

第三轮 修改意见回复

衷心的感谢各位专家对稿件的认真审阅，稿件也在专家们的建议中逐渐完善。我们对第三轮专家意见的回复具体如下：

给作者的意见：

1. 摘要首句提到“抑郁症与焦虑障碍共病现象普遍，而传统分类诊断方法存在局限性”。但并没有写清楚局限性是什么，请进行说明。

回应：感谢专家的宝贵意见！我们已根据您的建议对原文进行了修改，并在摘要的第一句中对局限性进行了简要阐述，同时进行了标红处理：传统的抑郁症与焦虑障碍共病诊断方法往往将这两种独立疾病简单叠加，忽视了症状间的动态交互作用及其个体异质性。希望这样的修改能更清晰地表达传统分类诊断方法在处理抑郁症与焦虑障碍共病时存在的不足。

2. 摘要中 AX-CPT、AY 与 BX、PBI 等简写，需要写出全称。

回应：感谢专家的建议。我们已在修改后的摘要中添加了 AX-CPT 和 PBI 的全称，并进行了相应标注。例如，“AX-CPT (The AX Continuous Performance Test, 简称 AX-CPT)”和“PBI (The Proactive Behavioral Index, 简称 PBI)”，以便更清楚地表达这些术语的具体含义。以上修改已在原文中标红。此外，关于 AY 与 BX，鉴于它们是 AX-CPT 范式中定义的特定试次类型，因此未进行全称补充。

3. 请完善摘要书写。目前的摘要中未出现被试人数，且需要按照目的方法结果结论的结构来写。

回应：感谢专家的反馈。在摘要中，我们已补充了被试的具体人数信息，具体包括抑郁症与焦虑障碍共病组 50 人、抑郁症组 50 人、焦虑障碍组 51 人和健康组 50 人。此外，摘要的结构已根据“目的、方法、结果、结论”的标准框架进行了重新调整，以确保内容更为规范、清晰。以上修改已在原文中标红。

4. 14 页“为更好的解决上述问题，双重认知控制理论(Dual Mechanisms of Control, 简称 DMC)将认知控制划分为主动性控制模式与反应性控制模式(Braver,2012)”此段过长，请重新调整完善，并具体写出使用 AX-CPT 范式的原因。

回应：感谢专家的宝贵意见。我们已根据您的反馈对第 14 页的段落进行了修改，以使

内容更加简洁，并更详细地阐述了使用 AX-CPT 范式的原因。修改后的内容更为简明地表达了双重认知控制（DMC）的核心概念，同时明确指出了 AX-CPT 范式在区分主动性控制与反应性控制方面的独特优势。修改后的段落已在原文中标红。

5. Cohen $f=0.25$, α error 等字母需要斜体的地方请使用斜体。请检查修改完善。

6. 文章中所有引用的格式请严格按照美国心理学会(APA)出版手册第 7 版格式检查，例如 16 页中(Dunstan&Scott,2019)的错误请修改。

7. 本刊字数要求。研究报告与论文以及综述一般不超过 12,000 字(包括中英文摘要及参考文献)。请检查目前的字数，并按照本刊要求进行修改。

回应：感谢专家的宝贵意见。**关于斜体的使用：**我们已经仔细检查了文章中所有需要斜体的地方，包括统计符号、误差字母等，并已按要求使用了斜体格式。相关内容已进行修正，确保符合格式要求。**关于引用格式：**我们已经严格按照美国心理学会（APA）出版手册第 7 版的格式检查了所有引用，特别是在 16 页中的（Dunstan & Scott, 2019）等格式错误部分，已做了正确修改。所有的引用格式现在已统一并符合 APA 第 7 版的标准。**关于字数要求：**我们已经检查了文章的总字数，当前字数为 11,694 字，符合本刊的字数要求（不超过 12,000 字，包括中英文摘要及参考文献）。感谢您的耐心指导，以上修改已在原文中完成。

此外，我们仔细通读了原文，对细节之处进行了核对。

最后，非常感谢专家在审阅过程中给予的细致和建设性意见。您的宝贵反馈帮助我们进一步提升了文章的质量。我们已经根据您的建议，仔细修改了文章中的格式、引用、统计符号及字数等方面，确保符合期刊的要求。每一项修改我们都认真检查，并已在原文中进行了标红。再次感谢您的耐心指导和支持！

第二轮修改意见回复

给作者的意见：

1、3.1.1 反应时/正确率的方法分析是否使用表格来展示更简洁清晰？

回应：感谢专家的宝贵意见。关于 3.1.1 节使用表格的建议，我们进行了认真考虑，也参考了类似的相关文献，觉得文字和表格同时出现的话会出现数据重复表达的情况，所以我们还是决定不放表格了。再次感谢您的建议。

