自闭症神经精神表型的数据科学与机器学习：综述与路线图

Peter Washington[1] 和 Dennis P. Wall[2]

[1]信息与计算机科学，夏威夷大学马诺阿分校[2]儿科学（系统医学）、生物医学数据科学、精神病学与行为科学，斯坦福大学医学院

摘要

自闭症谱系障碍（自闭症）是一种神经发育迟缓，至少影响每 44 名儿童中的 1 名。与许多神经系统疾病表型一样，其诊断特征可以观察到、能够长期追踪，并可通过适当的治疗和疗法加以管理，甚至消除。然而，自闭症及相关发育迟缓的诊断、治疗和纵向追踪流程中存在重大瓶颈，为开发新的数据科学解决方案创造了机会，以增强并改造现有工作流程，并让更多受影响家庭获得服务。众多研究实验室此前开展的多项工作，已推动自闭症儿童数字诊断和数字疗法取得长足进展。我们综述了运用数据科学量化自闭症行为的数字健康方法文献，并介绍了用于数字表型分析的病例对照研究和分类系统。随后，我们探讨了整合自闭症相关行为机器学习模型的数字诊断与治疗方法，包括转化应用必须解决的因素。最后，我们介绍自闭症数据科学领域面临的挑战和蕴藏的重要机遇。鉴于自闭症的异质性以及相关行为的复杂性，本综述所提供的见解也广泛适用于神经系统行为分析和数字精神病学。

引言

自闭症谱系障碍（自闭症）是一种复杂的神经精神疾病，可表现为多种表型表现。这些表现包括兴趣范围有限、社会交往缺陷、沟通延迟、无法表达和/或识别各种情绪、回避眼神接触，以及手部、头部和身体的特异性动作（图1）。

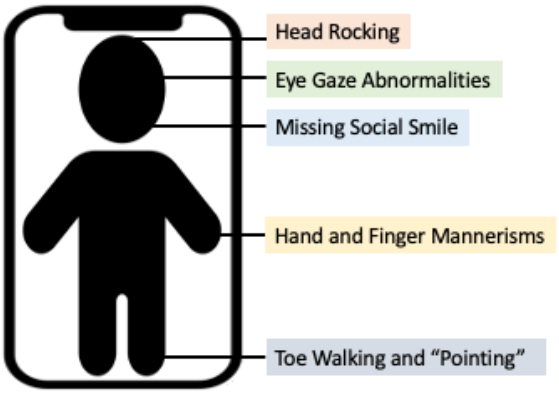

图1。一名2岁发育中儿童表现出自闭症迹象的部分可观察特征示例。这些特征可通过传统临床技术采集并录入结构化电子病历，也可通过智能手机上的游戏等替代方法采集。越是能够通过日益普及的设备被动地收集此类数据[3]，就越有可能捕捉到具有更高诊断精度的自然信号。这与在更为人工化的临床环境中捕捉到的信号形成对比，因为短暂性及其他因素可能导致误读或测量不准确。

据估计，美国每44名儿童中就有1名患有自闭症[1],，且患病率仍在持续上升。尽管发病率很高，据估计，美国83.86%的县缺乏获得诊断与治疗资源的渠道[2]。这种资源匮乏导致诊断时间延后至人生较晚阶段，往往造成较差的精神科结局[4-5]。

人工智能（AI），尤其是机器学习（ML），有潜力成为解决许多行为健康照护问题（如自闭症）的重要平衡力量。据皮尤研究中心称，美国97%的成年人拥有蜂窝设备，85%拥有智能手机；自2011 [6]以来，这一比例逐年上升。在墨西哥、委内瑞拉、哥伦比亚、南非、肯尼亚、印度、越南、菲律宾、突尼斯、约旦和黎巴嫩等新兴经济体中，大多数成年人都能使用某种移动电话，其中53%能够使用互联网和应用程序[7]。随着这些比例持续上升，支持互联网的设备日益普及，数字服务的获取可以在全球范围内实现民主化。尽管目前自闭症服务仅限于相对 privilegied 的人群，但由新兴数据科学方法论驱动的数字化解决方案可以使每个人都能获得孤独症治疗。

本综述聚焦于有助于定量理解神经精神表型的数字数据模态。自闭症行为数据科学领域涉及多种行为表现，范围广泛且内容丰富。自闭症表现具有异质性，因此需要从多个社会与沟通维度对自闭症进行量化，以便更好地理解自闭症表型与不同诊断队列之间的关系（图 2A），并开发用于自闭症分类的自动化方法（图 2B）。

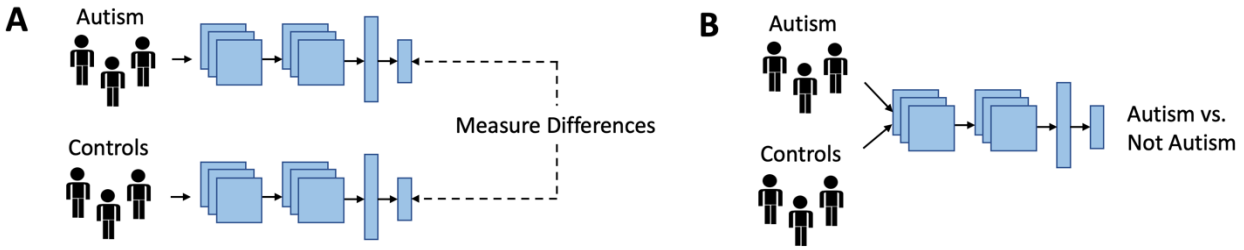

图 2。我们综述的两大类研究。(A) 病例对照研究（科学）：比较静态模型或计算分析方法在两个或更多组中的输出，通常是自闭症组与非自闭症组。(B) 机器学习研究（工程）：旨在训练模型，将自闭症与非自闭症分类。

