图像和文本数据驱动的两阶段区间决策方法

付 超[1]，崔凯旋[1]，倪文青[1]，薛 旻[1]，刘卫勇[2]

(1. 合肥工业大学 管理学院, 合肥 230009；2. 中国科学技术大学 第一附属医院，合肥 230001)

摘 要: 在图像和文本等多模态数据可用条件下, 利用单一深度网络进行融合决策, 存在可解释性和可靠性挑战. 针对挑战, 提出一种图像和文本数据驱动的两阶段区间决策方法, 包括准则预测和准则融合两个阶段. 在准则预测阶段, 依据各准则上的图像特征提取需求确定深度网络集合, 设计基于有放回采样的深度网络性能统计比较方法确定网络排序, 进而构建网络接续组合的选择方法确定可靠的最佳组合, 产生基于图像的区间数预测值. 在准则融合阶段, 通过文本训练集学习准则权重, 进而构建基于文本验证集的优化模型学习自适应权重函数, 最后利用准则权重和自适应权重函数融合各准则上的图像生成预测值, 产生可解释的总体预测值. 以安徽合肥某三甲医院超声部的乳腺彩超图像和文本数据为基础, 将提出方法用于乳腺病灶辅助诊断, 验证了方法的有效性.

关键词: 图像和文本数据驱动决策；两阶段决策；深度网络组合；自适应权重学习；乳腺病灶辅助诊断

中图分类号: C934 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.1285

引用格式: 付超, 崔凯旋, 倪文青, 等. 图像和文本数据驱动的两阶段区间决策方法 [J]. 控制与决策

The two-stage interval decision-making method driven by image and text data

FU Chao[1]，*CUI Kai-xuan*[1]，*NI Wen-qing*[1]，*Xue Min*[1]，*LIU Wei-yong*[2]

(1. School of Management， Hefei University of Technology， Hefei 230009， China； 2. The First Affiliated Hospital，University of Science and Technology，Hefei 230001，China)

Abstract: Under the condition that the multimodal data of image and text are available, the use of one deep network to make fusion decisions is a usual way. However, it faces the challenges of interpretability and reliability. To address these two challenges, this article proposes a two-stage interval decision-making method driven by image and text data, including the criterion prediction and criterion aggregation stages. In the criterion prediction stage, a set of deep networks is determined according to the requirements of extracting image characteristics, and then a statistical comparison method of deep network performance is designed based on sampling with replacement to determine the rankings of the deep networks in the set. With the resulting rankings, a method of selecting the sequential combinations of deep networks is constructed to identify the reliable optimal combination, which can be used to derive interval-valued predictions from images. In the criterion aggregation stage, criterion weights are learned from the training dataset of text, and then an optimization is constructed based on the verifying dataset of text to learn adaptive weight function. With the resulting criterion weights and weight function, the interval-valued predictions derived from images on each criterion are finally combined to generate the explainable overall predictions. Based on the image and text data collected from the ultrasonic department of a tertiary hospital in Hefei, Anhui, the proposed method is used to help diagnose breast lesions, which validates its effectiveness.

Keywords: image and text-data-driven decision-making； two-stage decision-making； combination of deep networks；adaptive weight learning；auxiliary diagnosis of breast lesions

0 引 言

随着人工智能技术的快速发展, 文本[1]、图像[2]等多模态数据处理能力不断提升, 推动决策范式从依赖单一、结构化的数据, 向综合利用多模态、异构数据的方向转变, 为实现更精细、更智能的决策提供了新的可能性[3].

单模态数据驱动的决策方法, 通过挖掘数据中蕴含的信息和决策偏好为决策者提供一致的决策支持. 例如, Liu 等考虑专家知识和数据的相关性, 提出了一种知识辅助与数据驱动融合的模糊决策方法;

收稿日期: 2025-12-12；**录用日期:** 2026-03-05.

基金项目: 国家自然科学基金项目 (72571090, 72171066)；安徽省研究生教学改革研究项目 (2023jyjxggyjY028).

以污水处理过程中历史操作记录与专家经验数据为基础, 运用所提方法帮助操作人员识别污泥膨胀的故障根源并提供可解释的抑制策略, 从而提升决策的可靠性与适应性[4]. 张海利等基于批次过程数据的图像化技术, 提出了一种面向间歇过程的卷积自编码故障监测方法; 以青霉素发酵仿真数据等文本为基础, 通过将三维时序数据矩阵转换为二维灰度图像, 运用所提方法提取纹理与结构特征, 提升了故障监测的准确性与鲁棒性[5]. 这些研究表明单模态数据对有效决策的贡献, 但单模态数据往往只能反映片面的信息表征, 在数据存在噪声、信息不完整或环境动态变化等复杂情境下, 依靠单模态数据的决策在准确性和可靠性等方面面临挑战[6].

应对挑战, 多模态数据驱动的决策分析应运而生, 并逐渐应用于疾病诊断[7]、商业投资[8]、工业过程控制[9] 等领域. 通过有效挖掘不同模态数据之间的内在关联与互补信息, 能够为复杂决策问题提供更全面、一致且互补的求解路径或方案[10]. 现有多模态数据驱动的决策方法多聚焦于图像和文本数据, 大体分为端到端决策方法和基于决策结构的融合推理方法.

端到端决策方法利用深度神经网络对多模态数据进行联合建模, 通过特征提取和对齐, 将不同模态数据映射至统一的表示空间, 进而融合输出最终结果. Lan 等针对智慧农业中果树病虫害视觉问答问题, 采用双线性融合与协同注意力机制融合图像和文本信息, 实现了高效的病害识别, 为智能农业中的精准病害管理提供了有效解决方案[11]. Fan 等基于车载摄像头得到的 RGB 图像与雷达数据, 提出了基于分布校正的跨模态融合架构, 在 Transformer 编码器中进行多层次特征对齐的基础上引入分布校正模块, 提升了自动驾驶系统在未知与恶劣场景下的泛化能力, 为工业 5.0 背景下的人机协同自动驾驶提供了可靠的多模态感知决策解决方案[12]. 这些方法依赖模型内部结构实现多模态信息的统一表征与融合决策, 可解释性较弱, 不符合以人为主体的决策预期, 难以在医疗辅助诊断等场景中应用.

基于决策结构的融合推理方法, 将决策结构中的准则等要素引入多模态数据驱动的决策过程, 在一定程度上提高了方法的可解释性. Chen 等以餐厅推荐的图像和文本在线评论为基础, 通过图像描述技术将图像评论转化为文本, 结合文本分析提取评价准则, 并根据图像和文本中的对象及内容的一致性进行模态信息融合, 为消费者提供了更丰富、一致的决策支持[13]. 此方法将准则引入多模态信息融合与推理过程, 提升了可解释性. 但是, 方法还是构建在深度模型框架下, 缺乏清晰的阶段划分, 其可解释性与决策者的多准则决策预期仍存在差距.

在多模态数据驱动决策的基础上, 有学者引入多阶段决策, 在相互关联的各个阶段对特定任务进行优化, 并通过阶段间的信息传递与反馈实现决策性能的整体提升[14]. 例如, He 等基于计算机断层扫描图像、临床报告、治疗方案等多模态数据, 引入多个分类器, 提出了一种基于多准则决策的分层融合模型, 将分类器输出量化为备选方案, 结合分层架构对模态内与模态间预测结果进行两阶段递进融合, 为预后分析提供了更可靠、层次化的决策支持[15]. 引入多阶段决策, 有利于更精细地处理不同模态数据的异质性与复杂性; 同时, 能够通过融合传统决策模型与深度网络等新模型, 增强不同决策维度上的稳定性和可靠性. 然而, 相关研究的阶段划分服务于网络结构设计, 并未在多准则决策框架下实现准则评价值的预测和融合, 仍不能完全满足决策者的多准则决策预期.

综上, 现有多模态数据驱动的决策方法在特征融合与预测性能方面取得了显著进展, 但在面向决策者的多准则决策预期的多阶段决策方面仍存在不足.

针对上述问题, 本文聚焦图像和文本两种模态, 提出一种图像和文本数据驱动的两阶段区间决策方法, 包括准则预测和准则融合两个阶段. 首先, 在准则预测阶段, 以图像数据为客观输入, 依据各准则上的图像特征提取需求选择可行的深度网络集合; 设计基于有放回采样的统计比较方法, 实现基于图像验证池的深度网络性能排序; 基于排序的可行深度网络, 构建网络接续组合的评价方法, 有效防止网络组合爆炸问题, 实现基于图像验证集的最佳网络组合选择; 基于此, 依据决策者的偏好由各准则上的图像数据产生文本评价值, 为后续融合阶段提供结构化、可解释的输入信息. 其次, 在准则融合阶段, 依据决策者的偏好, 通过图像数据构建文本训练集, 利用文本训练集学习各准则的权重; 在此基础上, 构建优化模型学习自适应权重函数, 通过文本验证集求得函数的最优参数. 给定图像测试集, 由第一阶段产生各准则上的可靠区间数预测值, 进而利用第二阶段学习的准则权重和自适应权重函数融合各准则上的预测值, 产生总体区间数预测值. 以安徽合肥某三甲医院超声部的乳腺超声图像和文本检查报告为基础, 利用提出方法进行乳腺病灶辅助诊断, 验证了方法的有效性.

