arXiv:2303.06966v1 [stat.AP] 13 Mar 2023

一种基于临床病理数据和相似肿瘤特征预测 Oncotype 评分的新方法

Zeina Al Masry[1], Romain Pic[2], Cl´ement Dombry[2], Christine Devalland[3]

[1]法国国家科学研究中心（CNRS）与 SUPMICROTECH-ENSMM 联合所属 FEMTO-ST研究所，勃艮第-弗朗什-孔泰大学，法国贝桑松 Savary 街 25 号[2]贝桑松数学实验室，法国国家科学研究中心 UMR 6623，勃艮第-弗朗什-孔泰大学，法国贝桑松[3]北弗朗什-孔泰医院病理解剖与细胞学科，法国特雷弗南 Moval 路 100 号，邮编 90400。

摘要

引言： Oncotype DX（ODX）检测是一种市售的乳腺癌分子检测，可为激素受体阳性、HER2阴性患者提供乳腺癌复发风险的预后和预测信息。本研究旨在提出一种新方法，协助医生进行决策。**方法：** 本研究为回顾性研究，纳入了 2012 至 2020 年间来自勃艮第-弗朗什-孔泰地区三家医院接受 ODX检测的 333 例病例。研究人员利用临床和病理报告收集数据，并基于 9 项临床病理特征开发了一种分布式随机森林方法。该方法尤其可用于识别训练队列中与新患者具有相似特征的患者，并预测 ODX评分分布的估计值。

结果： 参与者的平均年龄为56.9岁。我们正确分类了92%的低风险患者和40.2%的高风险患者。总体准确率为79.3%。低风险患者的正确预测值（PPV）比例为82%。高风险患者的正确预测值（NPV）约为62.3%。F1分数和曲线下面积（AUC）分别为0.87和0.759。

结论： 所提出的方法能够预测患者的ODX评分分布，并对预测评分作出解释。将该方法与病理学家对不同组织学和免疫组织化学特征的专业知识相结合，具有临床意义，有助于肿瘤科医生在乳腺癌治疗方面进行决策。

关键词： 乳腺癌、Oncotype DX、临床病理数据、机器学习、分布式随机森林。

1 引言

Oncotype DX（ODX）检测是Genomic Health推出的一种可商业获取的乳腺癌分子检测，可为激素受体阳性、HER2阴性患者提供乳腺癌复发的预后和预测信息。ODX检测基于1 A级证据，并被纳入主要国际临床指南的建议中

例如美国临床肿瘤学会（ASCO [3]）和美国国家综合癌症网络（NCCN）的指南，以及AJCC第8版 [12]最新分期指南。ODX检测是全球应用最广泛的分子检测。该检测通过实时定量聚合酶链式反应分析21个基因（16个癌症相关基因和5个管家基因），旨在通过提供0至100分的复发评分，预测10年复发风险并评估辅助化疗的获益。多项回顾性和前瞻性研究验证了该检测及其临床效用。Paik等人 [23] 基于NSABP B-14试验，发现接受他莫昔芬治疗的雌激素受体阳性/HER2阴性、淋巴结阴性乳腺癌患者的ODX评分与无病生存期相关。关于化疗获益，Paik等人 [24] 和Albain等人[1] 利用NSABP-B20和SWOG 8814相关研究评估了该检测。前瞻性III期试验TAILORx研究 [27] 调整了ODX评分的临界值（低风险 <11、中等风险11-25和高风险 >25），以避免癌症治疗不足。具体而言，低风险组接受激素治疗后，5年复发风险极低（$< 10\%$），证实无需加用化疗[26]。对于中等风险组，化疗仅对年龄小于50岁的女性有获益。高风险组则强烈建议接受化疗。然而，三分之一的激素受体阳性乳腺癌女性存在淋巴结受累。因此，前瞻性RxPONDER试验研究评估了ODX检测预测淋巴结阳性患者化疗获益的能力 [16]。RxPONDER研究显示，淋巴结受累且ODX评分为0至25的绝经后患者无法从化疗中获益；而淋巴结受累、受累淋巴结为1至3个且ODX评分为0至25的绝经前患者则能显著从化疗中获益。

尽管已证实 ODX检测具有价值，但由于成本高昂，目前尚未常规使用。因此，欧洲不到20%的患者能够接受 ODX检测。开展卫生经济学研究，旨在了解该检测对哪些患者最有用[2]。从卫生经济学角度出发，人们开发了许多替代工具来预测这一评分。这些工具基于临床病理数据，例如 Magee方程[18, 28]和 IHC4评分 [30]。事实上，许多研究已显示，后者工具的结果与 ODX评分 [11]之间存在相关性。少数研究运用机器学习技术处理特征，提出基于 ODX 的方法，将患者划分为癌症低风险或高风险类别 [17, 22, 5, 25]。

本文旨在提出一种新方法，以协助医生进行决策。该方法基于 Meinshausen [21]和 Athey 等人 [4]提出的分布回归随机森林。该方法依据临床病理特征，建立新患者与用于训练的队列之间的关联。这些关联尤其可用于识别训练队列中与新患者相似的患者，并预测ODX评分分布的估计值。临床医生可利用这些信息，更好地了解肿瘤可能的临床演变，从而优化治疗。

