arXiv:2304.00379v1 [eess.IV] 1 Apr 2023

利用辅助监督改进小型数据集的多模态融合

Gregory Holste[1,2], *Douwe van der Wal*[2], *Hans Pinckaers*[2], *Rikiya Yamashita*[2] *Akinori Mitani*[2], *Andre Esteva*[2]

[1] 德克萨斯大学奥斯汀分校，奥斯汀，得克萨斯州，美国[2] Artera公司，山景城，加利福尼亚州，美国

摘要

前列腺癌是全球男性癌症相关死亡的主要原因之一。与许多癌症一样，其诊断需要专家综合分析异质性患者信息，例如成像、临床风险因素等。因此，近年来出现了许多针对临床决策任务的图像与非图像数据深度多模态融合研究。其中许多研究提出融合来自各患者模态的学习特征的方法，并通过跨模态注意力门控、克罗内克积融合、正交性正则化等技术，显著改善下游任务的表现。尽管这些增强型融合操作可以优于特征拼接，但它们通常具有极高的学习容量，这意味着即使应用于小型或低维数据集，也很容易过拟合。我们不设计表达能力极强的融合操作，而是提出三种适用于小型数据集的简单多模态融合改进方法，这些方法通过在训练期间生成辅助监督来源来促进优化：**额外监督**、**临床预测**和**密集融合**。我们在基于配对的组织病理学成像和表格化临床特征进行前列腺癌诊断的任务上验证了所提出的方法。这些方法易于实现，可应用于任何配对图像与非图像数据的分类任务。

索引词— 多模态融合、组织病理学、前列腺癌、自动化诊断

1. 引言

前列腺癌是全球第三大最常确诊的癌症，也是男性癌症相关死亡的主要原因之一 [1, 2]。诊断和风险分层主要通过人工分析组织活检中的组织病理学成像，并结合患者病史以及年龄、前列腺特异性抗原（PSA）水平等风险因素来进行[3]。与许多癌症一样，诊断依赖于整合有关患者的多种异质信息。鉴于这一临床现实，以及深度学习技术在自动化诊断方面持续取得的成功，近期，研究者开始关注用于癌症风险自动分层的多模态融合；这涉及运用深度学习技术，联合学习患者成像数据（放射学、组织病理学或其他成像数据）、患者风险因素，以及实验室结果和基因组学特征等其他数据流 [4, 5, 6]。在这一研究方向中，一个核心问题仍未解决：融合来自各个模态的信息，最佳方式是什么？

在关于成像和电子健康记录（EHR）数据多模态融合方法的综述中，Huang 等人 [7]将这些方法分为三类：早期融合（拼接来自各模态的输入）、联合融合（拼接从各模态学习到的中间特征）以及后期融合（拼接来自各模态的预测概率）。大多数研究发现，某种形式的联合融合（通常称为**“后期联合融合”**）最为有效，即结合学习到的模态特异性特征向量，以生成单一的多模态表征 [7, 8]。例如，给定一个影像表征和一个临床数据表征，可以通过多种方式聚合这两个特征向量，其中最常见的是拼接和逐元素平均。鉴于仍有改进空间，研究人员提出了更复杂、更具表现力的融合操作 [9, 10]和架构 [11, 12]，以实现多模态表征的最佳结合。例如，Chen 等人 [9, 11] 使用跨模态注意力门控和克罗内克积融合，聚合基因组学与组织病理学成像特征，以预测癌症结局。Braman等人 [10] 在此基础上加入了正交性损失函数，促使单模态表示在融合过程中提供互补信息。

尽管这些融合方法在诊断任务上取得了显著改进，但由此产生的模型通常具有很高的学习容量（可学习参数数量庞大）。因此，将这些方法应用于小型数据集或低维输入时，更容易对训练数据中的虚假模式过拟合 [13]。我们假设，另一种提升泛化能力的途径是，在不大幅增加模型容量的情况下，借助多种监督来源简化模型优化。在这项工作中，我们提出了三种简单的方法，通过生成辅助监督来源来帮助优化：额外

