer; ESCC: Esophageal squamous cell carcinoma; JDPCA: Joint diagonalization principal component analysis; ME: Magnification endoscopy; WL: White light; AUC: Area under the curve; ECS: Endocytoscopy systems; DNN: Deep neural network; CE: Chromoendoscopy; ESD: Endoscopic submucosal dissection; ANN: Artificial neural network; GBS: Glasgow-Blatchford score.

Supplementary Table 2 Studies examining the lower gastrointestinal tract

Reference	Author (year)	Disease	Diagnostic Modality	Objective	Analytic model	Validation	Accuracy/AUC	Sensitivity/Spe
101	Kudo *et al.* (2020)	CRC	Colonoscopy (CE, NBI)	Detect CRC	EndoBRAIN	Internal/External	98.0% (CE mode) 96.0% (NBI mode)	96.9% / 100 mode) 96.9% / 94.3 mode)
99	Echle *et al.* (2020)	CRC	Histopathology (H&E)	Detect CRC by detecting MSI and dMMR		Internal/External	AUC: 0.96	95.0% / 67
100	Ito *et al.* (2019)	CRC	Colonoscopy (WL)	Detect deeply invasive (cT1b) CRC	CNN	internal	Accuracy: 81.2%, AUC: 0.871	67.5% / 89
119	Mori *et al.* (2020)	Polyps	Colonoscopy	Estimation of cost reduction	Compared diagnose-leave to resect all	Internal	N/A	93.3% / 95
73	Ayaru *et al.* (2015)	LGIB	EHR	Outcome prediction	Gradient booster, MLR	Internal + External	88%, 91%, 83% for recurrent bleed, therapeutic intervention and severe bleed, respectively;	57.0% / 91.0 recurrent ble 60.0% / 92.0 therapeu intervention, 89.0% for se bleedin
121	Rechling *et al.* (2020)	Cancer	Histopathology	Outcome prediction	LASSO algorithm-based DGMate score	Internal	AUC: 0.56	N/A
123	Skrede *et al.* (2020)	Cancer	Histopathology	Outcome prediction	CNN	Internal + External	Uncertain and poor prognosis group: 76.0%; Good and uncertain to poor prognosis group: 67.0%	Uncertain an prognosis g 52.0% / 78 Good and un to poor prog group: 69.0%
109	Gross *et al.* (2011)	Polyps	Colonoscopy (NBI+Mag)	Polyp Classification	Computer-based algorithm	Internal	93.1% vs 92.7 human experts	95.0% / 90.3% 93.4% / 91 Experts
113	Mori *et al.* (2015)	Polyps	Colonoscopy (endocytoscopy)	Polyp Classification	CAD	Internal	89.2%	92.0% / 79
114	Sanchez-Montes *et al.* (2019)	Polyps	Colonoscopy (HD WLI)	Polyp Classification	SVM	Internal	91.1%	92.3% / 89
115	Tischendorf *et al.*. (2010)	Polyps	Colonoscopy (NBI Mag)	Polyp Classification	SVM	Internal	86.6% (vs 90.9% human)	96.9% / 53.1 96.9% / 71
88	Bernal *et al.* (2017)	Polyps	Colonoscopy	Polyp detection	Comparison of 8 (D)CNN based	Internal/External	N/A	1. 9.6-69. /6.9-72.3

					end-to-end learning methods and/or handcrafted models			2. 16.7-71.4 13.6-93.5
89	Blanes-Vidal *et al.* (2019)	Polyps	Colorectal capsule endoscopy (CCE), conventional colonoscopy and histopathology	Polyp detection	DCNN	Internal	96.4%	97.1% / 93
90	Fernandez-Esparrach *et al.* (2016)	Polyps	Colonoscopy WL	Polyp detection	WM-DOVA maps	Internal	AUC: 0.79 (in high quality frames) vs 0.75 (in all frames)	70.4% / 72
91	Figueiredo *et al.* (2019)	Polyps	Colonoscopy WL	Polyp detection	CAD	Internal	91.1%	99.7% / 84
80	Gong *et al.* (2020)	Polyps	Colonoscopy (RT)	Polyp detection	DNN (ENDOANGEL)	Internal	95.2%	93.2% / 98.0%
81	Klare *et al.* (2019)	Polyps	Colonoscopy (RT)	Polyp detection	APDS	Internal	N/A	Model: PDR 5 ADR 29.1 Endoscopists ADR 30.9
93	Kominami *et al.* (2016)	Polyps	Colonoscopy (NBI+MAG - RT)	Polyp detection	SVM	Internal	93.2%	93.0% / 93
94	Lequan *et al.* (2017)	Polyps	Colonoscopy	Polyp detection	FCN	Internal	N/A	71.0% / 88
96	Misawa *et al.* (2021)	Polyps	Colonoscopy	Polyp detection	CAD	Internal		90.5% / 93
97	Mori *et al.* (2018)	Polyps	Colonoscopy (NBI ECS RT)	Polyp detection	CAD	Internal	NPV: 93.7 - 96.4%	CAD-NBI:92.0% / 89.8%-93 CAD-stain 92.0%-94.6 87.5%-93.
82	Repici *et al.* (2020)	Polyps	Colonoscopy (RT)	Polyp detection	CAD (GI Genius)	Internal/External	ADR 54.8% vs 40.4% in control	N/A
83	Su *et al.* (2020)	Polyps	Colonoscopy (RT)	Polyp detection	CNN	Internal	ADR 28.9% vs 16.5% in control	94.8-98.0 94.5-99.5
84	Urban *et al.* (2018)	Polyps	Colonoscopy	Polyp detection	CNN	Internal	96.4% (AUC 0.991)	93.0% / 93
85	Wang *et al.*. (2019)	Polyps	Colonoscopy	Polyp detection	CADe	Internal	ADR CAD: 29.%1 vs 20.3% human	N/A
86	Wang *et al.* (2020)	Polyps	Colonoscopy	Polyp detection	CADe	internal	ADR 34.1% vs 27.6% human	N/A
98	Wang *et al.* (2018)	Polyps	Colonoscopy	Polyp detection	CADe	Internal	AUC:0.984	94.4% / 95

95	Misawa *et al.* (2018)	Polyps	Colonoscopy	Polyp detection	CADe	Internal	76.5%	90.0% / 63
92	Hassan *et al.* (2020)	Polyps	Colonoscopy	Polyp detection	CADe (GI-Genius)	Internal	N/A	Sn: 99.7
107	Byrne *et al.* (2019)	Polyps	Colonoscopy NBI	Polyp differentiation	DCNN	Internal	94.0%	98.0% / 83
108	Chen *et al.* (2018)	Polyps	Colonoscopy	Polyp differentiation	DNN	Internal	90.1%	96.3% / 78
110	Horiuchi *et al.* (2019)	Polyps	Colonoscopy (RT)	Polyp differentiation	CAD-autofluorescence imaging	internal	91.5%	80.0% / 95
111	Misawa *et al.* (2016)	Polyps	Colonoscopy (NBI endoscytoscopy)	Polyp differentiation	EndoBRAIN	Internal	96.9%	97.6% / 95
112	Mori *et al.* (2016)	Polyps	Colonoscopy (endocytoscopy)	Polyp differentiation	CAD	Internal/External	89.0%	Diminutive le 92% / 89 Small lesions: 89%
87	Liu *et al.* (2020)	Colorectal polyps, adenomas	Colonoscopy	Polyps and adenomas detection (CADe)	CNN	internal	ADR in CON: 0.2389; in CADe: 0.3910	N/A
122	Kather *et al.* (2019)	Cancer	Histopathology	Predict Survival	CNN	Internal + External	94.0%	OS HR 1.63, C HR 2.29, relap survival HR
120	Thakkar *et al.* (2020)	Quality	Colonoscopy (RT)	Quality of examination metric development	CAD	Internal	N/A	N/A