2、实验设计部分显示采用混合实验设计，在 3.1.1 的方法分析时采用 4*4 的重复测量方差分析，为何忽略了组间差异及两者交互作用的分析？

回应：感谢专家的宝贵建议。原文的 3.2 “结果”部分统计分析描述中，已经分别从主效应与交互作用进行了分析（用红色字体标出）。

3、叶节点所需要样本量一般采用 1、5、10，本论文中采用的是 1，是否存在过少的情况，为什么这么设置？

回应：感谢您提出的深刻问题。关于随机森林模型中叶节点最小样本数（`min_samples_leaf`）设置为 1 的原因，我们是基于以下三点考虑：

（1）本研究的主要目的是利用反应时和正确率行为指标，构建一个灵敏的预测模型。将 `min_samples_leaf` 设置为 1，允许决策树生长到最细，有助于模型捕捉数据中可能存在的一些细微真实的个体差异。如果将该值设置得过大（如 5 或 10），模型会过早地进行剪枝，可能导致一些有预测价值的细微信号被覆盖，从而损失模型的表现力。

（2）通过其他参数进行有效规避过拟合风险。虽然将叶节点样本数设为 1 有增加模型复杂度和过拟合的风险，但是随机森林的泛化能力依赖于“集体决策”组合效应。在本研究中，通过以下多重措施进行控制：首先，是限制树的最大深度(`max_depth = 10`)，这是最关键的正则化手段。即使叶节点样本数很小，树的生长深度也被严格限制在 10 层。这从根本上防止了单棵决策树无限制地生长到完全拟合噪声的程度，在模型复杂度和简单性之间取得了平衡。其次，集成学习本身抗过拟合，随机森林通过自举聚合集成了数百棵（本研究为 500 棵）树的结果，其“少数服从多数”或“平均”的机制，这使得它比单棵决策树稳定得多。最后是特征采样(`max_features = 'sqrt'`)，每次分裂时只考虑部分特征，增加了树的多样性。

(3) 基于网格搜索和交叉验证的客观最优解。最重要的是“叶节点样本数为 1”并非我们的主观预设,而是基于网格搜索(GridSearchCV)联合 5 折交叉验证(5-fold CV)的客观结果。我们的参数调优过程对比了包括关键优化参数中决策树数量(`n_estimators`: 300, 500, 800, 1000)、树的最大深度(`max_depth`: None, 10, 20)、节点分裂所需最小样本数(`min_samples_split`: 2, 5, 10)在内的多种参数组合。最终,在 70%的训练集上,采用 `min_samples_leaf=1` 并结合 `max_depth=10` 等设置的参数组合,在 5 折交叉验证中取得了最低的平均均方误差(MSE)。这表明该配置在训练集内部具有最佳的拟合与泛化平衡能力。

综上所述,该参数设置是为了达成研究目标,在充分考虑了过拟合风险,通过严谨的自动化调优流程筛选出的最优解。同时,我们使用了完全独立的测试集(占总样本 30%)进行最终评估,报告的 MAE、RMSE、 R^2 等指标均表明模型具有良好的泛化能力,未出现严重过拟合现象。再次感谢您的审阅,让我们有机会更清晰地阐述模型构建的考量。

参考文献:

- Wu, C.-C., Chen, Y.-L., Liu, Y.-H., & Yang, X.-Y. (2016). Decision tree induction with a constrained number of leaf nodes. *Applied Intelligence*, 45(3), 673–685. [10.1007/s10489-016-0785-z](https://doi.org/10.1007/s10489-016-0785-z)
- Zeng, Y., Hao, Z., Cai, R., Xie, F., Ou, L., & Huang, R. (2020). A causal discovery algorithm based on the prior selection of leaf nodes. *Neural Networks*, 124, 130–145. [10.1016/j.neunet.2019.12.020](https://doi.org/10.1016/j.neunet.2019.12.020)
- Probst, P., Wright, M., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3), e1301. [10.1002/widm.1301](https://doi.org/10.1002/widm.1301)
- Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295–316. [10.1016/j.neucom.2020.07.061](https://doi.org/10.1016/j.neucom.2020.07.061)
- Qi, X., Wang, S., Fang, C., Jia, J., Lin, L., & Yuan, T. (2025). Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biology*, 79, 103470. [10.1016/j.redox.2024.103470](https://doi.org/10.1016/j.redox.2024.103470)

第一轮修改意见回复

审稿人 1

给作者的意见:

1.增补纯抑郁组 (MDD-only)、纯焦虑组 (AD-only) 与 HC 组四组比较, 验证共病特异性;

回答: 感谢专家的意见, 我们已经补充了纯抑郁组 (MDD-only)、纯焦虑组 (AD-only) 的实验数据; 已验证共病的特异性。

2.补充超参数调优 (如 GridSearchCV 确定最优树数量/深度);

回答: 感谢专家的意见。为获得最佳模型性能, 本研究采用网格搜索 (GridSearchCV) 与 5 折交叉验证对随机森林回归模型进行超参数调优, 关键优化参数包括决策树数量 (n_estimators: 300, 500, 800, 1000)、树的最大深度 (max_depth: None, 10, 20)、节点分裂所需最小样本数 (min_samples_split: 2, 5, 10) 等。为验证参数优化结果的稳健性, 我们进行了多随机种子稳定性分析, 使用三个不同随机种子 (0,1,2) 重复整个训练流程。模型的最优参数具体见红色修改之处。这些超参数范围的选择基于随机森林模型的文献惯例和计算资源的平衡。例如, n_estimators 的搜索范围 (300-1000) 确保了模型有足够的复杂度捕捉模式, 同时避免过度计算消耗。max_depth=None 允许树生长到最大深度, 让数据本身决定最佳复杂度。为确保结果不是偶然, 本研究通过多随机种子稳定性验证了结果的可靠性。三个独立运行的特征重要性排名高度一致, 这表明超参数优化发现的模式是稳定和可重复的, 而非偶然产物。