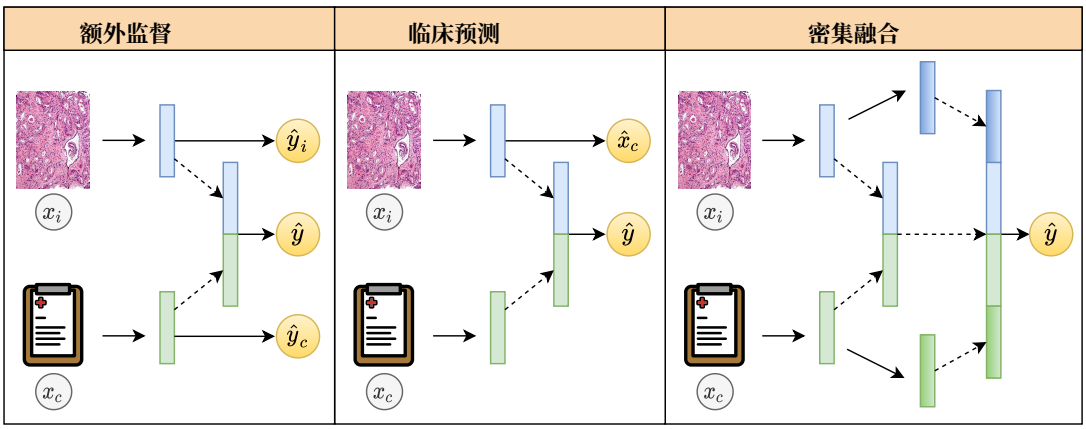

图 1。针对小型或低维数据集提出的多模态融合方法示意图。**额外监督**（左）生成额外的模态特定预测，**临床预测**（中）利用学习到的图像表征直接预测相关的非图像输入，而**密集融合**（右）通过整合网络不同阶段的信息，促进特征之间的密集交互。

监督、临床预测和密集融合。我们使用配对的组织病理图像特征和表格化临床风险因素，在前列腺癌分类任务上验证了这些方法。实验表明，这三种方法结合使用，效果优于后期联合融合（采用拼接或克罗内克积融合），且可轻松应用于任何具有配对图像和非图像数据的任务。

2. 材料与方法

2.1. 数据集描述

本研究使用了 NRG Oncology 开展的五项前瞻性随机 III 期局限性前列腺癌男性患者临床试验的数据：NRG/RTOG-9202、9408、9413、9910 和 0126 [14, 15, 16, 17, 18]；其中 4,581 名患者的数据可用于本研究中的机器学习研究。对于每位患者，我们使用配对的组织病理学成像和临床风险因素数据来预测远处转移（DM），即癌症从原发肿瘤部位扩散的二元事件。图像特征由自监督学习模型 [19] 预先提取，因此每张切片都表示为一个低维“包”，其中包含图像特征 $\boldsymbol{x_i} \in \mathbb{R}^{K\times 128}$，其中 K 为固定大小（256×256）图像块的数量。每个相应的临床输入向量 $\boldsymbol{x_c} \in \mathbb{R}^6$ 包含患者的六项表格描述变量：年龄、PSA、T分期，以及由病理学家确定的格里森评分和模式 [20]。在全部 4,581 名患者中，12.2% 发生远处转移。有关数据采集和图像特征提取的完整详情，请参阅 Esteva 等人 [19]。