1 方法基础

1.1 区间数

传统决策分析中, 方案评价值多采用确定值进行刻画. 然而, 在实际决策情境中, 伴随着信息的不完全或环境的不确定, 决策者倾向于采用区间数、模糊数、信念分布等不确定信息表达方式来刻画方案评价值. 考虑到区间数的简单易用, 本文采用区间数作为评价值表达方式. 作为提出方法的基础, 两个区间数之间的相似测度如下所示.

定义 1[16] 给定两个区间数 $a=[a^-,a^+]=\{r|a^-\leqslant r\leqslant a^+,a^-,a^+\in\mathbb{R}\}$和 $b=[b^-,b^+]=\{r|b^-\leqslant r\leqslant b^+,b^-,b^+\in\mathbb{R}\}$, 设 $\overline{a}=0.5\cdot(a^-+a^+)$, $\overline{b}=0.5\cdot(b^-+b^+)$, $l(a)=a^+-a^-$, $l(b)=b^+-b^-$, $d=[d^-,d^+]=a\cap b$, $l(d)=d^+-d^-$, 则 a 和 b 之间的相似度计算为

$$S(a,b)=1-\sqrt{(\overline{a}-\overline{b})^2+\frac{1}{3}\cdot((0.5\cdot l(a))^2+(0.5\cdot l(b))^2)-\frac{1}{6}\cdot(l(d))^2}. \tag{1}$$

定义 1 中, 相似度满足$0\leqslant S(a,b)\leqslant 1$.

1.2 基于图像和文本数据的决策基础

使用图像和文本两种模态数据构建两阶段区间决策方法, 相关的图像数据集和文本数据集如表 1 所示.

表1 图像和文本数据集的符号表示

符号	含义
$e_i(i=1,\ldots,L)$	准则集合
$(G_1^i,H_1^i)=((g_{l_1}^i)_{1\times M_1},(h_{l_1}^i)_{1\times M_1})$	准则e_i上的图像训练集, 包括图像及其对应的区间数标签
$(G_2^i,H_2^i)=((g_{l_2}^i)_{1\times M_2},(h_{l_2}^i)_{1\times M_2})$	准则e_i上的图像验证集, 包括图像及其对应的区间数标签
$(G_3^i,H_3^i)=((g_{l_3}^i)_{1\times M_3},(h_{l_3}^i)_{1\times M_3})$	准则e_i上的图像验证池, 包括图像及其对应的区间数标签
$(G_4^i,H_4^i,H_4)\,(i=1,\ldots,L)=((g_{l_4}^i)_{1\times M_4},(h_{l_4}^i)_{1\times M_4},(h_{l_4})_{1\times M_4})$	L准则上的图像测试集, 包括图像、准则上区间数标签、总体区间数标签
$(U_1,V_1)=((u_{m_1}^i)_{L\times N_1},(v_{m_1})_{1\times N_1})$	L准则上的文本训练集, 包括准则和总体区间数评价值
$(U_2,V_2)=((u_{m_2}^i)_{L\times N_2},(v_{m_2})_{1\times N_2})$	L准则上的文本验证集, 包括准则和总体区间数评价值
$C_b(b=1,\ldots,B)$	标签和评价值的区间数集合
$\tilde{H}_o^i\,(o=2,\ldots,4)=(\tilde{h}_{l_o}^i)_{1\times M_o}$	准则e_i上的区间数预测值
$\tilde{H}_4=(\tilde{h}_{l_4})_{1\times M_4}$	基于图像的总体区间数预测值

表 1 给出了多准则决策框架下的图像和文本两种模态数据. 图像数据包括训练集(G_1^i,H_1^i) $(i=1,\ldots,L)$、验证集(G_2^i,H_2^i)、验证池(G_3^i,H_3^i)、测试集(G_4^i,H_4^i,H_4), 用于实现准则预测阶段各准则上基于深度网络的区间数评价值预测, 产生预测值$\tilde{H}_o^i$ $(o=2,\ldots,4)$.

依据决策者的偏好, 由图像数据构建文本数据, 包括训练集(U_1,V_1)和验证集(U_2,V_2), 用于实现准则融合阶段的准则权重自适应学习. 构建基于自适应准则权重的决策模型, 融合各准则上的区间数预测值产生总体区间预测值$\tilde{H}_4$, 与总体标签H_4对比得到总体预测精度. 图像和文本数据中相关的评价值和预测值均使用集合$C_B=\{C_b\ (b=1,\ldots,B)\}$中的区间数表示.

图像数据和文本数据存在内在关联性, 使得基于文本数据构建的决策模型能够用于融合各准则上图像的预测值. 注意, 尽管文本数据源自图像数据, 这里的图像数据集和文本数据集不存在交集, 即产生文本数据集的图像数据不出现在准则预测阶段的图像数据集中, 以确保两阶段决策的合理性.

上述包含准则预测和准则融合的两阶段决策划分, 符合决策者的多准则决策预期, 使得提出方法具有较好的可解释性. 值得一提的是, 这里的可解释性并不涵盖由图像产生准则预测值的深度网络内部机制, 重在决策过程与决策者预期的一致性.

2 两阶段区间决策方法

2.1 方法框架

面向决策者的多准则决策预期, 以图像和文本数据为基础, 提出包含准则预测和准则融合的两阶段区间决策方法, 整体框架如图 1 所示.

由整体框架易知, 提出方法主要包含准则预测和准则融合两个阶段. 准则预测阶段使用训练集、验证集、验证池、测试集等 4 个图像数据集产生各准则上可靠的区间数预测值; 准则合成阶段使用训练集和验证集等 2 个文本数据集自适应地学习准则权重, 进而使用准则权重融合各准则上的预测值形成总体

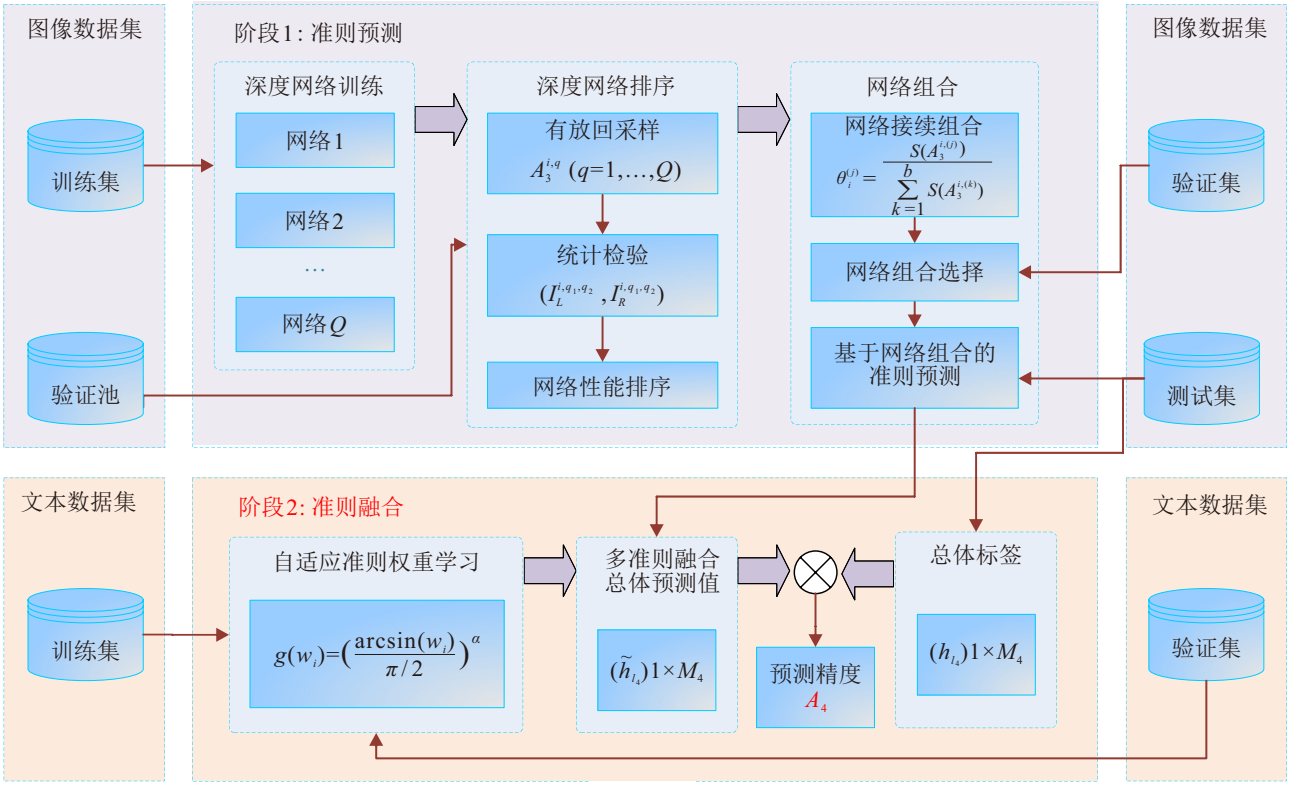

图1 两阶段区间决策方法的整体框架

预测值, 与图像测试集中的总体标签对比产生预测精度.

