Periodica Polytechnica 电气工程与计算机科学

天气悖论：为何降水无法预测大规模美国数据中的交通事故严重程度

Yann Bellec[3], Rohan KAMAN[1], Siwen CUI[1], Aarav AGRAWAL[2], Calvin CHEN[1]

[1] 加利福尼亚大学圣迭戈分校计算机科学系，9500 Gilman Dr，拉霍亚，CA 92093[2] 加利福尼亚大学圣迭戈分校数据科学系，9500 Gilman Dr，拉霍亚，CA 92093[3] 加利福尼亚大学圣迭戈分校神经科学系，9500 Gilman Dr，拉霍亚，CA 92093* 通讯作者，电子邮箱：rkaman@ucsd.edu，s4cui@ucsd.edu，cac028@ucsd.edu，a7agrawal@ucsd.edu，ybellec@ucsd.edu

摘要

本研究考察环境、时间和空间因素对美国交通事故严重程度的预测能力。我们使用涵盖2016–2023年的50万起美国交通事故数据集，训练了一个经过随机搜索交叉验证优化、并通过类别加权调整类别不平衡的XGBoost分类器。最终模型的总体准确率达到78%，在多数类（严重程度 2）上的表现出色，精确率和召回率均为87%。特征重要性分析显示，一天中的时段、地理位置以及天气相关变量（包括能见度、温度和风速）均是事故严重程度最有力的预测因素之一。然而，与最初的假设相反，降水和能见度的预测能力有限，这可能反映了驾驶员在明显危险条件下的行为适应。数据集中中等严重程度事故占多数，限制了模型学习极端案例有效模式的能力，这凸显了采用替代抽样策略、加强特征工程以及整合外部数据集的必要性。这些发现有助于推进循证交通管理，并为严重程度预测研究指明未来方向。

关键词

交通事故严重程度，机器学习，XGBoost分类器，天气预测，环境因素，时间模式，空间分析，类别不平衡，道路安全，特征重要性，能见度

1.1 背景与意义

交通事故是全球公共卫生与安全领域的一项重大关切。2016年，全球约有135万人因此死亡，也是5至29岁人群的首要死因 [1]。除了给事故相关人员造成直接的身体和经济损害外，交通事故还会严重扰乱交通网络、延长通勤时间，并带来更广泛的社会成本。因此，了解导致事故发生及其严重程度的条件，对于改善公共福祉、尽可能减少对日常生活的干扰至关重要。鉴于交通事故严重程度预测对公共安全和交通运输管理的影响，这一领域已受到广泛关注 [11, 13]。

1.2 相关研究

近期研究强调了地理背景在事故预测中的重要性。Zhi（2025年）运用聚焦于加利福尼亚州的可解释机器学习模型，证明风险因素因地点而异，且特定条件的重要性在很大程度上取决于地理背景[2]。这一发现凸显了风险因素的异质性，以及广义假设在安全建模中的局限性。

方法学的进步也致力于应对不平衡数据集带来的挑战。发表在 Scientific Reports 上的研究开发了用于处理事故数据中固有不平衡问题的神经网络方法；在这类数据中，严重事故的发生频率远低于轻微事故 [3]。这些研究表明，天气和能见度是影响严重事故的重要因素，同时也凸显了建模交通结果时了解数据分布的关键必要性。

约旦开展的一项补充研究运用地理信息系统、核密度估计和结合贝叶斯超参数优化的随机森林分类，对2016年至2021年间的177,378起事故进行了分析，准确率达到90% [4]。值得注意的是，该研究发现，时间因素是界定严重程度的第三大重要变量；事故高峰出现在星期一和星期五、六月和八月，以及尤其是下午2至4点这一时段。作者将这些规律归因于警觉性下降、疲劳以及交通流量增加。重要的是，研究还发现，恶劣天气条件与事故严重程度之间存在一种违背直觉的负相关，这可以用以下因素解释：

驾驶员在感知到风险时会更加谨慎并采取适应性行为。

来自伊朗的另一项研究采用层次网络分析法，对环境风险因素进行了排序，发现道路湿滑是最关键的环境子因素，其次是路面状况和车道类型 [5]。

1.3 研究问题与假设

本研究探讨以下研究问题：能否根据天气状况、能见度、降水和一天中的时段等环境、时间及空间因素，预测交通事故严重程度（以对交通流的干扰或影响程度来衡量）？此外，哪些因素对发生高严重程度事故的可能性影响最大?

我们假设环境、空间和时间状况，尤其是降水、能见度和一天中的时段，是交通事故严重程度的重要预测因素 [7,17, 18]。预计能见度差和降水量大将降低驾驶员感知危险因素的能力并延长反应时间，从而增加事故严重程度，使碰撞事故更难避免，并导致更大范围的交通中断 [19, 20]。

一天中的时段也预计会影响事故严重程度。据推测，发生在通勤高峰时段的事件影响更大，这是因为交通密度更高、机动空间有限且应急响应时间更长 [21，22]。相反，夜间或交通稀疏时段发生的事故可能涉及更高的行驶速度，但总体影响的车辆较少；根据道路类型和能见度不同，这类事故可能造成较短暂但更严重的交通中断 [23, 26]。

我们承认行为适应可能发挥作用：驾驶员在不利条件下提高警惕，从而减轻事故严重程度。此前关于恶劣天气下驾驶员适应行为的研究结论不一；有些证据表明，驾驶员会更加谨慎，从而降低事故发生可能性，但事故一旦发生，严重程度未必会因此减轻 [31,33]。通过考察不同条件下的这些关系，我们旨在检验观察到的规律是否始终成立，并确定最能预测高严重程度结果的因素组合[35, 36]。

2.1 数据集描述

本研究使用的数据集是 Moosavi 整理并托管于 Kaggle 的美国交通事故数据集（2016–2023）[6]。完整数据集包含超过 770 万条交通事故记录，涵盖 46 个变量，数据来自美国交通部、州交通机构和交通监测基础设施。为提高计算效率，我们使用了数据集整理者提供的 50 万条代表性样本，并保留全部 46 个原始变量。

