

# 《心理科学进展》审稿意见与作者回应

题目：机器学习在发展性阅读障碍儿童早期筛查中的应用

作者：卜晓鸥，王耀，杜亚雯，王沛

---

## 第一轮

### 审稿人 1 意见：

作者较好地总结了机器学习在儿童发展性阅读障碍研究中的应用，结合常用的机器学习和标准化心理教育量表、行为、脑影像等数据检测识别阅读障碍组和对照组或预测高风险组等领域的应用研究。这为发展性阅读障碍儿童早期识别和筛查具有很好地研究和应用价值。尚有以下问题需要作者给予补充或修改：

**意见 1：**机器学习尤其注重特征工程，首先在 3.2 特征提取章节，应补充标准化心理教育量表、行为、影像等技术提取了哪些具体特征并加以描述，建议以表格形式总结。其次，文章未能很好地总结并传递给读者哪些特征因子是可以很好地识别阅读障碍，应在 4.1 章节中给予总结补充。

**回应：**感谢评审老师的宝贵意见。根据所提建议，我们补充了标准化心理教育测试、眼动追踪、网络/手机游戏、脑成像技术、其他这五种类型的测试提取的具体特征，并以表格形式总结，修改部分在文中已用红色字体标出，具体修改内容见表 2。此外，我们根据纳入本篇系统综述的文献，总结了这些文章中具体表述的最具预测性的特征，修改内容见表 3。

**意见 2：**第 3.3 机器学习模型中，人工神经网络（ANN）不属于深度学习范畴。请注意机器学习和深度学习模型算法区别与归属。

**回应：**非常感谢老师的纠正和提醒。在新修正的文本中，我们添加了“ANN 是深度学习的基础，它模拟了大脑神经网络结构和功能...”以及“随着神经网络的发展，深度学习从浅层的 ANN 中衍生出来...”等相关内容，明确了算法的区别和归宿，并体现了递进发展的趋势。

**意见 3：**4.1 章节中“机器学习作为一种新的统计学方法”、以及文中多处提及“鲁棒性”等表述存在不严谨，不规范。机器学习与统计学存在诸多区别，不能简单地把机器学习归属为统计学方法；按照规范的机器学习术语“鲁棒性”应为“稳健性”或“健壮性”。

**回应：**感谢评审老师的指正。为了避免歧义，我们删除了“机器学习作为一种新的统计学方法”，并且将“鲁棒性”改为“稳健性”。

**意见 4：**表 1 的 25 篇文献罗列有哪些内在的逻辑顺序？当前看起来仅是简单的罗列，比较混乱。主要的指标罗列的比较笼统，建议这一列应稍微详细些。

**回应：**感谢评审老师的提问和提醒。在原有的版本中，文献罗列是根据在文文献在文中出现的先后次序来排列。在新修改的表格中，我们按照作者的姓名首字母顺序排列，并且对数据指标进行了补充。具体修改内容见表 1。

**意见 5：**作者在结论中提及“采集多个来源数据(如量表、行为、脑成像等)、混合多种模型以及考虑多模态的深度学习框架提高机器学习方法的预测力”，但该文纳入分析的 25 篇文献未有多元异构数据联合作为识别或预测 DD 人群。这应为作者的一般预测和展望，当前研究尚不能直接支持提升预测能力。

**回应：**感谢评审老师的宝贵建议。我们将本句话修改为“未来的研究需要采集多个来源数据(如量表、行为、脑成像等)、混合多种模型以及考虑多模态的深度学习框架提高机器学习的预测力”。由于先前我们对表 1 的概括并不够详细和精准，所以在表格的呈现中并没有让您直观地看到多模态数据，抱歉给您带来困惑。我们重新对表格进行了修改，在纳入的文献中有一项多模态数据：Skeide 等人(2016)的特征选择为基因和 MRI (白/灰质形态学特征)。基于前人的研究和我们对于未来中文阅读障碍的研究计划，我们非常想要构建多模态数据以多方位地精准识别阅读障碍儿童，因此将这一点体现在了我们对于未来的展望中。

---

**审稿人 2 意见：**

**意见 1：**题目和全文中都把“筛查”放到了“识别”的后面，我们一般是先对所有群体进行“筛查”，有个便捷的筛查系统，才能做到“识别”特殊群体，因此建议作者应当进行全文的调整。此外，文章中虽然对筛查和识别做了介绍，但是没有很好地将两者进行区分，建议文章对于两者给出明确的区分。