在准则预测阶段, 理论上, 一种深度网络即可用于由输入图像产生集合C_B上的预测分布. 然而, 选择一种合适的深度网络是一个挑战. 对于不同准则上的输入图像, 需要提取和识别的特征有所不同, 难以由一种深度网络完成. 因此, 深度网络选择可视为不同网络在多个准则上的性能折衷. 旨在充分考虑各准则上的图像特征提取和识别, 引入Q种不同的深度网络DN_q $(q=1,\ldots,Q)$, 进行网络性能排序和接续组合, 产生可靠的各准则预测值.

在准则融合阶段, 文本训练集用于学习准则权重, 以实现各准则上预测值的线性加权融合, 产生总体预测值. 旨在进一步提高总体预测精度, 引入文本验证集进行自适应准则权重学习.

2.2 准则预测阶段

准则预测阶段主要包括深度网络训练、深度网络性能排序、深度网络组合选择等 3 个核心工作.

(1) 深度网络训练

如前所述, 图像训练集(G_1^i, H_1^i)中的标签$h_{l_1}^i$使用集合$C_B=\{C_b\ (b=1,\ldots,B)\}$中的区间数表示, 即最多$B$种标签. 基于此, 使用准则$e_i$ $(i=1,\ldots,L)$上的图像训练集(G_1^i, H_1^i)进行Q种深度网络训练, 得到图像$g_{l_1}^i$和区间数标签$h_{l_1}^i$之间的最佳映射, 即

$$F_q^i: g_{l_1}^i \to h_{l_1}^i. \tag{2}$$

这一映射表明提出方法的可解释性并不涵盖由图像产生准则预测值的深度网络内部机制. 不同的深度网络擅长提取和识别不同的图像特征. 例如, CNN 擅长提取和识别图像中的局部信息, 而 Transformer 擅长提取和识别图像中的关联信息[17].

为了充分发挥不同类别深度网络的优势, 有必要在各准则上进行网络组合. 理论上, Q种深度网络存在2^Q-Q-1种有效组合. 使用图像验证集从2^Q-Q-1种有效组合中进行最佳组合选择, 是一件非常耗时、耗计算资源的任务, 尤其当Q值较大时. 聚焦高效的深度网络组合选择, 设计一种接续网络组合选择方法, 首先在各准则上利用图像验证池进行Q种深度网络的性能排序, 而后利用图像验证集比较Q种有序深度网络的$(Q-1)$种接续组合, 产生各准则上的最佳网络接续组合.

(2) 深度网络性能排序

旨在产生可靠的Q种深度网络性能排序, 使用准则e_i $(i=1,\ldots,L)$上的图像验证池(G_3^i, H_3^i)来统计比较其上不同深度网络的性能.

采用有放回采样策略从图像验证池(G_3^i, H_3^i)中选择一定比例的图像验证数据, 重复N_r次. 依据图像训练集中数据尺寸M_1, 确定有放回采样的选择比例N_p为 40% 或 50%. 为了确保通过统计比较得到可靠的Q种深度网络性能排序, 同时兼顾计算效率和

统计效果的平衡, N_r可以设为 400, 500 或 800.

设在图像验证池(G_3^i, H_3^i)上进行N_r次有放回采样, 得到的采样数据集为$(G_{3,x}^i, H_{3,x}^i)$ $(x = 1, \ldots, N_r)$. 使用Q种深度网络的训练后映射函数F_q^i, 将$G_{3,x}^i$中的所有图像映射为区间数预测值$(\tilde{h}_{l_3}^{i,q})_{1\times M_{3,p}}$, 这里, $M_{3,p} = M_3 * N_p$. 基于预测值$(\tilde{h}_{l_3}^{i,q})_{1\times M_{3,p}}$和其对应的标签$(h_{l_3}^{i,q})_{1\times M_{3,p}}$, 深度网络$DN_q$的预测精度计算为

$$A_{3,x}^{i,q} = \frac{\sum_{l_3=1}^{M_{3,p}} S(h_{l_3,x}^{i,q}, \tilde{h}_{l_3,x}^{i,q})}{M_{3,p}}. \tag{3}$$

这里, $S(\cdot,\cdot)$表示定义 1 中两个区间数之间的相似度.

已知两个深度网络DN_{q_1}和DN_{q_2}在准则e_i $(i = 1, \ldots, L)$上的预测精度向量$A_3^{i,q_1} = \{A_{3,x}^{i,q_1}$ $(x = 1, \ldots, N_r)\}$和$A_3^{i,q_2} = \{A_{3,x}^{i,q_2}$ $(x = 1, \ldots, N_r)\}$, 使用威尔科克森符号秩检验[18] 对两个网络的性能进行统计比较. 具体而言, 分别进行左尾检验和右尾检验. 左尾检验时, 零假设为$A_3^{i,q_1} = A_3^{i,q_2}$, 备选假设为$A_3^{i,q_1} < A_3^{i,q_2}$. 右尾检验时, 零假设为$A_3^{i,q_1} = A_3^{i,q_2}$, 备选假设为$A_3^{i,q_1} > A_3^{i,q_2}$. 设执行左尾检验和右尾检验得到$p$值, 分别为$p_L^{i,q_1,q_2}$和$p_R^{i,q_1,q_2}$. 依据$p_L^{i,q_1,q_2}$和$p_R^{i,q_1,q_2}$, 得到两个检验的指示$I_L^{i,q_1,q_2}$和$I_R^{i,q_1,q_2}$, 即

$$I_L^{i,q_1,q_2} = \begin{cases} 1 & p_L^{i,q_1,q_2} \leqslant 0.05 \\ 0 & p_L^{i,q_1,q_2} > 0.05 \end{cases} \text{和} \tag{4}$$

$$I_R^{i,q_1,q_2} = \begin{cases} 1 & p_R^{i,q_1,q_2} \leqslant 0.05 \\ 0 & p_R^{i,q_1,q_2} > 0.05 \end{cases}. \tag{5}$$

性质 1[19]　设执行左尾检验和右尾检验得到指示I_L^{i,q_1,q_2}和I_R^{i,q_1,q_2}, 则两个指示满足

$$I_L^{i,q_1q_2} + I_R^{i,q_1q_2} \leqslant 1. \tag{6}$$

依据性质 1, 指示对 $(I_L^{i,q_1,q_2}, I_R^{i,q_1,q_2})$ 可能的取值为 (0, 0)、(0, 1)、(1, 0), 其含义如下所示.

定义 2　设执行左尾检验和右尾检验得到指示对 $(I_L^{i,q_1,q_2}, I_R^{i,q_1,q_2})$, 则深度网络$DN_{q_1}$和$DN_{q_2}$在准则$e_1$ $(i = 1, \ldots, L)$上的性能优劣关系为

$$\begin{cases} DN_{q_1} \prec DN_{q_2} & (p_L^{i,q_1,q_2}, p_R^{i,q_1,q_2}) = (1,0) \\ DN_{q_1} \approx DN_{q_2} & (p_L^{i,q_1,q_2}, p_R^{i,q_1,q_2}) = (0,0) \\ DN_{q_1} \succ DN_{q_2} & (p_L^{i,q_1,q_2}, p_R^{i,q_1,q_2}) = (0,1) \end{cases}, \tag{7}$$

其中, 符号“$\prec$”、“$\approx$”、“$\succ$”分别表示“劣于”、“不可比”、“优于”.

