

一种用于环形交叉口事故严重性分析的降维可解释人工智能框架

Rohit Chakraborty
Ingram School of Engineering
Texas State University
San Marcos, Texas, USA
xuw12@txstate.edu

Subasish Das, Ph.D.
Ingram School of Engineering
Texas State University
San Marcos, Texas, USA
subasish@txstate.edu

Abstract

环形交叉路口减少了严重事故，但风险模式因条件而异。本研究使用两步、可解释的工作流分析了2017年至2021年俄亥俄州的环形交叉路口事故。聚类对应分析 (CCA) 识别共同出现的因素，并产生四种事故模式。然后通过 SHAP 解释基于树的严重性模型，以量化在不同模式内和跨模式内的伤害驱动因素。结果显示，在黑暗、湿滑路面以及较高的限速与固定物体或角度事件同时发生时，严重程度更高；而在清晰视野、低速设置下，则严重程度较低。特定模式的解释突出了入口处（未能让行、间隙接受）、多车道循环中（不正确的操作）和减速期间（追尾事故）的机制。该工作流将模式发现与案例级解释相结合，支持现场筛选、对策选择以及审计准备报告。对信息系统领域的贡献是为公共安全分析提供了一个实用的 XAI 模板。

关键词： 可解释人工智能 (XAI)，SHAP，聚类对应分析 (CCA)，可解释机器学习，环形交叉口安全，事故严重程度，决策支持系统，交通运输分析。

1. 介绍

环岛是一种特别的平面交叉路口，广泛被认为比传统交叉口更安全的选择，主要是因为它能减少冲突点并鼓励降低车辆速度 (Maji and Ghosh, 2025)。在美国各地，环岛已被证明能够降低与有信号灯的交叉口相比的死亡率。由于环岛具有降低事故严重性、减慢车速以及相较于带有交通信号或停

车标志的传统交叉口在路口减少冲突的能力，因此在美国变得越来越重要。尽管环岛通常能改善安全性，但仍会发生事故，需要持续的研究和安全改进措施。研究环岛的安全问题之所以必要是因为正在建设更多的环岛，并且交通状况继续发生变化。环岛主要通过减少可能发生碰撞的点、降低车速以及让驾驶员的决策更简单来提高交叉口的安全性。许多研究表明，在建成环岛后严重事故的数量减少了 (Polders et al., 2015; Royce, 2022; Ashqar et al., 2024)。然而，发生在环岛上与在常规交叉口发生的事故原因、类型以及涉及人员方面存在差异。这些事故通过消除许多高风险的碰撞类型来降低事故严重性。特别是，在环岛环境中消除了交叉路径和左转行动相关的碰撞。此外，环岛的几何设计自然鼓励以较低的速度行驶，这既减少了发生碰撞的可能性也降低了当碰撞确实发生时造成的伤害程度 (Gross et al., 2013)。尽管如此，事故严重性的降低并不能完全避免之前研究指出的由于驾驶行为、设计复杂性和非法转弯所导致在环岛上发生的大量事故 (Sussman, 2022)。正因为这些差异，仔细研究发生在环岛上的事故有助于识别特定的安全问题和解决这些问题的方法。

持续进行环岛安全研究是重要的，因为交通系统不断变化。新技术已经改变了驾驶员接近、进入和在环岛内行驶的方式。高级驾驶辅助功能，如自适应巡航控制和自动紧急制动，可以平滑或突然改变进近速度，这可能会减少一些追尾风险，但也可能在让行线处造成犹豫并缩短接受的空档时间 (Mehta

et al., 2023)。车道保持和车道居中可以帮助改善多车道环岛中的车道纪律,但当标记或标志令人困惑时也可能增加晚变更车道的情况 (Rojas et al., 2024)。使用路边单元的联网车辆功能可以广播几何形状和建议信息 (例如,推荐速度、排队警告),影响入口和出口处的空档接受情况和速度分布。低速自动穿梭车和高度自动化测试车辆与混合交通中的人类驾驶员互动,有时会做出超出必要的让行并引发意外的启停行为。实际上,机构也在环岛中试行技术增强措施,如动态速度反馈标志、LED 增强型让行标志以及进近处的高摩擦表面处理,以解决湿滑/低摩擦条件和夜间能见度问题。这些变化意味着环岛事故模式会随时间演变,强化了需要检测道路、环境、车辆类型、操作和驾驶员相关因素组合变化的方法的需求。

提高环岛安全包括物理改进和教育努力。物理改进可能涉及通过调整入口角度或车道设置、安装更好的标志和灯光,或为行人增设设施来改变环岛设计。对驾驶员、行人和骑行者进行的安全意识提升项目和宣传活动也很重要。这些措施协同工作以减少事故的数量和严重性。因此,研究环岛安全、分析事故并制定有效的安全措施仍然很重要。随着环岛的普及,需要进行全面的研究和定期的安全评估。了解为什么在环岛上会发生事故有助于当局做出提高安全性的明智决策,并使环岛对所有人更安全。根据联邦公路管理局 (FHWA) 报告 (Steyn et al., 2015), 2005 年至 2013 年间,在美国的 19 个不同州发生了 46 起致命事故。最近在密歇根州的一项研究表明,尽管死亡人数有所减少,但环岛发生的事故比其他传统交叉口多出 58% (Savolainen et al., 2023)。这些事件不仅挑战了环岛本质上是安全的观点,而且还强调需要对环岛的特定事故特征进行更仔细的研究。本研究将调查 2017 年至 2021 年期间在俄亥俄州发生的环岛事故,以探究导致此类事故的相关因素。在俄亥俄州,随着越来越多地实施环岛作为一项安全措施,了解这些事故的本质和原因对于进一步提高环岛的安全性以及为工程、执法和教育干预提供信息至关重要。

