

自然 > 转化精神病学 > 文章 > 文章

下载 PDF

文章 | [公开权限](#) | 发布时间：2023年10月10日

# 用人工智能预测主要青少年精神疾病的个别病例

[尼娜·德·莱西](#) , [迈克尔·J. Ramshaw](#)、[Elizabeth McCauley](#)、[Kathleen F. Kerr](#)、[Joan Kaufman](#)和[J.内森·库茨](#)

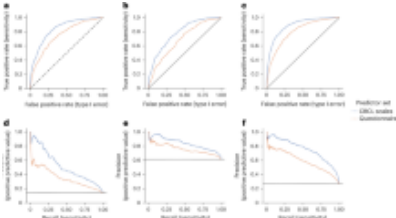
[转化精神病学](#)文章编号：314 (2023)

8006 访问 | 35个 引文 | [指标](#)

## 摘要

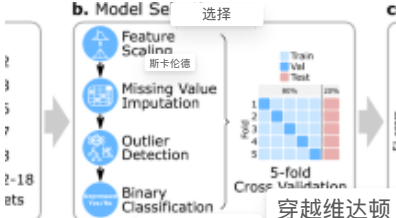
四分之三的终身精神疾病发生在24岁之前，但对如何有力地识别有风险的青年以针对已知改善结果的干预工作知之甚少。知识障碍包括获得可靠的预测，同时分析大量不同类型的候选预测因素。在一个新的、大型的、跨诊断青年样本和多域高维数据中，我们使用了160个候选预测器，包括神经、产前、发育、生理、社会文化、环境、情感和认知特征，并利用三种不同的机器学习算法，用一种新的人工智能元学习技术优化，来预测焦虑、抑郁、注意力缺陷、破坏性行为和创伤后应激的个别病例。我们的模型在看不见的、被保留的数据中进行了很好的测试 ( $AUC \geq 0.94$ )。通过利用大规模设计和先进的计算方法，我们能够以原则性的方式比较神经特征与心理社会特征的相对预测能力，并发现心理社会特征在对个别病例提供稳健预测的相对能力方面始终优于神经指标。我们发现，使用人工神经网络的深度学习和使用XGBoost的基于树的学习优于ElasticNet的逻辑回归，支持将精神疾病概念化为多因素疾病过程，在预测变量之间具有非线性关

系，可以用计算精神病学技术进行稳健建模。据我们所知，这是第一项测试这些黄金标准算法在青年多种心理健康状况的相对预测能力的研究，在同一研究设计中，在多领域数据中使用>100个候选预测因素。建议进一步研究，在纵向数据中探索这些发现，并在外部数据集中验证结果。



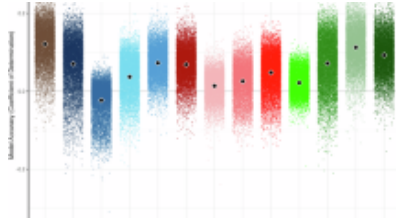
**青少年心理健康风险的预测**

第5条 | 2025年3月5日



**利用机器学习从ALSPAC的产前和儿童数据预测青少年抑郁症**

文章 | 开放获取  
2024年10月7日



**繪製心理病理學的神經發育預測因素**

2024年8月6日第6条 |

介绍

大多数终身原发性精神疾病在10至24岁（y）的临床诊断，发病率呈指数级增长，50%的终身疾病在14岁之前被诊断，75%在24岁之前被诊断[1]。重要的是，显性疾病通常先于~8-10年的症状[1,2,3,4]，为干预和预防提供了一个窗口。然而，目前很难轻易识别高风险的青年，而且诊断往往是追溯性的。由于跨疾病的证据表明，早期干预可以改善结果[5，6]并减少资源使用[7]，因此确定预测个别病例的具体因素可能会促进人口健康的进步，并为公共政策提供信息。在过去的十年里，人们对使用机器学习（ML）技术构建青少年精神疾病的预测模型的兴趣越来越大。在这里，我们使用机器学习中常用的“预测”一词：通过分析输入数据中的模式（“预测器”或“特征”）来训练判别分类器来预测目标变量的定量值。ML技术可以在专注于识别群体层面的统计关联的工作和临床相关性之间提供有用的桥梁，因为它们可以“从多维数据中生成单个预测，提供在单一主题水

平上有效的多变量签名”[8, 9]。此外，这些方法可以同时分析数十或数百个候选预测变量（尽管这可能会带来特征选择和计算复杂性的问题），并纳入一组预测变量之间的非线性关系。

历史上，几种类型的数据集被用来对青少年精神疾病进行预测分类。在美国以外，人口级注册数据可能可用，提供大样本量（ $n > 10,000$ ），但通常在数据类型方面受到限制。特别是，神经成像和/或心理测量测试通常不可用。此类示例包括在澳大利亚、中国、瑞典和英国完成的工作[10,11,12,13]。或者，一些参与者样本已经用更广泛的数据类型组装，包括神经成像，这些数据类型侧重于个人情况，如IMAGEN（抑郁症、焦虑症）或ENIGMA（ADHD）数据集[14,15,16]。为了促进大规模的比较发现，联邦和其他组织最近赞助了大型跨诊断或基于人群的样本的形成，这些样本收集了广泛的多模态数据类型和数百到数千个描述变量。在青春期，这种类型的旗舰举措包括正在进行的人口级ABCD和健康大脑网络研究，后者用于本研究[17,18,19]。这种大规模的多域数据存储库提供了比较各种表型、生理和神经候选预测因素的相对能力的机会，以预测更自然、基于社区的样本中多种精神疾病病例。近年来，新一代研究开始出现，在这些大型、多域、自然主义参与者样本中应用机器学习预测分类，这些样本提供了广泛的候选预测数据类型。迄今为止，这些专注于比较不同机器学习算法在抑郁症或破坏性行为等个别条件下的预测能力，并使用了<100个候选预测因素。在基于机器学习的预测工作中，某些常见的青少年心理健康状况没有得到较少的重视。特别是，很少有研究旨在使用ML分类器预测青少年的焦虑或创伤后应激（PTS）。

现有的研究使用机器学习技术来构建青少年精神疾病病例的个人级预测分类器，通常比较不同类别中多种算法类型的性能，以探索将模型与数据相匹配的不同方法，这是建议的做法[20]。这些通常是线性或决策树算法类，它们具有产生固有可解释的模型的优势：每个最终预测器集的身份是已知的，它们的相对重要性可以量化。通常，此类研究发现决策树算法实现了卓越的性能。较新的深度学习算法（例如人工神经网络）是强大的技术，非常适合高维数据。然而，它们较少用于构建精神病学中的判别分类器，是使用人类无法解释的中间特征的“黑匣子”方法。虽然所有ML算法都有超参数设置来管理需要“调整”（一种优化性能超参数设置的原则方法）的学习，但深度学习算法也具有调整的挑战性。迄今为止，很少有研究比较了深度学习与基于线性和决策树的算法在构建青

少年精神疾病病例的预测（分类）模型时的相对有效性。大多数（但不是全部）这样做的人发现深度学习的表现优于[12, 13, 16, 21]。此外，据我们所知，没有研究比较同一参与者队列和研究设计中青年多种心理健康状况中不同机器学习算法类别的相对预测能力。

焦虑、注意力缺陷、抑郁、破坏性行为和创伤后应激（PTS）是年轻人中最常见和最致残的精神疾病。本研究的重点是扩大我们对机器学习算法为青年中这些主要精神疾病提供可解释预测模型的能力的了解，方法是（1）分析多种数据类型（例如，心理社会、认知和神经）的更多候选预测因素（ $n = 160$ ）；（2）比较不同类别（深度、决策树和逻辑回归分类器）的黄金标准机器学习算法的预测能力；（3）在同一分析设计中构建跨多种精神健康状况的可解释模型；以及（4）使用类似于心理健康诊所的大型自然主义参与者样本，即丰富至少一个行为投诉。

我们假设，在构建青春期周围发育生活阶段主要精神疾病的个人层面精神病病例时，用人工智能（AI）优化的深度学习将优于决策树和逻辑（线性）回归。为了确定高维多域数据中不同算法类的相对性能，我们比较了黄金标准逻辑回归（LR）、决策树