依据定义 2, 得到准则e_i $(i = 1, \ldots, L)$上深度网络DN_q $(q = 1, \ldots, Q)$的性能排序$DN_{(b)}^i$ $(b = 1, \ldots, Q)$. 这里, $DN_{(1)}^i, \ldots, DN_{(Q)}^i$表示准则$e_i$ $(i = 1, \ldots, L)$上性能第 1、…、第Q的深度网络.

(3) 深度网络组合选择

确定各准则上DN_q $(q = 1, \ldots, Q)$的性能排序后, 在$(Q-1)$种接续组合中选择最佳组合. 这一组合策略, 一方面极大地减少了待比较的深度网络组合种类, 降低了时间成本和计算成本; 另一方面, 在各准则上按照Q种深度网络性能降序进行组合, 兼顾了深度网络组合的预测精度.

定义 3　设$DN_{(1)}^i, \ldots, DN_{(Q)}^i$为准则$e_i$ $(i = 1, \ldots, L)$上性能第 1、…、第Q的深度网络, 则由它们形成的$(Q-1)$种接续组合表示为

$$\{DN_{(1)}^i, DN_{(2)}^i\},\ \{DN_{(1)}^i, DN_{(2)}^i, DN_{(3)}^i\},\ \ldots,$$
$$\{DN_{(1)}^i, \ldots, DN_{(Q-1)}^i\},\ \{DN_{(1)}^i, \ldots, DN_{(Q)}^i\}.$$

进行接续组合, 需要确定不同性能网络间的相对权重, 即权重反映了不同深度网络间的性能差异. 基于此, 考虑使用$A_3^{i,(b)}$ $(b = 1, \ldots, Q)$来计算接续组合中不同深度网络的权重. 为便于计算, 参考文[18] 中计算区间数得分值函数的方式, 定义向量$A_3^{i,(b)}$的得分值函数为

$$S(A_3^{i,(b)}) = \mu(A_3^{i,(b)})(1 - \sigma(A_3^{i,(b)})), \tag{8}$$

其中, $\mu(A_3^{i,(b)})$和$\sigma(A_3^{i,(b)})$分别表示$A_3^{i,(b)}$的均值和标准差. 使用得分值函数$S(A_3^{i,(b)})$计算接续组合$\{DN_{(1)}^i, \ldots, DN_{(b)}^i\}$中各深度网络的权重为

$$\theta_i^{(j)} = \frac{S(A_3^{i,(j)})}{\sum_{k=1}^{b} S(A_3^{i,(k)})}. \tag{9}$$

易知, 式 (9) 中的权重满足$0 \leqslant \theta_i^{(j)} \leqslant 1$和$\sum_{j=1}^{b} \theta_i^{(j)} = 1$.

确定接续组合中各深度网络的权重后, 使用准则e_i $(i = 1, \ldots, L)$上的图像验证集(G_2^i, H_2^i)比较$(Q-1)$种接续组合, 发现最佳的组合. 基于深度网络$DN_{(j)}^i$ $(j = 1, \ldots, b)$训练产生的映射函数$F_{(j)}^i$, 由(G_2^i, H_2^i)中的输入图像$(g_{l_2}^i)_{1\times M_2}$得到区间数预测值$(\tilde{h}_{l_2}^{i,(j)})_{1\times M_2}$, 进而计算接续组合$\{DN_{(1)}^i, \ldots, DN_{(b)}^i\}$的总体预测值为

$$\tilde{h}_{l_2}^{i,(b)} = \sum_{j=1}^{b} \theta_i^{(j)} \cdot \tilde{h}_{l_2}^{i,(j)}. \tag{10}$$

比较综合预测值$(\tilde{h}_{l_2}^{i,(b)})_{1\times M_2}$和标签$(h_{l_2}^i)_{1\times M_2}$, 得到接续组合$\{DN_{(1)}^i, \ldots, DN_{(b)}^i\}$在图像验证集上的预测精度为

$$A_2^{i,(b)}=\frac{\sum\limits_{l_2=1}^{M_2}S(h_{l_2}^i,\tilde{h}_{l_2}^{i,(b)})}{M_2}. \tag{11}$$

依据$(Q-1)$种接续组合在图像验证集上的预测精度$A_2^{i,(b)}$ $(b=2,\ldots,Q)$, 确定最佳接续组合$\{DN_{(1)}^i,\ldots,DN_{(b^*)}^i\}$ $(b^*\in\{2,\ldots,Q\})$.

上述接续组合预测有别于传统的集成学习方法. Bagging、Boosting、Stacking 是三种典型的集成学习方法. Bagging 和 Boosting 方法聚焦基于样本的集成, 通过样本重采样或迭代加权提升预测精度; Stacking 方法引入元学习结构来融合多分类器以提升预测精度. 而接续组合方法从模型层面出发, 利用统计检验对深度网络性能进行鲁棒排序, 进而选择最佳的网络组合进行预测, 在提升预测精度的同时兼顾预测可靠性和效率.

2.3 准则融合阶段

准则预测阶段, 利用图像训练集、图像验证池、图像验证集确定各准则上的最佳深度网络接续组合,实现各准则上输入图像的区间数预测生成. 融合各准则上的区间数预测值, 将产生总体预测值. 如何确定融合所需的准则权重, 是准则融合阶段的核心问题.

针对这一问题, 提出一种自适应准则权重学习方法, 从文本训练集和文本验证集中学习准则权重. 首先, 从文本训练集(U_1,V_1)中学习准则权重.

假设 1[16] 给定文本训练集(U_1,V_1), 各准则上评价值与总体评价值的相似度与准则权重成正比.

遵从假设 1, 依据文 [16] 中的权重计算方法, 从文本训练集(U_1,V_1)中学习的准则权重为

$$w_i=\frac{\sum\limits_{m_1=1}^{N_1}w_{i,m_1}}{N_1},\ i=1,\ldots,L, \tag{12}$$

其中,

$$w_{i,m_1}=\frac{S(u_{m_1}^i,v_{m_1})}{\sum\limits_{k=1}^{L}S(u_{m_1}^k,v_{m_1})}. \tag{13}$$

为了进一步提高准则融合的预测性能, 在文本验证集上构建如下自适应优化模型, 拟合U_2和V_2之间的关系.

$$Max\frac{\sum\limits_{m_2=1}^{N_2}S(v_{m_2},\tilde{v}_{m_2})}{N_2} \tag{14}$$

$$\text{s.t. }\tilde{v}_{m_2}=\sum_{i=1}^{L}\overline{w}_i\cdot u_{m_2}^i, \tag{15}$$

$$\overline{w}_i=\frac{g(w_i)}{\sum\limits_{k=1}^{L}g(w_k)},\ i=1,\ldots,L, \tag{16}$$

$$w_i=\frac{\sum\limits_{m_1=1}^{N_1}w_{i,m_1}}{N_1}, \tag{17}$$

$$w_{i,m_1}=\frac{S(u_{m_1}^i,v_{m_1})}{\sum\limits_{k=1}^{L}S(u_{m_1}^k,v_{m_1})}. \tag{18}$$

这里, $g(\cdot)$表示一个抽象函数, 满足如下条件.

定义 4 设$g(\cdot)$为式 (16) 中的抽象函数, 满足以下条件:

(1) 有界性. $g(1)=1$, $g(0)=0$.

(2) 有界性. $0\leqslant g(w_i)\leqslant 1$.

(3) 单调性. 当$0<w_2<w_1<1$时, 有$g(w_2)<g(w_1)$.

(4) 连续性. $\lim\limits_{w_i\to w_i^0}g(w_i)=g(w_i^0)$.

本文中, 依据定义 4 中的相关条件, 选取$g(w_i)$的具体函数形式为$(\frac{\arcsin(w_i)}{\pi/2})^{\alpha}$, 包含一个参数$\alpha$, 作为变量参与式 (14)-(18) 的模型优化过程中. 在文本验证集上求解此模型, 得到最优参数α[8].

给定图像测试集(G_4^i,H_4^i,H_4), 基于准则预测阶段确定的最佳接续组合$\{DN_{(1)}^i,\ldots,DN_{(b^*)}^i\}$和准则融合阶段确定的准则权重$w_i$ $(i=1,\ldots,L)$和最优参数α^*, 进行输入图像的总体预测.