此外，通过反馈和分析患者病史，该方法能够积累知识。本研究的一个结果是对解剖病理学数据的变异性进行加权，使这一新方法能够适应特定队列的特点。

2 材料与方法

2.1 数据集描述

该队列为一项回顾性研究，纳入了2012年至2020年间来自勃艮第-弗朗什-孔泰⁄大区三家医院的333例接受ODX检测的病例，医院分别位于贝桑松、贝尔福和第戎。所有患者均患有ER阳性、HER2阴性的早期乳腺癌。研究人员根据临床和病理报告收集了诊断年龄、绝经状态、治疗、复发、肿瘤大小、淋巴结状态、组织学类型、诺丁汉分级、雌激素受体（ER）表达、孕激素受体（PR）表达、人表皮生长因子受体2（HER2）状态，以及蛋白质p53和Ki67增殖指数等数据。对用于ODX检测的肿瘤组织块进行免疫组织化学染色（Ventana Benchmark XT系统®，RocheTM），采用UltraView Universal DAB检测，使用ER抗体（克隆SP1；Roche/Ventana Medical Systems，美国图森）、PR抗体（克隆1E2；Roche/Ventana Medical Systems，美国图森）、HER2抗体（克隆4B5；Roche/Ventana Medical Systems，美国图森）、Ki67抗体（克隆Mib-1，Dako，丹麦格洛斯楚普）和p53抗体（克隆DO-7，Dako，丹麦格洛斯楚普）。HER2免疫染色结果依据2018年美国临床肿瘤学会/美国病理学家学会指南进行判读 [29]。Ki67增殖指数通过计数器手动计数至少200个肿瘤细胞进行评估。采用免疫组织化学方法评估蛋白质p53。阳性阈值为肿瘤细胞核中阳性比例超过10%。ODX检测由Genomic Health（美国加利福尼亚州雷德伍德城）完成，并对石蜡包埋肿瘤组织块中的21个基因进行实时定量聚合酶链式反应分析。ODX评分取自临床报告。三个ODX分类与采用TAILORx和RxPONDER的ODX检测中定义的分类相同：低风险（< 16）、中等风险（16-25）和高风险（> 25）。机构审查委员会批准了本研究。

该队列包含 50 多个特征，我们根据随机森林中的特征重要性以及医生评估，从中选出了最关键的特征。表 1 使用本研究选取的特征描述了肿瘤特征。

2.2 分布随机森林

随机森林 [7]是一种功能强大的机器学习算法，可用于各种情境下的预测，并已成功应用于医学领域 [9, 10, 31]。我们的目标是根据临床病理特征预测昂贵的 ODX 检测结果。我们提出使用分布随机森林，根据临床病理特征提供 ODX 评分的预测分布。下面将介绍随机森林和分布随机森林的方法。

标准回归将响应变量 Y 的均值与一组特征 X联系起来，依据的是由特征-响应对构成的训练样本中的观测值，例如 (X_i， Y_i)，其中$i = 1$，.....，n。随机森林（RF）预测是一种集成方法，由随机化分类与回归树（CART， [8]）的自助聚合 [6] 组成。预测均值可写为以下平均值

$$\hat{Y} = \frac{1}{B}\sum_{b=1}^{B} T_b(X), \tag{1}$$

		按类别划分的患者百分比			
		< 16	$16 - 25$	> 25	总计
人群		113	138	82	333
Age	≤ 50 年	14.41	10.51	5.71	30.63
	> 50 年	19.52	30.92	18.92	69.37
肿瘤大小	$<$ 1厘米	3.90	4.51	3.00	11.41
	1-2 厘米	15.62	22.52	15.62	53.76
	$>$ 2厘米	14.41	14.41	6.01	34.83
p53	$\leq 10\%$	18.62	23.12	12.01	53.75
	$> 10\%$	15.32	18.32	12.61	46.25
SBR分级	1	5.41	3.30	0.00	8.71
	2	21.02	24.03	10.81	55.86
	3	7.51	14.11	13.81	35.43
有丝分裂分级	1	12.61	14.42	4.50	31.53
	2	17.12	20.12	12.01	49.25
	3	4.20	6.91	8.11	19.22
ER状态	阴性	0.00	0.00	0.00	0.00
	阳性（$\geq 10\%$）	33.93	41.44	24.63	100
PR状态	阴性	2.10	7.51	8.11	17.72
	阳性（$\geq 10\%$）	31.83	33.93	16.52	82.28
Ki67阳性细胞	$< 10\%$	0.00	0.30	0.00	0.30
	$10 - 20\%$	16.22	15.92	4.80	36.94
	$> 20\%$	17.72	25.22	19.82	62.76
淋巴结状态	0	15.02	15.52	13.81	45.35
	1	10.81	14.11	4.20	29.13
	2	3.61	3.30	0.90	7.81
	3	1.80	2.10	2.10	6.00
	NA	2.70	5.41	3.60	11.71

表1：按患者和肿瘤特征划分的ODX评分分布。

其中， $T^1(X)\ldots,T^B(X)$对应于基于不同自助采样样本构建的各棵树的预测。每棵树的预测都是对树中 X 的一个邻域取平均，即

$$T_b(X)=\frac{1}{|R_b(X)|}\sum_{X_i\in R_b(X)}Y_i,$$

其中， $R_b(X)$是树 T_b 中特征空间中包含 X 的区域，而 $|R_b(X)|$表示落入该区域的观测值数量。因此，随机森林预测（1）可等价地写为

$$\hat{Y}=\sum_{i=1}^{n}w_i(X)Y_i, \tag{2}$$

其中，随机森林权重定义为

$$w_i(X)=\frac{1}{B}\sum_{i=1}^{B}\frac{\mathbb{1}_{\{X_i\in R(X)\}}}{|R_b(X)|},\quad 1\le i\le n, \tag{3}$$

这些权重均为非负数，且总和为 1（概率权重）。

分布式随机森林（DRF）的核心思想基于公式（2）：预测值 $\hat{}Y$ 是加权样本 Y_i 的样本均值，其权重为 $w_i(X)$；该加权样本可视为给定 X时 Y 条件分布的近似。因此，累积分布函数 $F(y|X)=\mathrm{P}(Y\le y|X)$可近似为

$$\hat{F}(y|X)=\sum_{i=1}^{n}w_i(X)\mathbb{1}_{\{Y_i\le y\}}. \tag{4}$$

这一思想最早由 Meinshausen [21] 提出，他通过用加权经验分布（4）的分位数近似给定 X 时 Y 的条件分位数，构造了分位数回归森林。

图1展示了 DRF 流程及其各个步骤的示意图：对原始样本进行子抽样，在每个子样本上构建树，计算待预测点的邻域/权重，并对不同树给出的权重取平均，最终得到预测分布。

随机森林权重 (3) 本身就很有意义，能够提供有关相似观测值/重要观测值的相关信息。给定一个新的特征 X，权重$w_i(X)$ 可解释为观测值 Y_i 在给定 X的情况下对 Y 的预测所作贡献的比例。权重最大的观测值可视为 X 的最近邻，其依据是在预测变量空间上为预测响应量量身定制的隐式度量，见 [19]。随机森林权重使我们能够识别训练数据中最相似/最重要的个体。将 X 与这些相似观测值进行比较，有助于理解 X 与 Y之间的关系。