虽然环形交叉路口被广泛认为可以减少冲突点并提高交叉口的安全性,但现有的研究大多依赖于传统的统计模型来单独考察各个因素。这些方法通

常无法捕捉车辆类型、道路条件、驾驶行为和环境因素之间的复杂相互作用,而这些都是导致事故严重程度因素。此外,很少有研究将无监督学习或可解释的机器学习方法应用于环形交叉路口的事故数据中,这限制了发现隐藏模式或解释模型决策的能力。本研究通过使用聚类对应分析 (CCA) 作为降维和聚类技术 (Velden et al., 2017),解决了这些空白问题。该方法能够在低维度空间中基于共享分类特征对事故进行分组。这有助于揭示可能在传统分析中不明显的潜在模式。此外,应用 SHapley 加性解释 (SHAP) 来解释选定特征对不同聚类中的事故严重程度的影响 (Lundberg and Lee, 2017)。从计算机科学的角度来看,本研究表明如何通过结合无监督聚类、降维和模型可解释性提供一个透明且可扩展的框架,用于分析高维度的事故数据集,并指导以数据为导向的安全策略。一些先前的研究探讨了可解释的人工智能框架来提高交通安全并确定伤害严重程度水平 (Amini et al., 2022, Aboulola et al., 2024, Khasawneh et al., 2025)。本研究还解决了以下研究问题:

- **研究问题 1.** 哪些道路、环境、车辆、操作和驾驶员因素倾向于在环岛事故中共同出现?
- **研究问题 2.** 这些因素组合与环岛处的伤害严重程度有何关联?
- **研究问题 3.** 在每个聚类中,哪些因素最强烈地影响严重程度的结果,以及这些影响在不同的聚类之间有何不同?

2. 方法论

2.1. 聚类对应分析

CCA 是一种稳健的分析技术,旨在通过分组相似观察值 (Rahman et al., 2022) 来探索分类数据集内的关系。CCA 结合了对应分析与 K-均值聚类,在降低维度的空间中促进最优聚类分配和量化分类变量类别 (Velden et al., 2017)。这种混合方法通过确保类别量化强调跨集群的差异来增强可解释性。

CCA 从随机分配观测值到聚类开始,建立初始的 Z_K 并计算列联矩阵 F 。对 F 进行对应分析会产生一个分类量化矩阵 B , 然后用它来推导出对象坐标矩阵:

$$Y = \frac{1}{q}(I_n - \frac{1_n 1_n'}{n})ZB \quad (1)$$

随后, 对 Y 进行 K 均值聚类以更新 Z_K , 这个过程重复进行直到收敛 (Velden et al., 2017)。最终结果包括一个簇中心矩阵 G 和优化的类别量化矩阵 B 。

为了可视化, 这些矩阵使用以下因子进行重缩放:

$$\gamma = \left(\frac{K}{Q} \cdot \frac{\text{Tr}(B^\top B)}{\text{Tr}(G^\top G)} \right)^{1/4} \quad (2)$$

2.2. SHapley 加性解释法

为了进一步解释机器学习模型预测的严重性结果, 本研究使用了 SHAP。SHAP 提供了一种一致且局部准确的方法, 将模型的预测分解为每个输入特征的贡献。它基于合作博弈理论中的夏普利值概念, 其中每个特征都被视为对模型最终输出做出贡献的一个“玩家”。

SHAP 值表示每个特征对预测的碰撞严重性增加或减少的贡献程度。一个正的 SHAP 值表明某个特征使预测结果倾向于更严重的后果, 而一个负值则意味着对该结果的轻微影响。SHAP 值的幅度反映了该特征对模型输出的影响强度。对于给定的模型 $f(x)$, 特征 x_i 的 SHAP 值 ϕ_i 是通过以下公式 (Lundberg and Lee, 2017) 计算得出的:

$$f(x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (3)$$

2.3. 数据描述和变量选择

数据集最初包含 2017 年至 2021 年间俄亥俄州与环形交叉口相关的 6448 起事故, 共有 21 个分类变量描述了车辆特征、道路状况、环境因素和驾驶行为。为了确定对环形交叉口事故严重性影响最大的变量, 进行了系统化的变量选择过程, 如图 1 所示。为了确保分析有意义且无偏见, 移除了分布偏向单一类别超过 85% 的变量。这一预处理步骤得到了一组精简后的 18 个变量。然后使用两种机器学习

算法对这些变量进行了变量重要性分析: XGBoost (Chen et al., 2016) 和 随机森林 (Breiman et al., 2002), 这两种方法都以其在处理高维度和分类数据方面的鲁棒性而著称。两个不同的特征筛选方法依赖于两种树集成度量。随机森林的重要性是所有分裂和树中的 Gini 不纯度平均减少值。XGBoost 的重要性是从所有变量的提升树上的分割中累计的目标 (损失) 增益。在交叉验证训练折上计算了这些重要性, 进行了归一化并取平均值, 保留了持续高于中位数得分的变量。