自闭症数据科学领域极为广泛，不可能在一篇综述的篇幅内面面俱到。为保持重点，我们不详述所介绍研究所采用的模型和机器学习技术，不过会作简要提及。机器学习和数据科学的前沿技术不断演进，相较于具体使用了哪些模型，数据内容更为重要。只有在模型设计对于理解研究贡献至关重要时，我们才会例外讨论。

我们首先介绍与人类社会行为相关的数字表型分析项目中常用的多媒体数据模态，包括采集自然情境数据的常见方案。随后，我们讨论用于识别自闭症队列与非自闭症队列之间统计学显著差异的病例对照研究。接下来，我们综述利用机器学习自动分类自闭症的相关研究。最后，我们着重介绍该领域面临的挑战、创新者可借此产生重大影响的机遇领域，以及基于成功研究和应用于其他行为及神经精神表现的方法所发现的“唾手可得的成果”。

数据模态与采集

有多种数据类型可对用于数字表型分析的富含行为信息的数据进行编码（表1）。需要人工参与记录的数据，例如关于儿童行为的问卷，最不易出现计算错误，也最接近正式的临床诊断流程。另一方面，通过全自动程序获取的多媒体数据流可能包含细微却复杂的行为迹象，而使用预设提示的人类观察者可能会忽略这些迹象。然而，这类多媒体数据流更难成功转换为预测性嵌入。本文将讨论自闭症机器学习领域中使用的这两类行为数据模态。

模态	**数据采集技术**	**特征提取技术（非深度学习）**
问卷/测量工具 家长填写的问卷	（在网页或智能手机上），临床医生填写的问卷	转换为李克特量表
眼动	眼动追踪仪、网络摄像头、视频相机	注视热图、注视轨迹
情绪诱发	网络摄像头、摄像机	方向梯度直方图，面部关键点
受限和重复动作	网络摄像头、摄像机，运动传感器、深度传感器	骨骼姿态、光流、活动识别
特异性言语	录音设备	许多音频专属特征，例如俯仰、梅尔频率倒谱系数、时间形状，时间特征、频谱形状，etc.

表 1。病例对照研究和机器学习自闭症表型研究中常用的数据模态。随着深度学习模型在行为数据科学（以及更广泛的数据科学整体领域）中的应用日益普及，许多特征如今都通过神经网络自动学习，而不再需要明确预先指定。

临床表型分析最常见的数据来源之一是多项选择问卷，这些问卷包含用于诊断或量化儿童自闭症表现程度的临床评分表，或由此类评分表衍生而来。最常用的自闭症诊断评估工具包括自闭症诊断观察量表-2（ADOS）[8], 自闭症诊断访谈修订版（ADI-R）[9], 儿童自闭症评定量表（CARS）[10], 以及Gilliam自闭症评定量表-2（GARS-2）[11]。其他评分表则用于筛查或量化自闭症相关行为。可方便获取的数据集包括由临床医生或家长填写此类评分表的孤独症遗传资源交换库（AGRE）[12], 自闭症联盟（AC）、国家自闭症研究数据库（NDAR）[13], Simons单纯型样本集（SSC）[14], 西蒙斯个体变异项目（SVIP）[15], Autism Speaks（MSSNG）以及自闭症基因组项目（AGP）[16]。数据驱动研究要么直接分析这些评分表的集合，要么开发与之性质相似的独立问卷。该数据模态的主要优势在于：与高维多媒体数据相比，其搜索空间受到更多约束；并且由于特征提取过程需要人工参与，数据具有较高的可靠性。

富含多媒体流越来越多地应用于基于数据科学的表型分析工作。与自闭症相关、富含人类社会行为信息的常见数据流包括单幅图像 [17], 视频 [17], 眼动追踪 [17], 运动记录 [17], 文本 [18], 和音频 [19]。这些数据类型的主要优势在于能够编码丰富且细致的行为信息。与数据科学的许多其他应用一样，处理此类庞大且异构的数据所面临的挑战，是从噪声中发现信号。

大多数收集富含表型数据的自闭症数据科学研究都在受控实验室环境中采集数据。这种做法能够获得标准化数据，便于在受试者之间进行比较。本综述将介绍此类研究。然而，近期的创新集中于在自然环境中收集这些数据，而在采集同质化、结构化的数据流方面仍有很大的创新空间。例如，斯坦福大学的猜猜看项目提供了一种框架，可通过猜词游戏录制患有自闭症儿童的视频，从而采集富含社交行为信息的视频 [20]。在这个游戏中，家长或看护者将智能手机放在额头上，孩子则表演猜词提示。提示被正确猜出时，家长将手机向前倾斜；否则则向后倾斜 [21]。游戏提示结合手机倾斜动作的加速度计和陀螺仪日志，可提供结构化的带时间戳元数据，用作情绪、身体动作及其他行为的自动生成标签 [22]。研究表明，采集自猜猜看游戏环节、包含情绪游戏提示的视频富含情绪信息 [23]。研究人员对由此生成的视频帧进行了标注，其表现优于通用面部情绪识别分类器 [24-27]。尽管猜猜看最初是作为数据整理设备开发的，但可行性测试表明，有自闭症儿童的家庭居家参与后，社会反应量表（SRS-2）和 Vineland适应行为量表-II（VABS-II）得分有所改善 [28]。

富含表型的数据也可以从最初并非为数据采集目的而开发的数字疗法中获取。社会辅助机器人已被开发出来，可通过社会游戏以及在与儿童的社交互动中发挥主导作用来提供治疗，

从而引发儿童的社会行为 [29]。将与机器人的结构化互动和被动传感器的测量相结合，可以提供可用于诊断的客观测量结果 [30]。可穿戴设备也被探索用作治疗工具。*Superpower Glass* 项目是一种部署于可穿戴头戴设备上的教育工具，旨在为自闭症儿童提供实时面部情绪反馈 [31-33]。该设备为儿童及其父母带来了愉快的体验 [34-35], ，改善了儿童在社会反应量表（SRS）上的表现 2 [36-37], ，并最终在一项随机对照试验中使 VABS 社会化分量表得到显著改善 [38]。尽管对 *Superpower Glass* 系统收集的视频和眼动数据进行了评估，以考察其在自闭症与神经典型分类任务中的判别能力，但该分类器未能优于仅使用年龄和性别特征的分类器 [39], ，这可能是由于用户之间互动缺乏标准化。因此，必须格外谨慎地确保从交互系统获取的数据在不同用户之间经过校准和标准化。