最后，我们要指出，权重 (2)-(3) 取决于用于预测的树的具体结构。树通过递归二元分割生长，并以最大化同质性准则为目标；其目的是将特征空间划分为尽可能同质的不同区域。在标准 CART算法中，方差被用作同质性准则，因此得到的划分适用于均值预测。

统计学文献中已提出多种不同的分割规则，着重于分位数预测 [4, 广义随机森林] ，或整体分布预测 [32, 分布式随机森林]。

对整个数据集拟合分布式随机森林。使用 R 软件及 grf（广义随机森林）包来计算随机森林及其相应的权重。如果未提供新的测试集，grf::predict 方法会对训练集进行袋外预测。这意味着，对于每个训练样本，都会找出训练过程中未使用该样本的所有树（该样本即为“袋外”样本），然后仅使用这些树对测试样本进行预测。

2.3 分布式随机森林的应用

DRF 是一种完全非参数、无模型的方法，可用于概率预测和分布回归。对于一组特征 X，它能够提供响应变量 Y 的完整预测分布，也就是说，在已知这些特征的情况下，全面呈现其可能的波动。该方法信息丰富且功能强大（见图2），因为它能够提供:

- （分布回归）为每个新样本提供一个可用直方图表示的预测分布；
- （均值或中位数预测）当需要点估计时，可使用预测均值或中位数——通常使用均值，而中位数对离群值更具稳健性；
- （不确定性评估）可通过直方图的形状（尖峰或平坦）进行图形化评估，也可使用预测的标准误或置信区间等数值统计量评估不确定性；
- （分类）可立即计算特别关注类别的类别概率——对于 ODX 评分，考虑 ODX ≤ 25 和 ODX > 25 这两个类别；
- （相似/有影响力的患者）可通过随机森林权重轻松识别队列（训练集）中与新病例最相似的患者；这些权重可解释为邻近度的度量。这里的邻近度是指模型学习得到的一种隐式距离，它会赋予相关特征更高的重要性。此信息有助于临床医生对新病例与队列中的患者进行有意义且富有信息的比较。

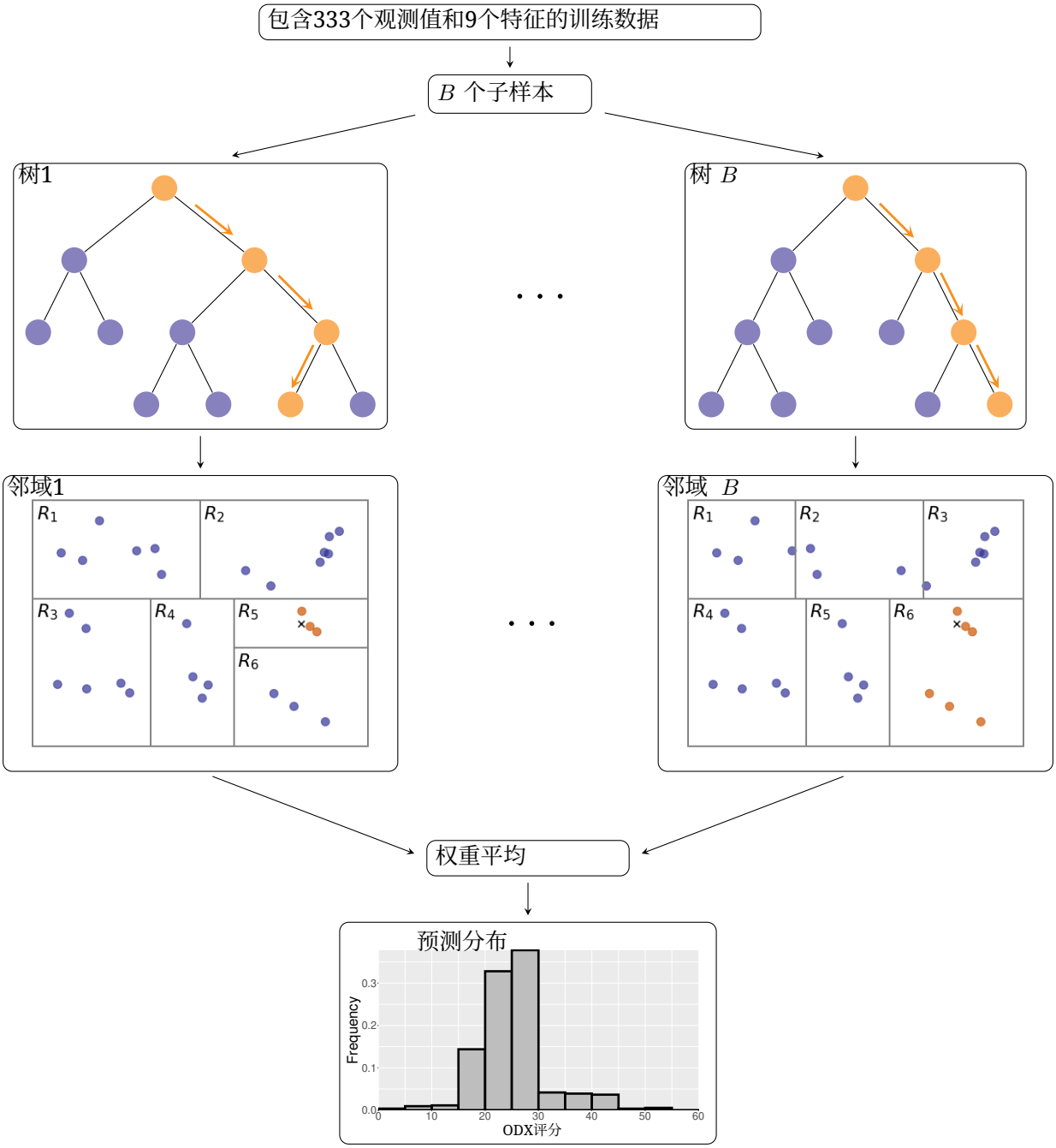


图1：分布式随机森林流程图。从训练数据中随机抽取大量子样本，并在每个子样本上构建二叉树；在每棵树中计算待预测点的邻域/权重，然后取平均值，从而得到森林权重；预测分布对应于使用这些森林权重对原始训练数据进行加权得到的加权样本。