基于两个模型的共识, 最终选定了用于 CCA 的 14 个变量。这些包括车辆类型、碰撞类型、促成因素、碰撞前行为以及道路和环境条件 (如天气、照明、路面轮廓、标示速度限制、车道数量和设施类型)。变量“严重性”被保留作为参考结果, 以支持基于 CCA 的碰撞模式的解释和讨论。这种数据驱动的方法确保了在聚类分析中仅使用相关且非冗余的变量, 从而在环形交叉路口碰撞的背景下实现更可靠的趋势检测和解释。图 2(a) 和图 2(b) 显示了分别使用 XGBoost 和随机森林模型进行的变量重要性分析。两个模型都选定了并突出了每个变量对碰撞严重性的得分。

3. 结果

本节展示了通过多步骤分析管道从环形交叉路口碰撞数据集中得出的关键发现。基于一组精炼的变量进行了 CCA 选择, 用于在数据集中发现不同的碰撞模式。随后对生成的聚类进行上下文和视觉解释。使用 SHAP 分析来说明各个特征如何影响每个聚类内的严重性预测, 提供模型级别的透明度并对碰撞动态有更深入的理解。

3.1. 聚类对应分析

本节包含对本研究发现的描述。表 1 提供了通过 CCA 从环形交叉口碰撞数据中识别出的聚类质心、变异性及聚类规模的汇总。质心 (Dim 1 和 Dim 2) 表示每个聚类在数据空间中的中心位置, 而组内平方和反映了各聚类内部事故间的变异程度。规模列显示分配给每个聚类的事故数量, 其中簇 1 包含最多事故 (2,453 起), 簇 4 最少 (1,298 起)。该表格说明了事故如何根据共享特征分组, 并展示了各

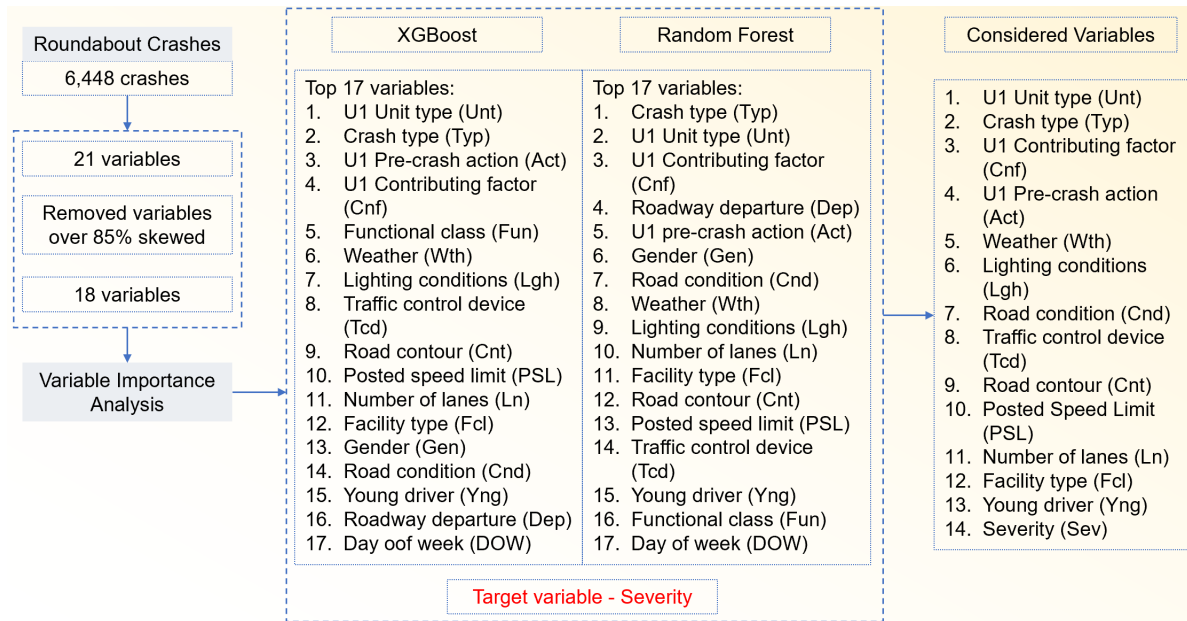


图 1: 考虑的变量列表

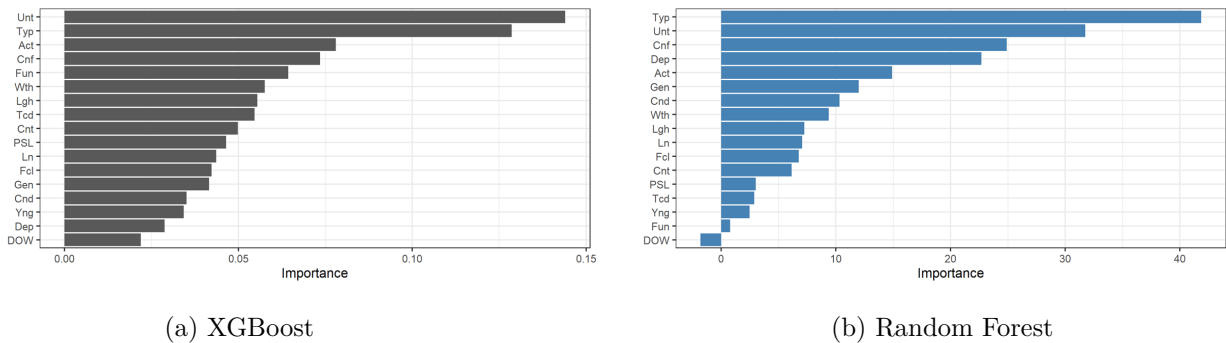


图 2: 变量重要性图

个聚类的相对内部一致性。

表 1: 聚类的中心点和大小

聚类	维度 1	维数 2	簇内平方和	大小
Cluster 1	-0.0096	-0.0053	0.0424	2453
Cluster 2	-0.0026	0.0036	0.0327	1396
Cluster 3	0.0048	0.0165	0.0396	1301
Cluster 4	0.0161	-0.0103	0.0408	1298