每条观测记录都代表一起经核实的交通事件，并包含详细的环境、时间和空间信息。与我们的研究问题相关的主要变量包括：Severity（1，4），用于表示事故影响程度的分类指标，其中 1 表示对交通影响最小，4 表示影响最大；Start_Time 和 End_Time，用于捕捉时间模式；Start_Lat 和 Start_Lng，表示地理坐标；以及环境变量，包括 Temperature（°F）、Visibility（英里）、Precipitation（英寸）和 Wind_Speed（mph）。其他背景指标包括日出日落和风向。_ _该数据集覆盖全国且样本量庞大，因此能够分析长期时间模式；不过，抽样方法可能会放大数据固有缺失的影响。

2.2 变量操作化

严重程度以 1 至 4 的等级量化事故影响对交通的影响，并以交通延误时长 [41] 衡量。地理变量包括街道、城市、县、州、邮政编码和时区，为每个事故地点 [42, 44]提供空间背景。天气_时间戳表示天气观测的当地时间。

温度和风寒指数_以华氏度为单位，低于 40°F 表示寒冷条件，高于 90°F 表示炎热条件 [45，46]。湿度以百分比表示，反映当前温度下相对于最大容量的含水量；低于 30% 表示湿度较低，高于 80% 表示湿度较高 [47, 48]。气压以英寸汞柱为单位测量大气压，高于 29.8 通常表示天气较晴朗，且海平面处空气稳定。

能见度以英里为单位，表示普通人眼能够辨认物体的水平距离[49]。风向_使用标准缩写（N、S、E、W），风速_以英里/小时为单位，将状况划分为从平静（~0 mph）、大风（40、50 mph）到风暴（>50 mph）等等级。降水量以英寸为单位衡量降雨量，低于 0.1 英寸归为小雨，高于 0.3 英寸则归为大雨。一天中的时段根据日出和日落时间，通过 Sunrise_Sunset 进行分类；民用_曙暮光、航海_曙暮光和天文_曙暮光则根据太阳位置提供其他分类。

2.3 数据质量评估与预处理

初步评估发现，严重程度分布存在显著的类别不平衡。严重程度为 2 的记录占据数据集的绝大多数（397,538 条，占 79.62%），而严重程度为 4（13,040 条，占 2.61%）和 1（4,260 条，占 0.85%）的记录则明显不足。我们在模型训练期间通过类别加权处理了这种不平衡，从而保留了足够的数据以确保统计效力 [50, 52]。

对连续变量的检查发现了一些不合常理的异常值，表明可能存在数据录入错误，例如温度高达 207°F、风速高达 822.8 mph [53]。极端异常值很容易识别并剔除，但此类数值的存在也引发了对系统性记录错误的担忧 [54, 55]。为减少这种潜在偏倚，我们使用基于 IQR 的异常值过滤方法，根据合理的物理限值对数据进行了验证 [56, 58]。

降水（142,563 个缺失值）和风_速（36,958 个缺失值）存在大量缺失数据 [59, 60]。鉴于我们的假设将天气状况视为预测因素，妥善处理这些缺失记录至关重要 [61]。此外，数据集汇总了来自多个应用程序接口和机构的数据，因此不同来源的变量标准可能存在不一致 [62, 63], ，这可能导致严重程度报告存在地理偏差。

依据既定阈值对异常值进行了检查：温度 < −50°F（阿拉斯加冬季最低气温）、能见度 >90 英里（美国环境保护署国家公园能见度范围）、风速 >130 英里/小时（按萨菲尔–辛普森等级，达到 4 级飓风阈值），以及降水 > 4 英寸/观测记录（基于世界气象组织的分类）。通过 GPS 坐标与历史天气记录交叉核对证实了若干极端值，包括纽约市一条 −77.8°F 的读数、≥100 英里的能见度测量值，以及在没有相应风暴活动的情况下记录的 131–142 英里/小时风速，均属于需要删除的记录错误。

2.4 统计分析与建模方法

探索性数据分析包括单变量分析（通过直方图和箱线图检查分布，以评估类别不平衡并检测异常值 [64]）、双变量分析（探究天气特征与事故严重程度之间的相关性 [65, 66]），以及时间/类别分析（调查基于时间的模式和类别特征 [67, 68]）。

为进行预测建模，我们采用了加权 XGBoost 分类器来解决类别不平衡问题。我们应用类别权重，优先检测罕见的高严重程度事故，而非最大化多数类的准确率。我们通过随机搜索交叉验证进行超参数优化。

3.1 单变量分布

严重程度分布分析证实类别不平衡现象显著：在总计 499,315 起事故中，严重程度 2 占 79.62%（n = 397,538），严重程度 3 占 16.92%（n = 84,477），严重程度 4 占 2.61%（n = 13,040），严重程度 1 占 0.85 %（n = 4,260）。

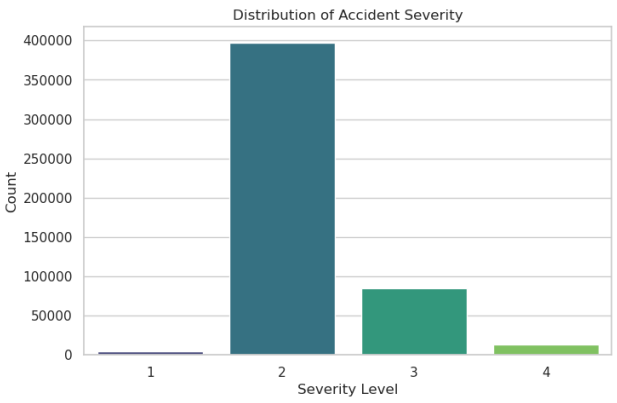

图 1. 数据集中事故严重程度等级（1–4）的分布。严重程度 2 占主导地位，占记录总数的 79.62%（n = 397,538）；严重程度 1（0.85%）和严重程度 4（2.61%）所占比例则明显偏低。