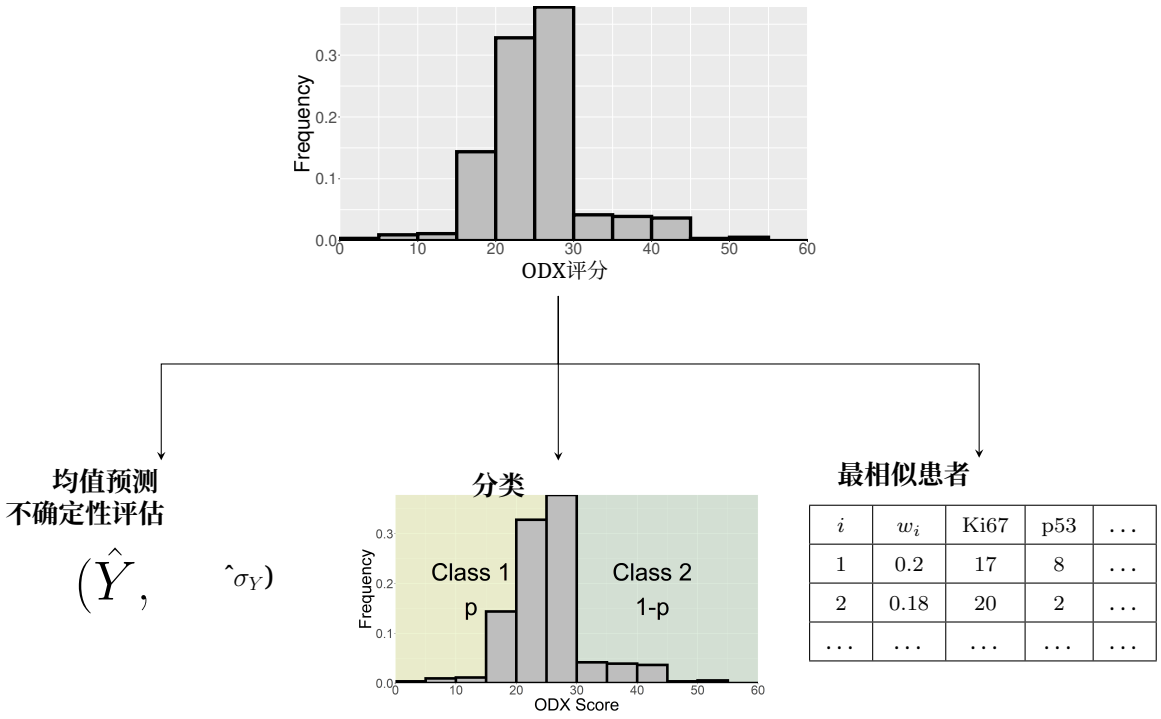


图2: 分布式随机森林的应用。训练好 DRF 后，可以轻松预测类别（分类）或条件均值、中位数（回归），并同时给出不确定性估计。此外，待预测点的权重可用于识别训练数据中与其最相似的邻居，并采用针对预测目的自适应调整的相似性概念。

2.4 预测性能评估

为评估分布式随机森林算法并将其与同期方法进行比较，我们采用适当评分规则 [14] 的理论。在概率预测中，评分规则用于比较预测分布 F 与结果 y。它的作用类似于衡量误差的指标，例如回归中的均方误差或分类中的误分类率。如果预测分布 F 与结果分布相匹配时的期望得分达到最小，则该评分规则称为严格适当评分规则。严格适当评分规则可用于评估概率预测和分布回归 [13]。

最常用的评分规则是连续排序概率评分（CRPS） [20]，其定义为

$$\mathrm{CRPS}(F, y) = \int_{\mathbb{R}} (F(z) - \mathbb{1}_{\{y \le z\}})^2 \mathrm{d}z.$$

当预测分布 F 对应于一个加权样本 $(y_i)_{1 \le i \le n}$ 时

其权重为（w_i）$_{1\leq i\leq n}$，CRPS 可轻松计算为

$$\mathrm{CRPS}(F,y)=\sum_{i=1}^{n}w_i|y_i-y|-\sum_{1\leq i<j\leq n}w_iw_j|y_i-y_j|.$$

第一项比较预测分布 F 与观测值 y （校准），第二项则评估预测的精确度（锐度）。该表达式还表明，CRPS(F， y) 与观测值 y 使用相同的单位，并且它将绝对误差推广开来；当 F 是确定性预报，也就是说是点测度时，该表达式便简化为绝对误差。

为了评估模型的泛化能力，也就是其在新样本上的预测性能，可以采用不同的验证方法来评估预测误差。简单验证使用训练集拟合模型，并使用独立的测试集计算误差（CRPS）。 K折交叉验证更为复杂，会将数据划分为 K 组，依次用作测试集。更具体地说，拟合 K 个不同的模型，每次使用除一折之外的所有折构成训练集，留出的一折在训练时不参与拟合，并用作测试集来计算 CRPS；由此得到 K 个不同的测试误差，将它们取平均后即可得到 K折交叉验证误差。对于装袋法（包括我们的随机森林方法），则可以采用袋外（OOB）方法。它通常能得到与 K折交叉验证相似的结果，但数值计算效率高得多，因为只需拟合一次模型。具体而言，由于重采样，某个观测值不会出现在所有子样本中，因此可以考虑由所有未使用该观测值进行训练的树所组成的子模型；然后在该观测值处评估此子模型并计算误差（CRPS）；对所有这些误差取平均即可得到 OOB 误差。

3 结果

我们将 DRF 应用于 333 位患者，利用表 1 所列的 9 个特征预测 ODX评分。为与文献进行比较，我们重点关注二分类（ODX ≤ 25 和 ODX > 25）。

在介绍 DRF 的结果之前，我们先评估模型。通过考察根据 DRF 得出的回归模型，可以进行简单的图形诊断。回归结果见图 3，其中绘制了预测均值（图 3a）和预测中位数（图 3b）与实际 ODX评分的关系。可以看出，拟合效果相当好，且相当一部分观测值落在各自的置信区间内。图 3a 中的灰色带半宽等于标准误，用于表示不确定性。图 3b 中的灰色带表示置信水平为 90% 的可信区间。

此外，为了评估 ODX 概率预测，我们针对每个观测值比较了 ODX 的 OOB 预测分布与实际观测值。预测误差采用第 2.4 节介绍的 CRPS 进行衡量。不同的得分见图4a。CRPS 越小，预报越准确。可以看到，大多数预测的 CRPS 较小或中等，表明总体预测质量良好。少数观测值的 CRPS 较大，说明对于这些个体而言，ODX 评分与我们根据总体人群所预期的值存在明显差异。除了 CRPS 之外，