图 3绘制了应用于 CCA 坐标的 k -means 的总簇内平方和 (WCSS), 跨越 $K = 1-10$ 。

WCSS 衡量簇内离散程度; 较低的值表示更紧凑、更连贯的簇。

曲线从 $K = 1$ 到 $K = 3$ 急剧下降, 并在 $K = 4$ (红色圆圈) 处显示出明显的斜率变化。超过 $K = 4$ 后的减少是逐步的且几乎呈线性, 表明增加聚类群组的收益递减。这种模式支持选择四个簇: $K = 4$ 捕获了数据中的主导结构, 而更大的 K 主要分裂现有的群体而不揭示新的结构。四簇解决方案还产生了均衡的大小和可解释的环形交叉碰撞模式, 因此保留了 $K = 4$ 。

3.1.1. 聚类 1 (C1) – 环形交叉口的进入和产量相关冲突: 图 4 (a) 中的聚类 1 占数据集中所有环岛碰撞事故的 38%, 主要突出了与不当进入和让行

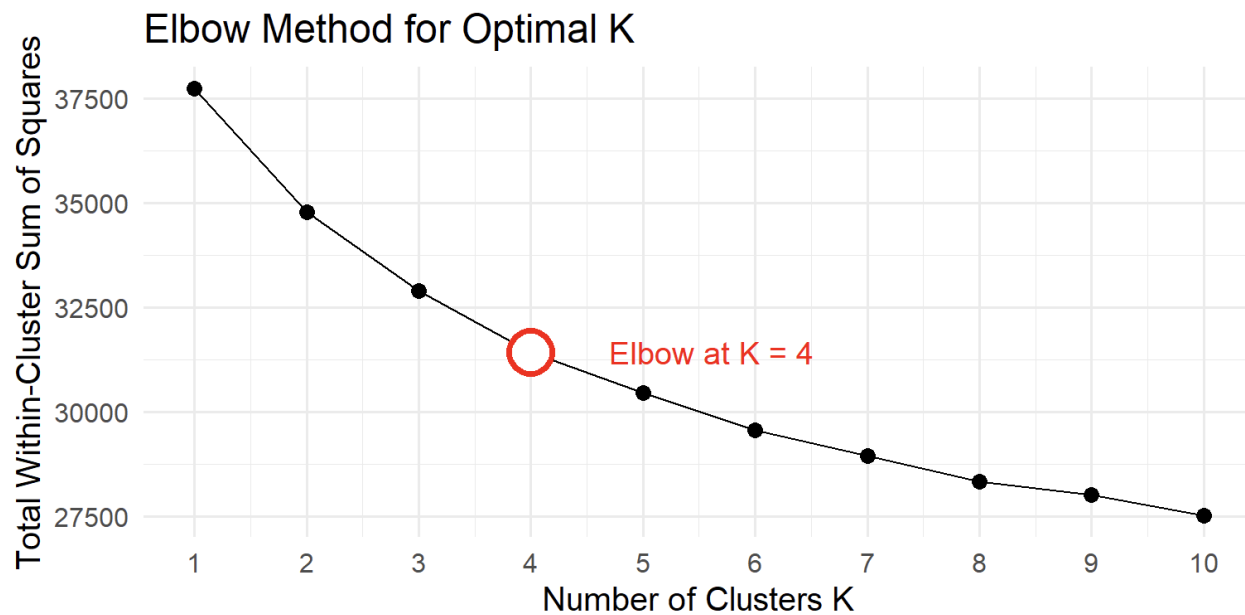
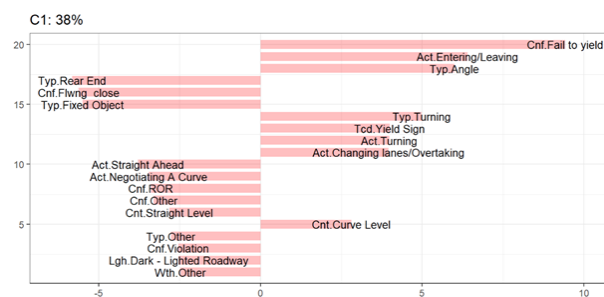
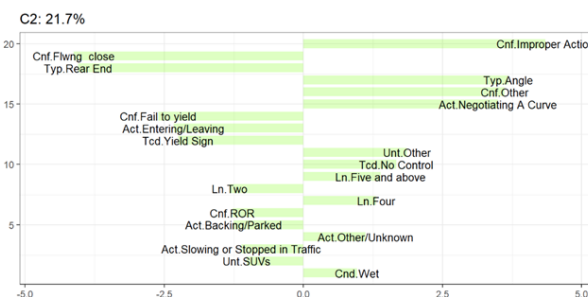