CRC: Colorectal cancer; CE: Chromoendoscopy; NBI: Narrow band imaging; MSI: Microsattelite instability; dMMR: Deficient mismatch repair; WL: White light; H&E: Hematoxyllin-Eosin; EHR: Electronic health record; CNN: Convolutional neural network; CAD: Computer-aided detection; ADR: Adenoma detection rate; CCE: Colorectal capsule endoscopy; NBI: Narrow band imaiging; DCCN: Deep convolutional neural network; FCN: Fully convolutional network.

Supplementary Table 3 Studies examining video capsule endoscopy

Reference	Author (year)	Disease	Diagnostic Modality	Objective	Analytic model	Validation	Accuracy/AUC	Sensitivity/Specificity
131	Ding et al. (2019)	All	VCE	Assist in evaluation of small bowel abnormalities	CNN	internal and external	N/A	99.9% / 99.9%
130	Aoki et al. (2020)	Bleeding	VCE	Bleeding detection	CNN	Internal	96.6% (0.9998)	99.9% / 99.9%
132	Fu et al. (2014)	Bleeding	VCE	Bleeding detection	SVM	Internal	94.0%	97.0% / 92.0%
133	Hassan et	Bleeding	VCE	Bleeding detection	SVM	Internal	99.2%	99.4% / 98.9%

	al. (2015)							
135	Leenhardt et al. (2019)	Bleeding	VCE	Angioectasia detection	CNN	Internal/External	98.0%	100% / 96.0%
136	Lv et al. (2011)	Bleeding	VCE	Bleeding detection	SVM	Internal	97.9%	97.8% / 98.0%
137	Noya et al. (2017)	Bleeding	VCE	Angioectasia detection	RUSBoost	Internal	96.6% (0.932)	89.5% / 96.8%
138	Pan et al. (2009)	Bleeding	VCE	Bleeding detection	Probabillistic Neural Network	Internal	87.4%	93.1% / 85.8%
139	Sainju et al. (2014)	Bleeding	VCE	Bleeding detection	MLP	N/A	93.0%	96.0% / 90.0%
140	Tsuboi et al. (2020)	Bleeding	VCE	Angioectasia detection	CNN	Internal	0.998	98.8% / 98.4%
141	Xing et al. (2018)	Bleeding	VCE	Bleeding detection	KNN	Internal	0.9922	95.5% / 99.5%
142	Yuan et al. (2016)	Bleeding	VCE	Bleeding detection	SVM	Internal	95.8% (0.9771)	92.0% % 96.5%
134	Iakovidis et al. (2014)	Bleeding/Ulcer/Polyp	VCE	Angioectasia/Bleeding/Polyp	SVM	Internal/External	94.0%	95.4% / 82.9%
152	Zhou et al. (2017)	Celiac	VCE	Celiac disease detection	GoogLeNet	Internal	N/A	100% / 100%
154	Wang et al. (2020)	Celiac	VCE	Celiac disease detection	CNN (InceptionV3, ResNet50 + SVM)	Internal	95.9%	97.2% / 95.6%
151	Tenorio et al (2011)	Celiac	VCE	Celiac disease detection	clinical decision-support system (CDSS)	Internal	84.2%	92.9% / 95.8%
153	Wimmer (2016)	Celiac	VCE	Celiac disease detection	CNN, SVM	Internal	92.5%	N/A
148	Charisis et al. (2016)	Crohn	VCE	Detect CD	SVM	Internal	93.8%	95.2% / 92.4%
147	Klang et al.	Crohn	VCE	Detect CD	CNN	Internal	96.7% (0.99)	92.5-97.1% /

	(2020)							96.0-98.1%
158	Wu et al. (2016)	Hookworm	VCE	Hookworm detection	CNN	Internal	87.3%	58.3% / 87.5%
156	Chen et al. (2013)	Hookworm	VCE	Hookworm detection	SVM	Internal	88.7%	84.5% / 93.0%
157	He et al. (2018)	Hookworm	VCE	Hookworm detection	CNN with online augmentation	Internal	88.5%	84.6% / 88.6%
144	Otani et al. (2020)	Multiple	VCE	Ulcer, Tumor, Angioectasia	CNN	Internal/External	0.996 for ulcers 0.950 for angioectasias, 0.950 for tumors	N/A
149	Saito et al. (2020)	Polyp	VCE	Polyp detection	CNN	Internal	98.6% AUC: 0.911	90.7% / 79.8%
150	Yuan et al. (2017)	Polyp	VCE	Polyp detection	CNN	Internal	98.0%	95.5% / 98.5%
160	Leenhardt et al. (2020)	Quality	VCE	Prep quality	CNN	Internal	95.7%	94.7% / 94.0%
159	Noorda et al. (2020)	Quality	VCE	Prep quality	CNN	Internal	95.2%	96.2% / 94.3%
128	Aoki et al. (2020)	Ulcers	VCE	Detect ulcers	CNN	Internal	0.958	N/A
143	Aoki et al. (2019)	Ulcers	VCE	Detect ulcers	CNN	Internal	90.8% (0.958)	88.2% / 90.9%
145	Fan et al. (2018)	Ulcers	VCE	Ulcer detection	CNN	Internal	95.3% (0.98)	96.8% / 93.7%
146	Wang et al. (2019)	Ulcers	VCE	Ulcer detection	CNN	Internal	90.1%, 0.9469	89.7% / 90.5%

VCE: Video capsule endoscopy; CNN: Convolutional neural network; SVM: Support vector machine; MLP: Multilayer perceptron; KNN: K-nearest neighbor; CDSS: Clinical decision-support system; CD: Crohn's disease.

Supplementary Table 4 Studies examining inflammatory bowel disease and disease subclasses.