（XGBoost: XGB）和深度学习（人工神经网络: ANN）算法的预测能力，以预测来自健康大脑网络（HBN）队列的5-21岁青年及其父母的大型（ $n = 1120$ ）、自然主义、跨诊断样本中的焦虑、注意力缺陷、抑郁和破坏性行为的个别病例。此外，我们添加了一个探索性分析，试图预测PTS，这在历史上不是ML技术的重点。构建了预测模型，同时调查和量化了160个候选预测因素的相对预测能力，包括神经、产前、发育、生理、社会文化、环境、情感和认知特征。我们没有对样本进行分层，以便在更自然的数据中测试该设计和算法的预测能力。通过将每个ML算法（LR; XGB; ANN）嵌套在创新的人工智能元学习方法中，我们通过共同学习超参数和执行自动化、有原则的特征选择来优化算法性能，同时使深度学习可解释翻译应用。所有模型都是在保留测试数据集中测试的，本稿中提出的所有结果都来自保留测试集的测试。据我们所知，这是第一项在多领域数据中在同一研究设计中测试不同类别青年多种精神健康状况的黄金标准算法的相对预测能力的研究，并将与累积的早期不良事件相关的PTS纳入基于机器学习的预测分类。

# 材料和方法

## 术语和定义

本稿通篇使用ML术语和惯例[20, 22, 23]。因此，“预测”是指通过分析用于预测目标变量值的输入数据（“候选预测因子”或“特征”）中的模式来预测目标变量的定量值。在这种情况下，输入数据在时间上与目标变量同步，因为这是一个横截面的参与者样本。我们既不打算使用先例特征来预测病例，也不打算预测未表现出症状的个体未来病例的发生。我们将所有输入数据集称为包含“候选预测因素”（特征），在最终优化模型中确定的实际预测因素称为“最终预测因素”。用于训练、组合和测试正在开发的模型的预测

下载 PDF

的、以前从未见过的数据的能力，尽管我们承认，推广在其他领域可能具有不同的含义。“精度”是指正确的正面预测的百分比；“召回”是指正确预测的真实正面比例；“准确性”是指正确预测的数量占总预测的百分比。我们提供接收器操作特性曲线（ROC曲线），显示分类模型在不同分类阈值下的性能，绘制真阳性与假阳性率，以及曲线下面积（AUC），定义为ROC曲线下从（0,0）到（1,1）的二维面积。

## 参与者样本

从正在进行的健康大脑网络（HBN）研究中选择了至少一次完整的静止状态fMRI扫描（365卷）和可用表型数据的参与者，以第8版[17]。HBN计划从纽约市地区至少有一个行为问题的5-21岁青年那里收集多领域数据。这是一个自然主义的、基于社区的横断面人口样本，而不是一个完全具有代表性的流行病学设计或纵向样本。在原始或当前研究中，没有尝试根据人口统计学或精神特征对样本进行分层，或确定任何精神健康状况的疾病发病点。排除标准是存在严重的安全问题、认知或行为障碍（例如，智商水平<66）或可能混淆大脑成像的医疗问题。整体和当前研究样本丰富了心理健康问题：大约2/3符合临床诊断标准。表1列出了该参与者样本的人口特征，以及参与者自闭症特征水平、手性、身体质量指数以及认知和执行功能核心指标表现的汇总指标。参与者样本

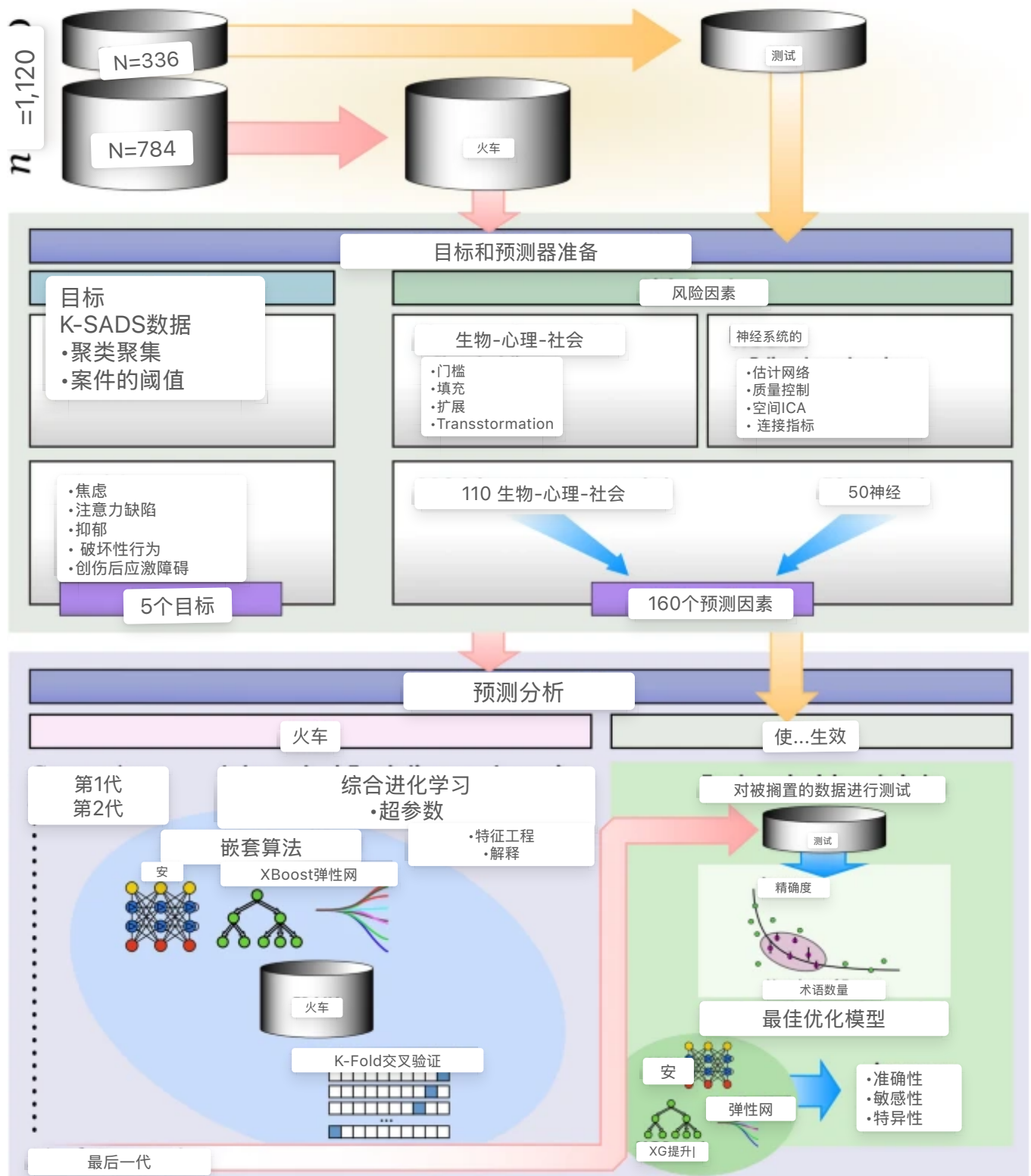


被随机分割，70%用作训练/验证集（784名参与者），30%（336名参与者）保留为未见的测试集。表2可以查看研究中每种心理健康障碍病例（预测目标）的受试者的数量和比例。数据准备和预处理管道分别应用于训练和测试集（图1）。HBN研究得到了切萨皮克机构审查委员会的批准。华盛顿大学机构审查委员会和犹他大学机构审查委员会认为本研究不是人体研究。

表1参与者样本的特征。

表2参与者样本中青年主要精神疾病的病例数。





研究计算管道用于分类实验。单个机器学习算法（人工神经网络、XGBoost和逻辑回归）嵌套在集成的进化学习架构中，通过跨超参数提供联合特征选择和优化来优化学习。该图改编自de Lacy等人，综合进化学习：一种人工智能方法，用于联合学习特征和超参数，用于优化、可解释的机器学习。《前线人工智能》（2022）。

## 数据和数据准备 数据收集

所有数据最初是由儿童心灵研究所收集的，作为正在进行的HBN研究的一部分。HBN项目的综合详细信息可以在项目数据门户网站上访问：

[https://fcon\\_1000.projects.nitrc.org/indi/cmi\\_healthy\\_brain\\_network/index.html](https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/index.html)。简而言之，在四次3小时的研究访问中收集了多领域数据，每项研究包括认知和语言测试、行为评估以及家庭结构、压力和创伤、身体功能和药物使用措施。第一次就诊时进行认知测试、生物采样和模拟扫描，第二次就诊时进行核磁共振扫描，第三次就诊时进行学习和语言测试，最后一次就诊时进行心理健康评估。所有数据均由儿童思维研究小组收集。有关数据收集协议的更多详细信息，请访问：