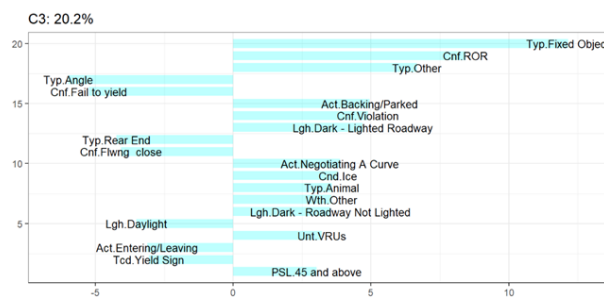
图 3: 肘部法用于识别最优聚类数量



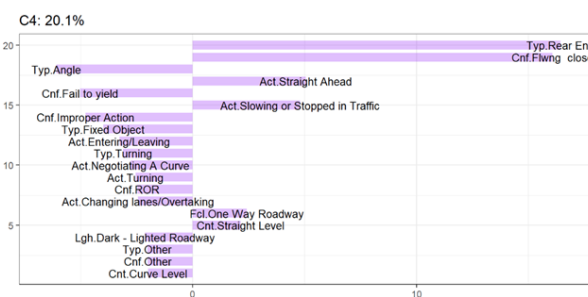
(a) Cluster 1



(b) Cluster 2



(c) Cluster 3



(d) Cluster 4

图 4: 环岛碰撞事故的聚类分析

行为相关的碰撞。该聚类中最突出的因素是“未让行”、“进出”和“角度型”碰撞，所有这些都与环岛独特的结构密切相关，要求驾驶员在进入前必须让行并保持对环形交通的关注。在这个聚类中，碰撞通常是由车辆未能适当让行给环岛内的其他交通而进入环岛所触发的。这些碰撞通常导致角度碰撞，这种情况常见于车辆合并时不当或误判了其他道路使用者的速度或间距 (Montella, 2011)。该聚类中“让行标志”交通控制变量的存在支持这一解释，表明即使存在让行标志，一些驾驶员要么没有注意到它们，要么选择不遵守通行权规则。

3.1.2. 聚类 2 (C2) - 在环形交叉口因不当操作和驾驶员误判导致的碰撞事故：图 4(b) 中的聚类 2 包括了所有环岛碰撞事故的 21.7%，主要由驾驶员行为如“不当操作”所塑造。该聚类还包括“角度”类型的碰撞和“处理弯道”，表明这些碰撞可能与在调整到环岛曲线时出现的困惑或突然动作有关。特别是在湿滑条件下，停车距离增加，轻微的误判可能导致碰撞。道路条件包括“潮湿”的情况支持这一解释，因为即使是轻微的抓地力丧失也可能将跟随距离上的错误变成事故 (Saccomanno et al., 2008)。该聚类中存在“五条及以上”和“四条”车道的道路也可能是原因，因为多车道环岛需要更全面关注车道位置及相邻车道车辆的行为，增加了复杂性。

3.1.3. 集群 3 (C3) - 环形交叉路口的固定物体和环境危害碰撞：聚类 3 在图 4(c) 中被突出显示。它占了所有环岛碰撞事故的 20.2%，主要特征是车辆撞击固定物体，尤其是在能见度差或存在环境危险的情况下发生的碰撞。在这个聚类中最显著的变量是“固定物体”类型的碰撞，与“驶出道路 (ROR)”的相关因素紧密相连，这表明驾驶员要么失去了对车辆的控制，要么错误地判断了环岛的曲率和几何形状，导致偏离预定路线。先前的研究指出，在环岛上，较高的接近速度和较大的交通流量也会导致 ROR 类型的碰撞事故 (Burdett et al., 2017)。这个聚类中的环境条件进一步支持了这一模式。在“黑暗 - 照明道路”和“黑暗 - 未照明道路”条件下发生的碰撞事故表明能见度有限，这可能影响驾驶员判断入口和出口点、车道位置或环岛内分隔带、标志

或其他基础设施的位置的能力。当这些情况与道路上的冰、动物存在或其他意外危险结合时，会变得更加严重。这些环境风险可能会分散驾驶员的注意力并延长反应时间，使得在圆形路径上保持控制变得困难。这个聚类中的另一个显著特点是涉及倒车或停放车辆的行为，在环岛上通常是不常见的。这可能表明存在混乱或违规行为，如非法操作或错误的车辆定位，特别是在较大的环岛或接近环岛的区域，比如分隔岛上或出口处。

3.1.4. 集群 4 (C4) - 环形交叉口流动中的追尾和跟车距离碰撞：图 4(d) 中的第 4 组代表了所有环岛碰撞事故的 20.1%，并与因跟车距离过近导致的追尾碰撞密切相关，尤其是在环岛内的常规或低速交通流中。这一组中最具有影响力的变量是“追尾”类型的碰撞和作为影响因素的“跟车太近”，这明显指向了由于驾驶员未能保持足够的安全距离而导致车辆相撞的情况。这种类型的事在环岛上很常见，因为那里经常有走走停停的移动。当车辆进入或让行时可能会突然减速，或者在出口车道变换过程中也可能减速。如果驾驶员分心或认为交通会持续流动而不间断，则可能无法及时作出反应。这个模式通过“交通中减速或停止”的变量得到了强调，表明许多此类事故发生在车辆减慢速度或暂时停顿的情况下，这些情况需要后车驾驶员保持持续的注意力。先前的研究还指出，在环岛上追尾碰撞很常见，尤其是在跟随前车时距离太近 (Burdett et al., 2016)。