由G_4^i中的图像, 使用最佳接续组合$\{DN_{(1)}^i,\ldots,DN_{(b^*)}^i\}$计算得各准则上的区间数预测值$\tilde{h}_{l_4}^{i,(b^*)}$. 与图像数据集中的标签$H_4^i$进行比较, 得到准则$e_i$ $(i=1,\ldots,L)$上的预测精度

$$A_4^i=\frac{\sum\limits_{l_4=1}^{M_4}S(h_{l_4}^i,\tilde{h}_{l_4}^{i,(b^*)})}{M_4}. \tag{19}$$

依据式 (15)-(16), 使用基于准则权重w_i $(i=1,\ldots,L)$和最优参数α^*计算得到图像$g_{l_4}^i$ $(i=1,\ldots,L)$的总体预测值$\tilde{h}_{l_4}$. 进而由总体预测值$\tilde{h}_{l_4}$和总体标签h_{l_4}得到总体预测精度为

$$A_4=\frac{\sum\limits_{l_4=1}^{M_4}S(h_{l_4},\tilde{h}_{l_4})}{M_4}. \tag{20}$$

3 乳腺病灶辅助诊断

3.1 实验数据

与安徽合肥某三甲医院超声部合作, 采集乳腺病灶诊断相关的图像数据集和文本诊断报告. 基于图像和文本数据, 运用提出方法进行乳腺病灶辅助诊断, 以检验方法的有效性.

依据诊断报告分析与合作医生的专业知识, 选取边界、轮廓、回声、钙化、血流等 5 个准则[19], 即 e_i $(i=1,\ldots,5)$, 进行辅助诊断分析. 聚焦准则预测阶段, 在边界、轮廓、回声、钙化等 4 个准则上使用灰阶图像, 而血流准则上则使用彩色多普勒血流 (CDFI) 图像. 5 个准则上图像训练集和验证池包含的图像数量为 (4 594, 4 447, 2 114, 1 498, 2 759) 和 (3 075, 2 964, 1 410, 999, 1 840). 其中, 边界、轮廓、回声、钙化这 4 个准则上的图像验证集包含 500 张灰阶图像, 而血流准则上的图像验证集包含 500 张 CDFI 图像. 5 个准则上的图像测试集也包含 500 张灰阶图像和 500 张 CDFI 图像, 与图像验证集不同的是, 500 张灰阶图像和 500 张 CDFI 图像一一对应, 表征相同的乳腺病灶.

聚焦准则融合阶段, 文本训练集 (U_1, V_1) 包含 882 份标注文本, 文本验证集 (U_2, V_2) 包含 220 份标注文本, 分别用于准则权重和自适应权重函数的学习. 在图像数据集和文本数据集中, 使用乳腺影像报告和数据系统 (BIRADS) 的 5 个类别, 即 {BIRADS 3, BIRADS 4A, BIRADS 4B, BIRADS 4C, BIRADS 5}, 对应的患癌风险区间来表示标签, 即 $\{C_b\ (b=1,\ldots,5)\}$= {[0, 0.02], [0.03, 0.1], [0.11, 0.5], [0.51, 0.94], [0.95, 1]}[20]. 同时, 文本数据集中各准则上的评价值也由病灶观察转化为风险区间, 具体过程见文 [19]. 实验中所有图像数据集和文本数据集的标签, 都由 8 位具有 6-10 年以上工作经验且年龄分布合理的超声医生协同给出, 为标签质量提供较好的保障.

3.2 深度网络选取与训练

超声医生依据边界、轮廓、回声、钙化、血流等 5 个准则进行乳腺病灶诊断, 在各准则上关注的图像特征有所不同. 在边界和轮廓两个准则上, 超声医生关注于乳腺病灶的形态学特征; 在回声和钙化两个准则上, 超声医生关注于病灶的声学特征; 而对于血流准则, 超声医生关注于病灶的位置信息特征. 对灰阶图像进行形态学分析, 侧重于病灶边界与轮廓的空间几何与形状连续性, 要求深度网络在局部边缘敏感、多尺度上下文建模及适度长程依赖等方面具备较强能力, 以抑制斑点噪声、稳定边缘识别. 灰阶图像的声学分析聚焦内部回声模式 (如等回声、低回声、混合回声等) 与高亮小目标钙化点, 需要兼顾局部纹理建模、多尺度统计特性以及对小目标的高分辨率刻画. 血流图像的位置信息分析则强调灌注范围与相对位置等空间分布特征, 对深度网络的空间注意力与跨区域依赖建模能力提出更高要求.

为满足 5 个准则上的形态学分析、声学分析、位置信息分析要求, 在准则预测阶段选取 10 种深度网络组成候选集合 $DN=\{DN_q\ (q=1,\ldots,10)\}$. 首先选取 4 种具有代表性的现有深度网络作为基础网络, 包括 ConvNeXt、EfficientNetV2、Swin Transformer V2 和 ViT-B, 分别记为 DN_1, DN_2, DN_3, DN_4. 其中, DN_1 和 DN_2 属于卷积神经网络家族, 依托局部感受野与多尺度卷积结构, 在病灶边缘、轮廓与局部纹理 (包括部分钙化小目标) 的稳定提取方面具有优势[20-21]; DN_3 和 DN_4 属于 Transformer 及其变体结构, 更擅长捕捉跨区域长程依赖与全局上下文信息, 对复杂轮廓形态、血流空间分布以及整体回声模式更为敏感[22-23].

为了进一步针对声学与位置信息特征强化通道级与空间级显著性, 在上述基础网络之上引入轻量级注意力机制, 构造 6 种改进网络. 具体而言, 在 ConvNeXt 与 EfficientNetV2 主干上分别嵌入高效通道注意力模块 ECA(Efficient Channel Attention)[24] 与空间注意力模块 SA(Spatial Attention)[25], 得到 DN_5 = ConvNeXt_ECA, DN_6 = ConvNeXt_SA, DN_7 = EfficientNetV2_ECA, 和 DN_8 = EfficientNetV2_SA. 同时, 基于卷积与 Transformer 混合架构 NextViT, 结合 ECA 与 SA 构造 DN_9 = NextViT_ECA 和 DN_{10} = NextViT_SA.

使用 5 个准则上的图像训练集, 在带 NVIDIA Tesla V100 GPU 的服务器上进行 10 个深度网络的训练. 相关训练参数设置如下: 优化器采用 AdamW, 权重衰减系数设为 1×10^{-4}, 初始学习率设为 1×10^{-4}, 最低学习率设为 1×10^{-5}, 训练批次设为 16, 最大训练轮次设为 200 轮; 引入早停机制, 当验证集上性能连续 25 轮次无提升时停止训练. 实验环境配置用于保障模型在现有图像数据规模上实现高效训练与稳定收敛. 通过训练, 得到由超声图像到区间数标签集合 $\{C_b\ (b=1,\ldots,5)\}$ 的映射函数 $F_q^i\colon g_{l_1}^i \to h_{l_1}^i$ $(i=1,\ldots,5,\ q=1,\ldots,10)$.

3.3 准则预测实验分析

基于 5 个准则上训练后的 10 种深度网络, 采用

有放回采样策略进行网络性能排序. 这里, N_p和N_r分别设为 50% 和 500. 在 5 个准则上, 从验证池中取 50% 的图像数据, 使用映射函数F_q^i $(i=1,\ldots,5,\ q=1,\ldots,10)$进行预测, 重复 500 次, 得到预测精度向量. 对任意两种深度网络的预测精度向量进行威尔科克森符号秩检验, 包括左尾检验和右尾检验, 得到相应的统计显著性p值, p_L^{i,q_1,q_2}和p_R^{i,q_1,q_2} $(i=1,\ldots,5,\ q_1,q_2=1,\ldots,10)$. 限于篇幅, 略去相关的$p$值.

依据式 (4)-(5), 由p_L^{i,q_1,q_2}和p_R^{i,q_1,q_2} $(i=1,\ldots,5,\ q_1,q_2=1,\ldots,10)$得到左尾检验和右尾检验的指示$I_L^{i,q_1,q_2}$和$I_R^{i,q_1,q_2}$. 以准则“边界”为例, 其上的检验指示对如表 2 所示. 相关结果印证了性质 1, 即$I_L^{i,q_1,q_2}+I_R^{i,q_1,q_2}\leqslant 1$. 依据定义 2, 可由 5 个准则上的检验指示对得到其上 10 种深度网络的性能排序, 如表 3 所示.