对关键数值特征的分析揭示了典型的分布模式。温度呈中度左偏，均值为 61.7°F。能见度、风速和降水呈现显著右偏，这与其物理特征相符：这些数值不可能为负，但可能出现极大的正值。鉴于数据存在偏度，采用中位数作为指标，结果显示能见度的中位数为 10 mi，风速为 7 mph，降水接近 0 in。能见度（2.72）和降水（91.76）的偏度系数较高，表明数据中可能存在不切实际的极端值，后续异常值分析也证实了这一点。

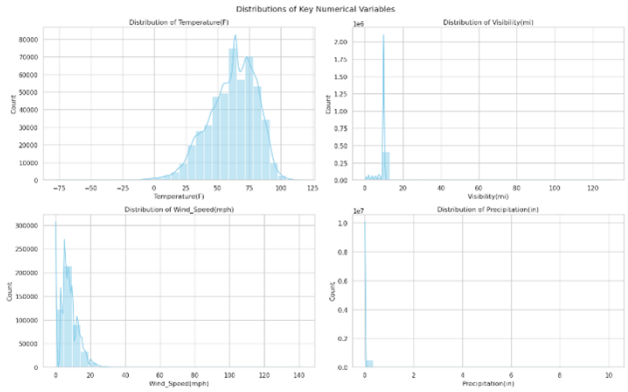

图 2。关键气象变量的直方图。温度呈中度左偏；能见度、风速和降水则呈明显右偏，这与物理约束相符。

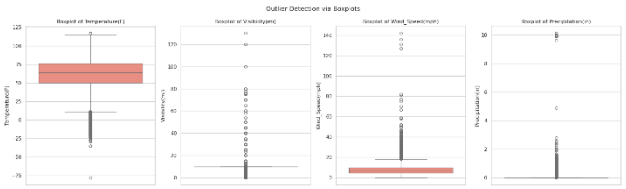

图 3。基于箱线图的气象变量异常值检测。使用四分位距标准识别异常值，并与历史天气记录进行交叉核对。

3.2 时间与类别模式

时间分析从事件时间戳中提取了月份、小时和星期几。严重程度呈现明显的季节性模式：春季逐渐上升，在盛夏（六月至八月）达到峰值，随后向冬季逐渐下降，这表明较暖月份可能与某些行为模式有关，并导致更严重的交通事故。

每日模式呈现出更明显的差异。严重程度在清晨（凌晨 2—5 点）降至最低，随后在上午晚些时候逐步上升，下午维持在较高水平，并在晚间约 7—8 点达到峰值，此时通常已过通勤时段。这一峰值可能反映了交通密度、下班后的疲劳以及能见度降低等因素的共同影响。相比之下，工作日变化很小，这表明相较于一天中的时段和季节效应，一周中不同日期对严重程度的影响有限。

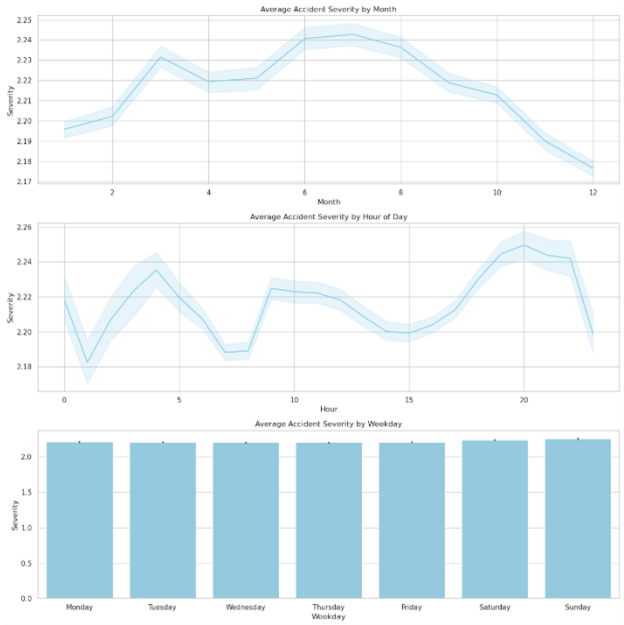

图 4。事故平均严重程度的时间变化。上图：季节性模式，夏季达到峰值（六月至八月）。中图：每日模式，傍晚达到峰值（~7–晚上 8 点）。下图：工作日之间的变化极小。

对风向和日出/日落进行分类分析后发现，不同风向下的严重程度略有差异，不过无风条件下的平均严重程度最高，这表明严重事故并非只与大风情形有关。日出/日落对比显示，白天事故与夜间事故的严重程度大致相同，这与普遍的

夜间事故因能见度限制而本质上更为严重的假设。

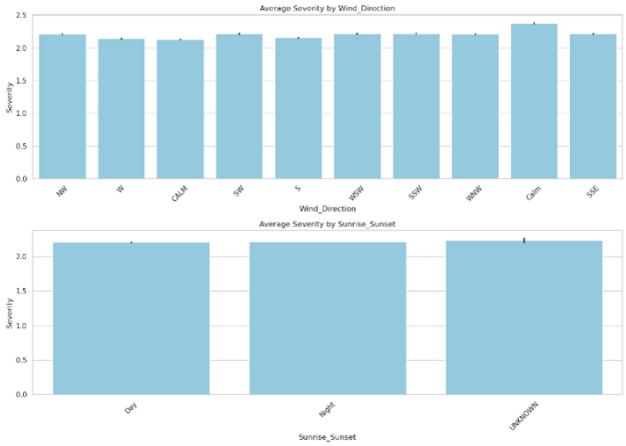

图 5。按风向（上）和日照条件（下）划分的平均事故严重程度。无风条件下事故严重程度最高；白天和夜间事故的严重程度等级相近。

3.3 双变量关系

对气象因素与事故严重程度之间关系的分析发现，两者关联极弱。箱线图显示，温度（分布范围较广，从负值到 >100°F，中位数相近）、能见度（集中在 10 mi 左右，低于 1 mi 的情况罕见）、风速（以 0–12 mph 的弱风为主）和降水（分布高度不对称，集中在零附近）在全部四个严重程度分层中的分布相似。

