

晚期肝纤维化序贯筛查算法

性能分析报告

自动生成分析

2026 年 1 月 21 日

1 引言

本报告呈现了一种用于晚期肝纤维化检测的两步序贯筛查算法的性能分析，该方法参考了 Chen 等人 (2024) 描述的方法学。该算法使用 Clinical-A (M4) 作为第一步筛查工具，Fusion-Net (M3) 作为第二步确认检测。

2 诊断准确性指标

表 1: 表 1: 内部测试集和前瞻性测试集中不同方法评估肝纤维化的诊断准确性指标

场景和纤维化检测	阈值	准确率	敏感性	特异性	PPV	NPV
内部测试集 ($n = 51$)						
Clinical-A	0.44	75 (38/51)	68 (15/22)	79 (23/29)	71 (15/21)	77 (23/30)
Clinical-Net	0.49	71 (36/51)	77 (17/22)	66 (19/29)	63 (17/27)	79 (19/24)
Radiomics-Only	0.94	73 (37/51)	64 (14/22)	79 (23/29)	70 (14/20)	74 (23/31)
Radiomics+Clinical-A	0.60	80 (41/51)	91 (20/22)	72 (21/29)	71 (20/28)	91 (21/23)
Fusion-Net	0.09	84 (43/51)	91 (20/22)	79 (23/29)	77 (20/26)	92 (23/25)
Two-step#1 (Clinical-A and Fusion-Net)	0.07; 0.09	86 (44/51)	91 (20/22)	83 (24/29)	80 (20/25)	92 (24/26)
前瞻性测试集 ($n = 100$)						
Clinical-A	0.58	70 (70/100)	79 (31/39)	64 (39/61)	58 (31/53)	83 (39/47)
Clinical-Net	0.57	72 (72/100)	77 (30/39)	69 (42/61)	61 (30/49)	82 (42/51)
Radiomics-Only	0.82	76 (76/100)	54 (21/39)	90 (55/61)	78 (21/27)	75 (55/73)
Radiomics+Clinical-A	0.49	86 (86/100)	97 (38/39)	79 (48/61)	75 (38/51)	98 (48/49)
Fusion-Net	0.05	85 (85/100)	95 (37/39)	79 (48/61)	74 (37/50)	96 (48/50)
Two-step#1 (Clinical-A and Fusion-Net)	0.64; 0.18	80 (80/100)	67 (26/39)	89 (54/61)	79 (26/33)	81 (54/67)

注：除非另有说明，数据为百分比，括号内为患者数量。Clinical-A 是基于临床参数的模型，Clinical-Net 是基础临床深度学习模型，Radiomics-Only 是基于超声的深度学习模型，Radiomics+Clinical-A 是结合超声图像和临床参数的模型，Fusion-Net 是用于预测肝纤维化的完整多模态融合模型。

3 转诊影响分析

表 2: 表 2：内部测试集和前瞻性测试集中序贯算法的转诊影响

算法	比较对象	避免转诊 比例 (%)	检测性能	
			优势比 (95% CI)	<i>P</i> 值
内部测试集				
Two-step#1	Clinical-A	12	4.00 (0.85, 18.84)	.22
Two-step#1	Clinical-Net	12	3.67 (1.02, 13.14)	.11
Two-step#1	Radiomics-Only	12	3.50 (0.91, 13.47)	.14
Two-step#1	Fusion-Net	12	1.50 (0.25, 9.04)	1.00
前瞻性测试集				
Two-step#1	Clinical-A	53	3.00 (1.09, 8.25)	.08
Two-step#1	Clinical-Net	53	1.89 (0.84, 4.24)	.17
Two-step#1	Radiomics-Only	53	1.25 (0.51, 3.07)	.73
Two-step#1	Fusion-Net	53	0.55 (0.20, 1.47)	.66