[https://fcon\\_1000.projects.nitrc.org/indi/cmi\\_healthy\\_brain\\_network/The%20Project%20Plan.html](https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/The%20Project%20Plan.html)。

## 表型候选预测器准备

本研究使用的候选预测因素可以在图中回顾。[2](#).表型候选预测因素包括多个指标，其中包含有关参与者的发育史、教育史、认知特征和功能、行为功能、青年和父母经验、青年和父母人口统计、青年社交技能和功能、生理特征和病史（包括使用的精神药物）的信息（当前和过去）。如上所述，这是一个自然主义的、基于社区的样本，因此，没有根据人口特征（例如种族/民族）进行分层。通常，类似研究设计中的机器学习预测分类通过将种族和性别/性别等人口特征作为候选预测因素来考虑。同样，在所有分析中，包括种族和性别在内的大量父母和青年人口特征被平等地纳入候选人预测因素。补充表[1](#)提供了有关每个候选预测因素的更多详细信息，包括每个特征的名称和描述，以及绘制该特征的原始心理测量或心理评估工具的名称。补充表[2](#)（由儿童心灵研究所提供）包括对每个原始心理测量或心理评估工具的更详细的描述，以及与每个工具相关的基础文献引用，包括其心理测量属性的指标（如适用）。丢弃了>40%缺失值的候选表型预测因子（ $n = 26$ ）。使用这个阈值，因为之前的研究表明，使用ML方法可以获得良好的结果，其归因高达50%缺失数据的阈值[24]。连续变量被修剪为平均值 $\pm 3$ 标准差，以消除



离群值。使用非负矩阵因数分解 (NNMF) 归因缺失值。NNMF是一种经过数学验证的归因方法，它最小化了缺失数据的成本函数，而不是假设零值。它有效地捕获了数据中的全局和局部结构，并且已被证明无论基本缺失模式如何，它都能表现良好 [25,26,27]。补充表3分别显示了训练和测试集的NNMF修剪和填充的观测数量和百分比。对于连续的衡量标准，我们选择了汇总或总指标。对于11个仪器（图2中斜体），我们通过应用特征聚集来递归合并单个项目并生成单个连续度量来计算汇总度量。然后使用MinMaxScaler对所有功能进行缩放。具有非正态分布的特征使用分位数和功率变换进行变换，并选择了最类似于正态分布的后变换特征版本。计算了每个候选表型预测因素与每个目标的单变量赔率和置信区间，可以在补充表4中查看。

**图。2： 人工智能适应性选择前后青年精神疾病的候选预测因素。**

同时评估了跨越人类功能多个领域的160个候选预测因素预测精神疾病的能力。该图改编自de Lacy等人，综合进化学习：一种人工智能方法，用于联合学习特征和超参数，用于优化、可解释的机器学习。《前线人工智能》（2022）。

## 神经连接功能

我们从功能性核磁共振（fMRI）中计算出黄金标准的大脑功能连接测量。多频段3T静止状态，眼睛睁开fMRI，包括365卷，在2个部位获取。删除前10卷后，使用SPM12中的标准算法（<https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>）对每个参与者的扫描

进行重新对齐、共同注册、归一化和平滑，全宽为6毫米，最大为一半。然后，通过计算与组掩码的相关性，将扫描提交质量控制，并淘汰了与该组掩码相关性<90%的21名参与者。使用DVARs指标[28]计算了每个参与者的头部运动，扫描部位和DVARs被包括为候选预测因素。我们使用既定的管道执行组空间ICA，并使用fMRI工具箱的组ICA提取代表15个功能成分的全脑分区方案[29]（GIFT：<https://trendscenter.org/software/gift/>）。ICA估计的组件被分为灰质内在功能网络与人工噪声组件，结合NdL的专家目视检查和低频波动和动态范围分数振幅的定量指标[29]。与大脑灰质或低光谱度量重叠不良的成分被丢弃，我们保留了一组10个功能性内在神经网络（IN）。我们按照既定的GIFT管道为每个IN构建了空间地图[29]。为了确定IN之间的功能连接特征，我们计算了每对可能的空间地图之间的Pearson相关性。随后，功能网络之间的皮尔逊相关性被费舍尔变换[ $z = \text{atanh}(k)$ ]平滑离群值。通过从fMRI ICA时间课程中描述稳定的动态全脑连接状态，并使用既定的滑动窗口方法将k均值聚类算法应用于连接窗口，计算了一组额外的动态连接度量[30]。滑动窗口长度是根据先前的工作选择的，表明窗口长度为40-60秒会产生合理和稳健的结果[30,31,32]。这些动态连接指标描述了参与者穿越大脑状态的流动性和范围，并且每个参与者都可以使用。具体来说，它们是穿越的大脑状态数量、受试者在大脑状态之间切换的次数、大脑状态和总距离（所有L1距离的总和）之间实现的最大L1跨度。在计算之前，使用通用线性模型从所有连接模型中回归了扫描仪站点、DVARs统计量和六个重新排列参数及其每个参与者的六个一导数的骚扰回归器。计算了每个候选神经预测因子与每个目标的单变量赔率和置信区间，可以在补充表4中查看。有关特定神经指标的更多详细信息，请参阅补充表1。

## 预测目标

精神疾病病例的预测目标是用从情感障碍和精神分裂症儿童时间表（KSADS-COMP）[33, 34]的计算机化版本中获得的数据计算的，这是一个由临床医生对青年参与者及其父母进行的标准化、半结构化的诊断访谈，广泛用于青年精神病学研究。诊断筛查评估的响应涵盖了通过临床医生共识获得的所有主要精神疾病的多种症状。对于本研究中的

五个疾病目标中的每一个，每个诊断类别可用的多个KSADS症状分数都使用特征聚集减少到单个指标。然后确定了每个疾病目标的分数分布，从而将参与者分为“病例”和“非病例”，形成一个二进制目标向量进行分类。具体来说，在KSADS症状上表现出零聚集得分的参与者组被认为是“非病例”。第二组显示的分数反映了每个目标的症状严重程度范围，该组成员被视为“病例”。鉴于该样本的自然主义、基于社区的性质，没有分层，在病例与非病例方面存在不平衡。使用SMOTEENN算法的合成过采样是在每个目标的参与者样本的训练/验证部分上进行的，但没有在测试分区上进行。表2可以看到培训集中合成过度抽样前后每类（病例与非病例）的参与者人数和比例。

## 预测分析概述

我们比较了三种领先的机器学习技术使用160个候选预测器（图2）预测每种类型精神疾病病例的能力：深度和基于决策树的学习和逻辑回归。所有使用的ML估计器（Adam；XGBoost；Logistic Regression）都是来自标准库（Tensorflow、Sci-Kit Learn）的“现成”算法。为了优化学习性能，每个ML模型都使用k倍交叉验证进行，并嵌套在集成进化学习（IEL）人工智能算法中。IEL是一种人工智能元学习遗传算法，它围绕每个ML算法来优化后者的性能。因此，每个ML算法和IEL都是独立但交互式的计算过程：ML算法适合一个模型并执行案例分类，而IEL共同学习每个ML算法的特征（预测器）和超参数。通常，IEL在每个优化过程中追求约40,000 ML模型拟合。为了减少偏见，纳入了一些方法和设计特征，这些功能详见上图和下文。简而言之，使用非负矩阵因数分解填充缺失数据；在训练和测试分区上分别进行预处理；通过原则性人工智能元学习架构执行大规模特征选择；探索数千个模型拟合以扩大解决方案空间；并根据使用信息理论适应性函数量化的原则基础终止训练。预测分析的代码可以在de Lacy实验室GitHub上访问：[https://github.com/delacylab/integrated\\_evolutionary\\_learning](https://github.com/delacylab/integrated_evolutionary_learning)。

## 使用人工神经网络进行深度学习

我们使用Adam算法训练人工神经网络，有3层，每层300个神经元，早期停止（耐心

=3, 度量=验证损失) 和Relu激活函数。最后一个输出层包含一个传统的softmax函数。选择Adam算法是根据其既定的计算效率和对大量参数问题的适用性, 如我们的研究[35]。学习参数(补充表5) 使用IEL深度学习模型进行调整, 并使用嵌入自定义Python代码的TensorFlow进行编码。