2.2. 小数据集的融合方法

多模态分类的标准“后期联合融合”方法包括学习模态特定表征，将这些特征向量融合为单一的多模态表征，然后使用该特征向量进行分类。具体而言，令 $\boldsymbol{x_i}$ 表示患者的组织病理成像，$\boldsymbol{x_c}$ 表示患者关联的临床信息。给定图像编码器 $f_i(\cdot)$ 和临床编码器 $f_c(\cdot)$，我们计算表征 $\boldsymbol{h_i} = f_i(\boldsymbol{x_i})$ 和 $\boldsymbol{h_c} = f_c(\boldsymbol{x_c})$，其中 i 指影像模态，c 指临床（或非成像）模态。随后通过拼接融合图像与非图像表征：$\boldsymbol{h} = \text{concat}([\boldsymbol{h_i}, \boldsymbol{h_c}])$。最后，使用分类器 $g(\cdot)$ 通过 $\hat{y} = g(\boldsymbol{h})$ 生成结局的最终预测。该模型通过最小化损失函数 $\mathcal{L} = \ell(y, \hat{y})$ 进行优化，其中 $\ell(\cdot)$ 是二元交叉熵损失函数。

在此基线的基础上，我们提出了三种通过在训练期间添加辅助监督来源来帮助优化的方法：**额外监督**、**临床预测**和**密集融合**（图1）。

2.2.1. 额外监督

虽然基线后期联合融合方法基于融合图像与非图像特征生成单个预测结果，但它并不会直接促使特定模态的表征能够预测结局。我们可以通过在每个特定模态表征之上添加分类头来解决这一问题。给定仅图像分类器 $g_i(\cdot)$ 和仅临床分类器 $g_c(\cdot)$，我们还分别仅基于图像特征计算 $\hat{y}_i = g_i(\boldsymbol{h_i})$，以及仅基于临床特征计算 $\hat{y}_c = g_c(\boldsymbol{h_c})$。这种方法，即**额外监督**，现在会生成三个结局预测结果，这些预测结果可以共同用于通过损失函数 $\mathcal{L} = \ell(y, \hat{y}) + \ell(y, \hat{y}_i) + \ell(y, \hat{y}_c)$ 优化整个网络。这样，来自特定模态特征的额外监督信号就可以反向传播，从而优化网络中的所有参数。类似的方法已经被应用于

见于 Kawahara 等人。 [21] 和 Holste 等人。 [6]。

2.2.2. 临床预测

在**临床预测**中，我们在标准的后期联合融合方法基础上，利用学习到的图像表征直接预测（回归）相应的非图像输入特征。具体而言，我们通过额外的分类头 $g_{i\to c}(\cdot)$计算$\hat{\boldsymbol{x}}_{\boldsymbol{c}} = g_{i\to c}(\boldsymbol{h_i})$。除了预测结局之外，这个新的预测临床特征向量$\hat{\boldsymbol{x}}_{\boldsymbol{c}}$还会在训练期间通过损失函数 $\mathcal{L} = \ell(y, \hat{y})+\text{sim}(\boldsymbol{x_c}, \hat{\boldsymbol{x}}_{\boldsymbol{c}})$充当辅助监督来源，其中 $\ell(\cdot)$是二元交叉熵损失，$\text{sim}(\cdot)$是针对连续输入的均方误差（MSE）损失。这种方法可以被解释为一种辅助预测任务，用于使学习到的图像表征与临床非图像特征保持一致。由于患者的图像与非图像“视图”预计存在相关性，这种模态对齐可能有助于实现图像与非图像信息的更优融合。