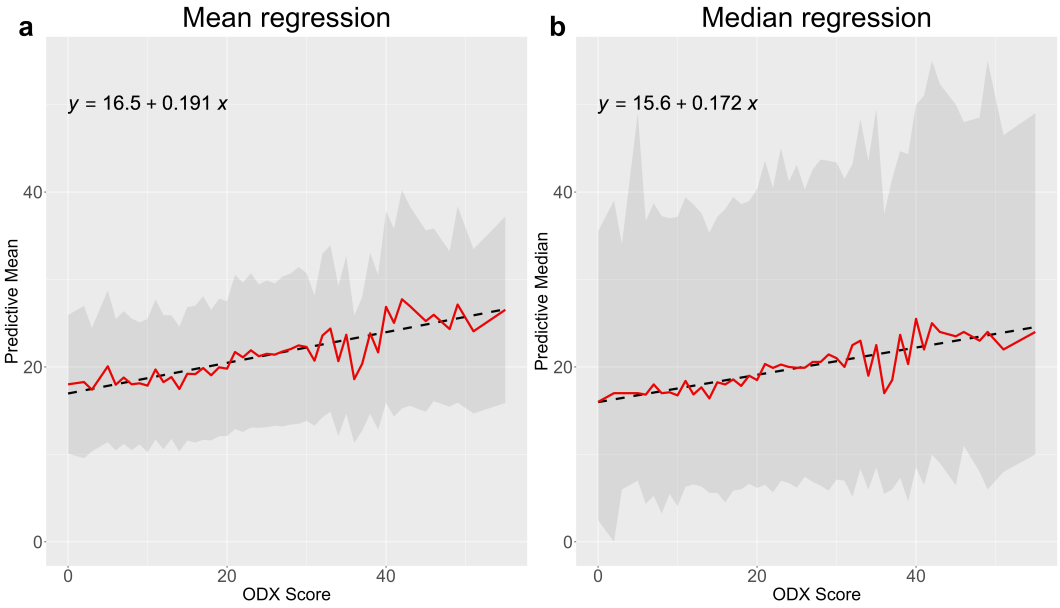


图3：使用回归诊断进行评估。均值回归（左图 - a）绘制了 ODX 观测值与其预测均值的关系；灰色带表示标准误。中位数回归（右图 - b）绘制了 ODX 评分与其预测中位数的关系；灰色带表示 90% 置信区间。

该图还给出了二分类任务（ODX $\leq$ 25 或 ODX $>$ 25）的结果：分类错误以橙色标示，正确分类的观测值则以蓝色标示。可以看到，分类错误与较大的 CRPS 高度吻合，这证实了 CRPS 评估预测质量的能力。随后，针对每位患者，DRF 会生成以直方图表示的预测分布，可将其与实际 ODX 评分进行比较。我们还标示了分别对应浅绿色/左侧类别和深绿色/右侧类别的两个类别概率。我们选取了三位患者，其 CRPS 分别较低（图4b，患者A）、中等（图4c，患者B）和较高（图4d，患者C）。对这些患者的预测分别可视为“良好”“一般”和“较差”。在图4b中，可以看到预测分布集中（直方图峰值明显），且 ODX 评分接近峰值。在图4c中，直方图较平坦，表明不确定性较高，而真实 ODX 评分位于高概率区域内。在图4d中，预测分布同样离散度较大，且 ODX 评分位于低概率区域内，这意味着两者吻合较差。需要强调的是，CRPS 较大并不一定意味着患者被错误分类，图4a中一些 CRPS 较高的取值便是如此。CRPS 考量的是分布回归，与此处所述的二分类并无明确关联。

由于 ODX 分类为 ODX $\leq$ 25 和 ODX $>$ 25 两类所产生的影响，我们将详细介绍基于 DRF 推导出的分类模型的评估结果（见表2）。该评估基于标准分类指标