参考文献:

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. [10.1016/b978-0-12-824271-1.00018-4](#)
- Probst, P., Wright, M., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3), e1301. [10.1002/widm.1301](#)
- Qi, X., Wang, S., Fang, C., Jia, J., Lin, L., & Yuan, T. (2025). Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biology*, 79, 103470. [10.1016/j.redox.2024.103470](#)

3.删除冗余的 RDoC 陈述, 聚焦 DMC 与行为指标的关联;

回答: 感谢您提出的宝贵意见。根据您的建议, 我们对引言部分的相关表述进行了修改, 使其聚焦 DMC 与行为指标的关联, 并删除了冗余的 RDoC 陈述。相关修改在引言部分用红色进行了标识。

审稿人 2

1. 研究立论基础较为薄弱。大量研究证据显示，抑郁症和焦虑障碍在行为学、神经病理学和生物学特征上存在差异和重叠。尤其是在认知模式上，无论是在回避或趋向动机，面孔识别、工作记忆特征和情绪调控策略上，两种疾病均表现差异或共享的脑特征。作者仅仅选择认知控制（主动和反应控制）这一单一角度，作为反映共病人群的特异认知模式关键指标，依据不足。作者需要补充：

（1）系统梳理既往研究中基于认知控制的研究结果，与基于其他认知特征的研究结果相比，有何优势和较强的证据指向，以至于作者在研究中只考虑认知控制这一指标。

回答：感谢专家的建议。我们在引言的第二段增加了既往研究中基于认知控制的相关研究，以进一步论证双重认知控制特征与传统认知控制特征相比所具有的优势。该部分内容用红色进行了标识。

（2）补充既往研究中在抑郁和焦虑障碍共病人群中在认知控制方面共性发现的相关研究证据，而不仅仅是在抑郁症或者焦虑障碍人群中的差异分析结果。跨诊断的思路是寻找不同疾病诊断人群中的共性指标，而非差异性指标；

回答：感谢专家的建议，我们已在文中相应的位置对“跨诊断”的表述进行了修改。

（3）机器学习建模选择随机森林方法的原因，对于线性和非线性特征，不同机器学习算法有着不同优势，作者仅采用单一方法建模，如果考虑结果的稳定性？是否考虑根据选择的特征，采取不同机器学习算法建模？

回答：感谢专家的疑问。我们对这一问题进行如下解释：本研究选择随机森林回归作为核心算法，是基于其特性与本研究数据特征和研究问题的吻合。首先，本研究中特征集包含两种行为指标：反应时（RT）和准确率（ACC），以往研究发现，这些特征与情绪状态间的关系是复杂且非线性的(Richter et al., 2024)。随机森林通过集成多棵决策树，擅长捕捉此类复杂的非线性关系和交互效应，这是线性模型（如线性回归、Ridge 回归）无法直接实现的。其次，本研究的核心目标不仅是预测，更重要的是理解与解释——即确定哪些认知行为特征对情绪预测的贡献最大。随机森林内置的特征重要性评估框架（SHAP 值）提供了模型可解释性的标准（Liu et al., 2025; Qi et al., 2025）。此外，尽管是单一模型，但其在独立测试集上取得了 $R^2 = 0.78$ 的优异表现，表明该模型已经能够稳定地捕捉数据中的大部分关键模式，能够满足本研究的需求。

关于专家对结果稳定性的顾虑，我们通过一个多层次、系统性的稳定性验证框架来确保研究发现的可重复性。内部稳定性即算法层面，本研究通过固定随机种子（`random_state=42`）确保了实验的完全可重复性（Williams et al., 2025）。其他研究者使用的代码和种子能够复现完全相同的模型和结果。外部稳定性即数据层面，我们采用了严格的 70/30 留出法，模型性能（MAE, R^2 ）全部基于从未参与训练的 30% 独立测试集进行评估，这确保了性能指标是可靠的，有效证明了模型的泛化能力；统计稳定性即结果层面。我们进行了多随机种子稳定性分析（`seeds=range(3)`）（Owen et al., 2024）。尽管改变了三次训练的初始条件，例如综合诊断模型的特征重要性排名显示出高度的一致性（平均斯皮尔曼 $\rho = 0.89$ ）。这也表明，本研究的核心发现（RT-BX、RT-AY 特征最重要）是数据驱动的稳健模式，而非某次特定随机运行的偶然结果。