2.2.3. 密集融合

受 Hu等人。 [22]，**密集融合**的启发，该方法通过在训练期间允许模态之间进行更密集的交互，改进了标准的后期联合融合。与后期联合融合一样，我们通过$\boldsymbol{h}_{\boldsymbol{i}}^{(1)} = f_{(1)\,i}(\boldsymbol{x_i})$ 和 $\boldsymbol{h}_{\boldsymbol{c}}^{(1)} = f_c^{(1)}(\boldsymbol{x_c})$ 学习纯图像和纯非图像表征，然后将这些表征融合以形成 $\boldsymbol{h}^{(1)} = \text{concat}([\boldsymbol{h}_{\boldsymbol{i}}^{(1)}, \boldsymbol{h}_{\boldsymbol{c}}^{(1)}])$；上标（1）表示这是网络的第一个中间层。然而，我们不再将分类器置于融合表征 $\boldsymbol{h}^{(1)}$之上，而是通过 $\boldsymbol{h}_{\boldsymbol{i}}^{(2)} = f_{(2)\,i}(\boldsymbol{x_i})$ 和 $\boldsymbol{h}_{\boldsymbol{c}}^{(2)} = f_c^{(2)}(\boldsymbol{x_c})$ 学习图像特征和非图像特征的更深层表征，其中 $f_i^{(2)}$和$f_c^{(2)}$是全连接层。随后，我们形成最终的融合表征；它不仅聚合纯图像特征和纯临床特征（如后期联合融合所做的那样），还纳入了网络前一阶段的融合表征：$\boldsymbol{h}^{(2)} = \text{concat}([\boldsymbol{h}_{\boldsymbol{i}}^{(2)}, \boldsymbol{h}_{\boldsymbol{c}}^{(2)}, \boldsymbol{h}^{(1)}])$。最后，使用分类器 $g(\cdot)$ 生成 $\hat{y} = g(\boldsymbol{h}^{(2)})$，并通过优化与后期联合融合中相同的二元交叉熵损失来训练模型。这使得来自每种模态的特征能够进行密集交互，并聚合网络不同阶段的信息。

虽然密集融合不像另外两种方法那样显式地产生额外的监督来源，但它可以与额外监督和临床预测结合，从而获得比单独使用任一方法更多的监督来源。密集融合提供多个中间特征向量；与额外监督结合时，这些特征向量可用于生成额外的 结局预测。此外，**这三种方法彼此互补，可以任意组合使用**。例如，图 2 展示了一个结合了所提出的全部三种方法的架构；该模型将损失函数最小化为 $\mathcal{L} = \ell(y, \hat{y})+\ell(y, \hat{y}_f)+\ell(y, \hat{y}_i^{(1)})+\ell(y, \hat{y}_i^{(2)})+\ell(y, \hat{y}_c^{(1)})+\ell(y, \hat{y}_c^{(2)})+\text{sim}(\boldsymbol{x_c}, \hat{\boldsymbol{x}}_{\boldsymbol{c}})$。对于所有模型，推理时的最终预测取为 $\hat{y}$，即从最终的融合多模态表示中学习得到的输出。

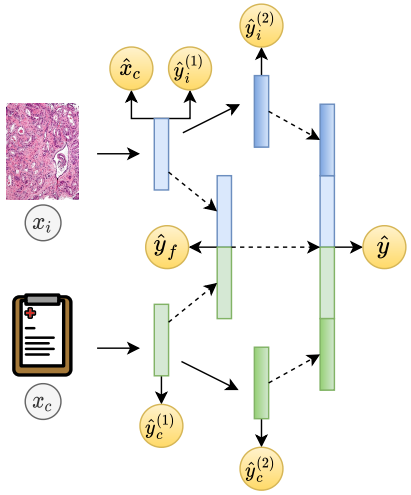

图 2。结合额外监督、临床预测和密集融合三种方法的多模态融合架构。模型通过优化六个交叉熵损失（每个预测的 $\hat{y}$ 对应一个）和一个均方误差损失（用于回归临床特征 $\hat{\boldsymbol{x}}_{\boldsymbol{c}}$）之和进行训练。

2.3. 实现细节

与 Esteva 等人 [19], 的做法相同，所有六个临床特征均被视为数值型并进行标准化。为给每个影像特征包 $\boldsymbol{x_i} \in \mathbb{R}^{K\times 128}$生成一个 128 维表示，通过学习注意力权重，以逐元素乘法调节某些特征，然后沿图块维度进行求和池化。所有模型均使用 AdamW 优化器 [23] 进行训练，学习率为 0.0075，批量大小为 128，分类头的随机失活率为 0.9。每个模型均训练 100 个训练轮次，并在每一折中采用类别平衡重采样以应对类别不平衡。在前 20 个训练轮次中，仅优化临床编码器 $f_c(\cdot)$ 和分类器 $g_c(\cdot)$；随后，冻结 $f_c(\cdot)$ 的参数，并在剩余的 80 个训练轮次中优化所有其他参数。

