

Doctor vs AI Comparison Report

with Radiologic Findings Stratification

Patients compared: 114 | With labels: 109

Doctor

57.0%

AI (Scalar Regression)

56.1%

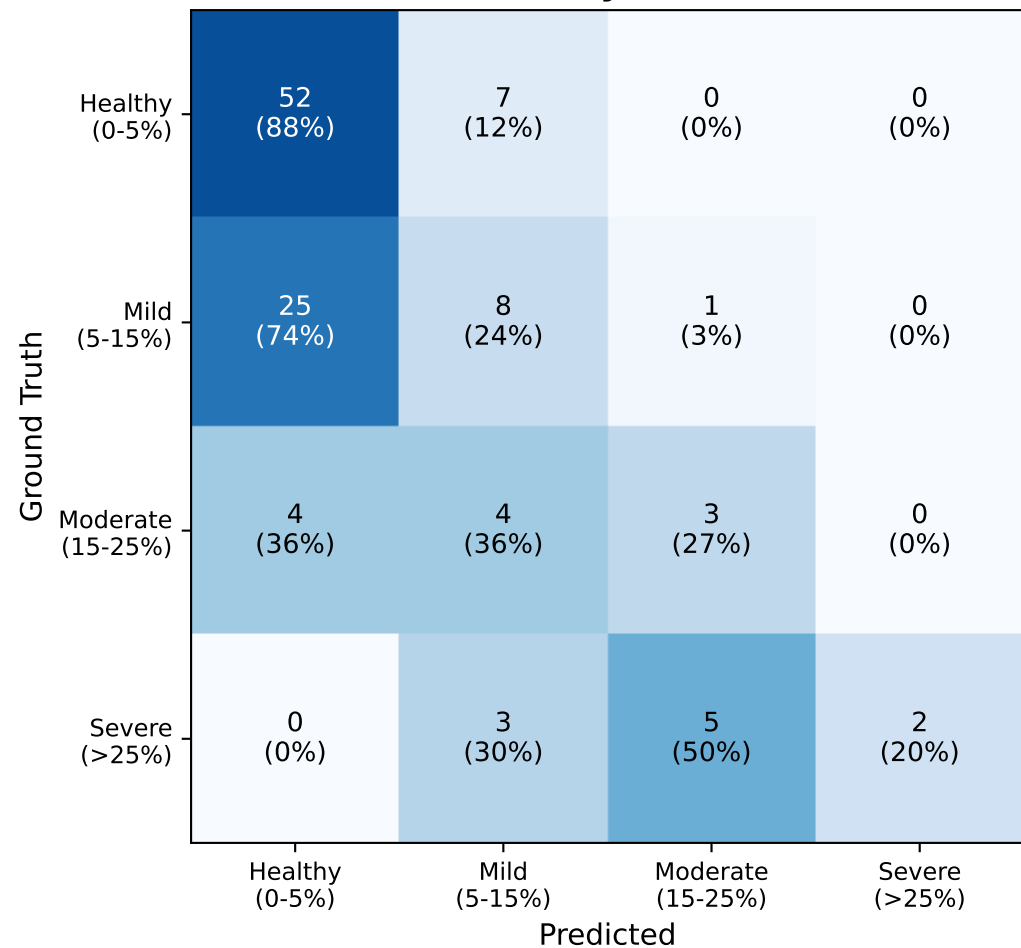
Finding prevalence (N=109 labeled):