表2 准则“边界”上的检验指示对

	DN_1	DN_2	DN_3	DN_4	DN_5	DN_6	DN_7	DN_8	DN_9	DN_{10}
DN_1	(0, 0)	(1, 0)	(0, 1)	(0, 1)	(1, 0)	(0, 1)	(0, 1)	(1, 0)	(1, 0)	(0, 1)
DN_2	(1, 0)	(0, 0)	(0, 1)	(0, 1)	(1, 0)	(0, 1)	(0, 1)	(1, 0)	(1, 0)	(0, 1)
DN_3	(1, 0)	(1, 0)	(0, 0)	(0, 1)	(1, 0)	(1, 0)	(0, 1)	(1, 0)	(1, 0)	(0, 1)
DN_4	(1, 0)	(1, 0)	(1, 0)	(0, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
DN_5	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 0)	(0, 1)	(0, 1)	(0, 1)	(1, 0)	(0, 1)
DN_6	(1, 0)	(1, 0)	(1, 0)	(0, 1)	(1, 0)	(0, 0)	(0, 1)	(1, 0)	(1, 0)	(0, 1)
DN_7	(1, 0)	(1, 0)	(1, 0)	(0, 1)	(1, 0)	(1, 0)	(0, 0)	(1, 0)	(1, 0)	(1, 0)
DN_8	(1, 0)	(1, 0)	(0, 1)	(0, 1)	(1, 0)	(0, 1)	(0, 1)	(0, 0)	(1, 0)	(0, 1)
DN_9	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 0)	(0, 1)
DN_{10}	(1, 0)	(1, 0)	(1, 0)	(0, 1)	(1, 0)	(1, 0)	(0, 1)	(1, 0)	(1, 0)	(0, 0)

表3 5 个准则上 10 种深度网络的性能排序

	DN_1	DN_2	DN_3	DN_4	DN_5	DN_6	DN_7	DN_8	DN_9	DN_{10}
边界	3	3	6	10	2	6	9	3	1	8
轮廓	1	7	5	4	2	3	9	6	8	10
回声	4	7	3	6	1	1	10	8	4	9
钙化	4	6	1	4	8	1	7	1	9	10
血流	8	3	5	10	7	2	5	1	3	8

基于表 3 中各准则上的网络性能排序, 依据定义 3 可得 5 个准则上的 9 种深度网络接续组合. 对于每一组合, 使用式 (8)-(9) 计算组合中各网络的权重, 进而基于由 500 张灰阶图像和 500 张 CDFI 图像组成的图像验证集, 使用式 (10)-(11) 计算得到预测精度, 如表 4 所示. 易知,5 个准则上的最佳网络组合分别为$\{DN_{(1)}^1,\ldots,DN_{(10)}^1\}$、$\{DN_{(1)}^2,\ldots,DN_{(5)}^2\}$、$\{DN_{(1)}^3,\ldots,DN_{(10)}^3\}$、$\{DN_{(1)}^4,\ldots,DN_{(10)}^4\}$、$\{DN_{(1)}^5,DN_{(2)}^5\}$, 对应的预测精度在表 4 中加粗表示.

基于由具有对应性的 500 张灰阶图像和 500 张 CDFI 图像组成的图像测试集, 使用 5 个准则上的最佳网络接续组合计算其上的区间数预测值, 进而依据式 (19) 与标签对比得到各准则上的预测精度为 (0.9376, 0.8666, 0.8123, 0.7997, 0.9111). 对比表 4 易知,10 种深度网络组合在回声和钙化两个准则上的测试精度明显高于验证精度, 而最佳网络接续组合在轮廓和血流两个准则上的测试精度也明显优于验证精度.

各准则上预测性能差异源自医学判读、图像特征、数据特性的综合作用. 灰阶图像中, 乳腺病灶与背景过渡、病灶边缘模糊度等具有较稳定的可视模式; 而回声和钙化受散斑噪声、细小高回声灶可见性等

表4 5 个准则上基于验证集的 9 种网络接续组合的预测精度

	组合1	组合2	组合3	组合4	组合5	组合6	组合7	组合8	组合9
边界	0.9122	0.9270	0.9270	0.9270	0.9273	0.9273	0.9269	0.9274	**0.9281**
轮廓	0.7849	0.7845	0.7758	**0.7923**	0.7871	0.7799	0.7804	0.7828	0.7807
回声	0.6630	0.6629	0.6637	0.6637	0.6676	0.6535	0.6664	0.6627	**0.6714**
钙化	0.7322	0.7322	0.7299	0.7299	0.7392	0.7326	0.7314	0.7394	**0.7426**
血流	**0.7916**	0.7799	0.7799	0.7749	0.7749	0.7544	0.7675	0.7564	0.7658

因素影响, 其判读展现出细粒度纹理和弱信号线索强关联的特性. CDFI 图像中, 血流分布和血供丰富度的结构化表现较好, 呈现出较为明显的特征. 综合以上分析, 边界、轮廓、血流准则上的图像特征较易为深度网络识别, 而回声和钙化准则上的图像特征则相对较难识别. 相应地, 回声和钙化准则上的预测性能更易受到数据波动和数据分布的影响. 另一方面, 边界和轮廓准则上的训练图像数量最大、准则和血流准则上次之、钙化准则上最少, 对各准则上的预测精度也会产生相应的影响. 综合医学维度和数据维度, 边界、轮廓、血流准则上的预测精度较高, 而回声和钙化准则上的预测精度较低, 这一结果是合理的.

为了凸显网络接续组合的性能优势, 选择 ConvNeXt、EfficientNetV2、Swin Transformer V2 和 ViT-B 等 4 种基础网络, 在相同实验设置和数据集条件下, 进行边界、轮廓、回声、钙化、血流等 5 个准则上的预测实验, 得到的预测精度如图 2 所示.

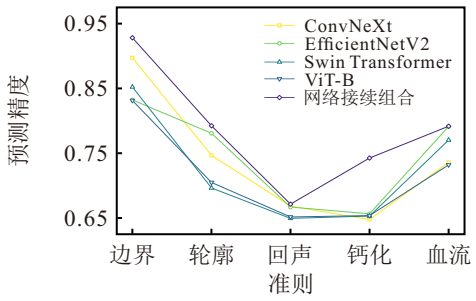

图2　准则预测阶段网络接续组合与 4 种基础深度网络的预测精度对比

由图 2 可知, 网络接续组合在 5 个准则上均存在性能优势, 阐释了多种深度网络组合对提升预测性能的重要作用. 边界准则上的预测精度最高, 说明网络组合有助于强化边缘过渡和形态特征; 钙化准则上的性能提升最为显著, 说明网络组合能够兼顾多尺度细节, 以应对钙化相关的信号弱、粒度细、噪音干扰大等特征带来的预测挑战; 轮廓、回声、血流等准则上也有一定的性能提升. 总体而言, 网络组合通过发挥不同网络的特征提取优势, 有效提高了 5 个准则上的预测性能, 为准则融合阶段提供了可靠的区间数预测值基础.

3.4　准则融合实验分析

遵从假设 1, 依据式 (12)-(13) 从文本训练集 (U_1, V_1) 中学习 5 个准则的权重为 $(w_1, \ldots, w_5)$ = (0.1896, 0.1371, 0.2264, 0.2195, 0.2274). 以学习的准则权重和文本验证集 (U_2, V_2) 为基础, 选择自适应权重函数 $g(w_i) = \left(\frac{\arcsin(w_i)}{\pi/2}\right)^{\alpha}$, 依据式 (14)-(18) 构建优化模型, 求解模型得到最优参数 $\alpha^* = 0.3916$.

基于带最优参数 α^* 的 $g(w_i)$, 依据式 (15)-(16), 融合图像测试集在 5 个准则上最佳网络接续组合的区间数预测值, 得到总体预测值. 进而依据式 (20) 与总体标签进行对比, 得到总体预测精度 $A_4 = 0.8440$.

为了凸显本文提出的决策方法的总体性能优势, 基于带 BIRADS 总体标签的图像数据集, 在相同的实验设置下, 选取 ResNet-50、ConvNeXt、EfficientNetB1、EfficientNetV2、Swin Transformer、ViT-B 等 6 种典型的深度网络进行端到端的预测实验, 得到的预测精度为 (0.7414, 0.7540, 0.7509, 0.7640, 0.7614, 0.7545). 对比 6 种网络的预测精度和提出方法的预测精度 (0.8440), 易知提出方法具有明显的性能优势. 这一优势来源于准则预测阶段的多个网络组合和准则融合阶段的自适应权重学习的综合作用, 同时说明提出的决策方法同时兼顾预测高性能、过程可解释、预测高可靠等优点.

由 3.1 节可知, 以上实验基于 8 位超声医生群体给出的标签完成. 当一位新的超声医生愿意给出所有图像数据集的标签, 则在准则预测阶段可以得到刻画此医生诊断偏好的最佳深度网络接续组合, 实现图像到区间数预测值的映射. 然而, 这项任务非常耗时, 不利于提出方法的高效利用. 在此情形下, 固化准则预测阶段并利用新超声医生自身的文本诊断报告进行准则融合, 是一个可行路径. 基于此思想, 以下使用 14 位超声医生 $R_d(d = 1, \ldots, 14)$ 的文本诊断报告进行准则融合实验, 分析他们对群体标签的拟合程度.