参考文献：

- Richter, T., Stahi, S., Mirovsky, G., Hel-Or, H., & Okon-Singer, H. (2024). Disorder-specific versus transdiagnostic cognitive mechanisms in anxiety and depression: Machine-learning-based prediction of symptom severity. *Journal of Affective Disorders*, 354, 473–482. [10.1016/j.jad.2024.03.035](https://doi.org/10.1016/j.jad.2024.03.035)
- Qi, X., Wang, S., Fang, C., Jia, J., Lin, L., & Yuan, T. (2025). Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biology*, 79, 103470. [10.1016/j.redox.2024.103470](https://doi.org/10.1016/j.redox.2024.103470)
- Liu, Y.-C., Liu, Y.-H., Pan, H.-F., & Wang, W. (2025). Unveiling new insights into migraine risk stratification using machine learning models of adjustable risk factors. *The Journal of Headache and Pain*, 26(1), 103. [10.1186/s10194-025-02049-5](https://doi.org/10.1186/s10194-025-02049-5)
- Williams, N. T., Hung, A., & Rudolph, K. E. (2025). Re: don't let your analysis go to seed: on the impact of random seed on machine learning-based causal inference. *Epidemiology*, 36(4), e12–e13. [10.1097/EDE.0000000000001860](https://doi.org/10.1097/EDE.0000000000001860)
- Owen, C. A., Dick, G., & Whigham, P. A. (2024). Using decomposed error for reproducing implicit understanding of algorithms. *Evolutionary Computation*, 32(1), 49–68. [10.1162/evco_a_00321](https://doi.org/10.1162/evco_a_00321)

（4）补充研究中采用的 AX-CPT 范式，相对于其他评估认知控制范式的优势所在。

回答：感谢您提出的宝贵意见。这部分内容已在引言的第三段进行了补充，并用红色字体进行了标识。

2. 在研究方法上:

(1) 样本选择上, 缺失符合抑郁症或焦虑障碍单一诊断、非共病的临床样本。因此, 与健康对照组的 AX-CPT 行为指标的差异分析结果, 不能说明是共病人群的特异性认知模式。

回答: 感谢专家建议, 我们补充了抑郁症和焦虑障碍单一诊断的被试, 以更好的说明相关指标分析是共病人群的特异性认知模式。

(2) 研究不适合使用“跨诊断”这一术语, 通常, 在跨诊断研究中, 纳入的临床样本诊断应该在三种或以上, 该研究样本不符合, 需要补充样本。

回答: 感谢专家建议, 我们已经补充了相关样本。目前临床样本有三组。

(3) 两组样本均混杂了未成年人群(<18岁), 尽管两组年龄最终没有显著差异。但未成年和成年病人的混合样本是不恰当的。对于临床研究, 以认知范式的行为学指标, 脑指标和生物学指标, 年龄会成为重要的混淆变量, 无法简单通过统计学方法进行处理。建议样本选择单一的成年样本。

回答: 感谢专家建议。我们删除了未成年人样本, 并进一步补充了成年样本, 以避免年龄成为混淆变量。

(4) 疾病的入排标准需要补充, 包括年龄、病程和用药情况等可能影响研究结果的变量。

回答: 感谢专家提出的宝贵建议。我们对年龄、病程和用药情况等方法部分进行了补充, 用红色进行了标识。我们对被试选择进行如下解释: 本研究纳入的被试均为门诊初次就诊、尚未接受任何系统性精神科药物治疗的患者。这一选择能够有效控制药物治疗这一混淆因素, 使本研究能更清晰地观察到疾病本身与认知特征之间的关联。选择“门诊初诊未用药患者”群体, 也是基于其对研究前期筛选的有利条件, 能最大程度地保证患者症状处于相对纯粹的自然状态, 避免长期病程复杂性和精神药物对认知功能的潜在影响, 从而更有利于揭示疾病与认知功能之间的内在联系。同时, 我们也在论文的【讨论】最后一段部分坦诚说明了由此可能带来的局限性, 并用红色进行了标识。

(5) 共病入组标准的“SDS 得分 > 52”有误, 根据划界分, 共病组 SDS 应该是 ≥ 53 , 健康对照组应该是 $SDS \leq 52$; 同样的问题也出现在 SAS 上, 共病组应该是 ≥ 60 。

回答: 感谢专家指出的这一错误, 我们已在论文中进行了修改。

(6) 机器学习算法的优势在于处理高维、线性和非线性特征的复杂数据。该研究所选择的特征只有 8 个。变量少, 采用传统回归分析即可, 使用机器学习算法的必要性和合理性不足。需要补充特征集, 可包含人口学特征、临床量表数据、与认知控制相关的行为学特征和神经生理特征。

回答: 感谢专家的建议。我们采用机器学习方法, 核心目的在于其能够更有效地处理小样本中特征与目标变量间的复杂非线性关系及交互效应, 而这正是传统线性回归的局限所在。本研究选用的 8 个行为学特征 (如 RT-AY, RT-BX 等) 理论上存在非线性关系, 随机森林算法无需预设公式, 可自动捕捉此类复杂模式, 避免主观构建多项式项带来的过拟合风险 (Kessler et al., 2016; Richter et al., 2021)。同时, 机器学习提供了更可靠的特征重要性评估 (如基于 SHAP 值), 能有效处理行为指标间的共线性, 客观量化各认知特征对症状的贡献, 这是传统回归系数难以实现的。此外, 我们严格遵循机器学习范式 (70/30 划分、交叉验证、独立测试集评估), 确保了模型具有良好的泛化能力与可靠性, 为后续纳入更多维度特征 (如生理、临床数据) 建立了可扩展的分析基础 (Patel et al., 2025; Salehinejad et al., 2023)。我们在论文的【方法】-【研究工具】中对这一问题进行了表述, 用红色进行了标识。