Reference	Author (year)	Disease	Diagnostic Modality	Objective	Analytic model	Validation	Accuracy/AUC	Sensitivity/Specificity
178	Takenaka *et al.* (2020)	IBD	Colonoscopy	Disease Severity	DNN	Internal	Accuracy: 90.1% for endoscopic remission, 92.9% for histologic remission	93.3%, 87.8%
171	Mossotto *et al.* (2017)	IBD	Endoscopic images, Histopathology	Diagnosis	SVM	Internal	Accuracy: 83.3%	83.0%, 86.0%
170	Mahapatra *et al.* (2016)	CD	MRI	Diagnosis	RF	Internal	Accuracy: 86.9%	N/A
168	Khorasani *et al.* (2020)	UC	Genetics	Diagnosis	SVM	Internal	AUC: 0.62	Specificity 62.0%
169	Kumar *et al.* (2012)	CD	VCE	Diagnosis	SVM	Internal	Accuracy: 96.5%	89.6,83.7%
183	Waljee *et al.* (2017)	IBD	Laboratory tests+Medications	Disease Course Prediction	RF	Internal	AUC: 0.87	74-80%, 80-82%
188	Wei *et al.* (2013)	IBD	Genetics dataset	Disease Risk	SVM	Internal	AUC: 0.864 for CD and 0.826 for UC	71.0%, 83.0%
187	Isakov *et al.* (2017)	IBD	Genetics dataset	Disease Risk	RF, SVM, Gradient boosting, elastic net regularized generalized linear model	Internal/External	Accuracy: 80.8%, AUC: 0.829	57.7%, 88.0%
174	Niehaus *et al.* (2015)	CD	Laboratory studies	Disease Severity	SVM, LR, RF	Internal	Accuracy: 68.7%	59.1%, 78.4%

172	Maeda *et al.* (2019)	UC	Colonoscopy + ECS	Disease Severity	CAD	Internal	Accuracy: 90%	74.0%, 91.0%
166	Ozawa *et al.* (2019)	UC	Colonoscopy	Disease Severity	CNN (GoogLeNet)	Internal	AUC: 0.86, 0.98 to identify Mayo 0 and 1	N/A
175	Reddy *et al.* (2019)	CD	EHR data	Disease Severity	Gradient boosting, RR, LR	Internal	AUC: 0.9282 in the GB model, 0.8270 in the RR, and 0.8112 in the LR	N/A
173	Matalka *et al.* (2013)	IBD	Histopathology	Disease Severity	N/A	Internal	N/A	98.3%, 98.3%
176	Stidham *et al.* (2020)	CD	CTE	Disease Severity	Semi-automated bowel measurement	Internal	Accuracy: 87.6%, AUC: 0.857	67.2%, 92.5%
177	Stidham *et al.* (2019)	UC	Colonoscopy	Disease Severity	CNN	Internal/External	AUC: 0.966	83, 96%
179	Yao *et al.* (2021)	UC	Colonoscopy	Disease severity	CNN	Internal/External	Accuracy: 87.6%	90.2%,87.0%
190	Firouzi *et al.* (2007)	IBD	EHR data	Other	WEKA (Waikato Environment for Knowledge analysis)	Internal	Accuracy: 86.2-89.8%	65.7-82.8%, 95.2-96.3%
189	Hou (2013)	IBD	Colonoscopy	Other	NLP ARC	Internal	Accuracy: 80%	77.0%, 0.88%
184	Waljee *et al.* (2018)	UC	Laboratory studies + demographics	Response to treatment	RF	internal	AUC: 0.73	72.0%, 68.0%
185	Waljee *et al.* (2017)	IBD	Laboratory studies + demographics	Response to treatment	RF	Internal	AUC: 0.79 (vs 0.49 6TGN)	N/A
182	Waljee *et al.* (2010)	IBD	Laboratory tests	Response to treatment	RF	Internal	AUC: 0.856 for non-responders; 0.594 for 6TGN	N/A

180	Doherty *et al.* (2018)	CD	Fecal microbiota	Response to treatment Ustekinumab	RF	Internal	AUC: 0.844	77.4%, 83.1%
186	Waljee *et al.* (2019)	CD	CRP, Albumin, demographics	Response to treatment	RF	Internal	AUC: 0.78 for 8w model, 0.76 for albumin/CRP at 6w model	8w model: 79.0% , 67.0%; albumin/CRP 6w: 77.0%, 68.0%
181	Douglas *et al.* (2018)	CD	Genetics from intestinal biopsy	Response to treatment/Diseas e severity	RF	Internal	Accuracy: 84.2%	N/A
143	Aoki *et al.* (2019)	CD	VCE	Ulcers	CNN	Internal	Accuracy: 90.8%, AUC: 0.958	88.2%, 90.9%

IBD: Inflammatory bowel disease; DNN: Deep neural network; SVM: Support vector machine; RF: Random forest; MRI: Magnetic resonance imaging; AUC: Area under the curve; CD: Crohn's disease; UC: Ulcerative colitis; VCE: Video capsule endoscopy; LR: Logistic regression; CNN: Convolutional neural network; CAD: Computer-aided detection; CTE: Computed tomography - Enterography; EHR: Electronic health record; NLP: Natural language processing.

Supplementary Table 5 Studies examining Hepatobiliary conditions

Reference	Author (year)	Disease	Diagnostic Modality	Objective	Analytic model	Validation	Accuracy(%) / AUC (dec)	Sensitivity/Specificity
251	Abajian et al. (2018)	HCC	MRI, Clinical data	Predict treatment response of TACE based on qEASL criterion	LR, RF	Internal	78.0%	62.5% / 82.1%
209	Ahmed et al. (2020)	HCV	Tagged MRI	Detect HCV-associated liver F1-F4 fibrosis from the heart-induced deformation in tagged MR images	SVM	LOO cross-validation	83.7%	81.8% / 86.6%
264	Ai et al. (2018)	DILI	Drug molecular fingerprints	To predict hepatotoxicity in early stages of drug development	SVM, RF, extreme gradient boosting, QSAR model	Internal/External	Training: 71.1% (0.764). Testing: 84.3% (0.904)	Training: 79.9%/60.3%. Testing: 86.9%/75.4%.
243	Andres et al. (2018)	PSC, LT	Clinical data	Predict individual survival after LT for PSC	Patient-specific survival prediction (PSSP) software	Internal	PSSP accurately estimates the survival probability over time	N/A
244	Ayllon et al. (2018)	LT	Clinical data	Validation of model for doner-recipient matching in liver transplantation. Outcome 1) graft survival 3 months 2) graft survival 12 months	ANN, MOEA	Internal/External	3-month:0. 94 (CCR and MS AUC) 12-month: 0.78, 0.82 (CCR, MS AUC)	N/A
265	Banerjee et al. (2018)	HCC	US	Extract LI-RADS scoring of HCC from structured and unstructured US reports	NN, combination with word semantics and rule-based LI-RADS coding	Internal	N/A	49.0% / 59.0%
245	Bertsimas et al.	LT	Clinical data	Predict death or unsuitable for LT	OCT model (decision tree	50% training, 20% validation,	0.859	N/A