2.4. 实验设置

在我们的第一项实验中，我们使用（i）拼接、（ii）克罗内克积融合、（iii）拼接 + 三种所提出的辅助监督方法，以及（iv）克罗内克融合 +三种辅助监督方法，训练一个后期联合融合模型（第 2.2 节）。通过比较这四个模型，我们旨在揭示其应用于低维数据集时模型复杂度与泛化能力之间的权衡。为了解每项所提出技术的贡献，我们还开展了一项消融式实验，

融合操作	辅助监督	# 参数	AUC
拼接		17.1K	0.781 ± 0.024
克罗内克		66.3K	0.770 ± 0.018
拼接	✓	43.1K	$\mathbf{0.792 \pm 0.014}$
克罗内克	✓	207.1K	0.781 ± 0.013

表 1。采用不同融合技术以及我们提出的辅助监督方法进行远处转移（DM）分类的结果。“辅助监督” =通过额外监督、临床预测和密集融合进行训练。基线后期联合融合方法以灰色突出显示。

逐步将各项技术添加到以后期联合融合与拼接为基线的模型中。在使用相同超参数和其他方面完全相同的架构的条件下，我们训练基线模型，并测试三种所提出方法的所有可能组合：额外监督（第 2.2.1 节）、临床预测（第 2.2.2 节）和密集融合（第 2.2.3 节）。我们采用五折交叉验证进行模型训练，并使用受试者工作特征曲线下面积（曲线下面积）进行评估。模型性能通过五个折叠上的曲线下面积均值和标准差进行汇总。

3. 结果

与标准的后期联合融合方法相比，我们提出的辅助监督方法显著提升了远处转移分类性能（表 1）。我们提出的模型（第 3 行）达到 0.792 ± 0.014 曲线下面积，而后期联合融合在采用拼接时达到 0.781 ± 0.024 曲线下面积，在采用克罗内克积融合时达到 0.770 ± 0.018 曲线下面积。我们发现，在我们的设置中，克罗内克融合使模型容量增加了 4-5× 参数数量，同时对泛化能力产生了持续的不利影响。尽管我们提出的方法也会增加模型容量，但辅助监督的益处体现在：我们提出的模型（第 3 行）优于后期联合克罗内克融合（第 2 行）。

在我们的消融研究中，我们发现，同时采用提出的三种辅助监督方法可取得最佳结果（表 2）。基线模型的曲线下面积为 0.781 ± 0.024，而额外采用额外监督、临床预测和密集融合的模型，其曲线下面积达到 0.792 ± 0.014。有趣的是，单独增加额外监督或密集融合并不能提升性能，但这两种方法的任意组合都优于基线。我们还发现，临床预测是三种技术中影响最大的一种；采用临床预测训练的三个模型，其跨折曲线下面积均值最高，且曲线下面积标准差最低。

额外监督	临床预测	密集融合	AUC
			0.781 ± 0.024
✓			0.778 ± 0.021
	✓		0.789 ± 0.013
		✓	0.776 ± 0.025
✓	✓		0.787 ± 0.015
✓		✓	0.785 ± 0.016
	✓	✓	0.790 ± 0.012
✓	✓	✓	$\mathbf{0.792 \pm 0.014}$

表2。采用所提出方法的所有组合进行的远处转移（DM）分类结果。最佳结果以粗体显示，次佳结果加下划线。基线后期联合融合方法以灰色突出显示。

4. 讨论与结论

总而言之，我们提出了三种简单的方法，以改进低维数据集上的深度多模态融合。与设计表达能力很强的融合或注意力操作不同，所提出的技术通过辅助监督来源帮助优化，而非显著增加学习容量，因此不易过拟合。我们在一项任务中验证了这三种方法——额外监督、临床预测和密集融合：联合利用数字化组织病理成像和表格临床数据，预测前列腺癌转移。结果显示，这三种方法均优于标准的后期联合融合方法。此外，所提出的方法比近期广受欢迎的克罗内克积融合表现更好，同时参数效率更高。尽管这些方法仅在一个特定应用和数据集上得到验证，但它们可以直接应用于任何使用配对的图像与非图像数据的多模态分类任务。