Steatosis: 18 Cirrhosis: 21 Cholestasis: 21 High Iron: 5 Metastasis: 34

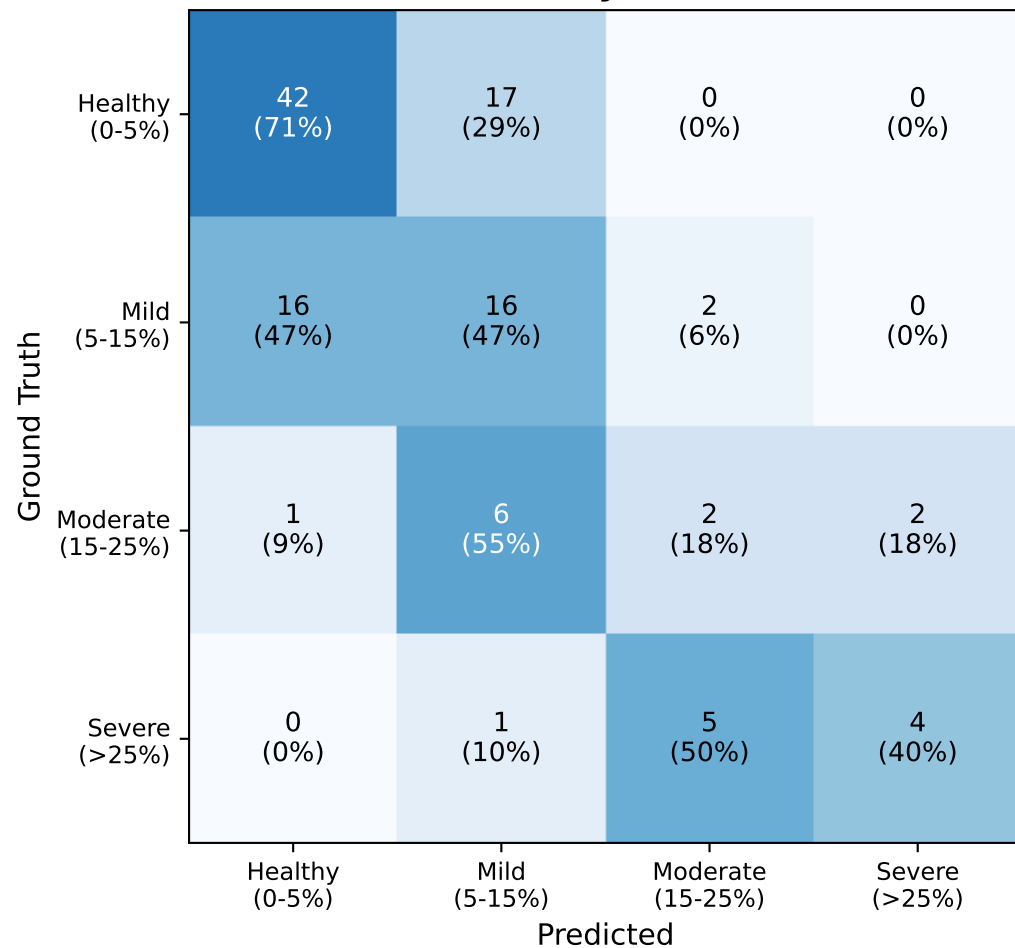
Clean (no findings): 26

Overall Confusion Matrices

Doctor
Accuracy: 57.0%



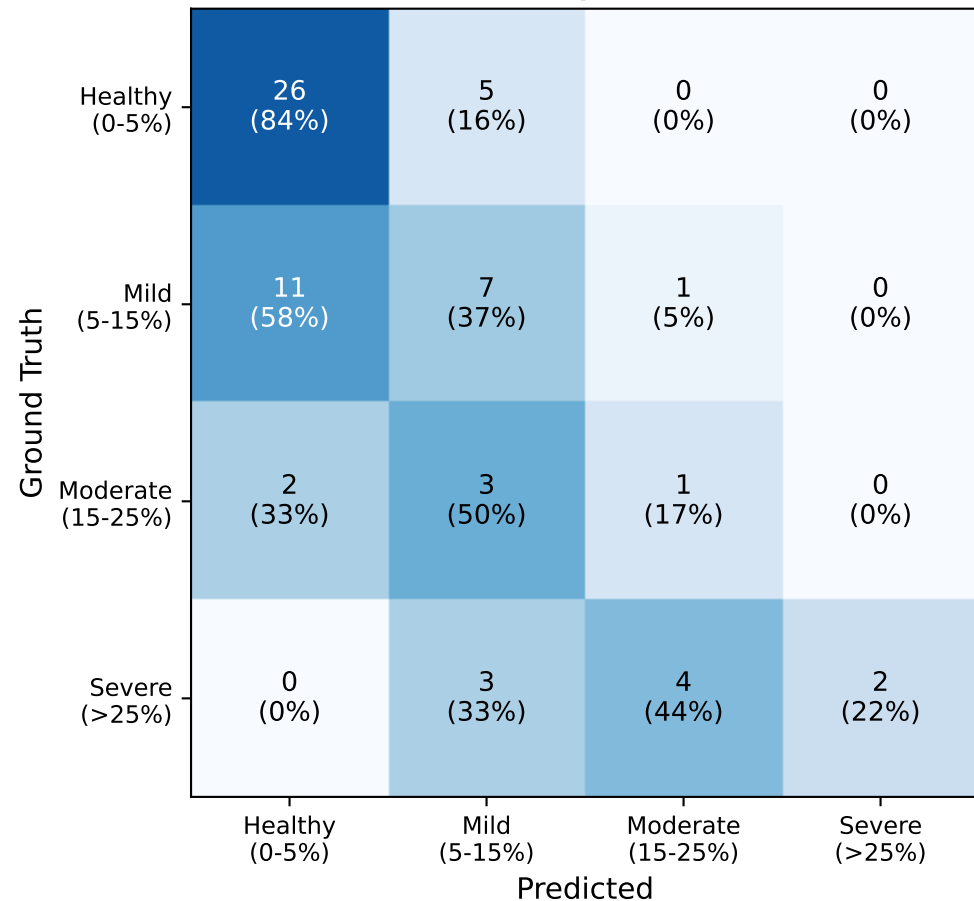
AI (Scalar Regression)
Accuracy: 56.1%



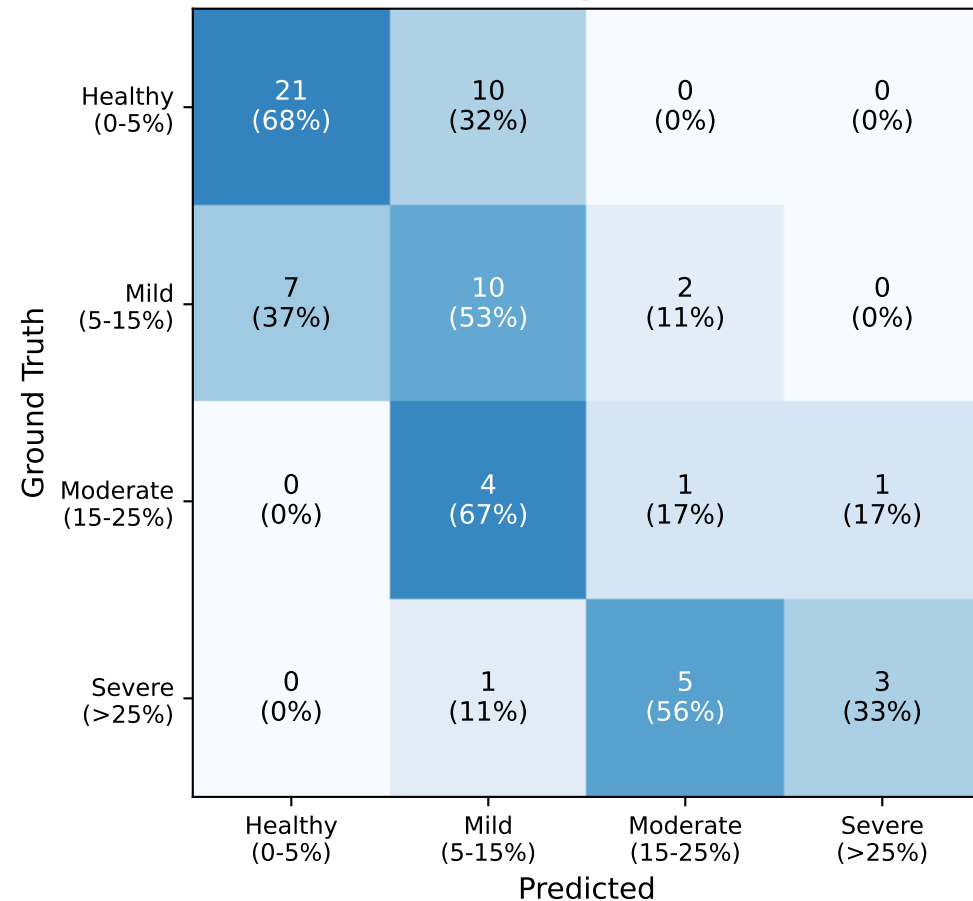
Filtered Confusion Matrices

Excluding cirrhosis, cholestasis, high iron (N=65, 44 removed)