	(2019)			within 3 months	based)	30% testing		
246	Bhat et al. (2018)	LT, NODM	Clinical data	Key predictors and survival outcome of NODM after LT.	RF, NN, LR, gradient boosting, SVM	70% training, 30% validation	N/A	Rater 1: 63%/ 77%; rater 2: 62%/ 74%
247	Briceno et al. (2014)	LT	Clinical data	Donor-recipient matching in LT and prediction of 3 month graft survival	ANN, LR, decision tree, SVM	Internal/External	NN-CCR: 90.79% (0.80). NN-MS: 71.42% (0.82)	N/A
210	Byra et al (2018)	NAFLD	US, B-mode	Assessment of steatosis level on ultrasound	D-CNN, SVM, Lasso regression method, ImageNet training	Internal	0.977	100% / 88.2%
235	Canbay et al. (2019)	NASH, NAFLD	Clinical data	Differentiate NAFLD from NASH	ensemble feature selection	Internal/External	0.730	N/A
211	Chang et al. (2016)	Cirrhosis	EHR	Improving the identification of cirrhosis patients by ICD-9 codes with addition of NLP.	n/a	Internal/External	N/A	NLP portion of algorithm: 90% / 98.98%
252	Chaudhary et al. (2018)	HCC	Multi-omics (RNAseq, miRNAseq, DNA methylation)	Survival prediction in HCC patients	SVM, autoencoders neural network	Internal/External	Concordance index of 0.69-0.77	N/A
212	Chen et al. (2017)	HBV, cirrhosis	US, elastography	Determine fibrosis stage of HBV or cirrhosis patients.	RF, k-nearest neighbor, SVM, naïve Bayes	Internal/External	Highest accuracy for RF: 82.9%	Highest values for SVM and naïve Bayes: 92.9% and 82.5%, respectively
193	Chen et al. (2020)	Gallbladder polyps	US	Differentiate diagnosis of neoplastic and non-neoplastic gallbladder polyps	Principal components analysis (PCA) and AdaBoost	Internal	87.5%	86.5% / 89.4%
213	Choi et al. (2018)	Liver cirrhosis, fibrosis	CT	Staging fibrosis	CNN	Internal/External	Significant fibrosis: 94.1% (0.96) Advanced fibrosis: 95%,	84.6-95.5% / 89.9-96.6%

							(0.97) Cirrhosis: 92.1% (0.95)	
214	Cui et al. (2021)	Liver fibrosis	CT, multiphase	Staging liver fibrosis on multiphase CT	gradient boosting	Internal/External	F>1: 0.650 F>2: 0.790 F4: 0.800	F>1: 39.6%/85.1%. F>2: 72.7%/73.2% F4: 78.2%/81.8%.
266	Dickerson et al. (2019)	End stage liver disease, hepatic encephalopathy	EHR	Assess pre-transplant cognitive impairment in patients with end stage liver disease through patient to provider messages.	19 NLP measures (Lexical, Lexico-syntactic, Syntactic, Lexico-semantic, Sentiment domains)	Internal	MELD≥30 decreased word length, fewer 6-letter words, increased sentence length	N/A
236	Docherty et al. (2021)	NASH	EHR	Predict NASH from NAFLD data of Optum EHR dataset with liver biopsy as gold standard.	extreme gradient boosting	Internal/External	14-feature model: AUC of 0.82. 5-feature model: AUC of 0.79	14-feature: Sn 81%. 5-feature: 80%.
204	Dong et al. (2019)	Cirrhosis, EV	Clinical data	Predict EV in liver cirrhosis patients	RF	Internal/External	EV: 0.82 VNT: 0.75	EV 92.3%/65.9%. VNT: 100%/49.3%.
256	Eaton et al. (2020)	PSC	Clinical data	PREsTo PSC risk estimate tool based on clinical and laboratory values to predict PSC outcomes (decompensation)	gradient boosting	Internal/External	PSC risk estimate tool predicts decompensation with C-statistic: 0.90 (higher then MELD or Mayo Score)	N/A
203	Abd El-Salem et al. (2019)	HCV cirrhosis, EV	Clinical data	Predict EV in HCV cirrhosis patients from clinical and laboratory data	ANN, naïve Bayes, decision tree, SVM, RF, Bayesian network	Internal/External	Bayesian network (highest performance): 68.9% (0.748)	65.3% / 72.0%
237	Fialoke et al. (2018)	NASH	EHR	Predict NASH from NAFLD	decision tree, LR, RF, extreme gradient boosting	Internal/External	76.2-79.7% (0.83-0.88)	74.3-77.4% / 77-80.8%

215	Forlano et al. (2020)	NAFLD	Histopathological slides	NASH, ballooning, and fibrosis	k-means	Internal	0.802	80.0% / 62.0%
257	Garcia et al. (2019)	ACLF	Clinical data	Predict mortality in patients with ACLF up to day 29.	extreme gradient boosting, LR	Internal	Day 1: 0.97. Day 29: 0.758.	N/A
273	Garcia-Carretero et al. (2019)	NASH	Clinical data	Prediction of NASH among patients with hypertension.	LASSO, RF	80% training, 20% testing	0.790	70.0% / 79.0%
218	Gatos et al. (2019)	CLD, liver fibrosis	US, SWE	Fibrosis staging	CNN	Internal	82.5-95.5%	N/A
216	Gatos et al. (2016)	CLD, liver fibrosis	US, SWE	Fibrosis staging	SVM	Internal	87% (0.85)	83.3% / 89.1%
217	Gatos et al. (2017)	CLD, liver fibrosis	US, SWE	Fibrosis staging	SVM	Internal	87.3% (0.87)	93.5% / 81.2%
194	Hamm et al. (2020)	Liver masses, HCC	MRI	Differentiate benign from malignant focal liver lesions	CNN	Internal	0.992	90.0% / 98.0%
267	He et al. (2019)	Liver cirrhosis, fibrosis	MRI, Clinical data	Stiffness estimation	Radiomics, SVM	Internal	75% (0.80)	63.6% / 82.4%
219	Heinemann et al. (2019)	NAFLD, NASH	ANIMAL - Histopathology slides	Define ballooning, inflammation, steatosis and fibrosis (features of NASH, Kleiner score) on histology slides	CNN (4x)	Internal	86.0–94.5%	N/A
205	Hong et al. (2011)	HBV cirrhosis, EV	Clinical, imaging data	Prediction of presence of esophageal varices in HBV cirrhosis patients based on clinical, laboratory and imaging variables	MLP-ANN (three-layered, feed-forward ANN model with three hidden nodes, with back propagation algorithm)	Internal	86.8%	96.5% / 60.4%
220	Huang et al. (2007)	HCV cirrhosis	Genomics	To predict cirrhosis risk in patients with chronic HCV (Cirrhosis Risk Score)	naïve Bayes	Internal/External	0.760	Low risk for cirrhosis: 87.9%/42.9%. High risk for cirrhosis: 53.6%/96.2%.