3.2. SHAP 解释

SHAP 值帮助解释机器学习模型中的每个特征如何影响预测结果，在本研究中，是环形交叉口碰撞的严重程度。蜂群 SHAP 图通过水平轴上基于每个特征的 SHAP 值显示个别碰撞实例。正 SHAP 值表示该特征有助于更高严重性的预测，而负值则表明该特征降低了严重性。每个点代表一次碰撞案例，并根据预测的碰撞严重程度进行颜色编码：红色表示死亡或严重受伤 (KA)，蓝色表示中度或轻微受伤 (BC)，绿色表示无伤害 (O) 的碰撞。特征按其在聚类中的总体影响垂直列出，这有助于识别哪些变量对影响碰撞结果最重要。

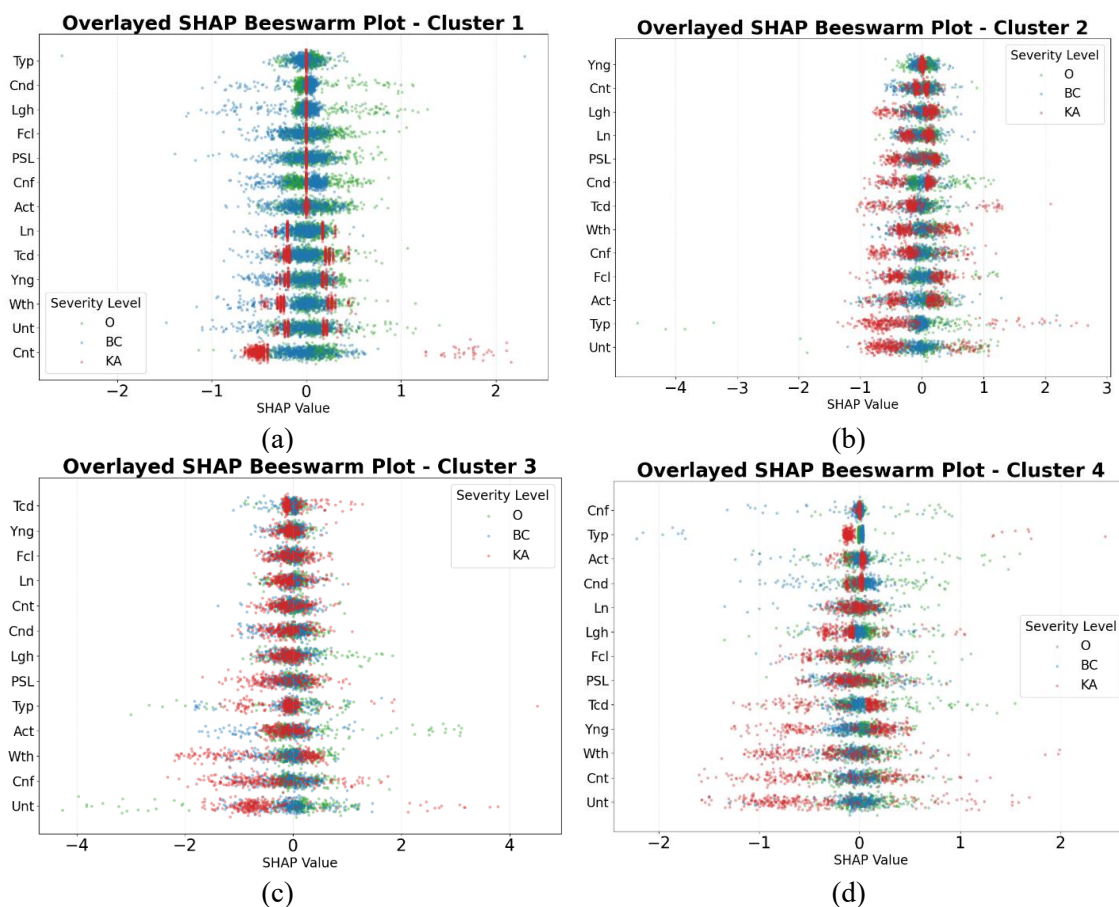


图 5: SHAP 值图显示了按聚类划分的碰撞严重性特征重要性。

3.2.1. 聚类 1: 聚类 1 包括通常在进入环形交叉路口时发生的碰撞，通常是由于未让行或并线错误导致的。图 5(a) 中的 SHAP 分析显示了哪些特征对模型预测此聚类碰撞严重性的影响最大。SHAP 分析确定单元类型是重要特征之一。涉及大型车辆（如卡车或公交车）的碰撞往往会带来更严重的后果，这可能是由于它们的质量和在并线时有限的操作灵活性。贡献因素也被认为很重要，尤其是在未让行的情况下。这些情况通常导致角度型碰撞，SHAP 值表明这种类型的碰撞更容易造成严重伤害。碰撞类型和道路轮廓也影响了严重性预测。

3.2.2. 聚类 2: 集群 2 包括与不当驾驶行为和环岛内误判相关的事故，特别是在湿滑或多车道条件下。图 5(b) 中的 SHAP 分析显示了各种特征如何影

响模型对该组事故严重性预测的影响。SHAP 分析确定车辆类型为重要特征之一。涉及特定车型的事故，特别是那些控制或能见度较低的车型，在该集群中与更严重的后果相关联。驾驶员行为也被发现会影响事故的严重程度。不当操作如突然变道或跨车道转弯在多车道环岛中较为常见，并可能导致更为严重的事故，这反映在 KA 结果的正 SHAP 值上。天气状况被视为一个重要环境因素。发生在雨天或湿滑路面上的事故与更严重的后果相关联，可能是由于轮胎抓地力降低和制动距离增加所致。