**回应：**感谢评审老师指出的这一重要问题。针对您提出的问题，我们重新确认了筛查相近含义的名词的概念，并进行了全文的调整。首先，我们遵循“筛查后才能识别”这一前提。其次，将题目中的“识别和筛查”换成“筛查”。最后，对文中引用的文献进行了追本溯源，在尊重原

文的基础上(确定文中是哪一种,如 identification/ screening/ detection/diagnosis)进行了相应的修改。具体修改情况如下:

(1)我们搜索出 2018-2023 年有关阅读障碍的 36 篇中文文章(心理学报/心理科学进展/心理科学/心理发展与教育/心理与行为研究/中国心理卫生杂志/中国临床心理学杂志),总结出中文关键字为“筛查/筛选/诊断/识别”。在此基础上我们进一步发现:①文章中对这些关键词存在相互混用、没有明确的概念界定的情况。②对 36 篇使用“筛查/筛选/诊断/识别”的中文文章进行了频次统计,发现“筛查”出现 26 次,“诊断”出现 18 次,“识别”出现 5 次。如,同一篇文章中同时有筛查和诊断,则各记 1 次。

(2)对我们纳入的 25 篇英文文章进行了总结,关键词为: Identification/ Screening/ Detection/ Recognition/ Prediction/ Diagnosis。在此基础上我们发现同样存在相互混用、没有概念界定的情况,80%的文章混用 3 种及以上概念。例如,“Diagnosis”出现 23 次,“Identification”出现 17 次,“Screening”出现 14 次,“Detection”出现 10 次,“Prediction”出现 3 次,“Recognition”出现 2 次。综合中文文章和英文文章,发现**筛查、诊断**这 2 个词使用最多。

(3)根据《特殊教育辞典》的概念界定(朴永馨,2014),筛查(screening)是在大量儿童中发现特殊儿童的一种检测活动,也就是发现高风险儿童的一种方式。一般而言,筛查的下一步是诊断(diagnosis),即对筛查出的高风险儿童进一步做全面细致的专业检查。正规诊断是由受过训练取得合格证书的专业人员进行,使用的测查工具更复杂,操作程序更严格和标准化,并附有书面结论(陈光华等,2022)。目前国内对于阅读障碍尚未有一个确切的诊断标准。根据机器学习分类阅读障碍儿童的特点,即从大量儿童中发现阅读障碍儿童,我们认为“筛查”一词可能更加合适,能体现出机器学习的特点。

综上,我们将题名改为《机器学习在发展性阅读障碍儿童早期筛查中的应用》。在列举的文献中,根据文章原本的描述(首选题目、摘要、结论)来选择使用哪一个词。修改部分在文中已用红色字体标出。

朴永馨.(2014). *特殊教育辞典* (pp. 3-4). 华夏出版社.

陈光华,陶冠澎,翟璐煜,白学军.(2022). 自闭症谱系障碍的早期筛查工具. *心理科学进展*, 30(04), 738-763.

**意见 2:** 这项研究纳入的是 2016-2022 年的文献,选择整个时间范围的标准是什么,通常这种类型的综述会纳入近 10 年左右的文章,或者是从本领域的第一篇文章开始纳入。需要在文章中交代清楚。

**回应:** 非常感谢评审老师的提醒。我们选择纳入从 2016 年开始的文献一方面是因为机器学习的技术迭代快, 另一方面是因为有关中文阅读障碍的机器学习的第一篇文章发表于 2016 年。虽然在纳入的 25 篇文献中, 对于中文阅读障碍的机器学习研究仅有 2 篇, 但本文的非常重要的目的是廓清阅读障碍的机器学习研究可能的发展路径与发展思路, 为将团队后续的中文阅读障碍的机器学习研究奠定基础。因此, 我们认为 2016 年的这篇中文阅读障碍的机器学习文章是具有里程碑意义的。从引用次数来看, 这篇文章获得了 110 次的引用, 也可见其重要性。综上, 我们对文章进行了修改, 添加了选择 2016~2022 年这一时间范围的原因: “考虑到机器学习技术的飞速发展和迭代, 并且第一篇中文阅读障碍的机器学习研究发表于 2016 年, 因此文献检索的日期范围设定为 2016 年 1 月 1 日~ 2022 年 10 月 1 日。”