258	Ibragimov et al. (2018)	Post SBRT liver injury	Clinical data, CT (3D dose plans)	Predict SBRT hepatoxicity on pre-treatment CT	CNN, SVM, RF, fully connected NNs	Internal	0.850	N/A
259	Jovanovic et al. (2014)	Choledocholithiasis	Clinical data, laboratory and extracted imaging features	Predict presence of biliary stones/ necessity for therapeutic ERCP	MNN, LR	Internal	0.884	92.7% / 68.4%
260	Kanwal et al. (2020)	Liver cirrhosis	Clinical data	Predict all-cause mortality in cirrhosis patients	LR with LASSO, extreme gradient boosting, partial path model	Internal/External	CiMM 0.780 vs MELD-Na: 0.670	Mean sensitivity of CiMM was 10/11% higher than MELD-Na score
248	Kazemi et al. (2019)	LT	Clinical data	Predict survival after LT	SVM, Bayesian network, decision tree, MNN, k nearest neighbor	Internal	0.900	Sensitivity: 81.0%
268	Khan et al. (2018)	HBV	Serum, Raman spectroscopy	Detect spectral differences between normal and HBV serum samples	SVM	Internal	98.0%	100% / 95.0%
195	Kim et al. (2021)	HCC	CT, multiphase	Detecting primary hepatic malignancies on CT	CNN, mask region based	Internal/External	84.8%	Sensitivity: 84.8%
196	Kim et al. (2004)	Liver cirrhosis (multiple etiologies)	Molecular gene analysis with cDNA microarray on surgical tissue	1) Determine molecular signature between two distinct groups of cirrhosis patient, low-risk vs high-risk s 2) Develop molecular gene signature for HCC	k-nearest neighbor, SVM	Internal/External	1) KNN: 78.0%, SVM: 86.0% 2) KNN: 79.0%, SVM: 89.0%.	N/A
222	Konerman et al. (2015)	HCV	Clinical data	Fibrosis prediction in HCV patients	LR, RF	Internal/External	Fibrosis progression: 0.78-0.79 Clinical progression:	Fibrosis progression: 85%/71-77% Clinical progression: 74-81%/70-78%

							0.79-0.86	
221	Konerman et al. (2019)	HCV	Clinical data	Prediction of cirrhosis in patients with HCV	Boosted survival tree	Internal/External	1 year: 0.830 3 year: 0.797 5 year: 0.787	1 year: 76%/77%. 3 year: 76%/73%. 5 year:73%/74%.
263	Konerman et al. (2017)	HCV	Clinical data	To predict fibrosis progression and clinical outcomes in HCV patients	RF	External	1 year: 0.78 3 year: 0.76	1 year: 80%/62% 3 year: 69%/65%
223	Kuppili et al. (2017)	NAFLD, liver fibrosis	US	Risk stratification for fatty liver disease on ultrasound images	SVM, extreme learning machine	Cross-validation	96.8% (0.97)	94.2% / 97.6%
224	Lara et al. (2014)	chronic HCV, liver fibrosis	Viral markers, HCV genetic assays	To identify patients with fast and slow fibrosis progression rates among patients undergoing liver transplant for chronic HCV infection.	k-nearest neighbor, linear projection, Bayesian networks	Internal/External	Split cross-validation: 90-95%. Validation: 85-90%.	70.0-71.4% / 92.3-100%
249	Lau et al. (2017)	LT	Clinical data	Predict graft failure or non-function after LT using donor and recipient factors	RF, ANN, LR	Internal	0.818	N/A
234	Lee et al. (2020)	Liver fibrosis	B-mode ultrasonography	Predict METAVIR score for liver fibrosis	CNN	Internal/External	Internal: 86.5% (detecting significant fibrosis, F2-F4). External: 88.3% (detecting cirrhosis, F4)	Internal: 91.3%/82.4% (detecting significant fibrosis/ F2-F4). External: 77.8%/93.7% (detecting cirrhosis/ F4)
250	Lee et al. (2018)	LT	Pre- and intraoperative variables by anesthesia and	Prediction of AKI after LT	decision tree, RF, gradient boosting machine, SVM, naïve Bayes, MNN,	Internal	84.0% (0.90)	N/A

			surgery		deep belief networks, LR			
225	Li et al. (2019)	HBV, liver fibrosis	US	Fibrosis staging in HBV patients	decision tree, RF, SVM, AdaBoost	Internal	AdaBoost, RF and SMV: 85.0%	AdaBoost: 87.5%/76.9%. RF 87.5%/76.9%. SVM: 93.8%/69.2%
269	Li et al. (2018)	HCC	CT	Automatic liver tumor segmentation	2D and 3D fully CNN (H-DenseUNet)	Internal	Effectively performs liver and tumor segmentation from CT volumes	N/A
206	Liu et al. (2020)	Liver cirrhosis	CT, contrast enhanced. MRI.	Identify clinically significant portal hypertension (CSPH) on 1) CT 2) MRI	CNN (pretrained-VGG19)	Internal/External	1. 91.1% 2. 88.9%	1. 91.4%/90.9%. 2. 92.0%/84.9%
274	Ma et al. (2018)	NAFLD	Clinical and laboratory data	NAFLD prediction	Bayesian network	Internal	83.0%	87.8% / 67.5%
207	Marozas et al. (2017)	portal hypertension	Clinical data	Predict presence of elevated HVPG	naïve Bayes, LR, decision tree, RF	Internal	89.7% (0.96)	83.0% / 92.0%
226	Meffert et al. (2014)	Liver steatosis	Clinical data	Steatosis score	Variable selection: boosting algorithm. Bayesian network.	Internal/External	0.876	N/A
275	Moccia et al. (2018)	LT, liver steatosis	Histopathology donor liver	Analysis of donor liver texture for hepatic steatosis		Internal	88.0%	95.0% / 81.0%
253	Morshid et al. (2019)	HCC	CT	Predicting TACE response of HCC patients	CNN, RF	Internal	74.2% (0.733)	N/A
270	Mueller-Breckenridge et al. (2019)	HBV	Genomics	Classify seroconversion to HBeAg from complete HBV genome of European and Asian	RF	Internal/External	97.0%	96.0% / 100%

				cohort				
227	Perakakis et al. (2019)	NASH, NAFLD	Serum (lipidomic, glycomic and free fatty acids)	NASH and fibrosis diagnosis	SVM	Internal	>90.0% NASH, NAFLD diagnosis >97.0% Fibrosis diagnosis	Multiple metrics reported in Fig. 6
276	Perveen et al. (2018)	NAFLD	EHR	NAFLD diagnosis risk and progression	Decision tree	Internal/External	76% (0.73)	Without random oversampling: 83.2-93.7% / 76.0%-78.0%
228	Piscaglia et al. (2006)	HCV, LT	Clinical data, laboratory	Predict post-LT fibrosis in HCV patients	ANN	Internal	83.3%	100% / 79.5%
208	Qi et al. (2019)	Liver cirrhosis, portal hypertension	CT angiography (virtual hepatic vein pressure gradient, HVPG)	Develop and validate computational model for non-invasive HVPG	Finite element analysis and computational fluid dynamics analysis	Internal/External	Training: 0.83. Validation: 0.89.	74.0% / 93.0%
229	Raoufy et al. (2009)	chronic HBV, liver cirrhosis	Laboratory data and age	Diagnose cirrhosis based on laboratory data and age	ANN	Internal	91.4% (0.898)	97.5% / 92.0%
230	Redman et al. (2017)	NAFLD	Radiology reports (US, CT, MRI)	Identify presence of fatty liver disease based on full-text radiology reports	CLAMP NLP software	Internal/External	US: 93.4%. CT: 98.8%. MRI: 100%	US: 90%/95.3%. CT: 93.5%/99.5%. MRI: 100%/100%
254	Saillard et al. (2020)	HCC	Histological slides, whole slide imaging	Prediction of survival after HCC resection	Pre-trained CNN	Internal/External	C-indices for survival prediction 0.75-0.78	N/A
197	Schmauch et al. (2019)	Liver masses	US	Detection and classification of focal liver lesions	CNN	Internal/External	Detection: 0.935. Characterization: 0.916.	N/A
255	Shan et al. (2019)	HCC	CT	Predict recurrence of HCC after resection or ablation based on	Radiomics, LASSO LR model	Internal	0.790	N/A