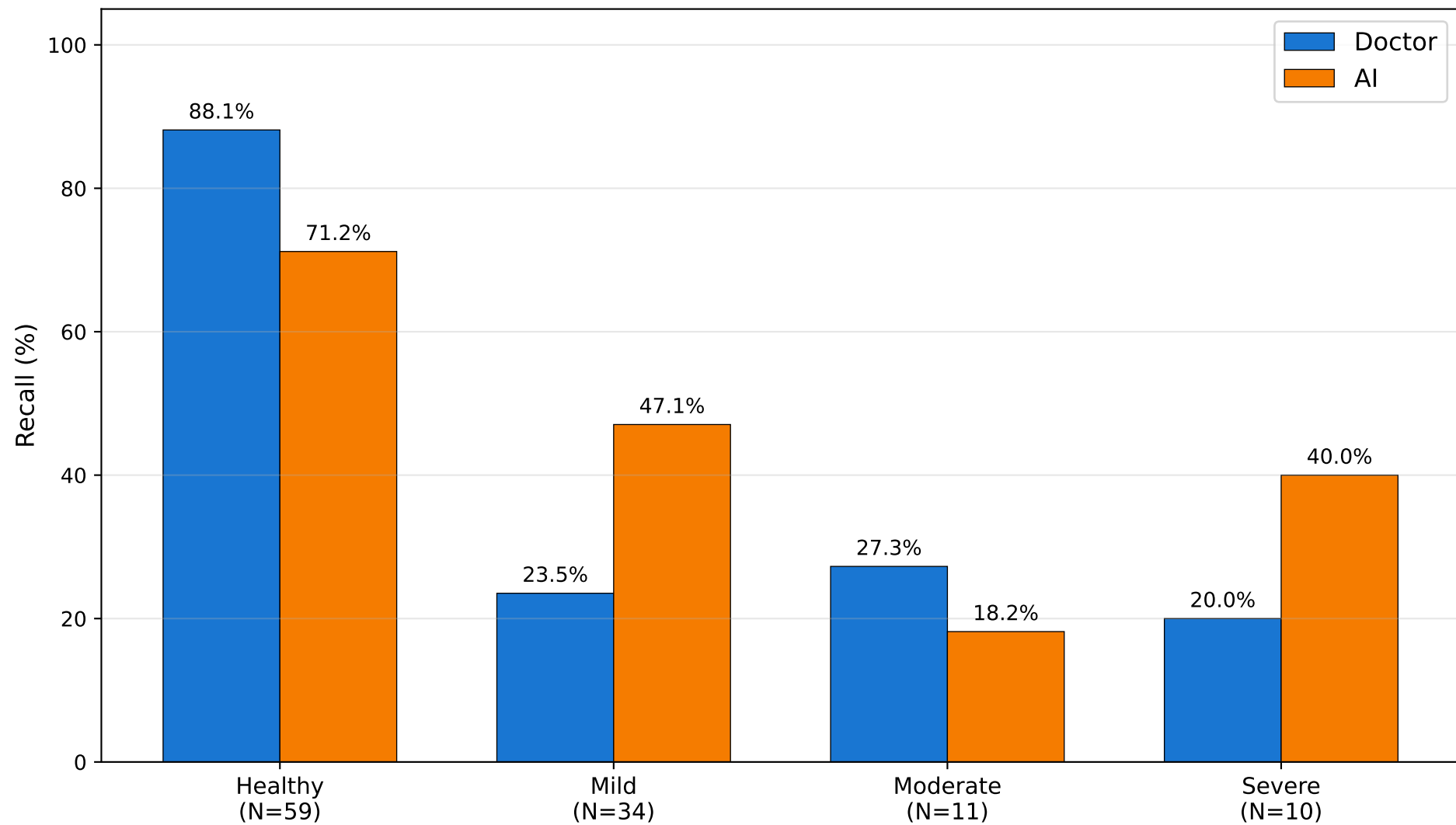
Doctor
Accuracy: 55.4%



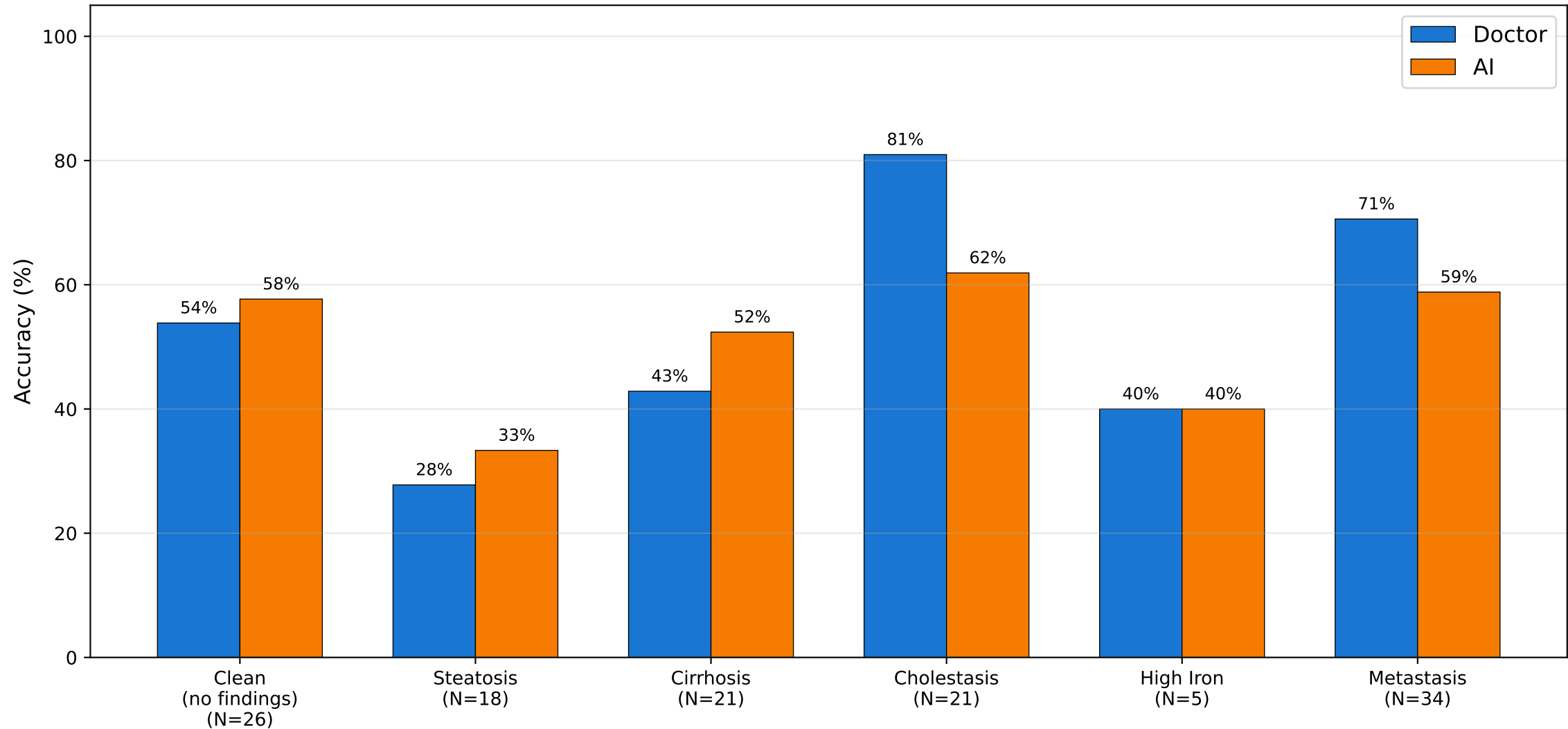
AI (Scalar Regression)
Accuracy: 53.8%



Per-Class Recall: Doctor vs AI



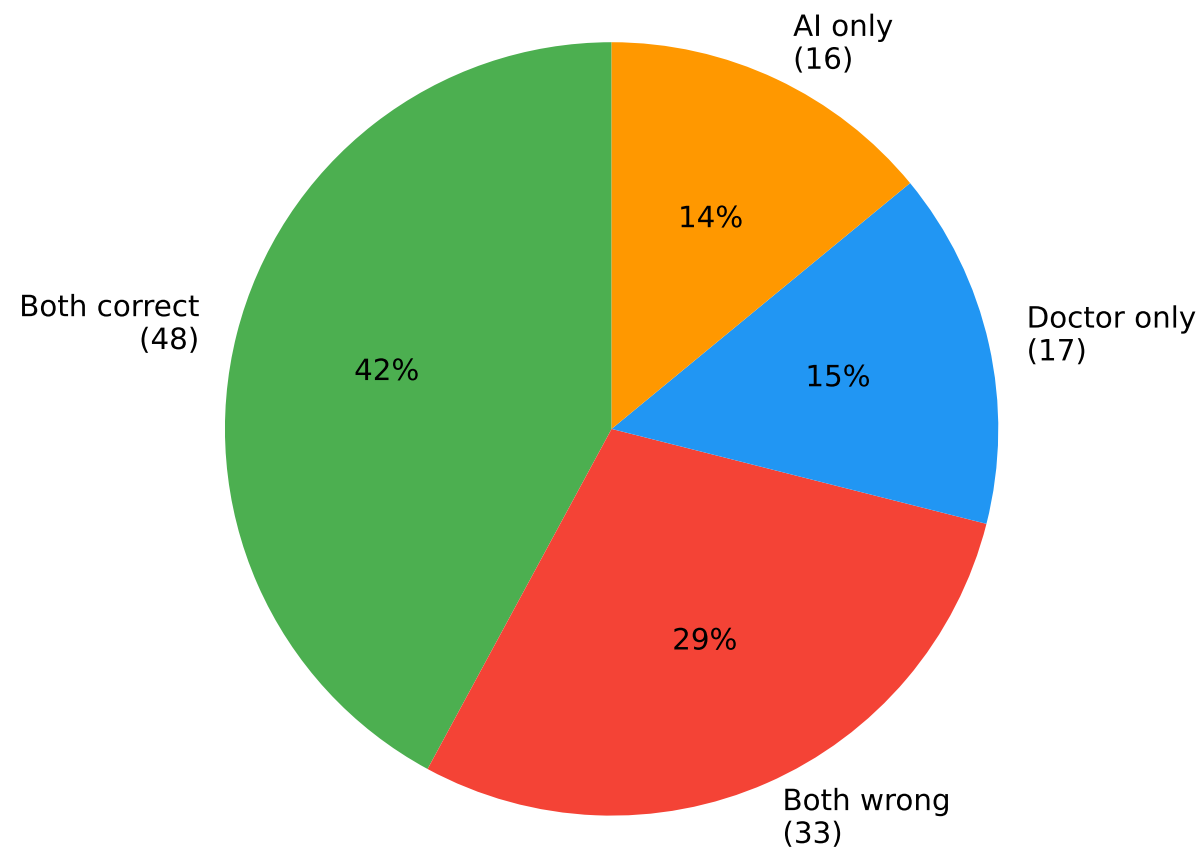
Accuracy by Radiologic Finding



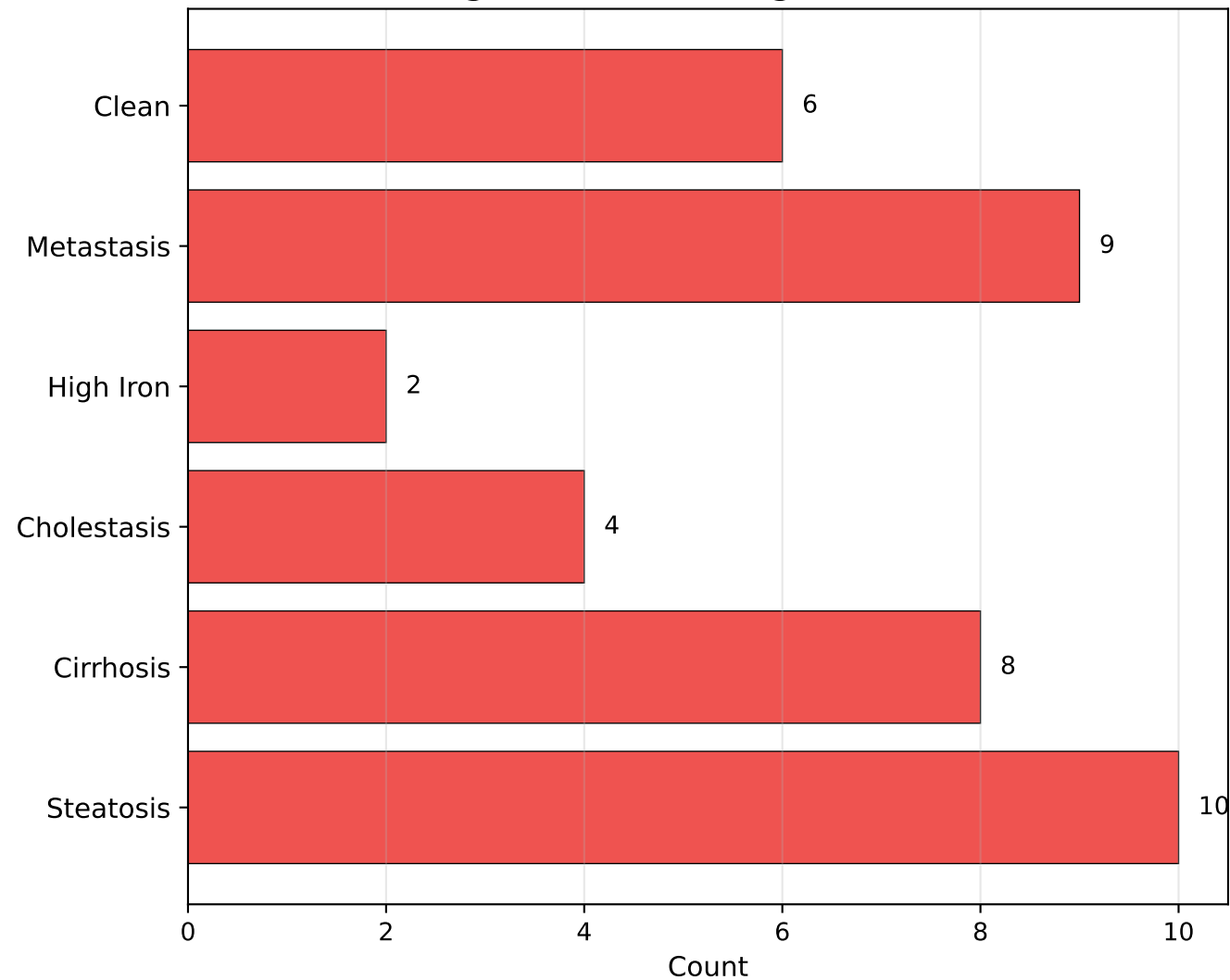
Detailed Accuracy by Radiologic Finding

Subgroup	N	Doctor	AI	Doc Hlthy	Doc Mild	Doc Mod	Doc Sev	AI Hlthy	AI Mild	AI Mod	AI Sev