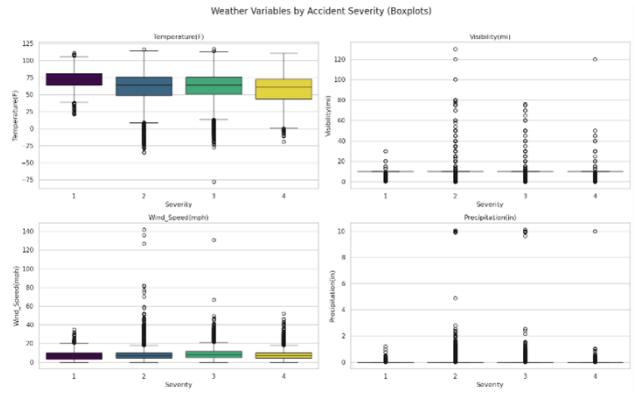

图 6。按事故严重程度划分的气象变量箱线图。各严重程度等级的分布相似，表明天气特征的区分能力有限。

核密度估计证实了这些模式：各严重程度等级的温度分布形状相似，中心密集且尾部较长；能见度密度主要集中在较高值范围，与严重程度没有明显关系；风速观测值呈现狭窄的核心分布并带有较长尾部，与严重程度无关；降水分布则高度集中在零值附近，无法揭示其与严重程度之间的关系。

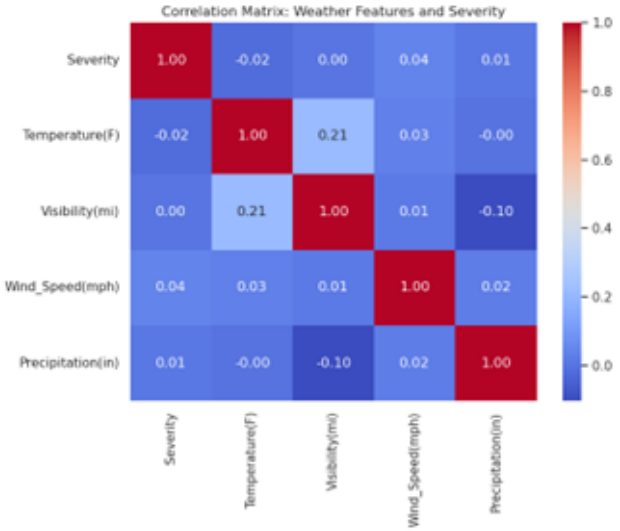

图 7。严重程度与气象变量之间的皮尔逊相关矩阵。相关性接近于零，证实天气状况与事故严重程度之间不存在任何线性关系。

相关性分析得出的结果显示，严重程度与所有气象变量之间的相关系数均接近于零，证实不存在可利用的线性关系。气象变量之间的相关性也较弱，只有温度与能见度之间存在中等程度的相关性，但这对严重程度没有影响。这些双变量分析结果表明，在该数据集中，所测量的气象变量与事故严重程度之间不存在可利用的关系；各严重程度类别的分布也呈现出稳定的模式。

3.4 XGBoost 模型性能

加权 XGBoost 分类器的总体准确率达到 78.2%，且在不同严重程度等级上呈现出不同的性能特征。对于严重程度 2（多数类），模型表现出较高的可靠性，精确率和

召回率均为 0.87，表明模型有效学习了标准的交通事故模式。对于严重程度 3，模型的召回率为 0.49，成功识别出近一半的严重事故，并证实特定的环境和时间特征是严重程度上升的有效预测因素。

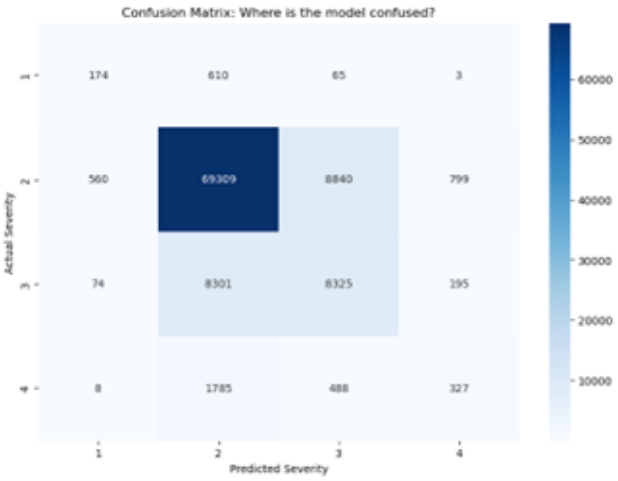

图 8。 加权 XGBoost 分类器的混淆矩阵。严重程度 2 显示出明显的对角线优势（精确率/召回率 =0.87）；严重程度 4 的案例大量被误分类为严重程度 2，反映了类别不平衡的影响。

对于严重程度 4（致命/严重事故），模型召回率为 0.13，灵敏度较低，但仍能对最严重的事件提供非零信号。混淆矩阵分析显示，2,608 起严重程度 4 事故中有 1,785 起被误分类为严重程度 2。这种在极端情况下降低的表现表明，致命事故可能取决于该数据集中未包含的变量（例如驾驶员行为、车辆速度、醉酒），而非仅仅取决于环境条件。

特征重要性分析确定了五个最强的预测因素：起始地点纬度、温度、起始地点经度、风速和一天中的时刻。这些结果在很大程度上支持了我们关于时间和空间因素预测价值的假设，同时表明地理坐标所提供的预测信息远比预期丰富。