				peritumoral radiomics				
272	Shousha et al. (2018)	HCV	Genetics	Discover predictors for advanced fibrosis in HCV patients	MNN, decision tree (REPTree)	Internal	MNN: 0.880	MNN 82.5% / 81.1%
198	Singal et al. (2013)	HCC	Clinical data	Predicting HCC development in cirrhosis patients	decision tree, RF	Internal	80.7%	80.7% / 46.8%
231	Sowa et al. (2013)	NAFLD	Liver serum parameters, hyaluronic acid and cell death markers	Fibrosis prediction in NAFLD	LR, decision tree, RF, SVM, k-nearest neighbor	Internal	79.0%	>60.0% / 77.0%
238	Sowa et al. (2014)	NALFD, ALD	Liver serum parameters, (adipo-)cytokines and cell death markers	Distinguish NAFLD from ALD.	LR, decision tree, SVM, RF	Internal	NAFLD from ALD non-cirrhosis: SVM (0.9118) – RF (0.8932)- DT 89.02%. ALD cirrhosis from non-cirrhosis: SVM (0.9058) – RF (0.8971) – DT 95.1%	Decision tree NAFLD from ALD non-cirrhosis: 74.2%/98.4%. ALD cirrhosis from non-cirrhosis: 94.1%/96.1%
261	Speiser et al. (2019)	ALF (acetaminophen)	Clinical data	Daily outcomes in acetaminophen induced ALF	decision tree (BiMM tree)	Internal/External	0.749	44.9-61.3% / 63.8-84.1%
262	Speiser et al. (2015)	ALF (acetaminophen)	Clinical data	APAP ALF prognosis prediction	decision tree (CART analysis)	Internal/External	72.0% (0.79)	71.0% / 77.0%
199	Sun et al. (2020)	Liver cancer	Histopathological image analysis	Classify liver histopathological images as normal or cancer	CNN	Internal	100%	100% / 100%

239	Taylor-Weiner et al. (2021)	NASH	Histopathological samples	Diagnose NASH on histopathlogy samples	CNN. Deep Learning Treatment Assessment (DELTA) Liver Fibrosis score	Internal/External	Concordance indices for inflammation, steatosis and ballooning: 0.57-0.67	N/A
240	Van Vleck et al. (2019)	NAFLD	EHR	Identifying patients with NALFD in EHR	CLiX clinical NLP engine (general-purpose stochastic parser, Clinithink)	Internal	N/A	NLP 93.0% / 89% NLP + ICD: 96.0% / 89.0%.
242	Vanderbeck et al. (2014)	NAFLD	Histopathological slides	Automatic classification of white regions indicative of NAFLD	SVM	Internal	89.0%	59.0-98.0% / 61.5-95.7%%
241	Vanderbeck et al. (2015)	NAFLD	Histopathological slides	Automatic quantification of 1) lobular inflammation and 2) hepatocyte ballooning	SVM	Internal	Lobular inflammation : (0.950) Hepatocyte ballooning (0.980)	Lobular inflammation: 49.0% / 70.0% Hepatocyte ballooning: 54.0% / 91.0%
200	Wang et al. (2019)	Liver masses, HCC	MRI, multiphase	Malignancy classification	CNN	Internal	N/A	82.9% / 76.5%
232	Wang et al. (2019)	HBV, liver fibrosis	US, elastography	Assess fibrosis in HBV	CNN	Internal/External	0.970-1.000	Multiple metrics reported in Table II
233	Wei et al. (2018)	HBV, HCV	Clinical data	Predict fibrosis in HBV patients	decision tree, RF, GB	Internal/External	0.918	Advanced fibrosis: 84.0% / 85.0% Cirrhosis: 85.0% / 78.0%.
271	Williams et al. (2020)	DILI	Hepatic safety assays	Predict DILI in compounds during drug development	Bayesian network	Bayesian approach (no cross-validation)	86.0%	87.0% / 85.0%
277	Wu et al. (2019)	NAFLD	Clinical data	Predict fatty liver disease	RF, LR, ANN, naïve Bayes	Internal/External	RF: 87.5% (0.925)	RF: 87.2% / 85.9%
201	Yasaka et al. (2017)	Liver masses	CT multiphase	Differentiate benign from malignant	CNN	Internal	84.0% (0.92)	Sensitivity of:

				liver masses				71-100%
202	Yasaka et al. (2018)	Liver masses	MRI	Differentiate liver masses	CNN	Internal	(0.84-0.985)	76-84% / 65-76%
278	Yip et al. (2017)	NAFLD	Clinical data	To predict NALFD combining laboratory values with presence of hypertension	decision tree, LR, RR,	Internal	87.0% (0.870)	92.0% / 90.0%

HCC: Hepatocellular carcinoma; MRI: Magnetic resonance imaging; HCV: Hepatitis-C virus; SVM: Support vector machine; LOO: Leave-one-out; DILI: Drug-induced liver injury; PSC: Primary sclerosing cholangitis; LT: Liver transplant; ANN: Artificial neural network; MOEA: Multi-objective evolutionary algorithm; NN: Neural network; US: Ultrasound; NODM: New-onset diabetes mellitus; LR: Logistic regression; NAFLD: Non-alcoholic fatty liver disease; NASH: Non-alcoholic steatohepatitis; NLP: Natural language processing; HBV: Hepatitis B virus; CT: Computed tomography; EHR: Electronic health record; MELD: Model end stage liver disease; EV: Esophageal varices; ACLF: Acute on chronic liver failure; CLD: Chronic liver disease; SWE: Shear-wave elastography; SBRT: Stereotactic body radiation therapy; HVPG: Hepatic vein pressure gradient; ALF: Acute liver failure.

Supplementary Table 6 Studies examining pancreatic conditions

Reference	**Author (year)**	**Disease**	**Diagnostic Modality**	**Objective**	**Analytic model**	**Validation**	**Accuracy (%)/AUC (dec.)**	**Sensitivity/Specificity**
294	Al-Haddad et al. (2010)	IPMN	EHR	Develop clinical registry of patients with surgically resected IPMN	Regenstrief EXtraction Tool (REX)	Internal	N/A	Sensitivity of 97.5%
283	Andersson et al. (2011)	AP	Clinical data	Predict severe acute pancreatitis	ANN	Internal	0.920	N/A
327	Blyuss et al. (2020)	PDAC	PancRISK, urine samples	Comparison of different AI algorithms for risk score of PDAC based on three urinary biomarkers	NN, RF, SVM, NF, LR	Internal	AUC: LR, NN and NF: 0.940, 0.930, 0.940	81.0% / 90.0% LR 81.0% / 90.0% NN 87.0% / 90% NF 82.0% /89.0% SVM 86.0% / 82.0% RF 96.0% / 96.0% LR +