正如我们将在下文各节中看到的，数据集规模和多样性是当前自闭症数据科学面临的最大局限之一。结构化数据收集、清理和标注流程的创新，可能有助于开展统计效能充足的研究，并构建稳健的诊断模型。

病例对照研究

我们介绍了运用数据科学方法量化自闭症队列与匹配的神经典型对照组之间差异的科学研究。这类实验旨在从多个群体收集等效数据，并确定是否存在统计学显著差异。

注视特征经常用于病例对照研究，因为回避眼神接触是有充分文献记载的自闭症行为。多项研究使用眼动追踪仪来测量儿童对结构化富含社交内容的视频场景的注意力。Jones 等人比较了59名自闭症高风险儿童（定义为自闭症儿童的同胞兄弟姐妹）与51名典型发育儿童在数月内的注视模式，发现后来被诊断为自闭症的儿童在2至6月龄期间的注视时间有所下降 [40]。Riby 等人比较了26名自闭症儿童和18名威廉姆斯综合征儿童；威廉姆斯综合征是一种罕见的发育障碍，其特征包括智力障碍、独特的面部特征和心血管问题。研究中的儿童观看卡通场景，同时研究人员测量了他们的注视。作者发现，自闭症儿童注视面部的时间少于典型儿童，而威廉姆斯综合征儿童注视面部的时间则长于典型水平 [41]。Chawarska 等人使用眼动追踪，比较了67名自闭症高风险儿童（自闭症儿童的同胞兄弟姐妹）与50名低风险儿童对社交场景的视觉注意，发现后来被诊断为自闭症的婴儿观看该场景的时间更短，注视场景中的女演员及其面部的时间也更少 [42]。Campbell 等人比较了22名自闭症幼儿与82名神经典型对照组对某一场景的视觉注意，发现自闭症参与者中仅有8%会在被呼唤姓名时转向，而神经典型参与者中这一比例为63% [43]。Sadria 等人将每个兴趣区视为一个节点，并将两个区域之间的每次扫视转换定义为一条边，从而对眼动追踪数据应用网络分析。他们发现，在涉及口部和右眼的区域，自闭症儿童与神经典型

对照组之间的度中心性存在统计学显著差异 [44]。Varma 等人突破了以往使用高度结构化互动和实验室环境收集数据的常规做法，利用在使用移动自闭症治疗应用期间收集的众包数据，通过网络分析测量不同队列之间的眼动差异，发现两组在单个兴趣区上存在统计学显著差异 [45]。Alvari 等人通过对62名自闭症儿童的数据进行无监督聚类，分析了不受约束的治疗师与儿童互动期间的眼神接触，并识别出由眼神接触动态定义的三个不同亚组 [46]。在一项规模最大的自闭症相关眼动追踪研究之一中，研究对象包括563名自闭症受试者和另外1,300名受试者；与其他儿童相比，自闭症儿童对动态几何图像的注视比例更高 [47]。在一项针对21名自闭症儿童和31名典型发育儿童进行的 Go/No-Go 游戏眼动追踪研究中，Putra 等人发现，自闭症儿童的视线调节更不稳定 [48]。

理解和唤起情绪的困难是自闭症儿童面临的另一项核心困难。因此，情感数据有助于将自闭症儿童与神经典型对照组区分开来。Guha 等人比较了20名高功能自闭症参与者与19名神经典型发育参与者的面部动态，发现自闭症组眼部区域动态的复杂性较低 [49]。在该领域一项规模异常庞大的研究中，Egger 等人比较了从1756个家庭通过基于 ResearchKit 的 iPhone 应用收集的视频中的情绪特征，发现与自闭症风险较低的儿童相比，自闭症风险较高的儿童在唤起中性情绪方面具有统计学显著性地更多，而唤起积极情绪方面则更少 [50]。

限制性和重复性的头部和身体动作是定义自闭症的另一项核心特征，这推动了相关研究对病例组和对照组在头部运动、身体运动及其他运动特征方面的量化差异进行分析。Martin 等人比较了有自闭症和无自闭症儿童的头部偏航、俯仰和翻滚情况（每组21名儿童），发现与神经典型对照组相比，自闭症儿童转头和倾斜的速度更快 [51]。Dawson 等人对104名幼儿的头部姿势控制进行了计算机视觉分析，其中22名被诊断为自闭症，结果发现自闭症儿童的头部运动速率是神经典型幼儿的2.22倍 [52]。

自闭症的另一项诊断标准是使用特异性言语，这催生了分析音频和文本特征的项目。Hudenko 等人在一项涉及15名自闭症儿童和30名典型发育儿童的研究中发现，自闭症儿童没有产生任何“无声”笑声；这类笑声大多没有音调，也无明显周期性，通常用于提供社会性肯定，而非表达积极情感。相比之下，典型发育儿童产生的笑声中有37–48%为“无声”笑声 [53]。Orlandi 等人在一项涉及12名对照组和2名高风险婴儿的研究中发现，高风险婴儿（已被诊断为自闭症的儿童的兄弟姐妹）与对照组的哭声音频平均基频在出生后10天、6周和12周时存在差异 [54]。

人类行为分析技术的进步必将带来新的分析方法，从而阐明新的行为差异。随着这些方法不断发展，将其与基因组数据和神经数据结合起来，有助于更精确地了解细微的生物学变化如何影响行为，并揭示自闭症的哪些方面是由生物学预先决定的，