ALL (labeled)	109	57.8%	55.0%	91%(55)	24%(33)	27%(11)	20%(10)	71%(55)	45%(33)	18%(11)	40%(10)
Clean	26	53.8%	57.7%	80%(15)	22%(9)	0%(2)	--	60%(15)	67%(9)	0%(2)	--
Steatosis	18	27.8%	33.3%	0%(1)	0%(1)	50%(6)	20%(10)	0%(1)	0%(1)	33%(6)	40%(10)
Cirrhosis	21	42.9%	52.4%	100%(8)	0%(9)	33%(3)	0%(1)	75%(8)	33%(9)	33%(3)	100%(1)
Cholestasis	21	81.0%	61.9%	100%(14)	25%(4)	67%(3)	--	79%(14)	25%(4)	33%(3)	--
High Iron	5	40.0%	40.0%	100%(2)	0%(2)	0%(1)	--	50%(2)	50%(2)	0%(1)	--
Metastasis	34	70.6%	58.8%	94%(18)	43%(14)	50%(2)	--	83%(18)	36%(14)	0%(2)	--
Multi-finding	13	46.2%	46.2%	100%(3)	17%(6)	67%(3)	0%(1)	100%(3)	17%(6)	33%(3)	100%(1)

Agreement Breakdown

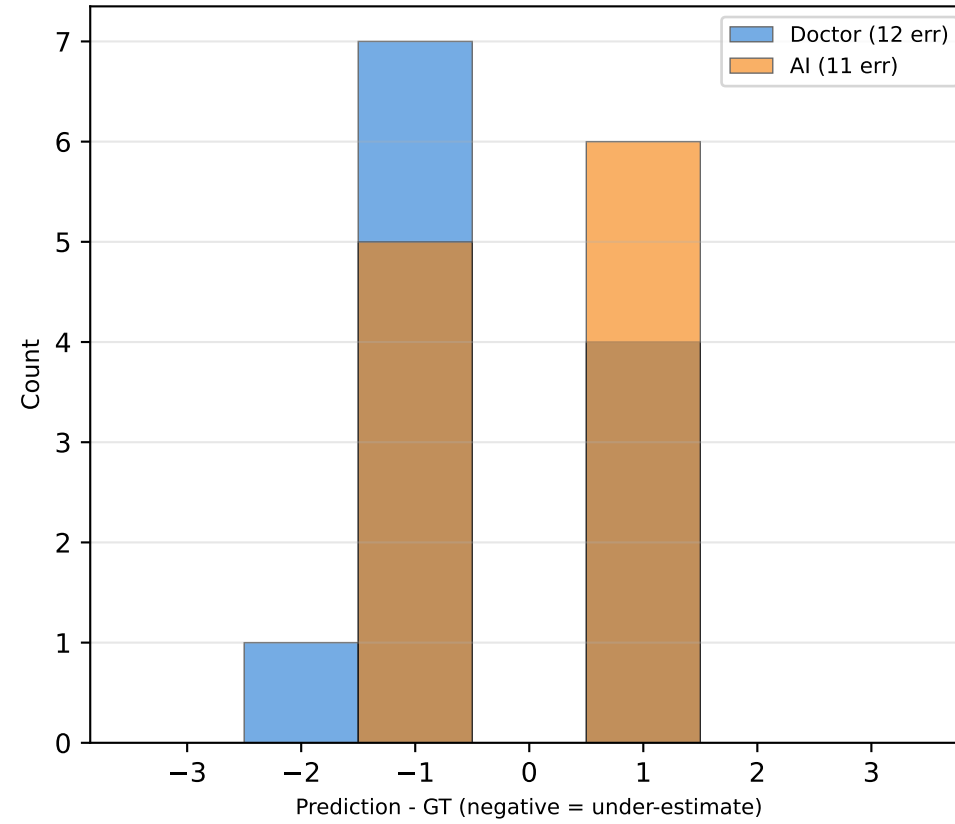


Findings in "Both Wrong" Cases (N=33)