依据文 [19], 将 14 位超声医生的诊断报告转化为他们各自的文本训练集 (U_1^d, V_1^d) $(d = 1, \ldots, 14)$ 和文本验证集 (U_2^d, V_2^d) $(d = 1, \ldots, 14)$, 包含的文本数据为 (484, 415, 194, 549, 347, 250, 334, 183, 243, 201, 387, 273, 438, 298) 和 (80, 69, 32, 91, 58, 42, 56, 31, 41, 33, 65, 46, 73, 50).

依据式 (12)-(13), 由 14 位超声医生的文本训练集计算得到他们对应的准则权重, 如表 5 所示. 以表 5 中的准则权重和各位医生的文本验证集为基础, 依据式 (14)-(18) 构建优化模型, 求解模型得到 14 位超声医生的自适应权重函数的最优参数为 (9.7938, 4.8141, 17.3988, 13.9333, 2.6825, 0.1564, 3.0874, 6.9488, 0.8532, 19.2706, 16.2190, 10.6378, 18.0212, 0.1952),以及各位医生在文本验证集上的最大预测精度, 如图 3 所示.

基于 14 位超声医生带最优参数的 $g(w_i)$, 依据

表5 14 位超声医生的准则权重

	w_1	w_2	w_3	w_4	w_5
R_1	0.2040	0.2083	0.1879	0.2106	0.1892
R_2	0.2004	0.2637	0.1931	0.1850	0.1578
R_3	0.2303	0.2257	0.1827	0.1701	0.1912
R_4	0.2015	0.2128	0.2047	0.1998	0.1812
R_5	0.1657	0.2373	0.1964	0.2142	0.1864
R_6	0.2110	0.1791	0.1998	0.2122	0.1980
R_7	0.2060	0.2358	0.1707	0.1814	0.2060
R_8	0.2321	0.1870	0.1951	0.2002	0.1856
R_9	0.2297	0.1819	0.1984	0.1886	0.2014
R_{10}	0.2094	0.1511	0.2097	0.2124	0.2173
R_{11}	0.2065	0.1847	0.2075	0.1990	0.2021
R_{12}	0.2294	0.1922	0.1932	0.1930	0.1922
R_{13}	0.2227	0.1704	0.2025	0.2018	0.2026
R_{14}	0.1594	0.2112	0.2791	0.1902	0.1601

式 (15)-(16), 融合图像测试集在 5 个准则上最佳网络接续组合的区间数预测值, 得到总体预测值. 进而依据式 (20) 与总体标签进行对比, 得到 14 位医生对群体标签的总体拟合精度, 如图 3 所示.

观察图 3 可知, 14 位超声医生在文本验证集上的预测精度普遍较高, 体现了各位医生自身偏好的一致性. 然而, 自身偏好一致性不同于各位医生对图像测试集标注群体的拟合程度, 表现为验证精度与测试精度的非同频变化, 即不成正比例变化. 这一现象源自乳腺病灶超声诊断的核心特征, 即医生的经验依赖. 在长期的超声诊断过程中, 不同医生逐渐沉淀出 5 个准则上超声图像到乳腺病征的识别偏好、

不同准则上乳腺病征的重视程度、超声图像到 BIRADS 类别的诊断偏好等. 由于超声医生独立问诊, 他们的偏好一般存在差异, 使得拟合他们偏好的模型相应地存在差异. 14 位医生中每一位的偏好与 8 位医生群体的偏好存在不同程度的差异, 导致每一位医生的验证精度与测试精度非同频变化.

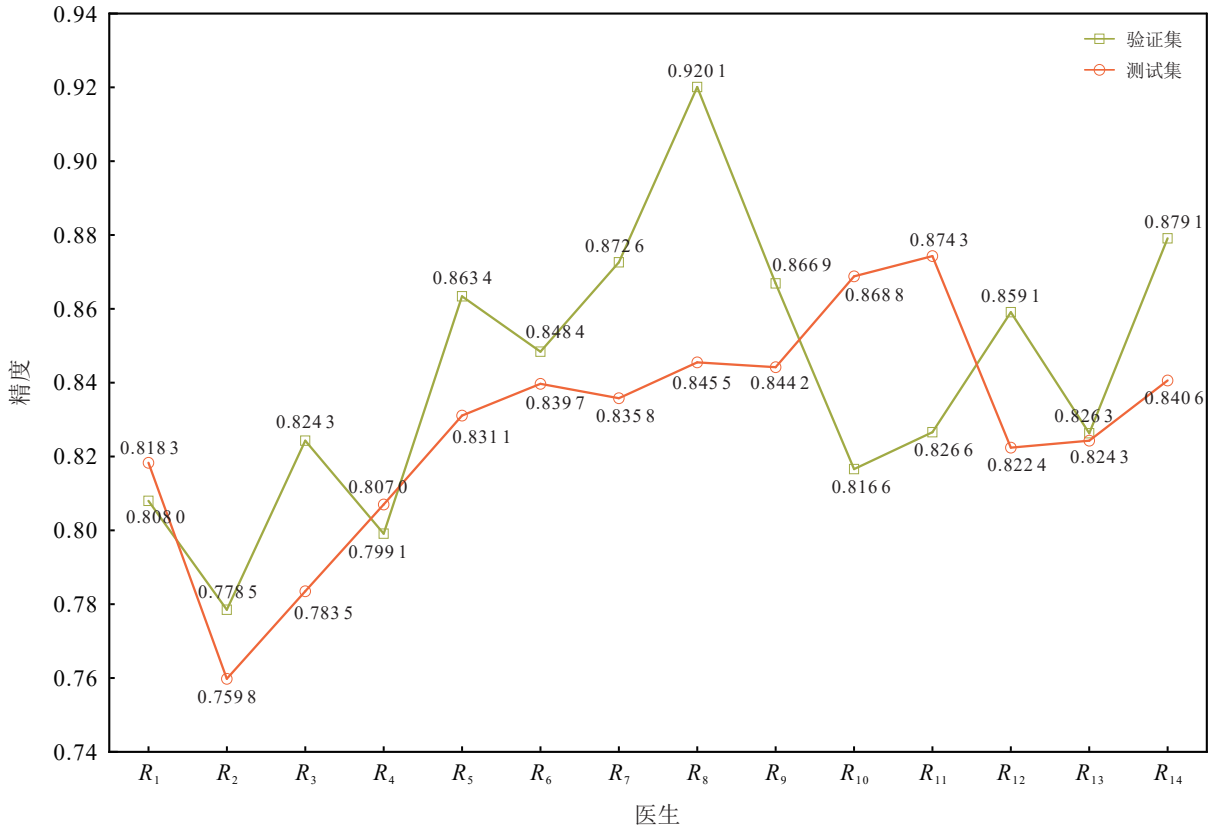

图3 14 位超声医生在文本验证集和图像测试集上的预测精度

4 结 论

以深度网络为代表的人工智能技术快速发展, 极大地提升了统一处理多模态数据的能力, 推动基于多模态数据的决策方法向各个领域快速渗透. 人们在庆幸多模态数据带来的更有深度、更全面决策支持的同时, 逐渐意识到其引发的新的决策难题, 即如何基于多模态数据产生可解释、可靠的决策支持. 应对这一难题, 充分考虑多阶段决策的优势, 提出了一种图像和文本数据驱动的两阶段区间决策方法. 该方法分为准则预测和准则融合两个阶段. 在准则预测阶段, 以图像训练集、图像验证池、图像验证集、图像测试集为基础, 依据各准则上的图像特征提取需求确定可行的深度网络集合; 设计基于有放回采样的统计比较方法得到两个网络间左尾和右尾检

验的指示对, 依此实现各准则上深度网络的性能排序; 进而构建网络接续组合的选择方法, 确定最佳的网络接续组合, 产生可靠的各准则预测值, 有效防止了基于深度网络集合的组合爆炸问题. 在准则融合阶段, 以文本训练集和文本验证集为基础, 学习准则权重, 在此基础上构建优化模型学习自适应权重函数. 利用第二阶段学习的准则权重和自适应权重函数, 融合第一阶段各准则上的预测值, 产生可解释的总体预测值, 进而与总体标签对比得到总体预测精度.

与安徽合肥某三甲医院超声部合作, 基于乳腺病灶诊断的彩超图像和诊断报告构建图像数据集和文本数据集, 开展实验研究. 依据各准则上的图像特征提取需求, 选取 4 种基础深度网络并改进 6 种网络, 展示了基于其上的详细实验过程, 验证了提出方法的应用性与有效性.