哪些方面则受到社会和环境因素的影响。未来病例对照研究最大的机遇之一，是大幅扩大研究规模。考虑到人类的多样性，每个队列仅包含几十名受试者的研究都会存在诸多偏差。

自闭症及相关行为分类

监督式机器学习是数据驱动领域中应用最广泛的数据科学技术之一 [55]。自然地，通过训练模型从数据中进行预测这一框架，可以助力创建自动化诊断系统。自动化诊断很容易被构建为监督学习问题：将我们讨论过的某种模态的数据作为模型输入，并由模型输出诊断结果。本文讨论了遵循这一通用框架、针对自闭症及其相关行为开展的研究工作。

尽管借助当前技术可以自动提取与自闭症相关的行为（例如，通过眼动追踪识别注意力，以及通过面部情绪分类识别情感），但诊断人员用于自闭症分类的许多行为仍超出当前技术的能力范围。例如，“向他人表达愉悦”、“分享兴奋之情”和“参与社交”都是自闭症的诊断标准；根据应用于临床评分表的特征选择算法，这些标准均是重要预测因子 [56]。对于此类行为，机器学习及其他自动化方法尚不可行。近年来，基于机器学习的人在回路自闭症诊断方法应运而生。这些方法要么使用众包工作者等未经培训的人员 [57-59] ，将其作为视听数据的人工特征提取者；要么由临床专家和受过培训的视频分析员进行面对面的儿童观察，例如结合家长填写的问卷 [60-64] ，Cognoa 这家儿科数字健康公司便采用了后一种方式。获得人工提取的特征后，研究人员会将其输入机器学习模型。Tariq 等人使用 116 段自闭症儿童视频和 46 段典型发育儿童视频，取得了 92% 的曲线下面积得分 [65] 。随后，研究人员在来自孟加拉国的 50 名神经典型儿童和 50 名自闭症儿童身上复现了这一流程，曲线下面积达到 85% [66] ，表明该方法有望推广至多元化的全球人群。

构建诊断模型时，一项重要的设计考量是识别较小的特征子集，以尽量缩短诊断所需时间。Wall 等人率先将特征选择方法应用于自闭症诊断评分表 [56]。随后，多项研究验证了使用机器学习特征降维方法，以算法方式从诊断和筛查工具中识别仅含 4 至 10 个问题的小型子集；仅凭这些子集即可预测自闭症 [67-69]。

将这些方法应用于现实世界诊断时，儿童隐私是另一个需要重点考虑的问题。Washington 等人将隐私保护变换应用于自闭症儿童的视频，包括通过逐帧人脸检测，用方框遮挡儿童的面部，对音频进行音高变换 [70], , 以及应用高斯模糊、像素化和光流等全局图像变换 [71]。这些机制均使人在回路模型在受试者工作特征曲线下面积和精确率-召回率曲线下面积等指标上的性能下降不到 7%。未来仍有充足空间探索精度更高的隐私保护机制。

尽管我们推测，这些方法能够更广泛地帮助包括注意缺陷多动障碍（ADHD）、焦虑症、抑郁症和其他发育迟缓在内的异质性行为障碍改善诊断实践，但这些人在回路诊断技术尚未应用于自闭症以外的行为障碍。接下来，我们将讨论利用此类丰富多媒体数据流进行分类。尽管（截至 2023 年）完全自动化的机器学习方法尚不能支持构成自闭症特征的多种复杂社交行为（例如：“孩子喜欢参与社交游戏和互动吗？”），但将机器学习方法应用于自闭症诊断的原始数据流，能够快速、客观地捕捉细微线索和表征。我们将讨论这些方法；它们利用了与病例对照研究中许多相同的模态。

与病例对照研究一样，注视模式是诊断自闭症的常见特征。Liu 等人利用面部扫描模式，在一个包含29名自闭症儿童、29名按年龄匹配的神经典型儿童以及29名按智商匹配的神经典型儿童的数据集上，以88.5%的准确率预测自闭症 [72]。Duan 等人创建了一个包含14名自闭症儿童和14名对照组的注视点图和扫描路径数据集 [73], ，此类数据集有助于支持自闭症预测模型的开发。Chang 等人使用决策树分类器，根据40名自闭症儿童和936名正常发育对照组观看在 iPhone 或 iPad 上播放的高度结构化且经过策略性设计的电影时的数据，以曲线下面积为

0 9 预测自闭

[74]

症。. Oliveira 等人利用眼动追踪仪获取的注视固定图，在一个包含76名自闭症受试者和30名正常发育受试者的数据集上预测自闭症，达到90%的精确率、69%的召回率和93%的特异度 [75]。

情感计算在自动化诊断方法中的应用不如在病例对照研究中普遍，不过已有一些初步研究。Drimalla 等人开发了一种工具，用于与儿童进行模拟社交互动，并记录受试者互动过程中的眼动、语音和面部表情 [76]。在一项纳入 37 名自闭症成人和 43 名神经典型对照组的研究中，Drimalla 等人利用从模拟社交互动中采集的面部表情和语音分析数据检测自闭症，准确率为 73%，特异度为 79%，敏感度为 67% [77]。

姿势与运动特征在区分自闭症病例和对照组方面相当成功。Khosla 等人利用正面图像中儿童眼睛、鼻子和嘴唇的排列，以 87% 的分类准确率对自闭症进行分类 [78]。Hashemi 等人通过在儿童的眼睛、耳朵和鼻子上使用边界框，追踪自闭症儿童的头部运动 [79]。Lidstone 等人使用微软 Kinect 的三维深度摄像机区分自闭症儿童和神经典型对照组；在包含 23 名自闭症儿童和 17 名神经典型对照组成员的样本中，受试者工作特征曲线下面积达到 0.94 [80]。Kojovic 等人使用卷积神经网络从姿势图像中提取视觉特征，并将这些特征输入长短期记忆网络，在由 68 名自闭症儿童和 68 名神经典型对照组成员组成的样本上完成自闭症二分类诊断任务，F1 分数达到 0.89 [81]。Cook 等人提出了一个新的数据集，其中的视频片段包含自闭症相关的 35 种非典型运动和 33 种典型运动的示例。他们使用该数据集训练决策树，输入特征包括从 OpenPose 提取的骨骼关键点，以及用于编码身体动作随时间变化的时间速度；各折的平均 F1 分数达到 0.71 [82]。Li 等人利用从 16 名自闭症儿童及其匹配对照组的手部动作模仿中提取的运动学特征对自闭症进行分类，总体准确率达到 70.5% [83]。