参考文献:

Kessler, R. C., Van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., ... Zaslavsky, A. M. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry*, 21(10), 1366–1371. [10.1038/mp.2015.198](https://doi.org/10.1038/mp.2015.198)

Richter, T., Fishbain, B., Fruchter, E., Richter-Levin, G., & Okon-Singer, H. (2021). Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research*, 141, 199–205. [10.1016/j.jpsychires.2021.06.044](https://doi.org/10.1016/j.jpsychires.2021.06.044)

Patel, S., Deng, J., Fung, M., Wang, B., Rubin, B., Anderson, T., & Subasri, V. (2025). Development and nested cross-validation of explainable machine-learning models for identifying obstructive coronary artery disease in patients presenting with angina. *Journal of the American College of Cardiology*, 85(12), 1934. [10.1016/S0735-1097\(25\)02418-0](https://doi.org/10.1016/S0735-1097(25)02418-0)

Salehinejad, H., Meehan, A. M., Rahman, P. A., Core, M. A., Borah, B. J., & Caraballo, P. J. (2023). Novel machine learning model to improve performance of an early warning system in hospitalized patients: A retrospective multisite cross-validation study. *Eclinicalmedicine*, 66, 102312. [10.1016/j.eclinm.2023.102312](https://doi.org/10.1016/j.eclinm.2023.102312)

(7) 机器学习建模的步骤需要完善, 尤其是三个模型建构过程中, 纳入特征集是什么, 结果变量是什么需要详细阐述。

回答: 感谢专家的建议。我们已在论文的【方法】-【机器学习建模】部分对建模步骤进行了详细补充, 并用红色进行了标识。具体内容为: 特征集方面, 所有模型均采用相同的 8 个 AX-CPT 行为学指标, 包括: 反应时(RT): AX、AY、BX、BY 试次的平均反应时(RT-AX, RT-AY, RT-BX, RT-BY) 正确率(ACC): AX、AY、BX、BY 试次的正确率(ACC-AX, ACC-AY, ACC-BX, ACC-BY)。综合诊断模型方面, 以综合情绪困扰得分(SDS 与 SAS 标准化总分)作为连续型结果变量。SDS 模型以抑郁症状严重程度(SDS 标准分)作为结果变量。SAS 模型以焦虑症状严重程度(SAS 标准分)作为结果变量。具体建模流程为: 数据集按 70/30 比例划分为训练集与独立测试集。在训练集上采用 5 折交叉验证的网格搜索(GridSearchCV)进行超参数调优, 以均方误差(MSE)为优化目标。使用最优超参数在全部训练集上重新训练最终模型。最终模型性能在未见过的 30%测试集上进行严格评估, 报告均方误差(MSE)、平均绝对误差(MAE)、决定系数(R^2)等指标。

(8) 数据表格不规范, 需要完善, 呈现均值和标准差的表格中未显示 F 值或事后检验的结果、以及 P 值。

回答: 感谢专家的建议, 表格中呈现的是四组被试在各实验条件下的平均反应时和错误率, 相关的统计检验我们以文字的形式进行描述。

(9) 从实验设置的 trial 数来看, 除了 AX 以外, 其余三个条件中每个条件的试次叠加不到 30 个试次, 结果是否可靠? 尤其是研究仅仅采用反应时和正确率的行为指标, 缺乏其他神经生理指标的情况下。

回答: 感谢专家的质疑。AX-CPT 范式是研究双重认知控制的标准范式(Braver, 2012)。该范式的试次比例设置是其精髓所在, 也是一种固定设置; 是该范式能够有效分离主动性控制和反应性控制的关键设计。该范式中, 被试会看到一系列连续的字母线索和探测刺激。基本规则是: 目标试次中, 如果线索是字母 A, 并且探测刺激是字母 X, 则要求做出目标反应(例如, 按“是”键)。非目标试次中, 所有其他组合(AY, BX, BY), 都要求做出非目标反应(例如, 按“否”键)。其中, 字母 B 代表除 A 以外的任何其他字母。在一个典型的 AX-CPT 实验中, 试次比例通常如下: AX 试次: 64% (高频, 目标试次); AY 试次: 12% (低频, 非目标试次); BX 试次: 12% (低频, 非目标试次); BY 试次: 12% (低频, 非目