3.2.3. 聚类 3: 集群 3 包括大部分涉及车辆撞击固定物体或在恶劣环境条件下失控的事故，例如照明不良或结冰路面。图 5(c) 中的 SHAP 分析突显了不同特征如何对这一组交通事故严重程度的模型预

测产生贡献。严重的后果通常与某些类型的车辆有关,这些车辆在滑溜或多雾环境下可能更难控制。红点的正值 SHAP 值表明这些车辆增加了发生更为严重事故的可能性。贡献因素在这个集群中也起到了重要作用。涉及道路偏离或失控的案例对严重性预测产生了强烈影响。这些行为与固定物体碰撞一致,在这种情况下,司机可能会因分心或能见度低而驶离道路。SHAP 分析进一步确认了天气和照明条件是重要的特征。在黑暗或多冰环境中发生的故事更可能为严重,这从红点 SHAP 值的集中可以看出。这些情况会降低反应时间和能见度,使与固定物体的撞击更加危险。

3.2.4. 聚类 4: 聚类 4 反映了主要由于环形交叉路口交通中跟车距离过近而引起的追尾碰撞。这些事故通常发生在常规流量中,车辆意外减速时发生。图 5(d) 中的 SHAP 分析确定了道路轮廓和天气条件为影响因素。在平直路段上以及恶劣天气下发生的事故更有可能导致更高的严重性,这可能是由于牵引力减少和制动延迟所致。贡献因素对形成严重程度起着重要作用。该聚类中的许多事故都与跟车距离过近有关,这种行为增加了交通突然减速时发生追尾碰撞的可能性。SHAP 值显示此类行为始终推动模型预测更严重的后果。其他特征,如事故类型和驾驶员操作也做出了有意义的贡献。“追尾”作为事故类型以及“在交通中减速或停止”作为操作经常与较高的 SHAP 值相对应,强化了这样一个模式:即在循环流量中的不注意或激进驾驶是严重程度的关键驱动因素。

4. 讨论

研究结果表明,环岛事故可以归类为少量一致的模式,并且伤害严重程度与特定的照明、路面状况、速度环境、事故类型和操作组合相关。对于公共机构而言,这支持了通过模式筛选地点、将对策与主导风险对齐(例如,在入口处设置让行控制和车道引导;在潮湿路面上提高摩擦力和排水功能;在黑暗和固定物体碰撞同时发生的区域升级照明)等实际行动,并使用相同的模式逻辑评估结果。从信息系统角度看,该研究展示了如何将可解释的分析管道(CCA 用于模式发现,随后由带有 SHAP 解释

的监督模型)嵌入决策工具中。两个设计要点尤为突出:(i) 将网络筛选的模式级别视图与地点诊断的案例级别解释配对;(ii) 与重要性排名一起展示支持计数和不确定性以维持适当的信任。这种方法适用于仪表板、年度安全报告和事后审查。

5. 结论

本研究提出了一种数据驱动的方法,利用机器学习和可解释数据分析技术来评估环岛交通事故模式及严重程度结果。通过应用 CCA 作为降维和无监督聚类方法,该研究能够揭示分类事故数据中的潜在分组。这些分组揭示了不同事故场景之间的有意义区别,如入口处未让行、操作不当、环境危害以及跟车距离错误,在四个独特的聚类中有所体现。为了进一步了解每个聚类中影响事故严重程度的因素,使用 SHAP 解释了在事故严重性预测模型中各个特征的相对重要性。

研究结果显示,涉及大型车辆的入口相关碰撞和未让行行为特别与较高的严重性结果相关。在车道环形交叉路口内部由于不当操作导致的碰撞,在潮湿条件下尤为危险。环境因素如照明不足、高速行驶和失去控制在固定物体碰撞中扮演了核心角色。在循环交通中因跟车距离过近而发生的追尾事故频繁发生,并且往往因为驾驶者注意力不集中和突然减速而加剧。这些见解为交通工程师和政策制定者提供了实际意义。增强入口点的标志和几何设计、强制车道纪律、安装更好的照明设备以及促进环形交叉路口内的安全行车距离是建议采取的一些策略。本研究直接回答了三个研究问题。对于 RQ1,CCA 揭示了在环形交叉路口出现的四个清晰共现因素模式,包括入口/让行冲突、潮湿条件下的多车道操作问题、在较高速限和黑暗环境中的冲出道路/固定物体事件以及循环交通中的追尾冲突。对于 RQ2,带有 SHAP 的监督严重性模型将这些模式与伤害结果联系起来,显示涉及黑暗、较高车速限制和固定物体碰撞的组合关联到更高的严重性,而晴朗天气和较低速度则关联到较低严重性。对于 RQ3,集群内部 SHAP 结果显示了每个模式中的最强严重性驱动因素,例如入口处未让行和角度类型;潮湿多车道条件下的不当操作;夜间照明不足和固定物体碰撞;以及在循环交通中跟车距离过近。综上所

述，这些发现将特定因素组合与严重性路径联系起来，并指出了实用步骤，包括改进入口的让行控制和车道引导、路面摩擦力和排水系统升级、更好的照明设备安装以及接近速度/排队警告，以减少环形交叉路口中的危害。

从方法论的角度来看，本研究强调了可解释的机器学习技术如何通过不仅识别模式而且还解释预测模型背后的逻辑来提高交通安全研究。使用 CCA 降低维度并揭示高基数分类数据中的结构，然后使用 SHAP 提供详细的解释，为交通和其他公共安全领域中的类似应用提供了有用的模板。