Anzulewicz 等人测量了 37 名自闭症儿童和 45 名典型发育儿童的平板触控及手势运动学，并利用这些特征以 93% 的准确率测量自闭症 [84]。Cavallo 等人根据 20 名自闭症儿童和 20 名神经典型儿童完成结构化抓握任务时的手部运动数据进行分类，准确率达到 83%，敏感度为 80%，特异度为 85% [85]。

音频和文本特征正越来越多地用于自闭症的机器学习诊断。Li 等人构建了一个分类器，用于预测与自闭症相关的非典型韵律和刻板的特异性言语；在包含接受 ADOS第2模块评估的118名儿童以及接受 ADOS第2模块评估的71名儿童的数据集中，其准确率分别为88.1%和77.8%。Chi 使用来自一款移动自闭症治疗游戏的众包视频，视频对象为20名自闭症儿童和38名神经典型儿童；通过卷积神经网络从频谱图图像进行预测，以79%的平衡准确率对自闭症进行分类 [87]。Lau 等人使用与韵律的节奏和音调方面相关的声学特征进行自闭症分类；在一个包含55名英语使用者自闭症患者、39名英语使用者对照组，以及28名粤语使用者自闭症患者和24名粤语使用者对照组的数据集中，使用节奏相关特征时曲线下面积达到0.90，使用语调相关特征时达到0.695 [88]。Maenner 等人利用发育评估中的词语和短语构建了一个自闭症随机森林分类器；在包含1,450名儿童的数据集中，其中754名符合自闭症诊断标准，该分类器的敏感度为84%，阳性预测值为89.4%，受试者工作特征曲线下面积为0.932 [89]。

近年来，人们开始结合多种数据模态来预测自闭症。Vabalas 等人结合眼动数据和运动数据，根据 22 名自闭症儿童和 22 名神经典型对照组儿童的数据预测诊断；使用两种数据模态时准确率达到 78%，仅使用运动特征时为 73%，仅使用眼动特征时为 70% [90]。Javed 等人利用儿童与机器人互动过程中记录的面部表情和上半身运动，预测 5 名自闭症儿童和 7 名典型发育儿童是否患有自闭症，其准确率、精确率和召回率均接近 90% [91]。

随着机器学习方法持续快速进步，且数据收集系统日益普及，自闭症自动化诊断必将不断改进。随着《精神障碍诊断与统计手册》的后续版本不断修订，自闭症的定义也会持续演变，因此分类器可能需要持续重新校准。我们预计，未来快速且免费使用的计算机辅助诊断将在全球范围内广泛普及。

诊断模型的转化

人工智能在自闭症（及其他疾病）可观察表型的自动评估和测量领域取得了蓬勃发展，有望在不久的将来转化为实用的临床和教育解决方案，以支持诊断、追踪和治疗。实现这一目标，需要审慎考虑设计、开展可行性测试，并采用有意义的终点指标检验临床有效性。本节仅关注已通过监管审查并获得市场准入许可的诊断和治疗工具。

自闭症诊断工具已开始进入市场。Cognoa [60-64] 拥有一款经美国食品药品监督管理局批准的数字诊断工具 Canvas Dx，该工具利用人工智能为临床决策提供支持，用于识别和排除18至72个月儿童的自闭症。这

款诊断设备适用于通过智能手机远程使用，也适用于初级医疗场景和更专科的场景。另一款诊断工具 Earlitec [40, 92-98], 采用单一的眼动追踪模态，为16至30个月儿童提供自闭症诊断支持，并获准专门用于专科临床场景。

已有多位学者指出，在医学人工智能中管理不确定性的重要性 [99-101]。由于对发育中儿童进行人工智能辅助诊断时需要精确测量，而短暂信号普遍存在且属于预期情况，因此管理不确定性尤为重要。Cognoa的canvasDx [60-64] 就是此类工具之一。CanvasDx会输出自闭症概率评分，但只有当该概率分别超过上限阈值或低于下限阈值时，才会作出“自闭症”阳性判定或“无自闭症”阴性判定。否则，分类器会放弃分类，表明模型的置信度不足以作出预测。弃权功能使模型在信息不足以作出任一方向判定时，能够避免作出诊断判定。这是一项重要的安全措施，但也可以根据新信息进行调整和优化，例如依据美国食品药品监督管理局近期关于纳入预定变更控制计划的指导意见进行调整。该指导意见已获准用于包括CanvasDx、美敦力、Caption Guidance和苹果房颤检测系统在内的全新及衍生设备。我们预计，美国食品药品监督管理局将要求所有使用人工智能进行诊断的医疗器械具备弃权功能，并且许多设备将纳入算法修改计划，以便结合真实世界使用数据，按照良好的机器学习实践进行审慎调校。

评估这些能够弃权、不作出预测的诊断模型时，需要审慎考虑评估指标。我们认为，阳性预测值（PPV）和阴性预测值（NPV）是评估能够弃权的模型的最佳指标，因为它们仅考虑模型作出的阳性和阴性判定，而敏感度和特异度等指标则会考虑所有阳性和阴性病例，无论分类器是否对这些病例作出判定。因此，我们建议将阳性预测值（PPV）和阴性预测值（NPV）作为诊断分类器的主要临床终点。

诊断系统的另一个关键终点是模型公平性。医疗机器学习系统的一个常见问题是，其在不同人口群体中的性能水平存在差异 [102-103]。近年来的机器学习研究提出了多种定量公平性指标，包括人口统计均等、均等机会率、机会均等、测试公平性和反事实公平性等 [103]。尽管该领域尚未标准化任何公平性指标，但我们认为，未来报告人口统计均等等标准指标以及标准性能指标将十分重要。大规模众包数据收集工作，例如西蒙斯基金会自闭症知识研究计划（SPARK）[104] ，以及可远程免费使用的智能手机应用程序 [20-23] ，都是实现公平、无偏数据收集的重要举措。