标试次)。这种看似“极度不平衡”的比例设置是经过精心设计的, 目的就是为了在被试身上诱发一种强烈的认知控制策略偏好, 从而使研究者能够观测和测量它。

该范式中这样设置的原因是: 一是, 可以建立强大的“A-X”行为模式与预期。将 AX 试次设置为高达 64%, 是为了让“看到 A 之后大概率会看到 X, 然后就要做目标反应”这一规则成为整个任务中最突出、最主导的优势反应。被试会学会将线索“A”与“准备做出目标反应”紧密联系起来。一旦看到“A”, 被试就会主动维持这个目标信息, 并预先准备好按下“是”键。这使得在 AX 试次上, 反应既快速又准确。二是, 创造关键的“认知冲突”条件。高频的 AX 试次为两个关键的冲突试次 (AY 和 BX) 创设了背景。这两种试次是探测不同认知控制模式的“探测刺激”。其中, AY 试次: 探测对主动性控制的抑制 (反应性控制的调用) 冲突来源: 看到线索“A”后, 被试已经主动准备好对即将出现的 X 做出“是”的反应。然而, 出现的却是“Y”。此时, 个体必须抑制由“A”引发的强烈的准备性冲动 (抑制优势反应), 并反应性地调用控制资源, 将反应改为“否”。若被试在 AY 试次上犯错 (错误地按了“是”), 表明个体过度依赖主动性控制, 无法在冲突发生后有效地进行抑制和调整 (反应性控制不足)。BX 试次中: 探测主动性控制的维持能力冲突来源。线索是“B” (非 A), 它不会引发任何准备反应。当探针“X”出现时, 由于没有前期的主动性准备, 个体需要完全依靠探针“X”本身来反应性地判断该做什么。此时, 如果个体的主动性控制很强, 他们会在整个任务期间始终主动地维持着“只有 A-X 才按‘是’”这个高级规则。因此, 即使看到 X, 但因为前面的线索不是 A, 他们会正确地说“否”。若被试在 BX 试次上犯错 (错误地按了“是”), 表明个体的主动性控制很弱。他们无法持续维持任务规则, 因此被最常见的刺激“X”所捕获, 做出了冲动性的错误反应。BY 试次是一种基线试次。BY 试次没有冲突, 线索是 B, 探测刺激是 Y, 两者都提示“否”反应。它通常表现很好, 用作一个基线测量, 确保被试理解基本规则。

参考文献:

[徐雷, 唐丹丹, 陈安涛. (2012). 主动性和反应性认知控制的权衡机制及影响因素. *心理科学进展*, 20(7), 1012-1022.]

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106-113. [10.1016/j.tics.2011.12.010](https://doi.org/10.1016/j.tics.2011.12.010)

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, 106(18), 7351-7356. [10.1073/pnas.0808187106](https://doi.org/10.1073/pnas.0808187106)

Braver, T. S., Kizhner, A., Tang, R., Freund, M. C., & Etzel, J. A. (2021). The dual mechanisms of cognitive control project. *Journal of Cognitive Neuroscience*, 1–26. [10.1162/jocn-a-01768](https://doi.org/10.1162/jocn-a-01768)

Lu, G., Leng, T., Chen, L.-J., Yu, M., Guan, T., Liu, Y., ... Wang, K.-P. (2025). Aging effects on control ability: An ERP study using the AX continuous performance task in middle-aged and older adults. *Current Psychology*. Advance online publication. [10.1007/s12144-025-07338-5](https://doi.org/10.1007/s12144-025-07338-5)

Masuyama, A., & Mochizuki, S. (2020). Induced sad mood affects context processing in cognitive control in mildly depressive undergraduates. *Current Psychology*, 39(4), 1476–1484. [10.1007/s12144-018-9854-2](https://doi.org/10.1007/s12144-018-9854-2)

Xu, M., Li, Z., Qi, S., Fan, L., Zhou, X., & Yang, D. (2020). Social exclusion modulates dual mechanisms of cognitive control: Evidence from ERPs. *Human Brain Mapping*, 41(10), 2669–2685. [10.1002/hbm.24970](https://doi.org/10.1002/hbm.24970)

3. 在机器学习建模上，主要问题如下：

（1）数据未分为训练集、测试集和验证集，即综合来看，所有数据各自在不同折交叉验证中均参与训练，模型生态效度未知。

回答：感谢专家的建议。本研究遵循了机器学习范式。在数据划分方面，通过 `train_test_split` 保留 30% 数据作为独立测试集，这些数据从未参与任何训练或调优。在 70% 的训练集内部，使用 5 折交叉验证（GridSearch CV）选择最优超参数，确保调优过程的客观性。最终评估所有报告的性能指标（如 R^2 、MAE）均基于独立测试集计算，直接反映模型在全新数据上的泛化能力。这种“70/30 留出法+训练集内交叉验证”的策略，在有限数据量下平衡了模型训练需求与评估严谨性。本研究的测试集指标（ $R^2=0.77$ ）证明，模型能够有效泛化到未参与训练的数据，生态效度良好。