								CA 19.9
316	Chakraborty et al. (2018)	IPMN	CT, clinical data	Investigate CT imaging features as markers for assessment of IPMN risk (low vs high).	Radiomics, RF, SVM	Internal	0.770 for imaging features alone. 0.810 with clinical variables.	N/A
301	Chu et al. (2019)	PDAC	CT	Differentiate PDAC from NP	Radiomics, RF	Internal	0.999	100% / 98.5%
317	Corral et al. (2019)	IPMN	MRI	Detect dysplasia in pancreatic cysts. Detect high grade dysplasia or cancer.	CNN	Internal	0.780	Detect dysplasia: 92.0% / 52.0% Detect HGD/cancer: 75.0% / 78.0%
302	Das et al. (2008)	PDAC, CP	EUS, radial scanning echoendoscopes	Differentiating PDAC from non-neoplastic tissue on EUS images	DIA (Image J) with PCA, ANN	Internal	0.930	93.0% / 92.0%
318	Dmitriev et al. (2017)	PCN	CT (2D axial 0.75mm), clinical data	Classification of four most common pancreatic cyst types (IPMN, MCN, SCA, SPN)	Bayesian combination of RF and CNN	Internal	83.6%	N/A
328	Facciorusso et al. (2019)	PDAC	EUS-CPN, clinical data	Prediction of pain response after celiac plexus neurolysis	ANN	Internal	0.940	N/A
285	Fei et al. (2019)	AP	Clinical data	Predict risk and severity of ARDS following severe acute pancreatitis	BP-ANN	Internal	84.4%	Sensitivity: 87.5%
286	Fei et al. (2017)	AP	Clinical data	Predict occurrence of portosplenomesenteric venous thrombosis (PSMVT)	BP-ANN	Internal	83.3%	80.0% / 85.7%
284	Fei et al. (2018)	AP	Clinical data	Predict acute lung injury in severe acute pancreatitis	BP-ANN	Internal	84.4%	87.5% / 83.3%
295	Gao et al. (2020)	N/A	MRI (T1-weighted	Differentiate pancreatic	CNN	Internal/Externa	79.5%	N/A

			contrast enhanced)	diseases on MRI	augmented by synthetic images from GANs	l	(0.9451)	
303	Gao et al. (2019)	PNET	MRI (T1-weighted contrast enhanced)	Predicting WHO grade of PNET	CNN augmented by synthetic images from GANs	Internal/Externa l	Cross validation: 85.1% External: 79.1-82.4%	N/A
329	Hayward et al. (2010)	PDAC	Clinical data	Prediction of clinical performance of patients with pancreatic cancer.	Rule-based, decision trees, k-nearest neighbor, Bayesian methods, LRA, MNN	Internal	96.2%	81.3% / 98.9%
287	Hong et al. (2013)	AP	Clinical data	Prediction organ failure in AP.	BP-MNN	Internal	96.2%	81.3% / 98.9%
330	Kaissis et al. (2019)	PDAC	MRI 1.5T (DWI)	Predict above vs below median OS in PDAC patients.	Radiomics, RF	Internal	0.900	87.0% / 80.0%
305	Kaissis et al. (2020)	PDAC	CT (portal-venous-phase)	Predict molecular subtype of pancreatic cancer (quasi-mesenchymal, QM vs non-quasi-mesenchymal , non-QM) on CT	Radiomics, RF	Internal	0.930	84.0% / 92.0%
304	Kaissis et al. (2019)	PDAC	MRI 1.5T (DWI)	Predict molecular subtype of pancreatic cancer (KRT81+) on MRI	Radiomics, Gradient boosted-tree algorithm	Internal	0.930	90.0% / 92.0%
288	Keogan et al. (2002)	AP	Clinical data (physical, biochemical, radiographic)	Prediction of hospital stay length in AP	ANN	Internal	0.830	100% / 29.0%
319	Kurita et al. (2019)	PCN	Clinical data, cyst	Differentiate benign	ANN	Internal	92.9%	95.7% / 91.9%

			fluid cytology and chemistry and extracted imaging features	from malignant pancreatic cystic lesions through cyst fluid obtained during surgery or EUS-FNA				
320	Kuwahara et al. (2019)	IPMN	EUS still images	Define benign from malignant IPMN	CNN (based on ResNet50)	Internal	94.0%	95.7% / 92.6%
331	Li et al. (2019)	PDAC	CT	Predict survival time in PDAC based on radiomics and HMGA2 and C-MYC gene expressions profile	Radiomics, SVM with k-fold	Internal	95.0% (0.900)	92.0% / 98.0%
335	Li et al. (2018)	PDAC	PET-CT	Pancreatic cancer CAD for PET-CT	SLIC with grey interval mapping for segmentation, DT-PCA for best feature selection, Hybrid SVM-RF to classify	Internal	96.5%	95.2% / 97.5%
321	Li et al. (2019)	PCN	MDCT pancreas protocol	PCN classification between four histopathologically confirmed subtypes (IPMN, MCN, SCN and SPN) with CAD	Densely connected CNN (Dense-Net), saliency maps	Internal	72.8%	N/A
306	Linning et al. (2020)	PDAC, AIP	CT, multiphase	Differentiate focal-type AIP from PDAC	Radiomics, RF	Internal	94.8%	93.3% / 96.1%
336	Liu et al. (2019)	PDAC	CT	Diagnose pancreatic cancer on CT faster than radiologists.	Faster R-CNN model	Internal	0.9632	N/A
315	Luo et al. (2020)	PNET	CT	To predict PNET grade on CT on arterial, venous and arterial/venous scans)	CNN	Internal/External	88.1% (0.820)	88.3% / 84.6%

280	Marya et al. (2020)	AIP	EUS video	Differentiate AIP from PDAC, CP and NP.	CNN, Occlusion heatmap analysis	Internal	75.6%	AIP from NP: 99%/98% AIP from CP: 94%/71% AIP from PDAC 90%/93% AIP from all 90%/85%
296	Mashayekhi et al. (2020)	CP, recurrent AP, functional abdominal pain	CT	Differentiate between functional abdominal pain, RAP and CP	Radiomics, ono-vs-one Isomap, SVM	Internal	82.1%, nonspecific abdominal pain AUC: 0.91, RAP AUC: 0.88, CP AUC of 0.90	Nonspecific abdominal pain group 79%, 100%; RAP: 95%, 78%; CP: 71%, 95%
337	Mehrabi et al. (2015)	PC	EHR	Identify patients with family history of pancreatic cancer	Unstructured Information Management Architecture (UIMA) with multiply analysis engines	Internal/External	N/A	75.3% / 91.3%
289	Mofidi et al. (2007)	AP	Clinical data	Identify severe acute pancreatitis and predict fatal outcome	ANN	Internal	Severity: 92.5% Death: 97.5%	Severity: 89%/96% Death: 88%/98%
307	Momeni-Boroujeni et al. (2017)	Solid pancreatic masses	EUS-FNA, cytology slides	Design computer model assisted diagnosis of solid pancreatic mass biopsy	MNN	Internal	Benign and malignant: 100% Atypical: 77.0%	80.0% / 75.0%
308	Norton et al. (2001)	PDAC, pancreatitis	EUS	Differentiate between CP and PDAC on EUS	n/a	Internal	80.0%	100% / 50.0%
322	Okon et al. (2001)	Intraductal proliferative lesions	Surgical specimen	Classification of pancreatic intraductal proliferative lesions based on nuclear	DIA	Internal	73.0%	N/A