数字疗法模型的转化

利用机器学习改善自闭症治疗服务提供及其质量，以及行为健康其他方面的数字化解决方案，具有巨大的发展前景。由于人工智能驱动的数字疗法是一个新兴的研究领域，人们对于患者

群体在可信度、可用性和可理解性等多个维度上对这些疗法的反应仍存在不确定性。

我们介绍两种用于儿童自闭症的人工智能赋能解决方案，它们都沿着这一路径实现了转化。SuperpowerGlass 系统就是这样一种治疗方式，通过 Google Glass 实施。在使用 SuperpowerGlass 进行治疗时，家长与自闭症儿童互动，而儿童佩戴 Glass 系统。该系统通过计算机视觉面部情绪检测系统，向儿童实时提供情绪提示。SuperpowerGlass 通过迭代式设计过程开发并不断完善。在早期原型设计阶段，研究人员通过与自闭症儿童及其家长开展协同设计会议，借助故事板设计和低保真原型设计，探索了可能的用户界面设计 [32-35]。在初步可行性测试之前进行了这些改动 [36-37], ，并据此进一步完善系统，直至正式开展随机对照临床试验 [38]。鉴于人工智能驱动的解决方案具有不确定性和新颖性，对于数字健康领域的新型治疗模式而言，这一过程至关重要。

Akili 推出的另一种治疗方式 EndeavorRx，是首个基于游戏的数字疗法，也是首批通过监管审查的数字疗法之一 [105]。该系统设计精密，内置软件模型，可供患有注意缺陷多动障碍（ADHD）的儿童独立进行游戏。注意缺陷多动障碍与自闭症在症状上存在诸多重叠，且常与自闭症共病。EndeavorRx 和 SuperpowerGlass 均已成功推向市场，并获得美国食品药品监督管理局授予的“突破性疗法”资格。

精神病学和行为科学领域的数字应用传统上采用“放之四海而皆准”的解决方案。然而，整合本综述中讨论的机器学习模型，可以实现自适应数字工具，分析用户的行为并相应地重新调整。自闭症症状的测量指标可以嵌入游戏化疗法中，使人工智能能够实时了解儿童对特定治疗“剂量”的反应。了解儿童的剂量-反应曲线后，数字化解决方案便可相应增加或减少治疗剂量。

挑战与机遇

自闭症行为数据科学是计算行为表型分析的一个特例。因此，随着计算行为分析工具变得更加先进、精细和稳健，这一领域很可能会持续发展。同样，自闭症数据科学也能受益于针对其他精神疾病开展的行为数据科学研究所取得的进展。我们简要探讨神经精神病学数字行为表型分析领域面临的挑战与机遇。我们讨论的许多机遇都涉及将应用于其他疾病的数据科学创新重新用于自闭症；而自闭症数据科学面临的许多挑战，也同样存在于更广泛的复杂人类行为分析数据科学领域。

自闭症数据科学领域最重要的难题之一，是研究中缺乏大型标准化数据集。不同论文报告的机器学习诊断系统的分类表现差异很大，但目前尚不清楚这些表现差异是否源于数据集本身存在的差异。因此，亟需建立标准化数据集。同样，病例对照研究和机器学习分析的样本量通常都很小，因为每组往往只有几十名儿童。

这些研究报告的数字限制了跨研究比较的能力，即使研究条件相同也是如此。我们迫切需要规模更大、更多样化的研究，因为这将有助于提高研究结果的普适性。众包领域特定数据采集平台 [20- 23] 有望帮助达到所需的样本量。然而，与受控实验室环境不同，这类系统面临数据一致性问题，导致严重的过拟合。模型个性化则是一种能够规避这一问题的替代方法：针对每个受试者的独特特征以及模型将要部署的环境，对模型进行优化。自闭症数据科学领域已发表了一些初步研究 [106] ，表明针对受试者对机器学习分类器进行个性化能够提升性能。这些方法需要增加儿童与受试者之间的互动，从而为新型人机交互创造机会。

自监督学习（SSL）是从无标签数据中自动学习特征表示的方法。当应用于单个用户的数据流时，SSL有助于加快模型个性化。Yu and Sano将SSL应用于可穿戴传感器数据进行压力检测，在3个独立数据集上，相较于监督学习基线，性能提升了7.7%至13.8% [107]。

区分自闭症与相关疾病是该领域面临的新兴挑战。越来越多来自自闭症遗传学、神经科学和精神病学的文献表明，自闭症并非由单一谱系构成，而是包含相互独立的亚型。已有一些研究尝试基于行为数据模态对自闭症进行定量分型，而这类研究有望验证并补充遗传学、神经科学和精神病学领域的相关工作。例如，Gardner-Hoag等人分析了854名自闭症儿童，并识别出7个不同的自闭症行为聚类，其中每个聚类都代表一种占主导地位的原型行为特征 [108]。

虽然此处报告的大多数分析都区分了是否患有自闭症，或是否为神经典型者，但迄今为止，很少有研究尝试区分症状重叠的多种精神疾病。如果已怀疑某种疾病，二元分类器会很有用；但理想情况下，应能将自闭症与注意缺陷多动障碍、精神分裂症、焦虑症、抑郁症和言语发育迟缓等状况区分开来。在这方面已有一些初步研究。Duda等人利用社会反应量表中的15个问题构建了一个分类器，用于区分注意缺陷多动障碍（N=174）和自闭症（N=248），受试者工作特征曲线下面积达到0.89 [109]。Wawer等人使用文本话语分别对自闭症（N=37）和精神分裂症（N=37）进行分类，发现对精神分裂症进行模型预训练后，再通过迁移学习预测自闭症，模型表现有所提升 [110]。Demetriou等人使用一组针对不同精神疾病设计的测试，对自闭症（N=62）、早期精神病（N=48）和社交焦虑障碍（N=83）进行分类，结果表明，整合针对不同重叠状况的输入可能是一种有效方法 [111]。Iakovidou等人利用可穿戴传感器采集的加速度计、血容量脉搏和皮电活动数据，在一组包含10名自闭症儿童和10名雷特综合征儿童的样本中，以95%的准确率区分了自闭症与雷特综合征 [112]。由于神经精神状况往往存在共病，未来的研究很可能需要将问题构建为多类别分类问题或类似的变体。