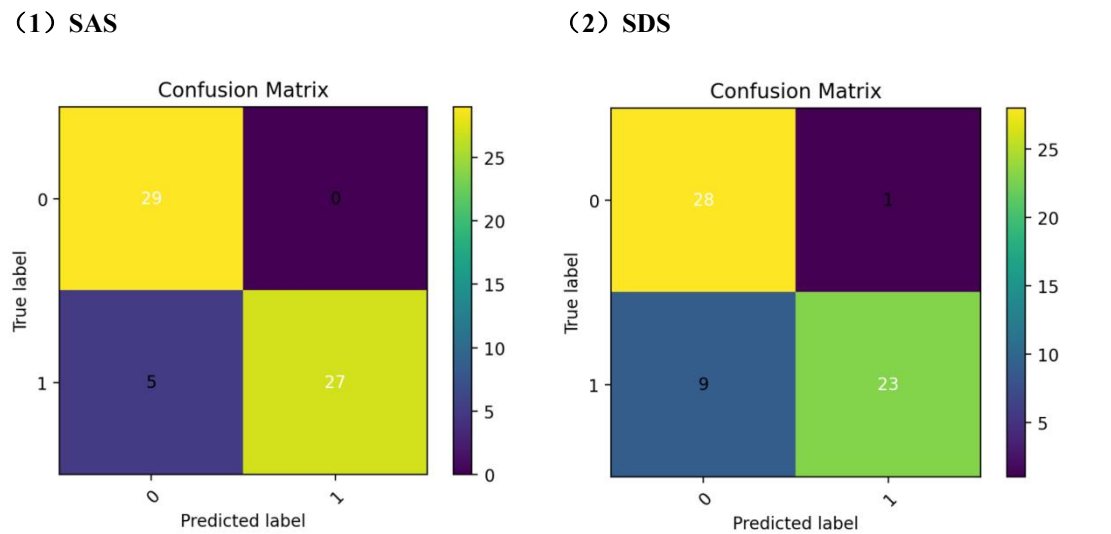
（2）将标准化后 SDS 得分与 SAS 得分平均值作为跨诊断得分，其尺度（0~100）与 SDS、SAS 分别的得分是否为同一尺度？如果不是，模型评估指标（RMSE 等）无法直接比较

回答：感谢您提出的深刻问题。SDS 得分与 SAS 得分与平均值作为综合诊断得分的尺度都为（0~100）。SDS 和 SAS 拥有不同的临床划界分，这源于它们发展时各自独立的诊断标准。然而，现代精神病理学越来越倾向于维度观，即抑郁和焦虑等症状并非独立的“全或无”的类别，而是一个连续的、共存的谱系。本研究的综合诊断得分正是基于这一理念，通过标准化后取平均值的方法，将两个量表的信息融合，试图创建一个单一的、连续的情绪困

扰严重程度指数。这个 0-100 的分数不再直接对应传统的 SDS 或 SAS 分级，而是代表了从“健康”（接近 0）到“严重困扰”（接近 100）的连续变化过程。这一转换虽然使其失去了与原临床分级的直接对应，但获得了作为机器学习回归模型的理想因变量。

（3）SDS、SAS 存在划界分，得分在划界分前后意义不同。以 SDS 为例，以得分可以划分为正常、轻度抑郁、中度至重度抑郁和严重抑郁 4 种程度，每个程度的得分分差为 9 分，以连续变量构建模型，表 3 显示模型评估存在约 10 分的误差，表明模型性能较低，由此得到的变量重要性结果可靠性存疑。

回答：感谢您从临床转化角度提出的深刻意见。本研究的核心目标是构建一个从行为特征中量化评估情绪困扰连续风险的预测模型，而非直接用于临床诊断分类。该模型的价值在于揭示认知行为特征与症状严重度之间的定量关联趋势，并为后续研究提供一个可扩展的基线模型。我们已遵循您的建议，补充了随机森林分类任务的分析：本研究以标准划界分（SAS $\geq$ 50；SDS $\geq$ 53）将被试划分为“健康”与“有症状”两组，使用相同的 AX-CPT 行为特征进行建模。分类模型取得了良好的性能：在区分是否焦虑的任务上，准确率达到 83%，AUC 值为 0.88；在区分是否抑郁的任务上，准确率达到 91%，AUC 值为 0.97（见图：分类模型性能评估混淆矩阵），再次证明本研究训练的数据集的有效性，尽管回归模型的绝对误差存在，但其揭示的核心认知特征具有稳定性，为理解共病机制提供了计算行为学基础。我们已在论文的【结果】、【讨论】部分增加了对模型误差临床意义的探讨，强调了当前模型更适用于趋势预测，未来的优化方向是开发更低误差、更高临床实用性的辅助诊断工具。



分类模型性能评估混淆矩阵

(4) 随机种子数设置不同，得到的模型评估指标和特征重要性排序可能存在较大变动，本文未报告如何解决。

回答：感谢您的宝贵意见。为消除这一潜在影响，本研究在模型构建与评估中采取了以下措施：**一是固定随机种子**：在模型训练、超参数调优以及数据划分（70/30 分割）过程中，本研究均已固定随机种子（`random_state=42`）。以确保研究的可重复性，包括最终的模型性能指标（MAE、RMSE、 R^2 ）和特征重要性排序。**二是采用稳健的评估方法**：本研究性能评估的核心依据是在一个固定划分的独立测试集（30%数据）上进行的。该测试集在研究初期即被随机划分并固定下来，所有模型的训练与调优均未使用该部分数据，从而保证了性能评估结果的客观性。通过上述方法，本研究最大限度地降低了随机性对结果的干扰，确保了模型性能与发现的可重复性。我们已在论文的【方法】部分（机器学习建模小节）补充说明了固定随机种子的具体做法。并用红色进行了标识。