4.1 研究结果解读

本研究表明，采用机器学习方法，可以根据环境、空间和时间特征，以中等准确率预测交通事故严重程度。模型在严重程度 2 事故上的出色表现（精确率和召回率均为 0.87）证实了其准确分类大多数交通事故的能力。纬度、经度、温度、风速和时刻被确定为主要预测因素，这在很大程度上支持了我们关于时间和天气因素重要性的假设，同时凸显了地理位置的重要贡献。

值得注意的是，降水和能见度这两个更直接反映危险驾驶条件的变量，并未成为显著的预测因素。尽管这一发现起初看似违反直觉，但它与先前关于行为适应的研究相符：在明显恶劣的条件下，驾驶员可能会更加谨慎，从而降低碰撞严重程度，即使事故发生率并未改变。相比之下，温度和风速较少被视为警示信号，却可能作为特定环境条件的替代指标，在不触发补偿性警觉的情况下影响严重程度。

地理坐标表现出的强预测作用，进一步印证了 Zhi（2025年）关于地点在事故风险评估中至关重要的发现。时间特征的重要性则与约旦道路事故研究的结果相符；该研究同样指出，一天中的时段是严重程度的关键决定因素。晚上 7—8 点左右严重程度达到峰值，这可能反映了交通拥堵、驾驶员疲劳和过渡性光照条件的叠加影响。

4.2 局限性

数据集中严重的类别不平衡构成了主要的方法学局限。严重程度 2 占观测值的 79.62%，而严重程度 4 仅占 2.61%（测试集中 n = 2,608），这意味着模型缺乏足够的极端案例样本，无法有效学习区分性模式。尽管采用了类别加权，模型在极端类别上的表现仍受限制，而且被识别为重要的特征可能无法推广到轻微或重大事故。

重大事故本身较为罕见，加上轻微事件可能存在漏报，这解释了这种分布偏斜。双变量分析进一步表明，该数据集中的天气变量无法充分捕捉严重程度预测所需的背景信息：降水和风速的数值几乎都接近零，因此无法恰当地研究极端条件；而能见度几乎总是良好。

这些发现与更广泛的文献一致：天气变量仅能解释交通事故死亡人数变异的约 6.9%，而 86.6% 的事故发生在正常条件下，且天气与严重程度之间没有明显关联 [5]。天气的影响存在异质性，有时甚至呈现悖论，并且在很大程度上取决于事故类型、驾驶员特征和道路网络配置。

4.3 伦理考量

有几个伦理层面值得考虑。首先，数据收集过程可能未能充分透明地说明是否已取得事故记录中相关人员的知情同意。其次，交通监测基础设施分布不均可能造成地理偏差，从而可能导致偏远地区或经济欠发达地区被低估。第三，使用交通影响持续时间作为严重程度代理指标，可能引入我们的建模方法未能捕捉的系统性偏差。

公开数据集不包含可直接识别个人身份的信息，所有位置、时间和环境数据均已预先匿名化。尽管如此，仍应保护模型输出并限制其仅用于学术用途，以防止其在法律或财务情境中被不当使用，例如造成保险定价或理赔评估中的偏倚。

4.4 未来方向

要解决类别不平衡这一局限，需要针对性地采集数据，重点关注极端严重程度的案例，采用先进的重采样技术，并整合被动传感器网络目前无法获取的行为变量。未来研究应探讨，在严重程度分布均衡的情况下，地理位置、温度、风速和时间的特定模式是否与更严重的事故相关。

从应用角度看，这些结果支持开发纳入时空风险模式的实时预测系统，以便实施动态干预，从自适应信号控制到有针对性的驾驶员警报，并在严重程度风险达到峰值的准确时间和地点发布。将实时天气数据整合到交通管理系统中，并结合动态标志和依情境而定的限速等工程措施，是降低严重程度的一个有前景的方向。

5 结论

本研究表明，利用机器学习方法，可以根据环境、空间和时间特征，对美国的交通事故严重程度进行中等程度的预测。针对类别不平衡进行优化的XGBoost分类器准确率达到78%，并识别出地理坐标、温度、风速和一天中的时段是主要预测因素；与直觉相反，能见度和降水的预测影响却很小。这一模式与新近证据相符：驾驶员会在明显危险的条件下调整行为，而较不明显的环境线索则未能引发补偿性警觉。

极端严重程度事件所占比例严重不足，限制了模型对后果最严重事故的泛化能力，凸显了观测性交通数据固有的一项根本局限。要弥补这一缺口，需要有针对性地采集数据、采用先进的重采样方法，并整合被动监测目前无法获取的行为变量。

除了方法学改进之外，这些结果还对循证交通管理具有直接意义。依据时空风险模式构建的实时预测系统，可在严重程度风险达到峰值的时间和地点精准实施动态干预。随着城市出行系统日益复杂，这类数据驱动方法为打造可衡量的更安全道路提供了一条可扩展的路径。

参考文献

[1] Al-Omari, A.，Alsaleh, M.，& Shatnawi, N.（2025年）。交通事故的时空分析与严重程度预测。《工程科学》。https://doi.org/10.30919/es1472

[2] Kang, X., Wu, D., Sha, W., Song, K., & Wang, S.（2025）。基于可解释空间机器学习的交通事故分析与预测：加利福尼亚州案例研究。《先进交通学报》，*2025*(1)。https://doi.org/10.1155/atr/3184284

[3] Shangguan, A., Feng, N., Hei, X., Fei, R., Jin, Y., Mu, L., & Li, Y.（2025）。利用变分自编码器注意力和图卷积网络，根据不平衡数据预测道路交通事故严重程度。*Scientific Reports*，*15*(1)，34372。https://doi.org/10.1038/s41598-025-17064-4

[4] Safari, M., Alizadeh, S., Bazargani, H. S., Aliashrafi, A., Shakerkhatibi, M., & Moshashaei, P.（2019）。采用层次网络分析法确定影响碰撞严重程度因素的优先级。《伤害与暴力研究杂志》，*12*(1)，11–19。https://doi.org/10.5249/jivr.v12i1.1229