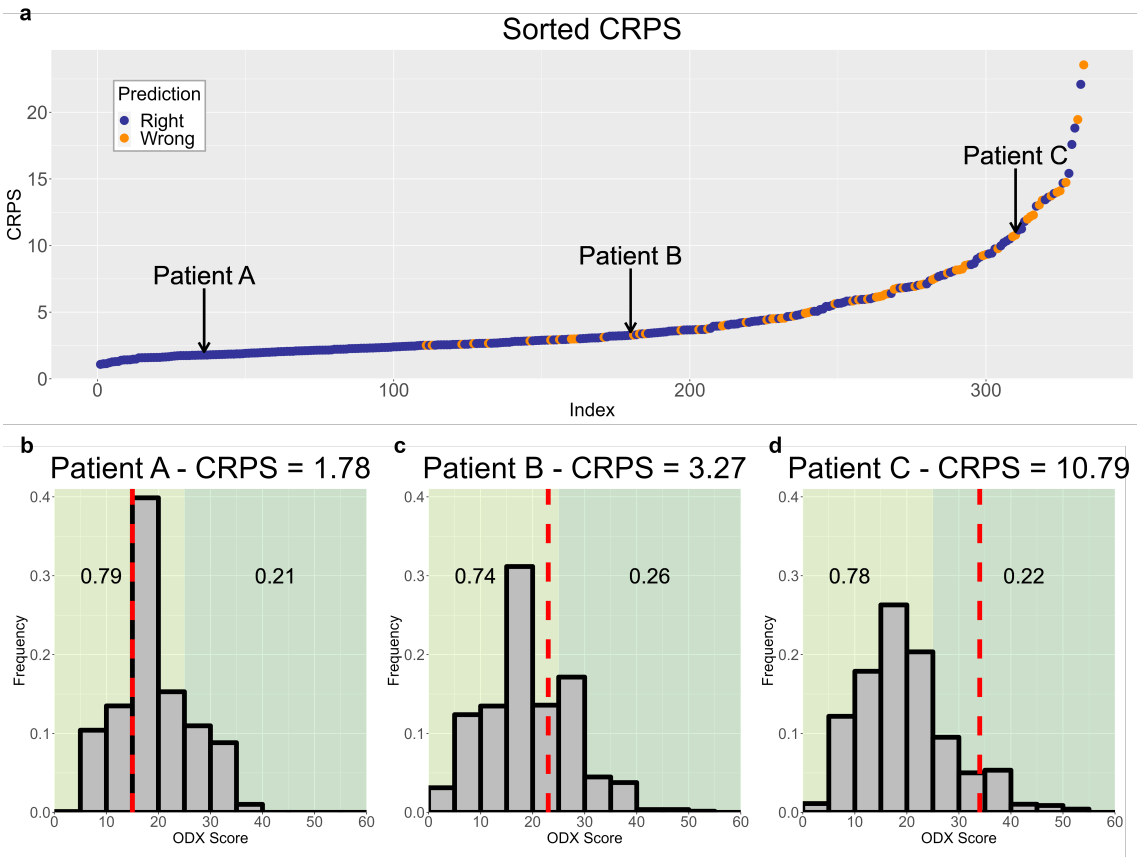


图4：使用 CRPS 对预测进行袋外评估——CRPS 越低，预报越精确。子图 b、c 和 d 分别对应从子图 a 所示的不同 CRPS 范围中选出的三位患者。在下方三个子图中，灰色直方图表示由 DRF 得到的 ODX 评分预测分布。红色虚线表示患者的真实 ODX 评分。两类（ODX ≤ 25 和 ODX > 25）以不同颜色的区域表示，并给出了每位患者属于各类别的预测概率。

		预测值	
		ODX ≤ 25	ODX > 25
True	ODX ≤ 25	231	20
	ODX > 25	49	33

准确率	79.3%
灵敏度	92.0%
特异度	40.3%
阳性预测值	82.5%
阴性预测值	62.3%
F1分数	0.870
曲线下面积	0.759

表2：分类诊断评估。混淆矩阵（左）以及标准指标（右）。

例如混淆矩阵和标准指标。标准指标如下 :

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}, \tag{5}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{6}$$

$$\text{Specificity} = \frac{TN}{FP + TN}, \tag{7}$$

$$\text{Positive Predictive Value} = \frac{TP}{TP + FP}, \tag{8}$$

$$\text{Negative Predictive Value} = \frac{TN}{FN + TN}, \tag{9}$$

$$\text{F1-score} = \frac{2 * \text{Positive Predictive Value*Sensitivity}}{\text{Positive Predictive Value+Sensitivity}} \tag{10}$$

其中，TP 是被正确分类为 ODX ≤ 25 的患者数量，FP 是被错误分类为 ODX ≤ 25 的患者数量，TN 是被正确分类为 ODX > 25 的患者数量，FN 是被错误分类为 ODX > 25 的患者数量。

我们在低风险组中正确分类了251名患者中的231名（92%），在高风险组中正确分类了75名患者中的33名（40.2%）。总体准确率为79.3%，p值小于0.05。低风险组正确预测值（PPV）的比例为82%。高风险组正确预测值（NPV）的比例约为62.3%。F1分数和曲线下面积（AUC）分别为0.87和0.759。DRF将提供更多信息，例如最近邻患者、ODX评分的分布以及不确定性预测（见图2）。现在，我们来考察这69名被错误分类的低风险和高风险患者。首先，我们注意到，这些患者中的大多数预测结果都接近决策边界（即接近 ODX $= 25$）。由于采用了二元决策，这些患者被错误分类；DRF方法提供的其他信息表明，患者的ODX评分接近决策边界，或者由于训练队列的局限性，患者所在的邻域并不符合实际情况。第一部分错误分类患者的CRPS可能较小，因为CRPS考虑了预测的离散度及其偏差。第二部分错误分类患者的ODX评分在我们的队列中属于极端值。因此，DRF为这些错误分类患者提供的最近邻患者信息较少，因为这些邻近患者来自一个无法代表这些离群患者的队列。在

为了给出更多定量结果，我们比较了69名错误分类患者的ODX评分、Ki67和p53与其邻域加权平均值之间的平均绝对差。错误分类患者的ODX评分与其邻域之间的平均绝对差为9.84，而正确分类患者的平均绝对差为6.29。就Ki67和p53而言，错误分类患者的平均绝对差分别为24.56%和5.77%；正确分类患者的平均绝对差则分别为Ki67的16.77%和p53的5.84%。

随后将这些分类结果与最先进的技术进行比较 [18, 15,17, 22, 5, 25]。详细比较见表3。

4 讨论

ODX是早期雌激素受体阳性/HER2阴性乳腺癌中最常用的乳腺基因组检测。它有助于确定哪些患者不太可能从化疗中获益。ODX评分基于6个基因组。这些基因组对应于病理报告中的标志物分析。一些研究将ODX评分与这些免疫组织学数据进行比较，并证实了它们与ODX评分之间的预测关系。已有多项研究采用不同方法，利用这些临床病理数据预测ODX评分（见表3）。本研究旨在使用分布式随机森林，根据一个包含333名患者临床和免疫组织学数据的特定区域队列预测ODX评分。这种预测一方面伴有预测误差，另一方面也能确定相似患者。所提出的DRF模型检出了82%的低风险患者（ODX $\leq$ 25）和62.3%的高风险患者（ODX $>$ 25）。

已有少数研究提出了一些用于预测 ODX评分的工具 [18, 15, 17, 22, 5,25]。每项研究均依据原始 ODX分类和 TAILORx 对患者进行特定分类（见表3中的 ODX预测阈值）。不同研究的预测结果相似，且均基于临床病理数据。我们当前的研究与所选的六项已发表研究一样，均使用肿瘤大小、肿瘤分级和孕激素受体。 [22] 和 [25]中未使用 Ki67。在本研究中，我们纳入了 p53。各研究所采用的 ODX评分阈值各不相同。我们的分布随机森林模型表现与其他预测工具相当。其创新之处在于为预测提供了额外信息（见图2），例如类别概率（低风险和高风险）、相似档案以及不确定性预测。