## 梯度增强基于树的学习

我们训练了基于决策树的模型, 以使用gbtree助推器使用XGBoost算法预测精神疾病病例[36, 37]。这是一种基于合奏的方法, 可以生成大量决策树, 对复合预测进行“投票”。它是准确的, 在适当调整时可以抵抗过度拟合, 并使用模型残差(实际预测值)来惩罚不能改善预测的叶子, 减少偏差和方差。从经验上看, 梯度增强技术非常成功。超参数是用IEL调整的, 可以在补充表5中查看。XGBoost在自定义Python代码中用Scikit-Learn包装器进行编码。

## 逻辑回归

我们训练了线性模型, 用ElasticNet [38]对逻辑回归进行分类。后一种正则化方法线性结合了LASSO(最不绝对收缩和选择算子)的L1惩罚和Ridge方法的L2惩罚。它在现实世界和模拟数据中产生了卓越的结果, 特别是单独使用LASSO。对于逻辑回归分析, 计算了优化模型的赔率和置信区间。使用IEL调整了L1和L2参数(补充表5)。我们使用嵌入在自定义Python程式中的Scikit-learn算法对Logistic回归模型进行编码。

## 用于机器学习优化的综合进化学习

ML模型, 特别是深度学习模型, 以难以“调整”而闻名, 即确定许多控制学习的算法超参数(设置)的正确值, 并可以对性能产生巨大影响。为了实现机器学习实验的最佳性能, 我们开发并在这里应用了一种我们称之为综合进化学习的元学习技术。我们之前已经证明, 与使用默认模型超参数和传统设计相比, IEL在可比数据中将ML预测算法的性能提高了20-25%[39]。

通常, 调整通常通过“拇指规则”手动完成, 并探索 $\leq 50$ 模型拟合, 引入了偏差和限制解



空间的可能性[40,41,42]。此外，使用人工神经网络进行深度学习等强大技术可以充当“黑匣子”，通过机器生成的中间特征获得人类无法解释的预测。对于生物医学中常见的翻译或机械应用，可解释的机器学习是一个优先事项。许多研究人员放弃了深度学习，专注于决策树或线性模型等固有的可解释技术。相比之下，IEL通过利用进化算法、高级计算元启发式、实例化计算机代码中的生物进化选择概念，为任何ML技术提供自适应、自动特征选择和原则性超参数调整。所有结果，包括从深度学习中获得的结果，都是完全可以解释的。IEL通过选择改进健身功能（在这里，贝叶斯信息标准），在数百代中自适应地“繁殖”了优化模型（图1）。

对于每种学习技术，我们用随机超参数值或“染色体”初始化了第一代100个模型。这些随后在连续几代中被重组、突变或消除。在重组中，将“父”超参数平均为“子”。在突变中，超参数设置被转移了。补充表5显示了可能值的范围。除了逻辑回归模型（在[0-1]中具有自然限制的超参数区间）外，这些模型被慷慨地设置，以便广泛探索潜在的解决方案集。在训练了最初的100个模型后，我们计算了每个解决方案的BIC。在80个最佳模型中，40个通过在中点枢轴点后平均超参数设置来重新组合，以产生20个“子”模型。20个通过突变转移值移动所需的超参数来突变，产生相同数量的子模型（补充表5）。剩下的20个被丢弃了。下一代模型是通过添加60个具有随机设置的新模型，并将这些模型添加到从第一代保留的40个子模型中形成的。之后，一个自动化过程继续以类似的方式重新组合、突变和丢弃每代100个模型，以尽量减少BIC，直到后者的健身功能停滞不前。

用人工智能分析多域“大数据”的另一个主要问题是特征选择：如何以原则的方式从正在考虑的更大集合中识别少量预测风险因素。从经验上讲，变量较少的模型更容易训练、运行、理解和推广。与调音一样，许多从业者仍然使用手动或半手动的方法。鉴于在培训期间筛选了大量功能，我们在IEL中结合了自动功能选择，以减少过度拟合的风险。在这里，在初始一代中为每个模型设置了范围内的随机数特征（2-50），并从160个可能的候选预测器中随机抽样。计算BIC后，将性能最好的60个模型的特征集单独分配给重组和突变的子模型。其他功能集被丢弃了。与超参数调谐一样，这个过程在后代重复，直到BIC稳定下来。

为了促进计算效率建模，IEL实现了递归学习。在训练模型后，直到BIC稳定下来，我们确定了绘制的健身功能的肘部与特征的数量。热启动后可用的功能集受限于功能子集，根据其重要性进行阈值，对应于健身功能肘部。在热启动后，学习通过在原始热启动特征重要性+2-10标准偏差的阈值增加时可供学习的阈值特征进行。此外，我们将每代模型数量减少到50个，重新组合20个模型，10个模型突变。否则，在暖启动阈值范围内重新启动训练过程后，随机初始化了第一代模型，并使用上述相同的原则完成了培训。

## 交叉验证

对于三种学习技术中的每一种，使用分层 $k$ 倍交叉验证来拟合单个模型，即IEL中每个学习一代的100个模型中的每一个都使用交叉验证进行单独训练和验证。由于每个模型拟合的特征数量在每一代IEL中每个模型中可能有所不同，因此 $k$ （拆分数）被设置为[样本大小/特征数]上方最接近的整数。使用scikit-learn StratifiedKFold函数实现了交叉验证。

## 确定预测因素的重要性

使用了两种技术来计算每个预测因素在预测中的相对重要性。首先，使用SHAP工具箱 (<https://shap.readthedocs.io/en/latest/>) 计算了Shapley加法解释 (SHAP) 值。SHAP是一种博弈论方法，可用于解释任何ML模型的输出，包括人工神经网络等“黑匣子”估计器[43]。它统一了LIME、Shapley采样值和树解释器等先前的方法。其次，采用了每种算法可用的固有方法。对于XGBoost，每个特征的重要性都是在算法中计算的，对于决策树估计器来说，这本质上是一种基于排列的方法。因此，为了提供可比性，我们还通过在IEL中嵌入eli5 (<https://eli5.readthedocs.io/en/latest/index.html>) 来确定人工神经网络的特征重要性[44]。在相关的情况下，通过计算其线性系数 (beta) 来确定每个风险因素对逻辑回归的相对重要性。

## 在保留的测试数据中进行测试

在完成每种精神疾病类型的培训后，在336名参与者的保留的未见测试集上测试了IEL过

程中表现最好的100个调整模型，并使用标准的Sci-Kit学习库确定了准确性、精度、召回和AUC。预测概率的阈值为0.5，图中为每种算法类型和每种目标精神健康状况提供了ROC曲线。[3.](#)为了测试，我们将从IEL训练/验证过程中性能最佳模型中获得的优化模型参数（超参数设置和选定特征）应用于看不见的数据。合成过采样没有用于制定预测目标（见：“预测目标”）。本文中报告的所有结果都来自对未见数据的测试。

**图。3：用三种机器学习算法预测青少年精神疾病病例的ROC曲线。**

ROC曲线显示给具有Adam估计器的深度学习；**b**具有XGBoost算法的决策树学习；以及具有ElasticNet正则化的**c**逻辑回归的预测模型，所有这些都用于预测青年的焦虑、注意力缺陷、抑郁、破坏性行为和创伤后应激的情况。

## 结果

使用IEL优化的深度和基于树的学习可以高保真度地预测主要精神疾病

我们在IEL架构中嵌套了三种具有交叉验证的黄金标准算法，以比较它们的预测能力：使用人工神经网络进行深度学习，使用XGBoost进行基于决策树的学习，以及使用ElasticNet进行正则化的逻辑回归模型。表3显示了每种技术在复制测试中预测青少年精神疾病病例的比较能力，包括准确性、精确性、回忆和AUC。可以在图中检查ROC曲线。[3](#).逻辑回归分析的赔率和置信区间可以在补充表6中查看。