除了病例队列分析和机器学习诊断系统，数据科学方法还有望实现自动化的纵向结局追踪和治疗反应测量。该领域已有一些初步研究，但仍有大量工作机会；进一步研究将有助于自闭症症状的数字化干预和治疗。McKernan等人使用自动语音处理方法测量儿童回应社交伙伴所需的时间，发现接受干预的儿童其延迟时间降幅大于未接受干预的儿童（每组105名儿童）[113]。Kolakowska 等人在游戏过程中测量平板电脑传感器的数据，发现这些传感器可用于预测40名自闭症儿童的治疗进展，准确率超过80% [114]。

自闭症研究涵盖多个学科领域，目前已有大量使用基因组学[115-118], 表观基因组学[119], 蛋白质组学[120-123], 宏基因组学[124-126], 和脑成像[127-130]开展的自闭症数据科学研究。虽然我们十分欣赏这些成果，但这些研究超出了本综述的范围。不过，我们注意到，生物学相关领域的研究工作未来有机会与本文所述类型的分析相融合，因为利用数字表型分析方法获得日益细致的行为表型，很可能有助于开展更稳健、更精确的生物学关联研究。我们还注意到，我们综述的许多数据科学方法可以更广泛地应用于神经系统疾病、发育障碍和行为障碍。

我们诚邀所有读者加入这一充满希望、蓬勃发展的领域。我们期待在未来二十年及更长时期内，医疗服对复杂的人类社会行为的科学认知必将不断提升。

参考文献

1. Maenner MJ、Shaw KA、Bakian AV、Bilder DA、Durkin MS等。2021年。2018年美国11个监 监测 e 摘要。70(11):1. 2. Ning M, Daniels J, Schwartz J, Dunlap K, Washington P, 等. 2019. 美国孤独症 21(7 DP. 2016. 量化大脑：一种基于移动设备评 Mar 与发育障 6(10):3281-94. 5. Fenske EC, Zalenski S, Krantz PJ, McClannahan LE. 1985年综合 儿研究中 http Tay 主按。反光研究中心。7. 8. Lord C, Risi S, Lan —通用版：一项用于衡量与孤独症谱系相关的社交与沟通缺陷的标准工具。《自闭症与发育障碍杂志》。30(3):205-23. 9. Rut 29(2 Marcus LM, Schopler E. 1992年. DSM-III-R与儿童孤独症评定 E. 199 C, Iver ack J. 等. 2001年. 孤独症遗传资源交换库：一项用于研究孤独症及相关神经精神状况的 (2):463-66. 13. Hall D, Huerta MF, McAuliffe MJ, Farl Fisc 系。《神 :192-95. 15. 西蒙斯VIP联盟。2012年. 西蒙斯个体差异项目（Simons VIP）：以遗传 》。73(6):1063-67. 16. Hu-Lince D, C 学杂志》。5(4) avero D. 2020. 计算机视觉在孤独症谱系障 18. 挖掘在精神病学中的应用：系统性文献 100. 19. Pusaroli R, Lambrecht A, Bang D, B 研究

20. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, 等. 2019. 猜猜看？《医疗保健信息 H, Jedoui K, Washington P, Wall DP. 2018. 用于图像情 Kalantarian H, Jedoui K, Wa 力。《医 等. 2 8. 用于众包视频以开展自闭症研究的游戏化移动系统。IEEE国际医疗保健信息学会议（ICHI）。50-3 能手机自 有绪检测分类器。medRxiv。25. Kalantarian H, Jedoui K, Dunlap K, S 量可行性 ent J, Husic A, K 儿科与育 ashington P, Kalantarian H, Kent J, Husic A, Kline A, 等. 2020. 使用来自 K, H 可行 -40. 29. Scassellati B, Admoni H, Matarić M. 2012年. 机器 如何帮助我们诊断、治疗并理解自闭症。《机器人研究》。552-63. 31. Haber N, Voss C, Wal 《IEE , 等. 201 etMobile：移动计算与通信》。23(2):35-38. 33. Voss C, Wa iels J, 等. 2016. Superpower Glass：在可穿戴系统中提供隐蔽的 Hab , 等. 2016. 面向孤独症儿童的可穿戴社交互动辅助设备。CHI计算系统中的 . 201 穿戴与 N, Voss C, Schwartz J, Tamura S, 等. 201 9(01 童社会情 与青少年精神病学会杂志》。56(10):S257. 38. Vos 童社 随机临床试验。《JAMA儿科学》。173(5):446-54.