4.论文中存在一些术语的常识性错误，需要逐一修改。例如：

(1) 现有研究多聚焦单一症状群体”“单一症状患者”表述错误。DSM 诊断标准中，所有疾病诊断的症状标准，都要符合多个症状，尤其是抑郁症和焦虑障碍。所以，只有符合单一诊断的患者，没有单一症状患者。

(2) “此外，(Chen et al., 2024) 发现。” 文献标注格式错误。

(3) “抑郁相关的神经资源耗竭可能削弱主动性控制”。什么是“神经资源耗竭”？

(4) “共病并非症状的简单叠加，而是通过 CCN 的交互失衡形成独特病理表现”。这一结论有待商榷。从大尺度脑网络的研究证据来看，不能得出聚焦于“CNN”的结论。

(5) “由于抑郁与焦虑障碍在共病率、躯体-情绪症状表现、诊断工具及治疗策略上均存在显著重叠 (Xie et al., 2022; Zhang et al., 2024)”，这句话不完整，语焉不详，而且“共病率”如何重叠？

回答：衷心感谢您对论文细节提出的宝贵意见。笔者已根据您的意见逐一进行了修改与完善：

(1) 关于“单一症状患者”表述错误：我们已对全文进行了检查，将所有相关表述（如“单一症状患者”、“单一症状群体”）更正为“单一诊断患者”或“符合单一诊断标准的患者群体”（例如：MDD-only, AD-only），以准确反映其临床诊断属性；

(2) 关于文献标注格式错误：感谢您指出的笔误。我们已对全文的参考文献格式进行了统一核查与修正，确保所有文献引用均符合标准格式；

(3)关于“神经资源耗竭”的表述,我们已对该结论进行了修正,不再将机制局限于“CCN”,删除冗余不妥表达。

(5)关于语句不完整与“共病率重叠”:我们对此深表歉意。该句旨在说明抑郁与焦虑障碍在多个层面存在高度关联性,但表述确实不当。我们已在论文的【引言】部分第四段进行了修改,使语义完整且准确。

最后,再次感谢您的细致审阅,这次审阅极大地提升了本研究的科学质量。我们已将所有修改完整地体现在论文中,再次为您付出的宝贵时间和精力致以诚挚的谢意!

附录

焦虑自评量表 (SAS)

感谢您的参与！本次为匿名填写，在开始评定前，请保持自己的状态，不受任何人的影响进行自我评定。本问卷包含 20 个项目，请您仔细阅读以下内容，根据最近一星期的情况如实填写，您的回答无对错与好坏之分，请填写最符合您情况的选项。

题目	没有或很少时间	小部分时间	相当多时间	绝大部分或全部时间
1 我觉得比平常容易紧张和着急				
2 我无缘无故地感到害怕				
3 我容易心里烦乱或觉得惊恐				
4 我觉得我可能将要发疯				
5 我觉得一切都很好，也不会发生什么不幸				
6 我手脚发抖打颤				
7 我因为头疼、头颈痛和背痛而苦恼				
8 我感到容易衰弱和疲乏				
9 我觉得心平气和，并且容易安静坐着				
10 我觉得心跳得很快				
11 我因为一阵阵头晕而苦恼				
12 我有晕倒发作或觉得要晕倒似的				
13 我呼气、吸气都感到很容易				
14 我手脚麻木和刺痛				
15 我因为胃痛和消化不良而苦恼				
16 我常常要小便				
17 我的手脚常常是干燥温暖的				
18 我脸红发热				
19 我容易入睡，并且一夜睡得很好				
20 我做恶梦				

抑郁自评量表（SDS）

感谢您的参与！本次为匿名填写，在开始评定前，请保持自己的状态，不受任何人的影响进行自我评定。本问卷包含 20 个项目，请您仔细阅读以下内容，根据最近一星期的情况如实填写，您的回答无对错与好坏之分，请填写最符合您情况的选项。

题目	偶尔	有时	经常	持续
1 我觉得闷闷不乐，情绪低沉				
2 我觉得一天之中早晨最好				
3 我一阵阵哭出来或想哭				
4 我晚上睡眠不好				
5 我吃得跟平常一样多				
6 我与异性密切接触时和以往一样感到愉快				
7 我发觉我的体重在下降				
8 我有便秘的苦恼				
9 我心跳比平时快				
10 我无缘无故地感到疲乏				
11 我的头脑跟平常一样清楚				
12 我觉得经常做的事情并没困难				
13 我觉得不安而平静不下来				
14 我对将来抱有希望				
15 我比平常容易生气激动				
16 我觉得作出决定是容易的				
17 我觉得自己是个有用的人，有人需要我				
18 我的生活过得很有意思				
19 我认为如果我死了别人会生活得更好些				
20 平常感兴趣的事我仍然照样感兴趣				