**表3三种算法类型在预测青年重大精神疾病的相对性能。**

在ML分类器中，没有单一的指标来确定最佳性能，因为如何判断性能可能取决于构建模型的用例。例如，在基因测试期间，精度（正预测值）可能被优先考虑，而召回（灵敏度）在大规模筛查试验中可能是最重要的。然而，AUC通常被用作所有可能的分类阈值中整体绩效的汇总衡量标准。在AUC方面，我们发现人工神经网络的深度学习在预测焦虑和注意力缺陷方面提供了卓越的性能，而决策树学习在预测抑郁、破坏性行为和PTS方面表现更好。逻辑回归在所有条件下表现一直不足，并达到0.5的AUC，这表明这些分类器的表现并不比随机机会好。就其他性能统计数据而言，深度学习在所有心理健康条件下都表现稳定，始终如一，实现≥85%的准确性、精度和回忆，尽管在PTS的情况下，鉴于AUC = 0.5，必须谨慎对待其性能。决策树学习在预测注意力缺陷、抑郁、破坏性行为和PTS方面在准确性和回忆方面表现普遍表现强劲，尽管不是焦虑，但除了注意力缺陷外，自始至终都表现出低回忆。逻辑回归技术在五个条件下实现了57到95%的合理精度，但除了注意力缺陷外，精度和回忆率显著较弱，鉴于所有条件的AUC为0.5，必须谨慎解释所有这些结果。有趣的是，我们的结果表明，注意力缺陷似乎是包括LR在内的所有方法中最容易处理的目标。



每种主要精神疾病都是通过最终预测因素的独特组合来预测的

在AUC测量的深度和决策树学习产生的最佳表现模型中，每个青年精神疾病都有独特的最终预测因素组合（表4）。逻辑回归模型只包含一个最终预测因素。除了探索性PTS模型外，深度学习模型通常是最复杂的，即包含最多的最终预测因素。除了抑郁症中的语音音素发音外，从功能核磁共振中得出的生理、认知和神经指标在优化模型中作为最终预测因素被不强调，其中大多数预测因素是心理社会或心理测量指标。就不同预测变量的相对重要性而言，Shapley和排列方法的重要性分数在深度和决策树模型中大致一致。

**表4 最终预测因素在预测青年重大精神疾病个体病例中的相对重要性。**

在性能更强的深度学习模型中，焦虑是由六个预测因素预测的：不良事件、外化特征、父母压力水平、内化特征、情绪化和社交技能能力。这些在排列或Shapley值测量的重要性相当相等。性能较弱的决策树算法只使用了其中两种（不良降临和内化特征），而逻辑回归模型只将特征内化作为最终预测因素。在注意力缺陷中，深度学习模型再次复杂，有六个最终预测因素（多动性和外化特征、亲子关系压力水平和父母对孩子的感受、手性和自闭症特征）。排列和Shapley都重视更强的加权多动和父母对孩子的感受。回忆率低和AUC的决策树模型还选择了多动和外化特征，但添加了“与同学相处融洽”和核磁共振部位的亲社会因素。逻辑回归模型只包含过度活动。在抑郁模型中，情绪和负面影响被决策树算法选择为同等权重的最终预测因素，以获得更高的AUC，其中深度学习模型增加了年龄、言语表达、父母对孩子的感受以及将特征水平外化到情绪，以达到更高的精度。逻辑回归模型只使用年龄。决策树模型的破坏性行为的预测因素由青年的行为障碍水平以及外化特征和多动程度主导。深度学习模型还加重了外化和多动特征水平，但增加了社会反应和情绪症状水平。逻辑回归模型仅使用外化特征。如上所述，鉴于其AUC = 0.05，我们的探索性PTS模型在深度学习或逻辑回归的情况下不能被认为是可靠的。然而，决策树模型表现良好（除了精度低）。在这里，该模型确定了日常生活功能水平、正常发展程度以及应对技能和情绪症状水平作为最终预测因素。

## 优化的模型是节俭的

IEL集成了强大的自动功能选择，根据健身功能自适应地选择性能最佳功能。我们的结果表明，深度或决策树学习只需要少量的最终预测器来生成高保真预测，这些预测在所有精神疾病中推广到看不见的、保留的数据，从最初纳入考虑的160个减少。在每种情况下，在训练期间，在拟合最终优化模型的过程中，探索了约40,000个模型拟合，该模型被选中用于在保留的、看不见的的数据中进行测试。没有最终模型（表4）需要超过六个最终预测因素来做出高质量的预测（焦虑和注意力缺陷的深度学习，以及抑郁症、破坏性行为 and PTS的决策树学习），通过准确性、精确度、回忆和AUC来衡量。

## 讨论

在青年的五个常见条件下使用相同的分析设计，我们表明，通过应用人工智能来优化性能，可以从自然主义参与者样本中的黄金标准ML分类算法中获得严格的个人案例预测（ $AUC \geq 0.94$ ）。

焦虑和抑郁是青春期青年的常见疾病，在现象学和流行病学上可能重叠，有时被归类为“内化”障碍。我们发现焦虑和抑郁具有不同的最终预测因素，有助于消除这些情况的歧义。XGBoost最好地预测了情绪和负面影响的抑郁（ $AUC = 0.94$ ；准确性=98%；精度=14%；回忆=88%），尽管鉴于后者算法的精度和回忆率更高，我们的深度学习结果（ $AUC = 0.77$ ；准确性=93%；精度=93%；回忆=98%）是有价值的。在这里，抑郁症是由年龄、情绪、说话发音、父母对孩子的感受和外化特征来预测的。在以前的研究中，我们的发现与最终预测因素（横截面和纵向）之间的共同点在于突出情绪症状水平和与父母或亲子关系相关的因素的预测作用。在澳大利亚青年的大型（ $n > 6000$ ）横截面样本中，Haque等人使用了62个最终特征（主要基于心理社会和症状）和几种类型的决策树算法。随机森林（RF）表现最好（ $AUC = 0.74$ ），最终预测11种基本临床抑郁症状中的任何5种（例如，无快感、易怒情绪、睡眠障碍、自杀倾向）[10]。Xiang等人比较了ABCD样本中的几种决策树算法，以预测抑郁症轨迹，发现梯度增强算法效果最好（ $AUC = 0.90$ ），睡眠障碍、父母心理健康负担和家庭经济逆境是最重要的预测因素

[45]。在IMAGEN队列中，Toenders等人使用LR ( $AUC = 0.70$ )，并确定了基线抑郁症状的严重程度、女性性别、神经质、先前的欺凌、不良生活事件和上回的表面积是后期青少年抑郁症的最重要预测因素[14]。最后，Huang等人发现，RF的表现优于CART和SVM算法 ( $AUC = 0.90$ )，自杀倾向、无快感、缺乏社会支持、童年时期的情感忽视、非自杀性自我伤害和不良的母性关系是青少年抑郁症最重要的最终预测因素[11]。

就基于机器学习的预测分类而言，焦虑的关注频率低于抑郁。我们发现，深度学习比XGB ( $AUC = 0.61$ ) 或LR ( $AUC = 0.5$ ) 产生明显更强的性能 ( $AUC = 0.94$ ；准确性=95%；精度=94%；回忆=95%)，焦虑是由不良生活事件史和内化、外化和情绪特征的水平以及亲子关系和社交技能的压力预测的。在情绪化、内化行为和家庭相关因素方面，我们的最终预测因素与之前使用基于机器学习的预测分类的青年焦虑研究相似。在这里，Chavanne等人使用27个特征（结构核磁共振的14个体积和13个心理社会）训练了一个由RF、SVM和LR ( $AUC = 0.68$ ) 组成的多数投票分类器，发现未来随访中集合焦虑目标（类似于我们自己的）的顶级预测因素是神经症、绝望、情绪症状、家庭、酒精消费和痛苦程度[15]。总体而言，我们分析中表现最好的模型区分了这些内化条件：在这个青年队列中，焦虑和抑郁之间没有重叠的最终预测因素。

注意力缺陷和破坏性行为在现象学和流行病学上可能同样重叠，通常集体被认为是“外化”障碍。与内化障碍一样，我们确定了这些条件的不同最终预测因素，尽管每种条件的最佳模型之间共享外化和多动行为水平。我们发现，在表现最好的算法方面存在分歧，注意力缺陷最好通过深度学习 ( $AUC = 0.95$ ；准确性 = 96%；精度 = 89%；回忆 = 90%) 预测注意力缺陷，使用多动性、父母对孩子的感受、双手、亲子关系中的压力程度、自闭症和外化特征，而XGB在预测破坏性行为 ( $AUC = 0.96$ ；准确性= 97%；精度 = 28%；回忆 = 91%) 方面表现最好，具有行为障碍、外化和多动特征。与抑郁症类似，XGB在预测破坏性行为方面的精度相对较低，因此通过深度学习 ( $AUC = 0.73$ ；准确性=85%；精度=85%；回忆=94%) 的外化和多动性状水平、社会反应和情绪症状水平的最终预测因素值得考虑。