[5] Lio, C.、Cheong, H.、Un, C.、Lo, I. 和 Tsai, S.（2019）。气象变量与道路交通伤害之间的关联：来自澳门的研究。PeerJ，7，e6438。https://doi.org/10.7717/peerj.6438

[6] Islam, N.、Iqra, S. A.、Huq, A. S. 和 Tasnim, A.（2023）。天气影响对孟加拉国道路交通事故严重程度的计量经济分析：来自达卡大都会区的证据。《可持续发展》，15(17)，12797。https://doi.org/10.3390/su151712797

[7] Zeng, Q.、Hao, W.、Lee, J. 和 Chen, F.（2020）。调查实时天气状况对高速公路事故严重程度的影响：贝叶斯空间分析。《国际环境研究与公共卫生杂志》，17(8)，2768。https://doi.org/10.3390/ijerph17082768

[8] Naseralavi, S., Soltanirad, M., Ranjbar, E., Lucero, M., Baghersad, M., Piri, M., Zada, M. J. H., & Mazaheri, A.（2025年）。基于驾驶员性别和交通事故类型对雨天交通事故严重程度进行建模。《未来交通》，5(4)，146。https://doi.org/10.3390/futuretransp5040146

[9] Morgan, A.、Mannering, F.（2011年）。路面状况、年龄和性别对驾驶员受伤严重程度的影响。《事故分析与预防》，43(5)，1852–1863。https://doi.org/10.1016/j.aap.2011.04.024

[10] Huang, L.（2016年）。台湾道路交通死亡的家庭负担风险因素识别：来自一项横断面调查的回归结果。《*BMC*公共卫生》，*16*(1)，1202。https://doi.org/10.1186/s12889-016-3813-3

[11] Obasi, T.、Mahmoud, M.（2023）。使用机器学习技术预测交通事故严重程度方面的有效性。《*Heliyon*》，*9*(8)，e18812。https://doi.org/10.1016/j.heliyon.2023.e18812

[12] Wang, X.、Su, Y.、Zheng, Z. 和 Xu, L.（2024）。基于元启发式算法优化的随机森林对机动车交通事故严重程度的预测与解释。《*Heliyon*》，*10*(16)，e35595。https://doi.org/10.1016/j.heliyon.2024.e35595

[13] Assi, K.，Rahman, S. M.，Mansoor, U.，和 Ratrout, N.（2020）。利用聚类技术协同机器学习算法预测碰撞受伤严重程度：一项前景可期的研究方案。《国际环境研究与公共卫生杂志》，*17*(15)，5497。https://doi.org/10.3390/ijerph17155497

[14] Xie, Z. 和 Yan, J.（2008）。网络空间中交通事故的核密度估计。《计算机、环境与城市系统》，*32*(5)，396–406。https://doi.org/10.1016/j.compenvurbsys.2008.05.001

[15] Läuter, H.（1988）。《统计与数据分析中的密度估计》。Chapman & Hall，伦敦—纽约，1986年，175页，£12.—。《生物计量学杂志》，*30*(7)，876–877。https://doi.org/10.1002/bimj.4710300745

[16] Yao, S., Wang, J., Fang, L., & Wu, J.（2018）。使用网络核密度估计和随机森林在微观层面识别车辆—行人碰撞热点：以上海（中国）为例。《可持续发展》，*10*(12)，4762。https://doi.org/10.3390/su10124762

[17] Jiang, Y.（2025）。基于机器学习的交通事故潜在因素预测分析。应用与计算工程，*99*(1)，112–120。https://doi.org/10.54254/2755-2721/99/20251788

[18] Adefabi, A.，Plaish, S.，Edmunds, T.，& Tella, E.（2023）。预测事故严重程度：使用随机森林模型分析影响事故严重程度的因素。*arXiv*（康奈尔大学）。https://doi.org/10.48550/arxiv.2310.05840

[19] Plainis, S.、Murray, I. J. 和 Pallikaris, I. G.（2006）。道路交通伤亡：了解夜间死亡人数。《伤害预防》，*12*(2)，125–138。https://doi.org/10.1136/ip.2005.011056

[20] JE, B. 和 O, M. O.（2020）。夜间降雨对昏暗道路停车视距的影响。《开放交通期刊》，*14*(1)，32–37。https://doi.org/10.2174/1874447802014010032

[21] De Souza, V. N., & De Oliveira Neto, F. M.（2023）。使用生存模型对高交通流量变异性的城市道路上的交通事故持续时间建模：福塔雷萨主干道案例研究。*Transportes*，*31*(2)，e2837。https://doi.org/10.58922/transportes.v31i2.2837

[22] Hossain, M., & Muromachi, Y.（2013）。利用高分辨率交通数据理解城市快速路上的事故机制。《事故分析与预防》，*57*, 17–29。https://doi.org/10.1016/j.aap.2013.03.024

[23] Stiles, J., Kar, A., Lee, J., & Miller, H. J. (2021年). 车流量减少，车速提高：新冠肺炎《居家政策》实施后城市道路交通事故类型、时间和严重程度的变化。《交通研究记录：交通研究委员会期刊》，*2677*(4), 15–27.https://doi.org/10.1177/03611981211044454

[24] Jägerbrand, A. K., & Sjöbergh, J. (2016年). 天气状况、光照条件和道路照明对车辆速度的影响。*SpringerPlus*，*5*(1), 505. https://doi.org/10.1186/s40064-016-2124-6

[25] Plainis, S., Murray, I. J., & Pallikaris, I. G. (2006b). 道路交通伤亡：了解夜间死亡人数。《伤害预防》，*12*(2), 125–138. https://doi.org/10.1136/ip.2005.011056