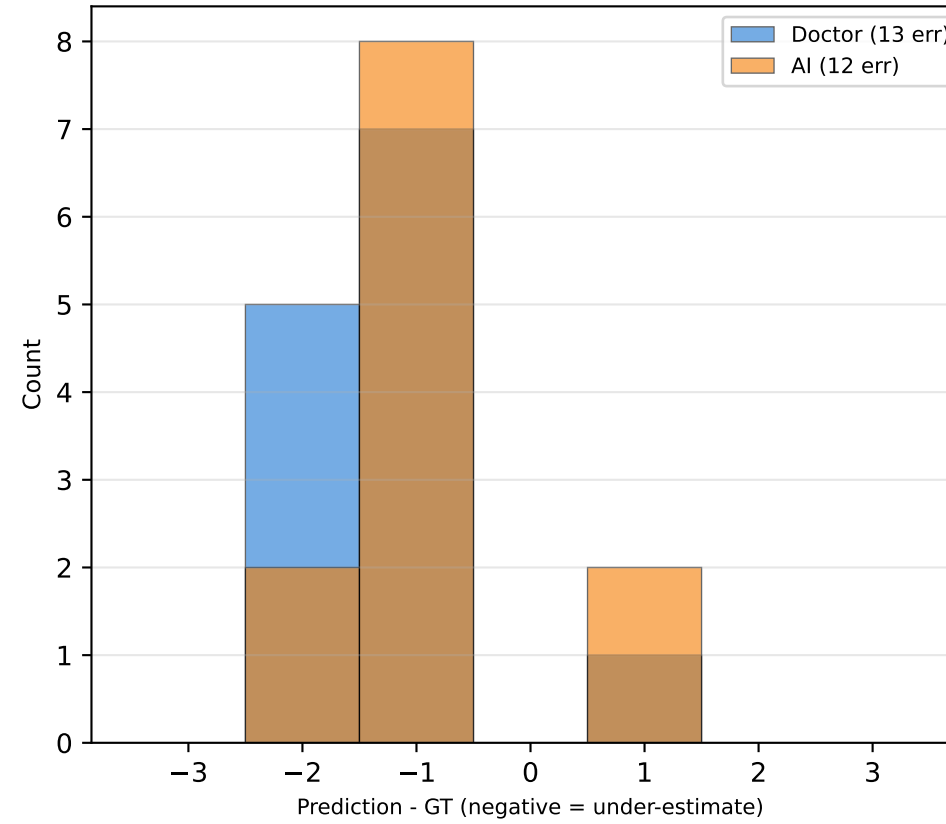


Error Direction by Finding (negative = under-estimation)