我们目前的研究确定了注意力缺陷和破坏性行为的最终预测因素，这些因素与之前基于机器学习的工作一致，突出了沟通缺陷和社会/情绪发展（在我们的实验中可能由自闭症

特征水平表示) 以及与家庭/家庭环境相关的因素。

多动症中基于机器学习的历史预测建模通常涉及较小的样本量, 许多研究缺乏保留的测试集[16]。然而, 最近, 在大型横截面样本中出现了几项关于多动症的研究, 包括神经成像的ENIGMA队列(为解决前一个问题而创建)。在瑞典注册处 ( $n > 200,000$ ) 中, Garcia-Argibay等人使用22个心理社会特征来预测多动症病例, 将深度学习与RF、XGB Naïve Bayes和LR与ElasticNet进行了比较。深度学习表现最好 ( $AUC = 0.75$ ), 有刑事定罪的亲戚、性别、学校表现、言语障碍和急性压力水平是最重要的预测因素[12]。Maniruzzaman等人使用日本国家调查 ( $n > 45,000$ ), 并将几种决策树算法与K最近邻居、SVM和两种深度学习算法(多层感知器(MLP); 卷积神经网络)使用19个心理社会预测器进行了比较, 将RF确定为最佳表现者 ( $AUC = 0.94$ ), 儿童年龄、性别、母亲年龄、过敏、哮喘、家庭结构和精神合并症作为最终预测因素[46]。在ENIGMA数据集中, Zhang-James等人比较了九种算法的预测性能, 包括多种类型的决策树和线性算法以及仅使用从维度减少主因子因子分析中得出的神经特征的MLP的两种变体。深度学习表现最好 ( $AUC = 0.64$ ), 尽管维度还原后的特征在常规上无法解释[16]。最後, Ter-Minassian等人使用英國Maudsley和National Pupil資料庫 ( $n > 55,000$ ) 將深度學習(MLP) 與RF、NB、LR和SVM以及68個心理社會特徵進行了比較, 確定了出勤率、社會/情感發展、寫作表現、男性和解決問題/推理[13]。这是为数不多的发现LR表现优于儿童心理健康的研究之一 ( $AUC = 0.72$ )。已经对破坏性行为进行了几项机器学习预测研究。Menon和Krishnamurthy在ABCD队列的1100名年轻人中使用了扩散、结构和静止状态成像, 并比较了个人与整体CNN, 在预测DBD时实现了0.74的AUC [47]。在一项特别有趣的研究中, Chan等人使用第一波ABCD研究(8-10岁的儿童)使用52个特征(心理社会、先前的多动症/ODD诊断、静止状态图神经指标)与神经网络算法预测行为障碍, 实现了 $AUC = 0.91$ 。在这里, 一个将心理社会与成像指标相结合的模式单独表现都优于其中, 最重要的最终预测因素是父母监控较低、家庭中的攻击性更高、收入较低、多动症和ODD症状较大、结晶认知和卡片排序性能较差以及皮质下和额顶网络的拓撲破坏[8]。

我们对创伤症状的分析更具探索性。绝大多数旨在预测创伤后应激障碍和创伤后应激障碍的研究都是在经历过军事、灾难或医疗相关创伤事件的成年人群中进行的。在2020年

的文献综合调查中，Ramos-Lima等人确定了49项此类研究。其中，只对青年参与者进行了一项研究，最近又发表了一项关于青年的研究[48,49,50]。两项现存的青年研究都侧重于为专门解决医疗伤害和/或住院后PTS而构建的队列。因此，我们的研究是第一个在自然主义队列中尝试预测的研究，在该队列中，PTS更有可能归因于不良生活事件的累积负担，包括情绪创伤。在我们的探索性分析中，我们发现PTS的预测略有困难。虽然性能最好的算法被证明是XGB（AUC = 0.97；准确性=100%；精度=5%；召回=94%），但这种技术在精度上很弱，我们也发现在其他条件下也是这种情况，事实上，它是表格数据中决策树技术的共同特征，可以在上述同行研究中看到。不幸的是，鉴于深度学习和逻辑回归模型（AUC = 0.5）的弱点，我们无法转向前者来访问其他四个条件的更精确的结果。XGB使用了日常生活功能水平、正常发展程度以及应对技能和情绪症状水平的最终预测因素。虽然之前没有基于机器学习的研究可供直接比较，但这些最终预测因素与之前使用相关技术在青年PTS中进行组比较的工作是一致的，这些工作发现较低的智力/发展水平、思想抑制（一种不适应的应对技能）和不良的生活/家庭功能具有中到大的影响大小[51]。创伤后应激障碍和创伤后应激障碍是复杂的现象，是一个非常活跃的研究领域。我们的结果很有希望，但模型的弱点可能反映了基础数据（例如，创伤的意向性；群体与个人创伤）或样本构建问题中缺乏PTS特定候选预测器。特别是，我们注意到，在当前分析中，用于分配案例来训练和测试分区的随机化过程在测试分区中产生了不可取的少数PTS案例。

据我们所知，这是第一项在同一分析设计中比较不同类别的机器学习算法在青少年多种主要心理健康状况的相对预测能力的研究，使我们能够进行一些进一步的观察。我们发现，在所有五种条件下，心理社会和心理测量结构是比大脑功能的神经指标更重要的最终预测因素。我们的经验在很大程度上与之前关于青年心理健康的研究一致，最近的大规模研究表明，大脑结构或功能个体间差异与复杂的认知或心理健康表型之间的关联效应大小比以前想象的要小，容易发生复制失败[52]。只有少数关于青年精神健康状况的预测研究同样纳入了多领域候选预测因素，包括神经指标和结果是混合的。在抑郁症方面，Xiang等人发现，心理社会预测因素在预测ABCD队列中轨迹严重程度方面的表现优于各种神经指标类型，而Toenders等人发现，两个结构性MRI指标在预测IMAGEN队列中的抑郁症方面很有用，但不如其他五个心理社会预测因素重要。在焦虑症方面，



Chavanne等人发现，结构性核磁共振体积测量并没有改善IMAGEN队列中集中焦虑目标的预测，尽管神经指标对预测广泛性焦虑症做出了部分贡献，但外向性以外的每个心理社会预测因素都超过体积。Chan等人使用第一波ABCD研究（8-10岁儿童）来预测行为障碍，并发现将心理社会与成像指标相结合，单独表现都优于其中之一。然而，预测分类的此类结果（成功时，通常识别出非常重要的最终预测因素）并不一定会降低神经指标在探测疾病机制中的作用，特别是当构建允许许多较小影响相互作用的专门模型时。此外，在包含多领域数据的青年预测分类方面的工作很稀缺，包括本研究在内的大多数现存工作都使用单一类型的神经成像（体积或功能）的指标。未来的研究将结合多种类型的神经指标与心理社会和认知候选预测因素，以改善我们对前者在青年心理健康预测分类中的潜在作用的理解。

在算法类型方面，我们发现使用ANN的深度学习或使用XGB的基于树的学习在所有五种心理健康条件下都优于逻辑回归。这一结果与之前对青少年的绝大多数可比研究一致，这些研究在单一疾病中比较了这些算法类别（见上文）。此外，我们发现，人口特征与相对严重程度和发病率的历史关联，如种族/族裔，特别是性别/性别，如果这些关联可能反映了病例确定的偏见，在表现最好的模型中，这些关联并没有被证明是青年主要精神疾病的共享或特定预测因素。除了零星的例外情况外，这一发现也与之前预测青年这些情况的研究一致[12, 13, 46]。此外，它提供了一些经验支持，即在确定案例中不存在可归因于种族/民族或性别/性别等特征的偏见。然而，我们也提醒说，这只是来自纽约市地区的基于社区的自然主义样本，因此可能无法得出更大的结论。