[26] Hossain, A., Sun, X., Thapa, R., & Codjoe, J. (2022). 应用关联规则挖掘研究不同光照条件下行人致死及受伤事故模式。《交通研究记录：交通研究委员会期刊》，*2676*(6), 659–672. https://doi.org/10.1177/03611981221076120[27] Retallack, A. E., & Ostendorf, B. (2020). 交叉路口交通流量与事故频率之间的关系。《国际环境研究与公共卫生杂志》，*17*(4), 1393.https://doi.org/10.3390/ijerph17041393[28] Wei, X., Shu, X., Huang, B., Taylor, E. L., & Chen, H. (2017). 分析不同光照条件下道路施工区域的交通事故严重程度。《先进交通学报》，*2017*，1–10.https://doi.org/10.1155/2017/5783696[29] Yasanthi RGN, Mehran B, Alhajyaseen WKM. 乡村公路上的速度行为建模：恶劣道路天气条件下驾驶的安全分析。PLoS One。2021年8月16日；16(8):e0256322。doi: 10.1371/journal.pone.0256322。PMID: 34398905；PMCID: PMC8367011。[30] Zeng, Q., Hao, W., Lee, J., & Chen, F. (2020). 研究实时天气状况对高速公路事故严重程度的影响：贝叶斯空间分析。《国际环境研究与公共卫生杂志》，17(8), 2768。https://doi.org/10.3390/ijerph17082768[31] Stiles J, Kar A, Lee J, Miller HJ. 较低流量，较高速度：新冠肺炎居家政策导致城市道路上的交通事故类型、时间和严重程度发生变化。Transp Res Rec。2023年4月；2677(4):15-27。doi: 10.1177/03611981211044454。在线发表日期：2021年9月23日。PMID: 37153167；PMCID: PMC10149494。[32] Stevens SE, Schreck CJ, Saha S, Bell JE, Kunkel KE. 降水与致命机动车事故：基于高分辨率雷达数据的大陆分析。《美国气象学会公报》。2019年8月1日；100(8):1453-1461。doi: 10.1175/BAMS-D-18-0001.1。PMID: 37293512；PMCID: PMC10248718。[33] Wei, F., Cai, Z., Liu, P., Guo, Y., Li, X., & Li, Q. (2021). 雾天和晴天单车事故中驾驶员受伤严重程度研究。《先进交通学报》，2021, 1–12。https://doi.org/10.1155/2021/9939800[34] Moosavi, S. (2023). 美国交通事故（2016年 - 2023） [数据集]。载于 Kaggle。https://doi.org/10.34740/kaggle/ds/199387[35] J. Liang and M. Ishihara,「利用驾驶模拟器在网格化道路布局上研究不同能见度条件下通过行为适应体现的驾驶经验」，2025年国际电路／系统、计算机与通信技术会议（ITC-CSCC），大韩民国首尔，2025年，第1-6页，doi: 10.1109/ITC- CSCC66376.2025.11137786。[36] Tafazzol, S., Sharif, H., Gholikhani, M. 等。得克萨斯州2006至2021年降雨期间的相对事故风险与道路安全。《科学报告》15, 36749 (2025)。https://doi.org/10.1038/s41598-025-20760-w [37] Zeng, Q., Hao, W., Lee, J., & Chen, F. (2020). 研究实时天气状况对高速公路碰撞严重程度的影响：贝叶斯空间分析。《国际环境研究与公共卫生杂志》，17(8), 2768。https://doi.org/10.3390/ijerph17082768[38] Ogungbire, A., & Pulugurtha, S. S. (2024). 天气相关事故严重程度分析中数据不平衡处理的有效性。《交通研究记录：交通研究委员会期刊》，2678(11), 88–105。https://doi.org/10.1177/03611981241239962[39] Wang, Y., Wei, F., Guo, Y., & Guo, Y. (2024). 影响乡村公路单车事故受伤严重程度因素的时间不稳定性。《交通快报》，17(4), 578–594。https://doi.org/10.1080/19427867.2024.2366731[40] Alshehri, A. H., Pervez, A., Hussain, M., Farooq, D., & Hussain, E. (2024). 使用广义线性模型评估事故严重程度时间稳定性相关因素——案例研究。《PLoS ONE》，19(4), e0299094。https://doi.org/10.1371/journal.pone.0299094[41] Zhang, Y., Pan, Z., Zhu, F., Shi, C., & Yang, X. (2024). 高速公路交通事故时空延迟效应的定量估计与分析。ISPRS《国际地理信息杂志》，13(11), 407。https://doi.org/10.3390/ijgi13110407[42] Sukonna, R. T., & Swapnil, S. I. (2023). 交通事故持续时间预测的双层框架：在实用最优流程中利用天气和道路状况数据。arXiv（康奈尔大学）。https://doi.org/10.48550/arxiv.2311.00634

[43] R. Chhabra, V. Khullar, S. Singh, R. Kaur and K. Singh, 「利用深度学习进行智能交通系统交通延误分析」，2023 IEEE第二届工业电子国际会议：发展与应用（ICIDeA），印度因帕尔，2023年，第312-315页，doi: 10.1109/ICIDeA59866.2023.10295168。

[44] Sun, X., Hu, H., Ma, S., Lin, K., Wang, J., & Lu, H.（2022）。基于统计分析和空间分布特征的道路交通事故持续时间影响研究：休斯敦的实证分析。《可持续发展》，*14*(22)，14982。https://doi.org/10.3390/su142214982

[45] Lee, W., Lee, H., Hwang, S., Kim, H., Lim, Y., Hong, Y., Ha, E., & Park, H.（2015）。温度是否会改变降雨和降雪对道路交通伤害的影响？《流行病学杂志》，*25*(8)，544–552。https://doi.org/10.2188/jea.je20140244

[46] Nazif-Munoz, J. I., Gilani, V. N. M., Rana, J., Choma, E., Spengler, J. D., & Cedeno-Laurent, J. G.（2025年）。热浪对魁北克五座具有不同热环境的城市交通安全的影响。《伤害流行病学》，*12*(1)，12。https://doi.org/10.1186/s40621-025-00564-2

[47] Lio, C., Cheong, H., Un, C., Lo, I., & Tsai, S. (2019b). 气象变量与道路交通伤害之间的关联：一项来自澳门的研究。*PeerJ*，*7*，e6438。https://doi.org/10.7717/peerj.6438