低风险和高风险的正确预测值分别为 82.5% 和 62.3%。我们使用 CRPS评分来区分最佳和最差预测。ODX特征低于 16 时，预测结果最佳。前十个最佳结果的 Ki67 平均值低于 14%，这与我们先前研究中的低风险特征相符 [5]。雌激素受体和孕激素受体的平均百分比分别为 93% 和 77%，符合相同的低风险特征。观察周围邻域和邻近患者的特征时，我们发现相似档案的 ODX评分在 0 到 25 之间。这些相似性均属于低风险特征。前十个结果中，相似档案的 Ki67 得分低于 25%。

et al. 20)	Pawloski et al. (2022)	Current study (DRF)
168)	(2,587, 1,293)	(333, OOB)
 5 84	– 62 56-69	56.9 58.0 30-84
✓	✓	✓
✓	✓	✓
	✓	
✓		✓
✓		✓
✓	✓	✓
✓		✓
		✓
cation	Classification	Distributional
18 30 30	≤ 25 > 25	≤ 25 > 25
ep Layer otron	Random Forest	Distributional Random Forest
5%	92.9%	82.5%
0%	65.1%	62.3%
%	96.3%	92.0%
%	48.3%	40.2%
3	–	0.759

score. For three

至于结果不一致的病例，它们属于高风险类别。其 ODX评分、Ki67 和孕激素受体的平均值分别为 46%、36% 和 22%。此外，可以观察到，在最佳和最差结果中，ODX 与孕激素受体呈负相关。这些病例的相似档案具有较高的孕激素受体水平。这种现象是由高风险类别中的病例数量较少所致。一个预测效果最差的例子是一位 ODX评分较高、属于高风险类别的概率为 50% 的患者。其实际 ODX评分为 49，预测的 ODX评分接近临界值。前 10 个相似档案的平均 ODX评分为 31，分布集中在 25 左右。相似档案分布非常分散，难以分析。大多数最近邻的 SBR分级为 3。预测效果不佳，但相似档案的 ODX评分较低而 SBR分级较高。队列规模以及训练和测试阶段可能会影响预测结果。此外，我们的研究队列并不均衡，因为高风险类别中的患者较少。在这种情况下，应考虑相似档案中影响 ODX评分的因素，例如孕激素受体、Ki67 和 p53。相同档案的分布使临床医生能够根据病程演变检索相似的既往病例。即使存在缺失数据，也可以应用所提出的模型。该模型能够以较高的确定性预测低风险类别，这意味着无需安排化疗。我们的研究与该数据集相关，因此难以推广至不同的队列，因为已知存在队列间变异，尤其是某些生物标志物（如 Ki67）的变异。

5 结论

本文提出了一种新的 Oncotype 评分预测方法。该方法基于分布式随机森林，并使用 9 项临床病理特征。它能够预测患者 ODX评分的分布，并通过计算患者属于低风险或高风险类别的概率、识别最相似的档案，解释预测评分。所提出的分布式随机森林模型检出了 82% 的低风险患者（≤ 25）和 62.3% 的高风险患者（> 25）。不过，DRF 也存在一些局限性。将 DRF 与病理学家对不同组织学和免疫组织化学特征的专业知识结合起来，可在临床上帮助肿瘤科医生就乳腺癌治疗进行决策。这一策略具有显而易见的医疗经济学价值。仍需开展更多研究，以进一步验证 DRF方法，并提升从病理数据中提取知识的能力。

参考文献

[1] Kathy S Albain、William E Barlow、Steven Shak、Gabriel N Hortobagyi、Robert B Livingston、I-Tien Yeh、Peter Ravdin、Roberto Bugarini、Frederick L Baehner、Nancy E Davidson 等。21基因复发评分检测在绝经后、淋巴结阳性、雌激素受体阳性乳腺癌患者化疗随机试验中的预后和预测价值。《柳叶刀·肿瘤学》，11(1):55，2010年。

[2]Joan Albanell、Christer Svedman、Joseph Gligorov、Simon DH Holt、Gianfilippo Bertelli、Jens-Uwe Blohmer、RomanRouzier、Ana Lluch 和 Wolfgang Eiermann。汇总分析：评估在雌激素受体阳性、

人表皮生长因子受体2阴性的早期乳腺癌女性中，使用21基因复发评分检测对临床决策的影响的欧洲前瞻性研究。《欧洲癌症杂志》，66:104–113，2016。[3] Fabrice Andre、Nofisat Ismaila、N Lynn Henry、Mark R Somerfield、Robert C Bast、William Barlow、Deborah E Collyar、M Elizabeth Hammond、Nicole M Kuderer、Minetta C Liu 等。使用生物标志物指导早期浸润性乳腺癌女性辅助全身治疗决策：ASCO临床实践指南更新——整合TAILORx研究结果。《临床肿瘤学杂志》，37(22):1956–1964，2019。[4] Susan Athey、Julie Tibshirani 和 Stefan Wager。广义随机森林。《统计年刊》，47(2):1148–1178，2019。ISSN 0090-5364。doi: 10.1214/18-AOS1709。URL https://doi.org/10.1214/18-AOS1709。[5]Aline Baltres、Zeina Al Masry、Ryad Zemouri、Severine Valmary-Degano、Laurent Arnould、Noureddine Zerhouni 和 Christine Devalland。利用深度多层感知机预测雌激素受体阳性、HER2阴性乳腺癌的Oncotype DX复发评分。《乳腺癌》，27(5):1007–1016，2020。[6] Leo Breiman。装袋预测器。《机器学习》，24(2):123–140，1996。ISSN 0885-6125。[7] Leo Breiman。随机森林。《机器学习》，45(1):5–32，2001。ISSN 0885-6125。doi: 10.1023/A:1010933404324。[8] Leo Breiman、Jerome H. Friedman、Richard A. Olshen 和 Charles J. Stone。分类与回归树。Wadsworth统计学／概率论丛书。贝尔蒙特，加利福尼亚州：Wadsworth国际集团（Wadsworth, Inc.旗下部门）。X，358页。29.25美元；18.95美元（1984年），1984。[9] Gaoxiang Chen、Qun Li、Fuqian Shi、Islem Rekik 和 Zhifang Pan。Rfdcr：利用级联随机森林和稠密条件随机场自动分割脑病灶。《神经影像》，211:116620，2020。[10] Carlos Fernandez-Lozano、Pablo Hervella、Virginia Mato-Abad、Manuel Rodr´ıguez- Y´a˜nez、Sonia Su´arez-Garaboa、Iria L´opez-Dequidt、Ana Estany-Gestal、Tom´as Sobrino、Francisco Campos、Jos´e Castillo 等。基于随机森林的卒中结局预测。《科学报告》，11(1):1–12，2021。[11] Melina B Flanagan、David J Dabbs、Adam M Brufsky、Sushil Beriwal 和 Rohit Bhargava。组织病理学变量可预测Oncotype DX复发评分。《现代病理学》，21(10):1255–1261，2008。[12] Armando E Giuliano、James L Connolly、Stephen B Edge、Elizabeth A Mittendorf、Hope S Rugo、Lawrence J Solin、Donald L Weaver、David J Winchester 和 Gabriel N Hortobagyi。乳腺癌——美国癌症联合委员会第八版癌症分期手册的重大变化。《*CA*：临床医师癌症杂志》，67(4):290–303，2017。[13]Tilmann Gneiting 和 Matthias Katzfuss。概率预测。《统计学及其应用年度评论》，1(1):125–151，2014年1月。doi: 10.1146/annurev-statistics-062713-085831。