本文提出的决策方法遵循后期融合思想, 利用基于图像数据的准则预测和基于文本数据的准则融合两个阶段模拟决策者的多准则决策过程, 虽具有较好的可解释性, 但并未对图像数据和文本数据进行统一表征与对齐, 使得他们有所割裂, 不利于决策精度的提升. 例如, 乳腺超声诊断中, “边缘毛糙”、“后方声影”、“点状强回声”等病征描述与超声图像中病灶的局部特征和信号线索等存在强关联性, 考虑这些关联性有助于提升决策精度. 另一方面, 本文实验基于 10 种深度网络和单中心的乳腺诊断图像和文本数据展开, 尽管验证了提出方法的有效性, 但在网络多样性和多中心数据方面仍有待提升. 鉴于此, 下一步将引入中后期融合机制, 通过跨模态注意力等深层交互方式将图像关键区域和文本语义进行对齐与融合, 实现语义引导的特征强化与不确定性校准, 进而提升决策精度. 同时, 运用提出方法解决多中心的疾病辅助诊断问题, 并依据不同疾病辅助诊断场景下的需求进一步探索新型深度网络的选取和改进路径, 不断提升提出方法的适用性. 持续关注带有准则标签和总体标签的公开数据集, 利用其进一步验证提出方法的有效性与可推广性.

参考文献 (References)

[1] Baltrušaitis T, Ahuja C, Morency L P. Multimodal machine learning: A survey and taxonomy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(2): 423-443.

[2] Chen X L, Xie H R, Tao X H, et al. Artificial intelligence and multimodal data fusion for smart healthcare: Topic modeling and bibliometrics[J]. Artificial Intelligence Review, 2024, 57(4): 91.

[3] Duan Y Q, Edwards J S, Dwivedi Y K. Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda[J]. International Journal of Information Management, 2019, 48: 63-71.

[4] Liu Z, Han H G, Yang H Y, et al. Knowledge-aided and data-driven fuzzy decision making for sludge bulking[J]. IEEE Transactions on Fuzzy Systems, 2023, 31(4): 1189-1201.

[5] 张海利, 王普, 高学金, 等. 基于批次图像化的卷积自编码故障监测方法[J]. 控制与决策, 2021, 36(6): 1361-1367.
(Zhang H L, Wang P, Gao X J, et al. Fault detection of batch image-based convolutional autoencoder[J]. Control and Decision, 2021, 36(6): 1361-1367.)

[6] Zhao F Y, Cao X G, Duan Y, et al. ML-LQI: A multi-modal learning method for low-quality imbalanced modality data with interpretability[J]. Knowledge-Based Systems, 2026, 333: 114991.

[7] Chu X L, Sun B Z, Zou H, et al. Multi-modal incomplete label information three-way bidirectional decision-making: Applications of disease assessment[J]. Information Fusion, 2025, 113: 102615.

[8] Yuan W K, Lin T, Jiang Z R, et al. Insights into the impact of visual and textual information on investment decision-making: A multimodal business plan analysis *via* deep representation learning[J]. Expert Systems with Applications, 2026, 296: 128911.

[9] Liu C L, Wang Y L, Yang C H, et al. Multimodal data-driven reinforcement learning for operational decision-making in industrial processes[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(1): 252-254.

[10] Xu S X, Chen Y F, Ma C, et al. Deep evidential fusion network for medical image classification[J]. International Journal of Approximate Reasoning, 2022, 150: 188-198.

[11] Lan Y B, Guo Y Q, Chen Q Z, et al. Visual question answering model for fruit tree disease decision-making based on multimodal deep learning[J]. Frontiers in Plant Science, 2023, 13: 1064399.

[12] Fan L L, Wang Y T, Zhang H, et al. Multimodal perception and decision-making systems for complex roads based on foundation models[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 54(11): 6561-6569.

[13] Chen Z Y, Chai N J, Wang J Q, et al. Restaurant recommendations under multimodal online reviews: A novel method based on image captioning and text analysis with multi-criteria decision-making[J]. Information Processing & Management, 2026, 63(1): 104308.

[14] Shao Z M, Dou W B, Pan Y. Dual-level Deep Evidential Fusion: Integrating multimodal information for enhanced reliable decision-making in deep learning[J]. Information Fusion, 2024, 103: 102113.

[15] He Q, Li X, Nathan Kim D W, et al. Feasibility study of

a multi-criteria decision-making based hierarchical model for multi-modality feature and multi-classifier fusion: Applications in medical prognosis prediction[J]. Information Fusion, 2020, 55: 207-219.

[16] Fu C, Chang W J, Liu W Y, et al. Data-driven group decision making for diagnosis of thyroid nodule[J]. Science China Information Sciences, 2019, 62(11): 212205.

[17] Wang N, Meng X J, Meng X C, et al. Convolution-embedded vision transformer with elastic positional encoding for pansharpening[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5413809.

[18] Lovric M. International encyclopedia of statistical science[M]. Berlin: Springer-Verlag, 2011.

[19] Fu C, Wang D Y, Chang W J. Data-driven analysis of influence between radiologists for diagnosis of breast lesions[J]. Annals of Operations Research, 2023, 328(1): 419-449.

[20] Liu Z, Mao H Z, Wu C Y, et al. A ConvNet for the 2020s[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 11966-11976.

[21] Tan M X, Le Q V. EfficientNetV2: Smaller models and faster training[J/OL]. 2021, arXiv: 2104.00298.

[22] Liu Z, Hu H, Lin Y T, et al. Swin transformer V2: Scaling up capacity and resolution[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 11999-12009.

[23] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J/OL]. 2020, arXiv: 2010.11929.

[24] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 11531-11539.

[25] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]. Computer Vision – ECCV 2018. Cham: Springer, 2018: 3-19.

作者简介

付超 (1978−), 男, 教授, 博士, 博士生导师, 主要研究方向为数据驱动的决策方法、面向医疗辅助诊断的智能决策, E-mail: chaofu@hfut.edu.cn;

崔凯旋 (1996−), 男, 博士研究生, 主要研究方向为数据驱动的决策方法、面向医疗辅助诊断的智能决策, E-mail: ckx888111@163.com;

倪文青 (1999−), 女, 博士研究生, 主要研究方向为数据驱动的决策方法、面向医疗辅助诊断的智能决策, E-mail: wenqing8726@163.com;

薛旻 (1990−), 女, 副教授, 博士, 硕士生导师, 主要研究方向为数据驱动的决策方法、面向医疗辅助诊断的智能决策, E-mail: 2019800012@hfut.edu.cn;

刘卫勇 (1977−), 男, 副主任医师, 博士, 硕士生导师, 主要研究方向为超声疾病诊断、面向超声辅助诊断的智能决策, E-mail: weiyongliu@ustc.edu.cn.

科研团队简介

付超教授科研团队依托合肥工业大学过程优化与智能决策教育部重点实验室, 聚焦智能制造工程管理和智慧医疗健康管理的重大需求, 面向复杂产品开发过程管理、全生命周期过程优化与风险评估、生产过程动态自组织管理、多源异构信息情形下的不确定决策、考虑一致可靠约束的群体决策、基于多模态数据的可靠可解释决策等重要问题, 围绕多模态数据驱动的智能决策方法、基于数据与知识融合的智能决策方法开展了基础理论研究工作, 成果应用于轿车整车开发流程优化与过程管理、高铁齿轮箱风险管控与维护管理、智能工厂和无人生产线自组织重构、关键零部件的生产过程管理、群体医疗辅助诊断、智能超声辅助诊断等实际工程场景中, 取得了较好的工程管理成效.

课题组负责人付超教授是国家自然科学基金优秀青年基金获得者, 中国优选法统筹法与经济数学研究会计算机模拟分会副理事长、中国自动化学会智能推理与决策专业委员会副主任委员, 《管理学报》和《Journal of Control and Decision》编委. 课题组现有教授 1 人, 副教授 3 人, 博士生 5 人, 硕士生 13 人, 承担国家自然科学基金项目 8 项, 科技部重点研发计划课题 1 项, 获安徽省科学技术奖一等奖 1 项、教育部科技进步奖一等奖 1 项、教育部自然科学奖二等奖 1 项、安徽省教学成果特等奖 2 项、安徽省教学成果一等奖 2 项. 课题组在 SCIENCE CHINA Information Sciences、 IEEE TSMC、 IEEE TFS、 IEEE TCYB、EJOR、仪器仪表学报、系统工程理论与实践、中国管理科学等信息和管理领域国内外著名学术期刊上发表论文 100 余篇, 出版学术专著 2 部, 国际、国内会议学术报告 20 余次. 课题组与英国曼彻斯特大学、美国堪萨斯大学、英国女王大学等多所国际知名高校保持密切科研合作.