**意见 3:** 摘要: 第一句多了一个“的”。摘要第二句不是很合适, “机器学习因其性能和原理方面的优势...”, 机器学习是一种机器算法, 可以有性能的优势, 但“原理”是无法比较的。

**回应:** 感谢评审老师的指正。第一句多出的“的”已经删除, 第二句修改为: “机器学习因其强大的数据处理和挖掘能力逐渐被应用到阅读障碍儿童的早期筛查中”。

**意见 4:** 一些语句的通顺度和主语与谓语等需要注意。如“目前, 儿童通常在 2 年级或更高年级在掌握阅读技能的过程中才有可能被识别出阅读障碍(Sanfilippo et al., 2020)。”建议修改为“目前, 儿童通常在 2 年级或更高年级在掌握阅读技能的过程中才有可能被识别出存在阅读障碍(Sanfilippo et al., 2020)。”“在经济发展落后的国家, 识别出贫困儿童阅读障碍的年龄更晚(Ballester et al., 2021)。”建议修改为“在经济发展落后的国家, 贫困儿童被识别出存在阅读障碍的年龄更晚(Ballester et al., 2021)。”“近年来, 因能够提供更高的检测精度和更好的预测结果,”修改为“近年来, 因其能够提供更高的检测精度和更好的预测结果, ”。“为此, 通过整合机器学习在阅读障碍的识别与筛查的最新进展”, 建议修改为“为此, 本研究通过整合机器学习在阅读障碍的识别与筛查的最新进展”。“本文的第一作者和第二作者依据上述标准进行独立筛查, 最后确定纳入本次系统综述的文献数量为 25 篇(表 1)。”建议删除““本文的第一作者和第二”。“眼动行为已经成为基于机器学习进行阅读障碍分类的常用指标”, 改为“眼动特征”,

**回应:** 感谢评审老师宝贵的修正建议。我们已经对您提及的语句通顺度问题进行了修改并进行了全文检查, 以避免上述错误。

**意见 5:** “然而，由于阅读障碍儿童的症状具有巨大的个体差异性”，这句话建议再补充下，比如阅读障碍儿童分不同的类型，这样能够展现这个群体的“巨大的个体差异性”，现在仅仅一句话，且给的结论比较大胆，不能够向读者展现这种差异性。

**回应:** 感谢评审老师的建议。我们对这一段语言表述做了如下修改：“然而，阅读障碍儿童的症状具有巨大的个体差异性，比如不同的病源因素会导致不同的阅读障碍亚类型(Aaron et al., 1999)。传统的阅读障碍检测技术低效耗时，敏感性和特异性指标不明确，难以满足大规模并快速筛查阅读障碍儿童的需求(Usman et al., 2021)”。

**意见 6:** 表 1 中，第 4 条文献漏掉了“岁”，此外第 4 条文献和第 17 条文献年龄的表达是 av 平均年龄，为什么会有两个平均年龄。第 23 条文献中 227 是什么类型的被试。

**回应:** 感谢评审老师的仔细审查。首先，我们对漏掉的“岁”进行了补充。其次，关于两个 av 平均年龄，抱歉给老师带来困惑。两个 av 对应的是样本量中的两个群体的分别的平均年龄，比如原第 4 条(现第 9 条)文献中的“av15.52 岁;av14.78 岁”，代表的是 DA 群体的平均年龄是 15.52 岁，non 群体的平均年龄是 14.78 岁。在原来的版本中，我们用“av15.52 岁, av14.78 岁”来表示两个群体的平均年龄，并不够直观。因此在新修正的版本中，我们用“;”代替“,”，改为“av15.52 岁;av14.78 岁”。最后，关于第 23 条文献中 227，抱歉给老师带来困惑。原文中对于被试的描述仅为“对 227 名学龄儿童(8-14 岁)的阅读数据和迷宫学习结果。。。”，因此 227 对应的是学龄儿童。为了更方便的理解，我们将 227 更改为 227SAC，并在表注中标注 SAC(school-aged children): 学龄儿童。

**意见 7:** 正文在罗列前人研究时，应当先陈述时间早的后时间晚的，并且尽量是能够说明白这些研究之间如果递进研究问题的。如“Ileri 等人(2022)记录”和“Bhargavi 和 Prabha (2020)收集了”，这两个研究对调顺序。“Rauschenberger 等人(2022)”和“Rello 等人(2020)设计了”对调。“Yu 等人(2022)基于”“Asvestopoulou 等人(2019)开发”。“Rauschenberger 等人(2022)设计了”“(Radford et al., 2021)。”这几项研究的顺序建议调整。