[14] Tilmann Gneiting 和 Adrian E. Raftery。《严格适当评分规则、预测与估计》。美国统计学会期刊，102(477):359–378，2007年。ISSN 0162-1459。doi: 10.1198/016214506000001437。URL https://doi.org/10.1198/016214506000001437。[15] Yanjun Hou、Gary Tozbikian、Debra L Zynger 和 Zaibo Li。使用修正的 Magee 方程识别不太可能从21基因复发评分检测（Oncotype DX 检测）中获益的患者。美国临床病理学杂志，147(6):541–548，2017年。[16] Kevin Kalinsky、William E Barlow、Julie R Gralow、Funda Meric-Bernstam、Kathy S Albain、Daniel F Hayes、Nancy U Lin、Edith A Perez、Lori J Goldstein、Stephen KL Chia 等。21基因检测用于指导淋巴结阳性乳腺癌的化疗获益。新英格兰医学杂志，385(25):2336–2347，2021年。[17] Isaac Kim、Hee Jun Choi、Jai Min Ryu、Se Kyung Lee、Jong Han Yu、Seok Won Kim、Seok Jin Nam 和 Jeong Eon Lee。基于机器学习、使用 Oncotype DX复发评分预测高/低风险组的模型。欧洲外科肿瘤学杂志，45(2):134–140，2019年。[18] Molly E Klein、David J Dabbs、Yongli Shuai、Adam M Brufsky、Rachel Jankowitz、Shannon L Puhalla 和 Rohit Bhargava。Oncotype DX复发评分预测：使用通过线性回归分析得出的病理学生成方程。现代病理学，26(5):658–664，2013年。[19]Yi Lin 和 Yongho Jeon。随机森林与自适应最近邻。美国统计学会期刊，101(474):578–590，2006年。ISSN 0162-1459。doi: 10.1198/ 016214505000001230。URL https://doi.org/10.1198/016214505000001230。[20] James E. Matheson 和 Robert L. Winkler。连续概率分布的评分规则。管理科学，22，1976年6月。doi: 10.2307/2629907。[21] Nicolai Meinshausen。分位数回归森林。机器学习研究杂志，第983–999页，2006年。[22] Amila Orucevic、John L Bell、Megan King、Alison P McNabb 和 Robert E Heidel。根据TAILORx 临床试验结果更新列线图——可使用临床病理学数据预测 Oncotype DX 乳腺癌复发评分。乳腺，46:116–125，2019年。[23] Soonmyung Paik、Steven Shak、Gong Tang、Chungyeul Kim、Joffre Baker、Maureen Cronin、Frederick L Baehner、Michael G Walker、Drew Watson、Taesung Park 等。用于预测接受他莫昔芬治疗的淋巴结阴性乳腺癌复发的多基因检测。新英格兰医学杂志，351(27):2817–2826，2004年。[24] Soonmyung Paik、Gong Tang、Steven Shak、Chungyeul Kim、Joffre Baker、Wanseop Kim、Maureen Cronin、Frederick L Baehner、Drew Watson、John Bryant 等。淋巴结阴性、雌激素受体阳性乳腺癌女性的基因表达与化疗获益。《临床肿瘤学杂志》，24(23):3726–3734，2006年。[25] Kate R Pawloski、Mithat Gonen、Hannah Y Wen、Audree B Tadros、Donna Thompson、Kelly Abbate、Monica Morrow 和 Mahmoud El-Tamer。用于预测50岁以上患者 Oncotype DX 风险类别的监督式机器学习模型。乳腺癌研究与治疗，191(2):423–430，2022年。

[26] Joseph A Sparano、Robert J Gray、Della F Makower、Kathleen I Pritchard、Kathy S Albain、Daniel F Hayes、Charles E Geyer Jr、Elizabeth C Dees、Edith A Perez、John A Olson Jr 等。乳腺癌21基因表达检测的前瞻性验证。新英格兰医学杂志，373(21):2005–2014，2015年。[27] Joseph A Sparano、Robert J Gray、Della F Makower、Kathleen I Pritchard、Kathy S Albain、Daniel F Hayes、Charles E Geyer Jr、Elizabeth C Dees、Matthew P Goetz、John A Olson Jr 等。根据21基因表达检测指导乳腺癌辅助化疗。新英格兰医学杂志，379(2):111–121，2018年。[28] Maher Sughayer、Rolla Alaaraj 和 Ahmad Alsughayer。应用新的 Magee 方程预测 Oncotype DX复发评分。乳腺癌，25(5):597–604，2018年。[29] Antonio C Wolff、M Elizabeth Hale Hammond、Kimberly H Allison、Brittany E Harvey、Pamela B Mangu、John MS Bartlett、Michael Bilous、Ian O Ellis、Patrick Fitzgibbons、Wedad Hanna 等。乳腺癌人表皮生长因子受体2检测：美国临床肿瘤学会/美国病理学家学会临床实践指南重点更新。《病理学与实验室医学档案》，142 (11):1364–1382，2018年。[30] B Yeo、L Zabaglo、M Hills、A Dodson、I Smith 和 M Dowsett。免疫组化4+ c评分在雌激素受体阳性早期乳腺癌中的临床效用：一项前瞻性决策影响研究。《英国癌症杂志》，113(3):390–395，2015年。[31] Alaa Zare、Lynne-Marie Postovit 和 John Maringa Githaka。使用非参数随机森林分析构建稳健的炎性乳腺癌基因特征。《乳腺癌研究》，23(1):1–6，2021年。[32] Domagoj ´Cevid、Loris Michel、Jeffrey N¨af、Nicolai Meinshausen 和 Peter B¨uhlmann。分布式随机森林：异质性调整与多变量分布回归，2021年。URL https://arxiv.org/abs/2005.14458。