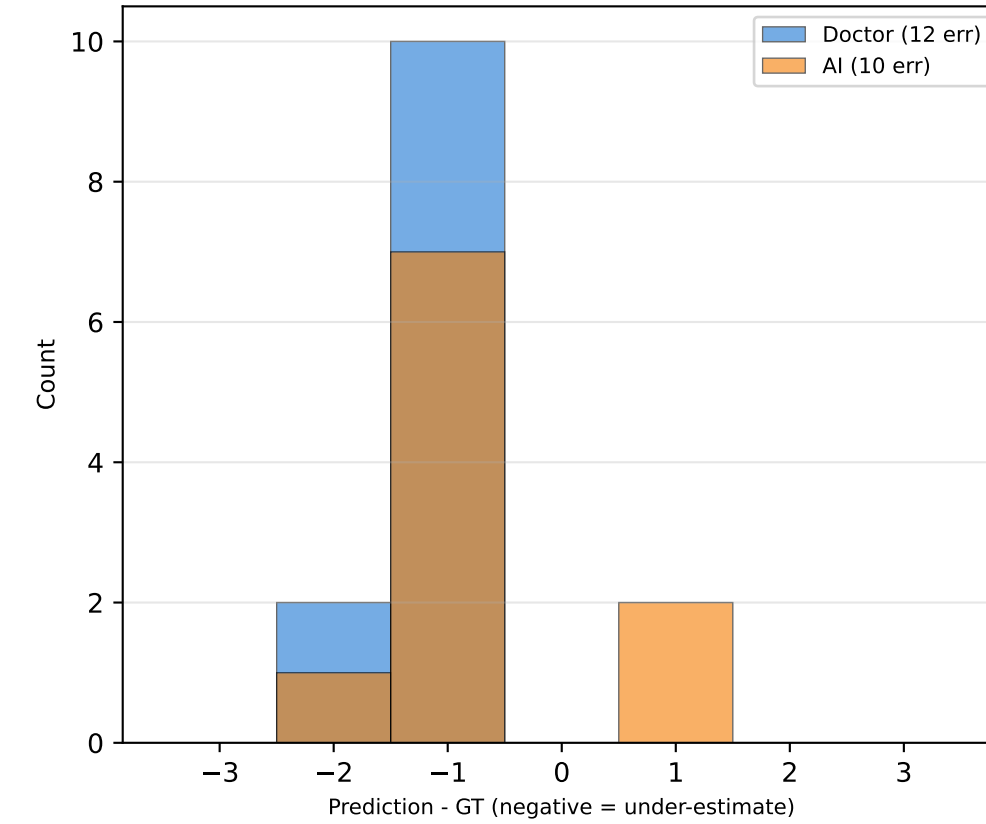
Clean (N=26)
Doc 54% | AI 58%



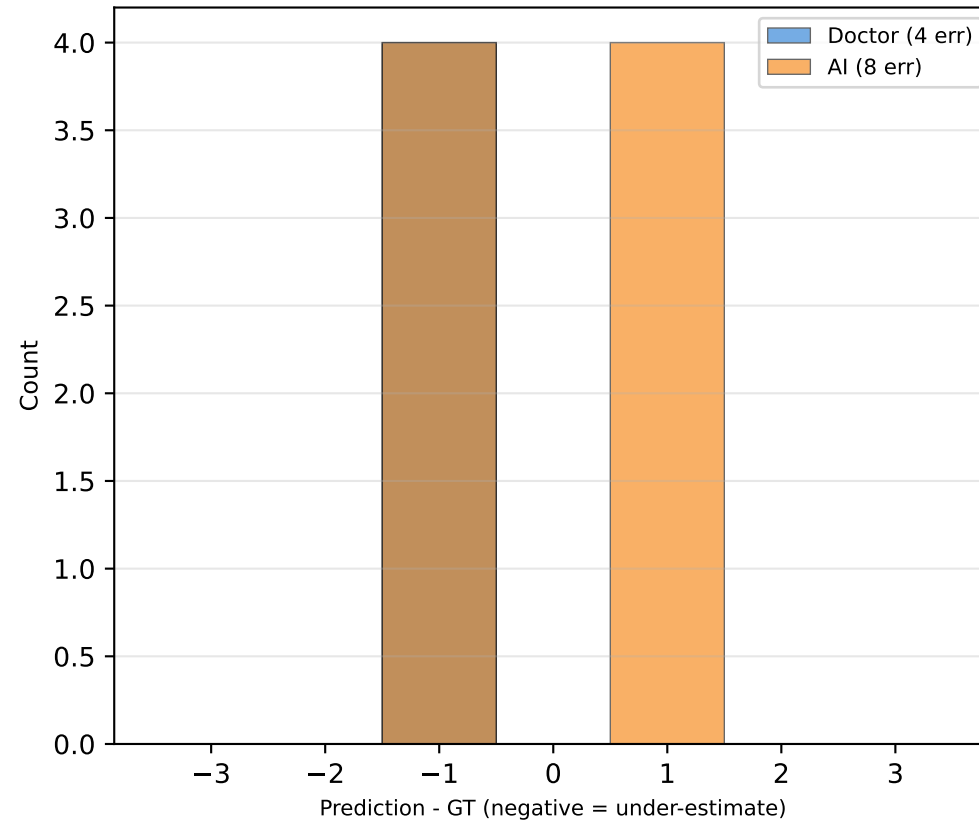
Steatosis (N=18)
Doc 28% | AI 33%



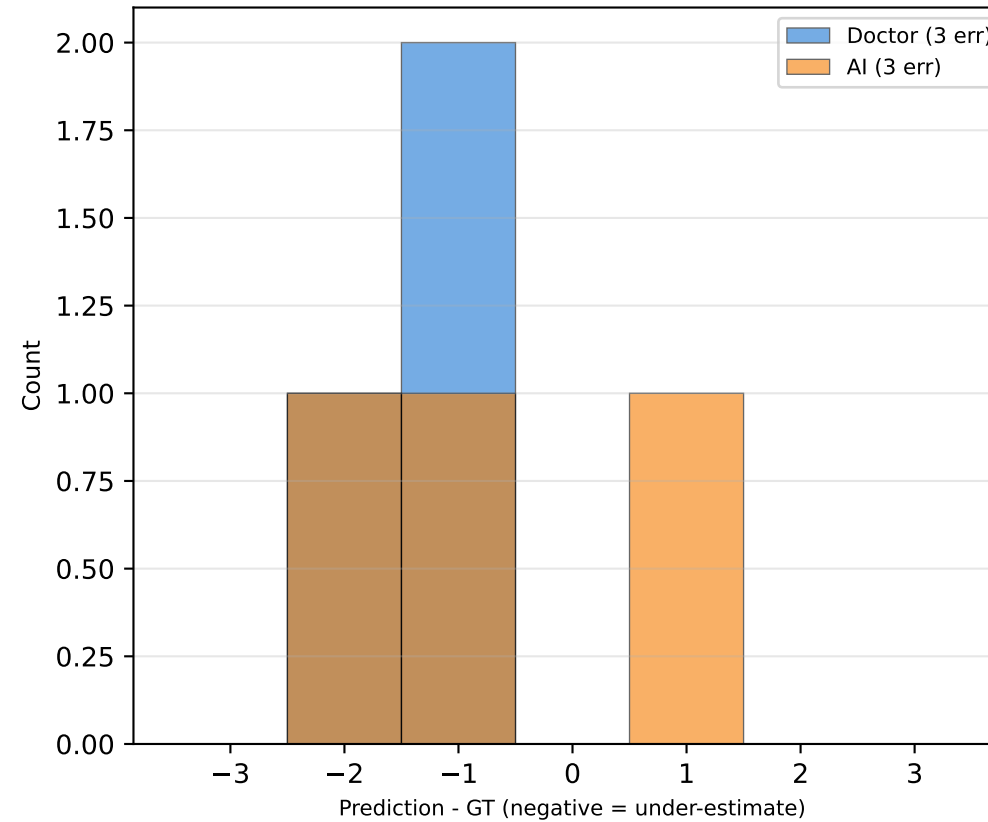
Cirrhosis (N=21)
Doc 43% | AI 52%



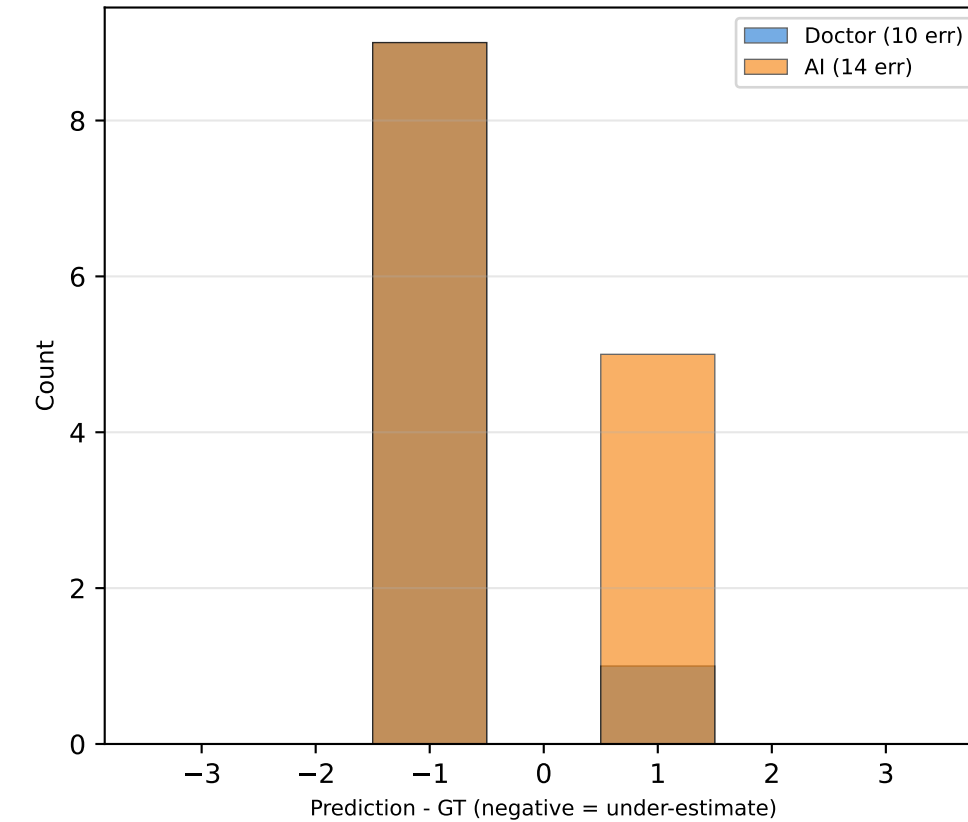
Cholestasis (N=21)
Doc 81% | AI 62%



High Iron (N=5)
Doc 40% | AI 40%



Metastasis (N=34)
Doc 71% | AI 59%



Misclassified Patients with Findings (66 total) (1/2)

Patient ID	GT	Doctor	AI	Steat	Cirrh	Chol	Iron	Meta	GT FF%
9149357101	Healthy	Healthy	Mild	.	.	.	X	.	4.2
9156553101	Healthy	Mild	Healthy	2.8
9178896301	Healthy	Healthy	Mild	.	.	X	.	.	4.4
0000577883	Healthy	Mild	Healthy	2.9
9179162001	Healthy	Healthy	Mild	.	X	.	.	.	3.2
9162557902	Healthy	Healthy	Mild	4.8
9179078701	Healthy	Healthy	Mild	3.5
9168240101	Healthy	Mild	Mild	X	3.2
9149855001	Healthy	Healthy	Mild	X	3.4
9163130001	Healthy	Healthy	Mild	X	3.2
9164358601	Healthy	Healthy	Mild	.	X	.	.	.	2.6
9168679501	Healthy	Mild	Mild	4.8
9181438501	Healthy	Mild	Mild	X	3.1
0000895710	Healthy	Healthy	Mild	.	.	X	.	.	3.9
9157103201	Healthy	Mild	Mild	3.4
9170309401	Healthy	Healthy	Mild	3.4
9165056801	Healthy	Mild	Mild	3.2
9171502101	Healthy	Healthy	Mild	4.0
9172900201	Healthy	Healthy	Mild	.	.	X	.	.	3.2
0000626275	Mild	Healthy	Mild	6.6
9176936101	Mild	Healthy	Healthy	5.1
9163424902	Mild	Healthy	Healthy	.	X	.	.	X	5.7
9163998901	Mild	Healthy	Healthy	X	6.4
9190810401	Mild	Moderate	Mild	5.9
0000807291	Mild	Healthy	Healthy	X	7.3
9186845701	Mild	Healthy	Healthy	6.8
9177343901	Mild	Healthy	Mild	X	8.4
9166885301	Mild	Healthy	Mild	8.1
0001409274	Mild	Healthy	Mild	5.8
9154307701	Mild	Healthy	Healthy	.	X	.	.	.	8.2