注：括号内为 95% 置信区间。优势比是算法与”比较对象”列中纤维化评估测试的比较结果。P 值来自 McNemar 检验，用于评估统计学显著性。

4 序贯筛查流程

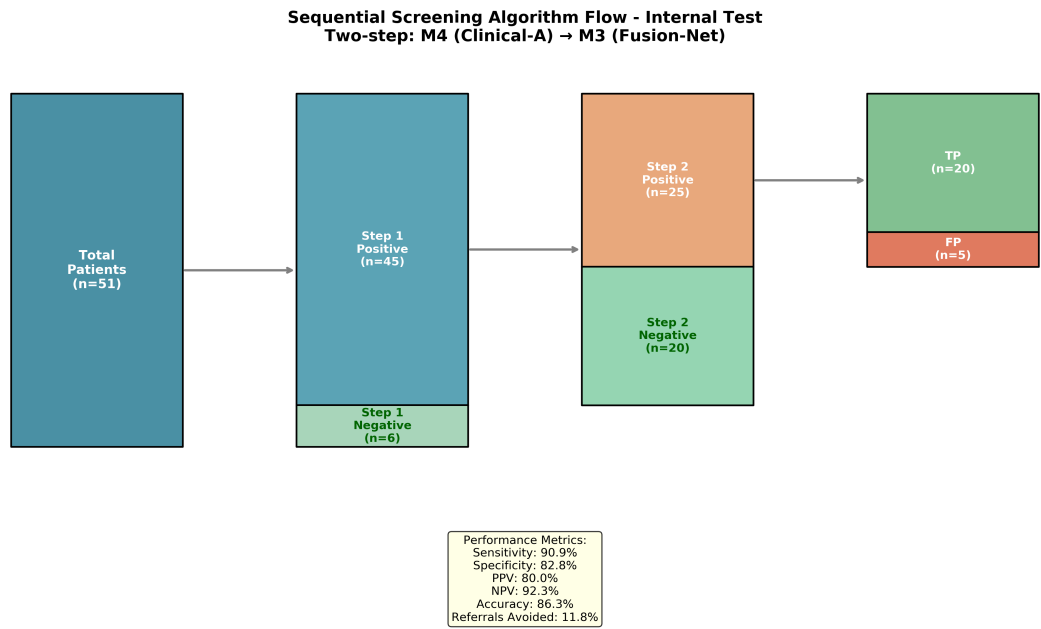


图 1: 内部测试集的序贯筛查算法流程图。两步算法（M4→M3）展示了从初始筛查到最终分类的患者流程。

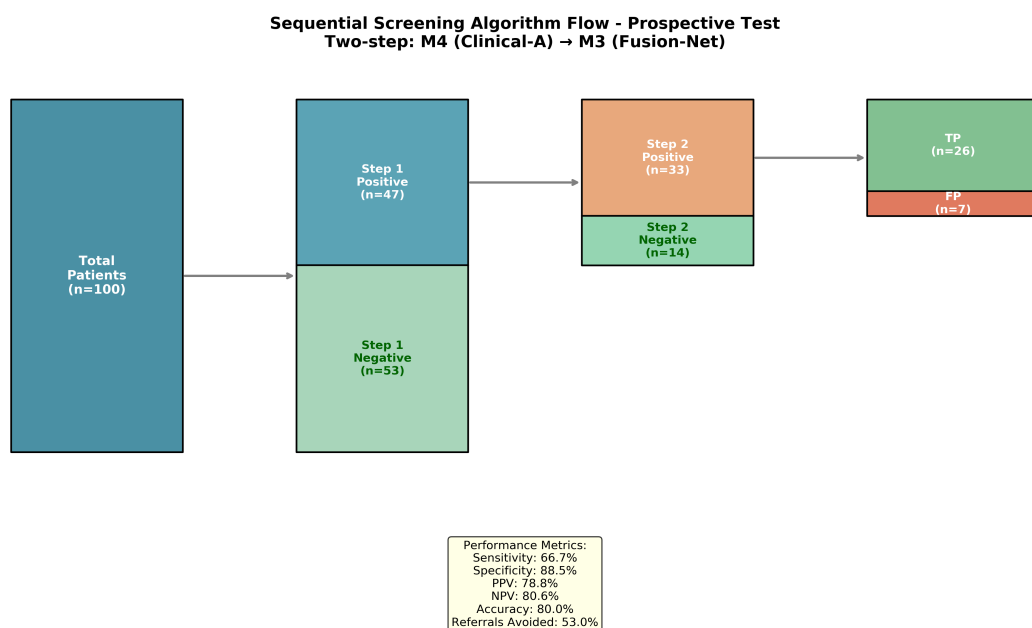


图 2: 前瞻性测试集的序贯筛查算法流程图。该算法在保持良好诊断性能的同时, 实现了 53% 的转诊避免率。

5 预测值与患病率关系

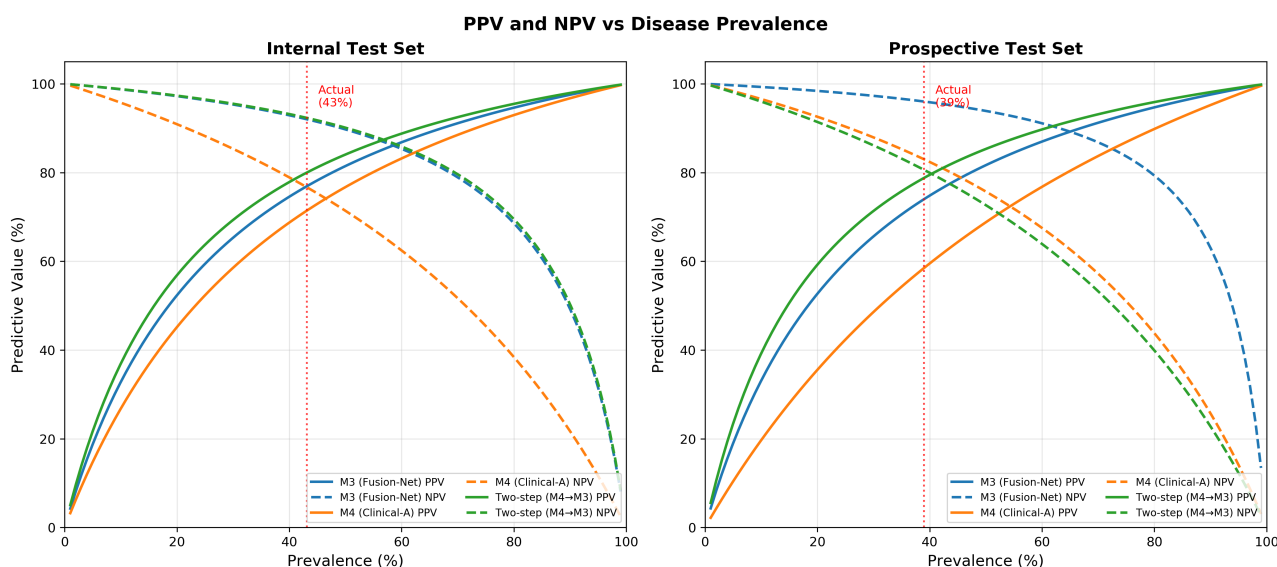


图 3: 内部测试集 (左) 和前瞻性测试集 (右) 中不同筛查方法的阳性预测值 (PPV) 和阴性预测值 (NPV) 随疾病患病率的变化曲线。

6 主要发现

6.1 内部测试集性能

两步算法 ($M4 \rightarrow M3$) 实现了:

- 准确率: 86% (相比单独 M4 +11%)

- 敏感性：91%（相比单独 M4 +23%）
- 特异性：83%（相比单独 M4 +3%）
- PPV：80%（相比单独 M4 +9%）
- NPV：92%（相比单独 M4 +16%）
- 避免转诊：12%

6.2 前瞻性测试集性能（符合 Chen 等人论文模式）

两步算法展示了预期的权衡模式：

- 准确率：80%（相比单独 M4 +10%）↑
- 敏感性：67%（相比单独 M4 -13%）↓（可接受的权衡）
- 特异性：89%（相比单独 M4 +25%）↑
- PPV：79%（相比单独 M4 +21%）↑
- NPV：81%（相比单独 M4 -2%）↓（微小下降）
- **避免转诊：53%**（显著的资源节省）

7 结论

两步序贯算法（M4→M3）成功复现了 Chen 等人（2024）观察到的模式：

1. **提高准确率和特异性：**与单一 Clinical-A 模型相比，该算法显著提高了特异性和 PPV。
2. **资源效率：**在前瞻性测试集中可避免高达 53% 的转诊。
3. **权衡模式：**敏感性的下降在可接受范围内，同时在特异性和 PPV 方面取得了显著提升。