**回应:** 感谢评审老师的宝贵建议。针对这一问题，我们对全文进行了调整，按照时间顺序对前人研究进行罗列。

**意见 8:** “通过脑电图通道之间的相位同步探究大脑的连通性，然后提取鉴别特征用于阅读障碍的识别。”“由此产生令人瞩目的预测准确性”，多了一个”的“。

**回应：**感谢评审老师在细节上的纠正。针对这类错误，我们就全文进行了审查并修正。

**意见 9：**”使用 SVM 来预测儿童是否具是高家庭阅读风险儿童”，不通顺。

**回应：**感谢评审老师的宝贵建议。现将此句修改为：使用 SVM 来区分有家庭风险的阅读障碍儿童和典型发展儿童。

**意见 10：**研究发现，有家族阅读障碍风险的识字前儿童在颞顶叶和颞枕叶区域已经显示出功能和结构上的改变，类似于在被筛查出阅读障碍的患者中观察到的变化(Hosseini et al., 2013; Kraft et al., 2015)。建议修改为“研究发现，有家族阅读障碍风险的识字前儿童的颞顶叶和颞枕叶区域已经显示出功能和结构上的改变，类似于被筛查出阅读障碍的患者中观察到的变化(Hosseini et al., 2013; Kraft et al., 2015)。

**回应：**感谢评审老师建议。我们根据您的建议进行了修改，同时对文中的“家族阅读障碍风险”这个名词进行了全文修正，改为“有家庭风险的阅读障碍儿童”。本句最终修改为：“这个阶段的有家庭风险的阅读障碍儿童的颞顶叶和颞枕叶区域已经显示出功能和结构上的改变，类似于被筛查出阅读障碍的患者中观察到的变化”。

**意见 11：**此外，与识字相关的重要区域的皮质可塑性可能是由基因调节的(Becker et al., 2017)。基于这些发现，Skeide 等人(2016)使用多变量非参数遗传模型将 NRSN1 基因(与识字相关)和视觉词形式区(VWFA)的灰质体积联系起来。这个逻辑是有问题的。

**回应：**感谢评审老师的建议。我们将此句修改为：Skeide 等人(2016)认为与识字相关的重要脑区的神经可塑性可能受到遗传变异的调控，从而预先限制了儿童的读写能力。为此，他们采集了 4~8 年级儿童和幼儿园~1 年级儿童的灰/白质体积以及与识字相关的基因信息，利用 LSVM 构建了阅读障碍的预测模型，其准确率分别达到了 73%和 75%。

**意见 12：**建议 4.2 和 4.3 互换下顺序。

**回应：**感谢评审老师的建议。我们互换了 4.2 和 4.3 的顺序，并对内容进行了调整。修改部分在文中已用红色字体标出。

**意见 13：**建议作者增加一个表，把目前纳入研究中提到的研究数据类型具体指标列清楚。

**回应：**感谢评审老师的建议。我们确实缺少了研究数据类型的具体指标，针对这一问题，首先，我们对表 1 中的数据指标进行了适当补充。其次，我们增加了表 2，将标准化心理教育测试、眼动追踪、网络/手机游戏、脑成像技术、其他这五种类型的测试提取的具体特征以表格形式总结。详细修改内容见表 1 和表 2。

**意见 14：**在参考文献部分，将用于分析的 25 条文献标识出来。

**回应：**感谢评审老师的建议。我们已在参考文献部分，将用于分析的 25 条文献以“\*”标识。

---

## 第二轮

**审稿人 1 意见：**

同意发表。

**审稿人 2 意见：**

**意见 1：**修改稿中，关键词建议“系统综述”去掉，发在心理科学进展上的期刊都是综述文章，系统综述也不是个方法。

**回应：**感谢评审老师的建议。根据您的建议，我们已经将关键词中的“系统综述”删除。

**意见 2：**参考文献部分，在第一条文献前加一句“\*号文献为纳入综述分析 25 篇文献”用以帮助读者理解\*号文献的意义。

**回应：**感谢评审老师的建议。我们已经在第一条文献前加上“\*为纳入系统分析的文献”，并用红色字体标出。

---

## 第三轮

**编委 1 意见：**同意发表

**编委 2 意见：**同意发表

**主编意见：**同意发表