人工智能技术非常适合在多域生物医学“大数据”中优化机器学习预测模型，这些模型的基本机制定义不明确，可能是多因素和/或非线性或非分层的。尽管其在自动化和大规模方面享有盛誉，但大多数学术机器学习研究都是以特殊方式在小规模上进行，探索了≤50个模型适合[23]。这限制了科学搜索空间，增加了偏见和噪声的可能性，并降低了统计能力。在这里，我们应用IEL来解决机器学习领域的可解释性、特征选择和超参数调整的突出和当前问题，并在训练过程中为每个条件拟合约40,000个模型，以生成严格的预测，这些预测将预测推广到看不见的、保留的数据，这些数据很容易作为开源代码。我们的结果凸显了在日益成为精神病学关注的大型多领域开放科学数据集中投资模型调整的原则方法（如IEL）的价值。未来的方向可能会将本文中的方法和发现扩展到正在进

行的大型纵向参与者样本，如ABCD研究，以减轻目前对横截面评估的依赖。我们自己的这种研究目前正在进行中。

## 限制

本研究将AI/ML预测分析应用于5-21岁青少年的横截面样本，在整个术语中，“预测”在ML意义上用于指对给定输入数据示例的类标签的预测或判别分类器的构建。既不是因果关系，也不是同一示例受试者未来的类标签的预测，对于尚未表现出症状的受试者未来病例的预测，应该或可以从本分析中推算。本研究的参与者样本是自然主义的、以社区为基础的、未分化的青年及其父母人群。本作者没有收集原始数据，无法控制数据收集集中存在的任何偏差。没有尝试根据人口统计学或其他特征对样本进行分层，所有参与者都来自纽约市地区。因此，结果可能不适用于来自不同地区或分层样本的其他类型的人口样本。此外，没有尝试收集原始数据来确定疾病发病点，这是一个横截面样本。因此，候选预测因素与疾病发病的时间关系未定义。我们承认，本研究不包括详尽的连接组或基因组数据，尽管在没有这些数据的情况下获得了高保真预测。未来的工作可能会探索包括其他数据类型可以逐步提高边缘的准确性、灵敏度和特异性，并应验证外部数据集中的发现。

## 结论

通过将人工智能优化的ML应用于通用分析架构中的跨诊断、多域数据集，我们能够用在看不见的、持有的测试数据中推广良好的模型（ $AUC \geq 0.94$ ）来划定青年中五种最常见的精神疾病状况的离散最终预测因素集。鉴于对ANN和XGB算法的偏好，我们的结果支持了目前的直觉，即青春期前精神病是多因素疾病过程，预测因素之间具有非线性关系。在目前的数据集中，我们发现心理社会和心理测量预测因素优于神经功能指标。

## 参考文献

1. Kessler RC、Berglund P、Demler O、Jin R、Merikangas KR、Walters EE。国家合并症调查复制中DSM-IV疾病的终身流行率和发病年龄分布。Arch Gen精神病学。2005； 62： 593–602。

---

2. Christiana JM、Gilman SE、Guardino M、Mickelson K、Morselli PL、Olfson M等。焦虑症和情绪障碍患者从发病到获得初始治疗的时间：对心理健康患者倡导团体成员的国际调查。心理医学。2000;30:693–703。

---

3. Olfson M, Kessler RC, Berglund PA, Lin E.美国和安大略省的精神疾病发作和第一次治疗接触。我是精神病学。1998;155:1415–22。

---

4. Wang PS、Angermeyer M、Borges G、Bruffaerts R、Tat Chiu W、DEG G等。世界卫生组织世界心理健康调查倡议中精神障碍首次发作后寻求治疗的延迟和失败。世界精神病学。2007;6:177–85。

---

5. Csillag C、Nordentoft M、Mizuno M、McDaid D、Arango C、Smith J等。精神病的早期干预：从临床干预到卫生系统实施。早期精神病学。2018年； 12： 757–64。

---

6. Conus P, Macneil C, McGorry PD。双相情感障碍的公共卫生意义：对早期干预和预防的影响。躁郁症。2014;16:548–56。

---

7. Hamilton MP、Hetrick SE、Mihalopoulos C、Baker D、Browne V、Chanen AM等。确定可能提高青年心理健康服务系统成本效益的护理属性。Med J Aust。2017;207:S27–S37。

---

8. Chan L, Simmons C, Tillem S, Conley M, 巴西IA, Baskin-Sommers A。使用生物心理社会模型和机器学习方法对行为障碍进行分类。Biol Psychiatry Cogn Neurosci神经成像。2023;8:599–608。

---

9. Poldrack RA、Huckins G、Varoquaux G. 建立预测证据的最佳做法：回顾。JAMA精神病学。2020;77:534–40。

---

10. Haque UM、Kabir E、Khanam R。使用机器学习方法检测儿童抑郁症。PLoS ONE。2021; 16: e0261131。

---

11. Huang Y、Zhu C、Feng Y、Ji Y、Song J、Wang K等。预测中国青少年自杀念头和抑郁的三种机器学习模型的比较：一项横断面研究。J影响不和谐。2022;319:221–8。

---

12. Garcia-Argibay M、Zhang-James Y、Cortese S、Lichtenstein P、Larsson H、Faraone SV。预测儿童和青少年注意力缺陷/多动障碍的发病：一种全国性的深度学习方法。Mol精神病学。2023;28:1232–9。

---

13. Ter-Minassian L、Viani N、Wickersham A、Cross L、Stewart R、Velupilhai S等。使用对相关教育和医疗保健数据的回顾性队列研究，评估机器学习以公平预测学生多动症。BMJ开放。2022;12:e058058。

---

14. Toenders YJ、Kottaram A、Dinga R、Davey CG、Banaschewski T、Bokde ALW等。根据临床、认知、环境和神经生物学数据预测年轻人的抑郁症发病。Biol Psychiatry Cogn Neurosci神经成像。2022; 7: 376–84。

---
15. Chavanne AV、Paillere Martinot ML、Penttila J、Grimmer Y、Conrod P、Stringaris A等。青少年焦虑发作：机器学习预测。Mol精神病学。2023; 28: 639–46。

---
16. Zhang-James Y、Helminen EC、Liu J、Group E-AW、Franke B、Hoogman M等。青少年和成人注意力缺陷/多动障碍中类似结构性大脑异常的证据：机器学习分析。翻译精神病学。2021;11:82。

---
17. Alexander LM、Escalera J、Ai L、Andreotti C、Febre K、Mangone A等。儿科心理健康和学习障碍的跨诊断研究的开放资源。科学数据。2017;4:170181。

---
18. Bjork JM、Straub LK、Provost RG、Neale MC。神经发育的ABCD研究：确定预防和治疗青少年药物滥用的神经回路靶点。Curr治疗选择精神病学。2017;4:196–209。

---
19. Karcher NR, Barch DM。ABCD研究：了解身心健康结果风险的发展。神经精神药理学。2021; 46: 131–42。

---
20. Kuhn M, Johnson K.应用预测建模。纽约：斯普林格；2013年。

---



21. Zhang J、Li X、Li Y、Wang M、Huang B、Yao S等。使用结构核磁共振对行为障碍进行基于三维卷积神经网络的分类。脑成像行为。2020; 14: 2333–40。

---
22. 里普利BD。模式识别和神经网络。剑桥：剑桥大学出版社；1996年。

---
23. Russell S, Norvig P.人工智能：一种现代方法。第3版。新泽西州上马鞍河：皮尔逊教育；2010年。

---
24. Jager S, Allhorn A, Biessmann F. 数据归因方法的基准。前端大数据。2021;4:693674。

---
25. Dhillon IS, Sra S。带有布雷格曼发散的广义非负矩阵近似。在：第18届神经信息处理系统国际会议记录。2005年。p.283–90。

---
26. Tandon R., Sra S.稀疏非负矩阵近似：新公式和算法。马克斯·普朗克生物控制论研究所技术报告编号193.2010

---
27. Xu J、Wang Y、Xu X、Cheng KK、Raffery D、Dong J。基于NMF的方法，用于质谱代谢组学数据的缺失值归因。分子。2021;26:5787。

---
28. Power JD、Mitra A、Laumann TO、Snyder AZ、 Schlaggar BL、Petersen SE。在静止状态下检测、表征和消除运动伪影的方法fMRI。神经图像。2014;84:320–41。