本研究可进一步扩展，在其他分类任务中对我们的融合方法进行更全面的验证，甚至将其应用于组织病理学以外的领域。尽管结合所有三种提出的方法取得了最佳表现，但对于不同任务和数据集而言，这是否是最优的方法组合，仍有待验证。最后，采用对类别不平衡具有鲁棒性的指标（如平均精确率或平衡准确率）进行更全面的评估，也可能有助于改进本研究。我们选择曲线下面积作为指标，以便与 Esteva 等人 [19]的评估保持一致，并避免选择决策阈值，不过类别不平衡可能会导致曲线下面积虚高 [24]。将我们的方法与其他基线进行比较也很有价值，例如人类专家表现。此外，通过特征重要性分析找出在这种多模态设置下最能预测远处转移的临床特征，或可提升研究发现的临床价值。

5. 遵守伦理标准

NRG Oncology 已通过 IRB00000781 获得机构审查委员会批准。由于所有数据均已匿名化，因此免除了知情同意。

6. 致谢

本项目获得美国国家癌症研究所资助，资助编号为 U10CA180822（NRG肿瘤学组数据统计中心）、U10CA180868（NRG肿瘤学组运营部门）和 U24CA196067（NRG标本库）。

7. 参考文献

[1] Freddie Bray、Jacques Ferlay、Isabelle Soerjomataram 等，《2018年全球癌症统计：GLOBOCAN 对 185 个国家 36 种癌症全球发病率和死亡率的估计》，*CA Cancer J. Clin.*，第 68 卷，第 6 期，第 394–424 页，2018 年。[2] Elizabeth M Ward、Recinda L Sherman、S Jane Henley 等，《美国癌症状况年度报告：聚焦 20–49 岁男性和女性的癌症》，*J. Natl. Cancer Inst.*，第 111 卷，第 12 期，第 1279–1297 页，2019 年。[3] Edward Schaeffer、Sandy Srinivas、Emmanuel S Antonarakis 等，《NCCN 指南解读：前列腺癌，2021 年第 1 版——NCCN 指南重点更新》，*J. Natl. Compr. Canc.Netw.*，第 19 卷，第 2 期，第 134–143 页，2021 年。[4] Pooya Mobadersany、Safoora Yousefi、Mohamed Amgad 等，《利用卷积网络，根据组织学和基因组学预测癌症结局》，*Proc. Natl. Acad. Sci.*，第 115 卷，第 13 期，第 E2970–E2979 页，2018 年。[5] Anika Cheerla 和 Olivier Gevaert，《利用多模态表征进行泛癌预后预测的深度学习》，*Bioinformatics*，第 35 卷，第 14 期，第 i446–i454 页，2019 年。[6] Gregory Holste、Savannah C. Partridge、Habib Rahbar 等，《端到端学习融合图像与非图像特征，以改进基于磁共振成像的乳腺癌分类》，载于 *Proc. IEEE Int. Conf. Comput. Vis. Works.*，2021 年，第 3294–3303 页。[7] Shih-Cheng Huang、Anuj Pareek、Saeed Seyyedi 等，《利用深度学习融合医学影像与电子健康记录：系统综述与实施指南》，*NPJ Digit.Med.*，第 3 卷，第 1 期，第 1–9 页，2020 年。[8] Can Cui、Haichun Yang、Yaohong Wang 等，《疾病诊断与预后中图像和非图像数据的深度多模态融合：综述》，*arXiv preprint 2203.15588*，2022 年。[9] Richard J Chen、Ming Y Lu、Jingwen Wang 等，《病理组学融合：融合组织病理学与基因组学特征以进行癌症诊断和预后的集成框架》，*IEEE Trans. Med.Imaging*，2020 年。[10] Nathaniel Braman、Jacob WH Gordon、Emery T Goossens 等，《深度正交融合：整合放射学、病理学、基因组学和临床数据的多模态预后生物标志物发现》，载于*Med. Image Comput. Comput. Assist. Interv.*施普林格，2021 年，第 667–677 页。[11] Richard J Chen、Ming Y Lu、Wei-Hung Weng 等，《用于吉像素全切片图像生存预测的多模态协同注意力变换器》，载于*Proc. IEEE Int. Conf.Comput. Vis.*，2021 年，第 4015–4025 页。