9154812401	Mild	Healthy	Healthy	.	X	X	.	.	8.1
9155037401	Mild	Healthy	Moderate	X	12.7
9180519901	Mild	Healthy	Healthy	.	X	.	.	.	5.7
9151660401	Mild	Healthy	Mild	10.6
9175532401	Mild	Healthy	Healthy	.	X	.	.	X	5.9
9160992501	Mild	Healthy	Healthy	.	X	.	.	X	6.1
9160625301	Mild	Healthy	Healthy	.	.	.	X	.	5.5
9177972101	Mild	Healthy	Mild	.	.	.	X	.	8.6
9176992301	Mild	Healthy	Healthy	.	.	X	.	X	5.6
9161101201	Mild	Healthy	Healthy	X	6.3

Misclassified Patients with Findings (66 total) (2/2)

Patient ID	GT	Doctor	AI	Steat	Cirrh	Chol	Iron	Meta	GT FF%
9167804601	Mild	Healthy	Mild	.	X	.	.	.	9.7
0000759325	Mild	Healthy	Healthy	.	.	X	.	.	9.2
9166229901	Mild	Healthy	Mild	.	X	.	.	.	11.0
9188883501	Mild	Healthy	Mild	.	X	.	.	.	5.2
9154212001	Mild	Mild	Moderate	X	14.8
9188752901	Mild	Mild	Healthy	8.4
9146422901	Mild	Mild	Healthy	X	5.1
0000024502	Mild	Healthy	Mild	7.2
9159113501	Moderate	Healthy	Mild	.	X	.	.	.	16.6
9152842601	Moderate	Healthy	Mild	X	15.5
9172655501	Moderate	Healthy	Healthy	X	X	.	X	.	24.8
0000470234	Moderate	Healthy	Mild	15.8
9176982801	Moderate	Moderate	Mild	X	.	X	.	X	23.7
9149561702	Moderate	Moderate	Mild	X	18.2
9167712201	Moderate	Mild	Mild	17.9
9150700001	Moderate	Mild	Severe	.	.	X	.	.	20.1
9172794101	Moderate	Mild	Severe	X	24.1
9161261301	Moderate	Mild	Moderate	X	24.0
0000322494	Severe	Mild	Mild	X	27.9
9181938701	Severe	Mild	Moderate	X	34.4
9176337201	Severe	Mild	Moderate	X	27.3
9176798701	Severe	Moderate	Moderate	X	29.3
9165595401	Severe	Moderate	Moderate	X	29.9
9180922001	Severe	Moderate	Severe	X	X	.	.	.	27.5
9155208901	Severe	Moderate	Severe	X	36.8
9145643301	Severe	Moderate	Moderate	X	38.9