[48] Chan, T., Pai, C., Wu, C., Hsu, J., Chen, R., Chiu, W., & Lam, C. (2022). 空气污染和天气因素与交通伤害严重程度的关联：一项在台湾开展的研究。《国际环境研究与公共卫生杂志》，*19*(12)，7442。https://doi.org/10.3390/ijerph19127442

Liu, H 基于随机森林模型的道路交通事故风险预测。《理论与自然科学》，*134*(1)，62– 72。https://doi.org/10.54254/2753-8818/2025.ad26486

[50] Chen, J.、Liu, P.、Wang, S.、Zheng, N. 与 Guo, X.（2025年）。基于不平衡交通碰撞数据，使用机器学习进行碰撞严重程度预测与解释。《安全研究杂志》，*93*，185–199。https://doi.org/10.1016/j.jsr.2025.02.018

[51] Fiorentini, N. 与 Losa, M.（2020年）。通过机器学习算法处理道路碰撞严重程度预测中的不平衡数据。《基础设施》，*5*（7），61。https://doi.org/10.3390/infrastructures5070061

[52] Jeong, H.、Jang, Y.、Bowman, P. J. 与 Masoud, N.（2018年）。机动车碰撞伤害严重程度分类：一种针对不平衡数据的混合方法。《事故分析与预防》，*120*，250–261。https://doi.org/10.1016/j.aap.2018.08.025

[53] Li, Z., Liao, H., Tang, R., Li, G., Li, Y., & Xu, C. (2023). 缓解异常值对交通碰撞分析的影响：一种稳健的贝叶斯回归方法及其在隧道碰撞数据中的应用。《事故分析与预防》，*185*，107019。https://doi.org/10.1016/j.aap.2023.107019[54] Moosavi, S., Samavatian, M. H., Parthasarathy, S., & Ramnath, R. (2019). 全国交通事故数据集。*arXiv*（康奈尔大学）。https://doi.org/10.48550/arxiv.1906.05409

[55] Ma, L 基于GC气象传感器网络中滑动窗口预测的温度时间序列异常值校正。《*Information*》，*8*(2)，60。https://doi.org/10.3390/info8020060

[56] *Dash, C .S., Behera, A .K., Dehuri, S ., & Ghosh, A . (2023).* 数据挖掘分类任务中的异常值检测与消除框架。《决策分析期刊》。

[57] Li, L.、Prato, C. G. 和 Wang, Y.（2020）。山区高速公路交通事故影响因素排序：基于不完整数据集的链式方程多重插补与随机森林分类器序贯方法。《事故分析与预防》，*146*，105744。https://doi.org/10.1016/j.aap.2020.105744

[58] Ardakani, S. P., Liang, X., Mengistu, K. T., So, R. S., Wei, X., He, B., & Cheshmehzangi, A. (2023). 使用机器学习赋能的数据分析进行道路交通事故预测。《可持续发展》, *15*(7), 5939. https://doi.org/10.3390/su15075939

[59] Qi, H., Zhao, X., Yao, Y., Yang, H., Chai, S., & Chen, X. (2023). 基于BGCP的交通数据插补及其在国家干线公路事故检测中的应用。《事故分析与预防》，*186*，107051。https://doi.org/10.1016/j.aap.2023.107051

[60] Deb, R., & Liew, A. W. (2016). 为分析不完整交通事故数据而进行的缺失值插补。信息科学,*339*, 274–289. https://doi.org/10.1016/j.ins.2016.01.018

[61] Gutierrez-Osorio, C., & Pedraza, C. (2020). 道路交通事故分析与预测的现代数据源和技术：综述。交通运输工程学报（英文版），*7*(4), 432–446. https://doi.org/10.1016/j.jtte.2020.05.002

[62] Shaw, R. J., Harron, K. L., Pescarini, J. M., Pinto, E. P., Junior, Allik, M., Siroky, A. N., Campbell, D., Dundas, R., Ichihara, M. Y., Leyland, A. H., Barreto, M. L., & Katikireddi, S. V. (2022). 关联行政数据用于流行病学研究时产生的偏倚：从登记到分析的概念框架。《欧洲流行病学杂志》, *37*(12), 1215–1224. https://doi.org/10.1007/s10654-022-00934-w

[63] Bassett E、Broadbent J、Gill D、Burgess S、Mason AM. 英国生物样本库中不同数据来源的事件定义不一致及其对偏倚和可推广性的影响：以静脉血栓栓塞症为例。《美国流行病学杂志》. 2024 年 5 月 7 日；193(5):787-797. doi: 10.1093/aje/kwad232. PMID: 37981722; PMCID: PMC11074710.

[64] Li, G., & Jung, J. J. (2021年). 用于多条气象时间序列异常值检测的动态图嵌入。*PLoS ONE*, *16*(2),e0247119. https://doi.org/10.1371/journal.pone.0247119[65] Endalie, D., & Abebe, W. T. (2023). 通过数据挖掘技术分析和检测道路交通事故严重程度：埃塞俄比亚亚的斯亚贝巴案例研究。《工程中的数学问题》, *2023*(1)。https://doi.org/10.1155/2023/6536768

[66] Shahsavari, S., Mohammadi, A., Mostafaei, S., Zereshki, E., Tabatabaei, S. M., Zhaleh, M., Shahsavari, M., & Zeini, F. (2022). 伊朗道路交通事故伤亡分析：双变量回归方法。*BMC* 急诊医学, *22*(1), 130. https://doi.org/10.1186/s12873-022-00686-6

[67] *Sabel, C.E., & Bartie, P.J. (2012).* 基督城（新西兰）道路交通事故的时空地理可视化与数据挖掘。

[68] Ambunda, R., & Neliwa, M. (2025年). 纳米比亚驾驶员致命及重伤道路交通事故的时间与人口统计分析。*Engineering Headway*, *28*, 25–35. https://doi.org/10.4028/p-s97sfl