[12] Shih-Cheng Huang、Liyue Shen、Matthew P Lungren 等，《GLoRIA：用于标签高效医学图像识别的多模态全局-局部表征学习框架》，载于*Proc. IEEE Int.Conf. Comput. Vis.*，2021 年，第 3942–3951 页。[13] Trevor Hastie、Robert Tibshirani 和 Jerome H Friedman，《统计学习要素：数据挖掘、推断与预测》，第 2 卷，施普林格，2009 年。[14] Christopher U Jones、Daniel Hunt、David G McGowan 等，《局限性前列腺癌的放射治疗与短期雄激素剥夺》，*N Engl. J. Med.*，第 365 卷，第 2 期，第 107–118 页，2011 年。[15] Jeff M Michalski、Jennifer Moughan、James Purdy 等，《中危前列腺癌患者接受标准放射治疗与剂量递增放射治疗的效果比较：NRG肿瘤学 RTOG 0126 随机临床试验》，*JAMA Oncol.*，第 4 卷，第 6 期，第 e180039–e180039 页，2018 年。[16] Thomas M Pisansky、Daniel Hunt、Leonard G Gomella 等，《局限性前列腺癌放射治疗前雄激素抑制的持续时间：放射治疗肿瘤学组 9910 随机临床试验》，*J.Clin. Oncol.*，第 33 卷，第 4 期，第 332 页，2015 年。[17] Eric M Horwitz、Kyounghwa Bae、Gerald E Hanks 等，《放射治疗肿瘤学组方案 92-02 的十年随访：局部晚期前列腺癌辅助雄激素剥夺持续时间的 III 期试验》，*J. Clin. Oncol.*，第 26 卷，第 15 期，第 2497–2504 页，2008 年。

[18] Colleen A Lawton、Michelle DeSilvio、Mack Roach III 等 .，《一项比较全盆腔与仅前列腺放射治疗，以及新辅助与辅助全雄激素抑制的 III 期试验最新进展：RTOG 94-13 的更新分析，重点探讨意外的激素与放射治疗相互作用》，国际放射肿瘤学、生物学与物理学杂志，第 69 卷，第 3 期，第 646–655 页，2007 年。

[19] Andre Esteva、Jean Feng、Douwe van der Wal 等，《通过在随机 III 期临床试验中应用多模态深度学习实现前列腺癌治疗个体化》，*NPJ Digit. Med.*，第 5 卷，第 1 期，第 1–8 页，2022 年。[20] Donald F Gleason 和 George T Mellinger，《通过组织学分级与临床分期相结合预测前列腺腺癌的预后》，*J. Urol.*，第 111 卷，第 1 期，第 58–64 页，1974 年。[21] Jeremy Kawahara、Sara Daneshvar、Giuseppe Argenziano 等，《七点检查表与使用多任务多模态神经网络进行皮肤病变分类》，*IEEE J.Biomed. Health Inform.*，第 23 卷，第 2 期，第 538–546 页，2018 年。[22] Di Hu、Chengze Wang、Feiping Nie 等，《用于分层联合表征的密集多模态融合》，载于 *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE，2019 年，第 3941–3945 页。[23] Ilya Loshchilov 和 Frank Hutter，《解耦权重衰减正则化》，载于*Proc. Int. Conf. Learn. Repr.*，2019 年。[24] Alberto Fern´andez、Salvador Garc´ıa、Mikel Galar、Ronaldo C Prati、Bartosz Krawczyk 和 Francisco Herrera，从不平衡数据集中学习，第 10 卷，施普林格，2018 年。