39. Nag A, Haber N, Voss C, Tamura S, Daniels J, 等. 2020. 迈向连续社会表型分析：通过可穿戴智
能眼 10.
40. 注意力会
逐渐 Hancock PJ. 2009. 观看电影和下通片：来自威
廉姆
Mac 发注意减少。《生
物精 S,
等. 2 觉分析捕捉到孤独症幼儿的非典型注意。《自闭症》。23(3):619-28. 44. Sadria
M, I in
biol dline A,
Leb 与视
觉扫 lvari G, Coviello L
Fur 稳健
的眼 ri J,
Xiao 型早期眼动追踪生物标志物的大规模验证。《科学报告》。
12(1
中的 》。
Gro
9(1)
中的 学》。
1(1) B, Hanlihan Z, Kerr G, Conn JF, Cassell J, 等. 2018. 对自闭症谱系障碍儿童与非
自闭
Can 孤独症谱系
障碍
A. 2
障碍
儿哭
295 Jordan MI. Mitchell TM. 2015. 机器学习：趋势、观点与前景。《科学》。
349
习缩 闭症筛查与诊断时间。《转化精神病学》。2(4):e100-. 57. Washington P,
Leb 众包工
作者

58. Washington P, Leblanc E, Dunlap K, Penev Y, Varma M, 等. 2020. 为儿童自闭症谱系障碍的远程医
59. J, Dunlap K, 等. 2019. 自闭症在线筛查的有效性
H, C 合问卷和家庭视频筛查，利用机器学习实现自闭症早期检测
201 数据国际会
利用 Daniels J, Duc e视频内容加速自闭症早期检测的潜力。《P 系障碍风险
11(7 过对家
15(1 加拉国 的机器学习模型检测发育迟缓和自闭症：开发与验证研究 《医学互联网研究杂志》。
21(4 评估。《自 68. Levy S, Duda M, Haber N, Wall DP. 2017. 稀疏化机器
Wa 的特征
Chr 用于机器学习自闭症谱系
Klir 《智能医学 机器学 . 自闭症
Ma , Baker J, Carpenter K, 等. 2021. 测量孤独症谱系障碍幼儿注视模式的计算方法。《JA J, 等. 202 辅助自闭症诊断。《科学报告》。11(1):1-1. 76.
Dri 自闭症

77. Drimalla H, Scheffer T, Landwehr N, Baskow I, Roepke S, et al. 2020年. 探索孤独症谱系障碍社会
Y, R 面部图像和深度学习检测孤独症个体。IEEE计算系统与
可持 er A,
Mor 标志的计算机视觉方法。IEEE发展与学习及
表观
Vid 动模仿计
算机
Nat 儿
的孤 019年.
探索 童的典型与非典型行为进行自动筛查。IEEE数据科学与高级分析国际会议（DSAA）。
504 仿识
别孤 索性研究。《PLOS One》。12(8):e0182652. 84. Anzulewicz A, Sobota K,
Del 症儿童。
《科 识
别孤
Zhu 异性短语的自动
化框 ,
et a 斗与
育儿 语音
韵律 MJ,
Yea 监测孤
独症 。11(12):e0168224. 90. Vabalas A, Gowen E,
Poli 预测自闭症
诊断 rk CH. 2020年. 通过儿童—机器人交互进行基于行为
的孤
Ken 场景受到遗
传控 in
A. 2 动儿不偏好注视定近的成年人眼睛，这可预测其社交障碍程度。《普
通精 现
看自
59(9 A, Lin DJ, Gorrindo P, Ramsay G, Jones W. 2009年. 两岁孤独症幼儿会注意
非社

96. Rice K, Moriuchi JM, Jones W, Klin A. 2012年. 解析孤独症谱系障碍的异质性：学龄儿童对动态社会场景的[illegible]ones W. 201[illegible]科学院院刊》。108(52):21270-75. 98. Shu[illegible]趋势》。22(5[illegible]杂志》。47(5[illegible]as D, Meert W, Davis J. 2021年. 带拒绝选项的机器学习[illegible]传达不确定性。《[illegible]1.6. 102. Char DS, Shah NH, Magnus D. 2018年. 在医疗保健中实施机器学习[illegible]三医学杂志》。378(11):981. 103. Menrabi N, Morstatter F, Saxena N, Lerman [illegible]54(6[illegible]ont A, Camba A, et al. 2018. SPARK：一个由50.000个[illegible] SH, DeLoss DJ, [illegible]严重程度的新型数字干预（STAR[illegible]78. 106. Rudovic O, Lee J, Dai M, Schu[illegible]。《科学·机[illegible]境中进行瞬时压力检测的半监督学习[illegible]s E, Dixon D, e[illegible]以体系基于聚类的孤独症谱系障碍儿童治疗效果：一项回顾[illegible]els J, Wal[illegible]。《转化精神病学》。7(5):[illegible]awer A, Choinicka J, Okruszek L, Sarzynska-Wawer J. 2022年. 自闭症与精神分裂症的单一与跨[illegible]Demetriou EA, Park SH, Ho N, Pepper KL, Song YJ, et [illegible]精神病和社交焦虑障[illegible]P. 2020年. 利用[illegible]学杂志》。9(9):2842. [illegible]rtino A, Shulman L, Kolevzon A, et al. 2022年. 主题内潜伏期作为[illegible]akowska A, Landowska [illegible]7(1)[illegible]组学：拥抱复杂性。《人类分子遗传学》[illegible]、智力障碍与[illegible]

117. Vorstman JA, Parr JR, Moreno-De-Luca D, Anney RJ, Nurnberger Jr JI, 等. 2017年。自闭症遗传学:临床
Wang B, Su W. 2022年。基因组学、超同神经科学与理解自闭症谱系障碍的进展。Nature Reviews Neuroscien
型传子：综述。Psy
clinical neurosciences。72(4):228-44。120. Abraham J, Szoko N, Natowicz MR. 2019。自闭症谱系障
dies in Psychiatric
ders。235-52。121. Anderson GM. 2015。自闭症生物标志物:挑战、陷阱
2013。蛋白
cians Surg. Pak。137-43。123. Ristori MV, Mortera SL
症谱系障碍功能
ational journal of molecular sciences。21(17):6274
肠道微生物组与自闭症谱系障碍。Nature Reviews Gastroente
肠道微生物组:自闭
A, Grace T. 2
atric and Neurodege
tive Disorders。253-69。127. Khodatars M, Shoeibi A, Sadeghi D, Ghaasemi N, Jafari M, 等
关于神经影像的诊断和康复的深度学习：综述。Computers in Biology and
19。机器学习算
象研究的系统综述与荟萃分析。JMIR mental health。6(12):e1410
器学习（机
学习）。Reviews in the Neurosciences。31(8):825-41。130. Paglioni AM, Conti E, Calderoni S, Fripp J, F
Inte