				features				
338	Ozkan et al. (2016)	PDAC	EUS	Detect pancreatic adenocarcinoma in 1. all patients 2. under 40 3. 40-60 years 4. 60+ years	ANN with age-based MLP	Internal	1. 87.5% 2. 92.0% 3. 88.5 % 4. 91.7 %	1. 83.3% / 93.3% 2. 87.5% / 94.1% 3. 85.7% / 91.7% 4. 93.3% / 88.9%
290	Pearce et al. (2006)	AP	Clinical data	Predict severity acute pancreatitis with APACHE II variables and CRP	KLR method	Internal	0.820	87.0% / 71.0%
291	Pofahl et al. (1998)	AP	EHR	Predict LOS in patients with AP	BP-ANN	N/A	N/A	Sensitivity 75.0%
292	Qiu et al. (2019)	AP	Clinical data	Predict MOF in moderately severe AP	SVM, LR and ANN	Internal	SVM 79.9% LR 77.9% ANN 71.1%	SVM 75.0% / 81.7% LR 79.2% / 77.5% ANN 86.1% / 65.5%
293	Qiu et al. (2019)	AP	Clinical data	Predict intra-abdominal infection in moderately severe AP	MNN	Internal/External	0.923	80.9% / 89.4%
309	Qui et al. (2019)	PDAC	CT	To predict histopathological grades of PDAC	SVM	Internal	86.0%	78.0% / 95.0%
339	Roch et al. (2015)	PCN	EHR	To implement NLP based pancreatic cyst identification system	Unstructured Information Management Architecture (UIMA) with novel negation algorithm DEEPEN	Internal	N/A	99.9% / 98.8%
297	Roth et al. (2018)	N/A	CT	Pancreas localization and segmentation	HNNs	Internal	N/A	Sensitivity of nearly 100% for all scans (except for two cases ≥ 94.54%)
310	Saftoui et al.	PDAC, CP	EUS elastography	Differentiate benign from malignant patterns	MLP-NN	Internal	91.1% training / 84.3%	87.6% / 82.9%

	(2012)			in focal pancreatic masses.			testing	
311	Saftoui et al. (2008)	N/A	EUS elastography, hue histograms	Differentiate between normal and diseased tissue on EUS elastography.	MLP-NN	Internal/External	89.7%	91.4% / 89.7%
312	Saftoui et al. (2015)	PDAC, CP	CEH-EUS, EUS-FNA	Validate parameters from TIC analysis with ANN model	BP-ANN	Internal/External	94.6%	94.6% / 94.4%
323	Song et al. (2013)	PCN	Cytology slides	Automate diagnosis between SCA and MCN.	Bayesian Classifier, k-Nearest Neighbors, SVM, ANN	Internal	Bayesian 79.0% k-NN: 78.0% SVM: 85.0% ANN 85.0%	Bayes: 93%/65% k-NN: 84%/75% SVM 86%/85% ANN 84%/86%
324	Springer et al. (2019)	PCN	Multimodality (Clinical, imaging and cyst fluid genetics and biochemical markers)	Classify patients with pancreatic cysts to surgery, monitoring or no further surveillance	CompCyst	Internal	69.0%	Discharge: 46% 100% Surgery: 91%/54% Surveillance: 99%/30%
332	Walczak et al. (2017)	PDAC	Clinical data	To predict survival likelihood in PDAC	ANN	Internal	N/A	91.0% / 38.0%
325	Wei et al. (2019)	SCN	CT	To differentiate SCN from MCN on MDCT	Radiomics, SVM	Internal/External	Cross validation AUC: 0.767 Independent validation: 0.837	Cross validation: 0.686, 0.709. Independent validation: 0.667, 0.818.
333	Xu et al. (2013)	PDAC	EUS images	Score the texture features of PDAC on EUS images and evaluate its prognostic value in patients with unresectable pancreatic cancer treated by EUS brachytherapy	DIA, fuzzy classification method	Internal	N/A	N/A

326	Yang et al. (2019)	PCN	CT, contrast-enhanced	To distinguish SCA from MCA.	Radiomics, RF, LASSO	Internal	2mm group: 74.0% (0.66) 5mm group: 83.0% (0.75)	2mm group: 86%/71% 5mm group: 85%/83%
313	Yeaton et al. (1998)	PDAC, CP	ERCP brush cytology	Distinguishing CP from PDAC on ERCP brush cytology	Decision tree method, production rule system.	Internal	88.9%	80.0% / 80.0%
298	Zhang et al. (2020)	PDAC	CT	To predict survival in PDAC	CNN	Internal/External	Index of prediction accuracy: 11.8% (from traditional radiomics method: 3.8%)	N/A
340	Zhang et al. (2010)	PDAC	EUS	Diagnose pancreatic cancer on EUS images	DIA, SVM	Internal	97.9%	94.4% / 99.5%
334	Zhang et al. (2020)	N/A	EUS images & video	DCNN1: Identify WL/EUS images and activate downstream models. DCNN2: filter unqualified images. DCNN3: recognize pancreas stations during scanning. DCNN4: segment landmarks and monitor pancreatic vision loss.	DCNN, BP MASTER system (station recognition RF classifier)	Internal/External	Station classification: 82.4% Segmentation: 90.0% Trainee recognition: 78.4% from, 67.2%	N/A
299	Zheng et al. (2020)	PDAC	MRI	Pancreas segmentation on MRI in the presence of PDAC	DCNN, 2D UNET, SE blocks, shadowed sets framework	Internal/External	99.9% local dataset 99.9% NIH dataset	Local dataset:64.4%/86.1% NIH dataset: 86.3%/83.1%
300	Zhu et al. (2015)	AIP, CP	EUS	Differentiate AIP from	SVM with LTPV	Internal	89.3%	84.1% / 92.5%

				CP on EUS	descriptor			
[314]	Zhu et al. (2013)	PDAC, CP	EUS images (enhanced/contrast)	Feasibility of CAD to differentiate between CP and PDAC	SVM	Internal	94.2%	96.3% / 93.4%

IPMN: Intraductal papillary mucinous neoplasm; EHR: Electronic health record; AP: Acute pancreatitis; ANN: Artificial neural network; PDAC: Pancreatic ductal adenocarcinoma; NN: Neural network; RF: Random forest; SVM: Support vector machine; LR: Logistic regression; AUC: Area under the curve; CT: Computed tomography; MRI: Magnetic resonance imaging; CP: Chronic pancreatitis; EUS: Endoscopic ultrasound; MCN: Mucinous cystic neoplasm; SCA: Serous cystadenoma; SPN: Solid pseudopapillary neoplasm; PSMVT: Portosplenomesenteric venous thrombosis; PNET: Pancreatic neuroendocrine tumor; DWI: Diffusion-weighed imaging; PET: Positron emission tomography; AIP: Autoimmune pancreatitis; CAD: Computer-aided diagnosis; RAP: Recurrent acute pancreatitis; CEH-EUS: Contrast-enhanced harmonic endoscopic ultrasound.