---

29. Allen EA、Erhardt EB、Damaraju E、Gruner W、Segall JM、Silva RF等。静止状态网络多变量比较的基线。前部系统神经系统。2011; 5: 2。
- 
30. Allen EA、Damaraju E、Plis SM、Erhardt EB、Eichele T、Calhoun VD。跟踪静止状态下的全脑连接动态。小脑皮层。2014;24:663–76。
- 
31. Damaraju E、Allen EA、Belger A、Ford JM、McEwen S、Mathalon DH等。动态功能连接分析揭示了精神分裂症中断开的瞬态状态。神经图像临床。2014;5:298–308。
- 
32. de Lacy N、Doherty D、King BH、Rachakonda S、Calhoun VD。控制网络功能的中断与更广泛的自闭症谱系中动态连接的改变相关。神经图像临床。2017;15:513–24。
- 
33. Kaufman J、Birmaher B、Brent D、Rao U、Flynn C、Moreci P等。学龄儿童情感障碍和精神分裂症时间表-在场和终身版本（K-SADS-PL）：初始可靠性和有效性数据。J Am Acad儿童儿童精神病学。1997; 36: 980–8。
- 
34. Townsend L、Kobak K、Kearney C、Milham M、Andreotti C、Escalera J等。开发情感障碍和精神分裂症儿童精神病学诊断访谈儿童时间表的三个基于网络的计算机化版本：初步有效性数据。J Am Acad儿童儿童精神病学。2020;59:309–25。
- 
35. 金马DP, Ba J。Adam：一种随机优化的方法。arXiv:14126980v9 [预印本] 2017。可从<https://arxiv.org/abs/1412.6980>获取。
-

36. Chen T, Guestrin C. XGBoost: 一个可扩展的树提升系统。arXiv: 1603.02754v3 [预印本] 2016。可从<https://arxiv.org/abs/1603.02754>获取。
- 
37. 克莱因伯格·E。一种用于模式识别的抗过度训练的随机建模方法。安·斯塔特。1996;24:2319-49。
- 
38. Zou H, Hastie T. 通过弹性网进行正则化和变量选择。J R Stat Soc Ser B (Stat Methodol) 。2005;67:301-20。
- 
39. de Lacy N, Ramshaw M, Kutz JN。综合进化学习：一种人工智能方法，用于联合学习特征和超参数，以进行优化、可解释的机器学习。正面Artif Intell。2022;5:832530。
- 
40. Goodfellow I、Bengio Y、Courville A.深度学习。剑桥：麻省理工学院出版社；2016年。p.775。
- 
41. 主教CM。模式识别和机器学习。纽约：斯普林格；2006年。p.738。
- 
42. Bouthillier X, Varoquaux G. NeurIPS 2019和ICLR 2020的机器学习实验方法调查。2020年。<https://hal.science/hal-02447823>。
- 
43. Lundberg SM, Lee S.解释模型预测的统一方法。编辑：Guyon I、Luxburg UV、Bengio S、Wallach H、Fergus R、Vishwanathan S等。神经信息处理系统的进步 30.库兰协会公司；2017年。p.4765-74。
-

44. 布雷曼·L。随机森林。马赫学习。2001; 45: 5–32。

---

45. Xiang Q、Chen K、Peng L、Luo J、Jiang J、Chen Y等。在青少年大脑认知发展（ABCD）研究中，使用机器学习方法预测儿童抑郁症状的轨迹。J影响不和谐。2022;310:162–71。

---

46. Maniruzzaman M, Shin J, Al Mehedi Hasan M。使用行为活动预测多动症儿童：机器学习分析。应用科学。2022;12:2737。

---

47. Menon SS, Krishnamurthy K.多模态综合深度学习预测儿童的破坏性行为障碍。前神经信息。2021;15:742807。

---

48. Ramos-Lima LF, Waikamp V, Antonelli-Salgado T, Passos IC, Freitas LHM。机器学习技术在创伤相关疾病中的使用：系统回顾。J Psychiatr Res。2020;121:159–72。

---

49. Saxe GN、Ma S、Ren J、Aliferis C。预测儿童创伤后应激障碍的机器学习方法：概念验证研究。BMC精神病学。2017;17:223。

---

50. Zhang J、Sami S、Meiser-Stedman R.暴露在创伤中的儿童和青少年的急性应激和创伤后应激障碍：计算预测和解释。J焦虑症。2022;92:102642。

---

51. Trickey D、Siddaway AP、Meiser-Stedman R、Serpell L、Field AP。对儿童和青少年创伤后应激障碍风险因素的荟萃分析。Clin Psychol Rev. 2012;32:122–38。

---

52. Marek S、Tervo-Clemmens B、Calabro FJ、Montez DF、Kay BP、Hatoum AS等。可重复的全脑关联研究需要数千人。自然。2022；603：654–60。

---

## 致谢

---

作者要感谢B. Ernesto Johnson在数字方面的帮助和儿童思维研究所在获取数据方面的帮助。NdL感谢美国国家心理健康研究所赠款编号K99MH118359的支持。JNK部分感谢国家科学基金会动态系统人工智能研究所赠款编号为2112085的支持。犹他大学高性能计算中心的支持和资源对此表示感谢。

## 作者信息

---

### 作者和隶属关系

亨茨曼心理健康研究所，盐湖城，犹他州，84103，美国

Nina de Lacy和Michael J.拉姆肖

犹他大学精神病学系，盐湖城，犹他州，84103，美国

Nina de Lacy和Michael J.拉姆肖

美国华盛顿州西雅图华盛顿大学精神病学和行为科学系

伊丽莎白·麦考利

美国华盛顿州西雅图华盛顿大学生物统计系

凯瑟琳·F·克尔

KSADS-COMP, LLC, 麦迪逊, WI, 美国

琼·考夫曼

美国华盛顿州西雅图华盛顿大学应用数学系

J. 内森·库茨

人工智能动态系统研究所，美国华盛顿州西雅图

J. 内森·库茨

## 贡献

NdL负责设计研究、进行分析和撰写论文。EMcC为论文的设计和写作做出了贡献。KK为设计特征选择算法和撰写论文做出了贡献。JK为目标指标的准备和论文的撰写做出了贡献。MR为分析和撰写论文做出了贡献。JNK为设计研究、进行分析和撰写论文做出了贡献。

## 通讯作者

给[Nina de Lacy](#)的通信。

## 道德声明

---

### 利益冲突

JK收到了辉瑞和大冢制药的咨询费，并在计算机管理的KSADS（KSADS-COMP, LLC）中拥有专有财务利益。所有其他作者都声明没有利益冲突。

## 其他信息

---

出版商注：Springer Nature对已出版地图和机构隶属关系中的管辖权主张保持中立。

## 补充信息

---

### [补充表1](#)

### [补充表2](#)

## [补充表3](#)

## [补充表4](#)

## [补充表5](#)

## [补充表6](#)

## 权利和许可

---

**开放获取** 本文根据知识共享署名4.0国际许可证获得许可，该许可证允许以任何媒介或格式使用、共享、改编、分发和复制，只要您向原作者和来源提供适当的信用，提供知识共享许可证的链接，并注明是否进行了更改。除非材料的信用行中另有说明，否则本文中的图像或其他第三方材料包含在文章的知识共享许可证中。如果材料未包含在文章的创作共用许可证中，并且您的预期用途不受法定法规允许或超出允许的使用范围，您将需要直接获得版权所有者的许可。要查看此许可证的副本，请访问 <http://creativecommons.org/licenses/by/4.0/>。

## [转载和许可](#)

## 关于这篇文章

---



## 引用这篇文章

de Lacy, N., Ramshaw, M.J., McCauley, E. 等。用人工智能预测主要青少年精神疾病的个别病例。 *翻译精神病学* 13, 314 (2023) 。 <https://doi.org/10.1038/s41398-023-02599-9>

### 收到

2022年6月3日

### 修订的

2023年9月10日

### 接受的

2023年9月20日

### 已发布

2023年10月10日

### 记录版本

2023年10月10日

## DOI

<https://doi.org/10.1038/s41398-023-02599-9>

## 共享这篇文章

与您共享以下链接的任何人都可以阅读此内容：

获取可共享的链接

由Springer Nature SharedIt内容共享计划提供

主题 [神经科学](#) • [精神障碍](#)

## 这篇文章被引用

### [人工智能在管理精神病住院患者中的作用：评估](#)

Md SamiunMoustaq, Karim Khan RonySadia ... Islam Nilima

*发现公共卫生* (2025)

### [多动症的个性化预测模型：系统回顾和元回归](#)

Gonzalo Salazar de PabloRaquel, IniestaSamuele ... Cortese

分子精神病学 (2024)

转化精神病学 (转化精神病学) | ISSN 2158-3188 (在线)