本研究还通过展示一条实用路径，为信息系统提供了透明且可审计的公共部门安全决策支持。CCA 和 SHAP workflow 将无监督模式发现与有监督、可解释的严重性建模联系起来，产生非专家可以阅读并采取行动的结果。它展示了如何将模型逻辑呈现给最终用户（哪些因素重要，在何种条件下以及为什么），这些解释如何与领域机制一致，以及它们如何推动关于对策和项目管理的特定选择。

本研究还承认了一些局限性。CCA 技术虽然在揭示广泛模式方面有效，但并未纳入时间动态因素，例如碰撞特征如何随时间演变。此外，SHAP 解释基于模型训练的关联，并不一定能捕捉因果关系。而且，当前的研究未纳入实时或车辆轨迹数据，这些数据可能会增强对导致碰撞的驾驶员行为的理解。也承认了地理上的局限性，因为本研究仅关注俄亥俄州的数据。未来的研究可以通过整合时间序列建模、多模式传感器或远程信息技术数据以及用于时间序列或基于图形的碰撞表示的高级可解释技术来解决这些问题。扩展框架以包含空间映射和仿真环境可能也有助于支持更安全的环岛设计层面的决策制定。

References

Aboulola, O. I., Alabdulqader, E. A., AlArfaj, A. A., Alsubai, S., & Kim, T.-H. (2024). An automated approach for predicting road traffic accident severity using transformer learning and explainable ai technique. *IEEE Access*, 12, 61062–61072.

- Amini, M., Bagheri, A., & Delen, D. (2022). Discovering injury severity risk factors in automobile crashes: A hybrid explainable ai framework for decision support. *Reliability Engineering & System Safety*, 226, 108720.
- Ashqar, H. I., Alhadidi, T. I., Elhenawy, M., & Jaradat, S. (2024). Factors affecting crash severity in roundabouts: A comprehensive analysis in the jordanian context. *Transportation Engineering*, 17, 100261.
- Breiman, L., Cutler, A., Liaw, A., & Wiener, M. (2002, April). randomForest: Breiman and Cutlers Random Forests for Classification and Regression [Institution: Comprehensive R Archive Network Pages: 4.7-1.2].
- Burdett, B., Alsghan, I., Chiu, L.-H., Bill, A. R., & Noyce, D. A. (2016). Analysis of rear-end collisions at roundabout approaches. *Transportation research record*, 2585(1), 29–38.
- Burdett, B., Bill, A. R., & Noyce, D. A. (2017). Evaluation of roundabout-related single-vehicle crashes. *Transportation research record*, 2637(1), 17–26.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., & Cortes, D. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Gross, F., Lyon, C., Persaud, B., & Srinivasan, R. (2013). Safety effectiveness of converting signalized intersections to roundabouts. *Accident Analysis & Prevention*, 50, 234–241.
- Khasawneh, M. A., Umar, I. K., & Khasawneh, A. A. (2025). Explainable

- artificial intelligence model for accident severity modeling. *Asian Journal of Civil Engineering*, 26(6), 2433–2445.
- Lundberg, S., & Lee, S.-I. (2017, November). A Unified Approach to Interpreting Model Predictions [arXiv:1705.07874 [cs]].
Comment: To appear in NIPS 2017.
- Maji, A., & Ghosh, I. (2025). A systematic review on roundabout safety incorporating the safety assessment methodologies, data collection techniques, and driver behavior [Publisher: Elsevier]. *Safety Science*, 181, 106661.
- Mehta, A. A., Padaria, A. A., Bavisi, D. J., Ukani, V., Thakkar, P., Geddam, R., Kotecha, K., & Abraham, A. (2023). Securing the future: A comprehensive review of security challenges and solutions in advanced driver assistance systems. *IEEE Access*, 12, 643–678.
- Montella, A. (2011). Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. *Accident Analysis & Prevention*, 43(4), 1451–1463.
- Polders, E., Daniels, S., Casters, W., & Brijs, T. (2015). Identifying crash patterns on roundabouts. *Traffic injury prevention*, 16(2), 202–207.
- Rahman, M. A., Das, S., & Sun, X. (2022). Using cluster correspondence analysis to explore rainy weather crashes in louisiana. *Transportation Research Record: Journal of the Transportation Research Board*, 2676(8), 159–173. <https://doi.org/10.1177/03611981221082582>
- Rojas, J. F., Patil, P., Masterson, A. M., Bradley, T. H., Ekki, A. R., & Asher, Z. D. (2024). Automated vehicle lane centering system requirements informed by resilience engineering and a solution using infrastructure-based sensors. *IEEE Access*, 12, 97605–97620.
- Royce, B. (2022). 'fix crash corner'-a roundabout story. *Journal of road safety*, 33(4), 61–67.
- Saccomanno, F. F., Cunto, F., Guido, G., & Vitale, A. (2008). Comparing safety at signalized intersections and roundabouts using simulated rear-end conflicts. *Transportation Research Record*, 2078(1), 90–95.
- Savolainen, P. T., Gates, T. J., Gupta, N., Megat-Johari, M.-U., Cai, Q., Imosemi, S., Ceifetz, A., McArthur, A., Hagel, E. C., Smaglik, E., Michigan State University, Michigan Department of Transportation, & Federal Highway Administration. (2023, January). Evaluating the Performance and Safety Effectiveness of Roundabouts – An Update (Digital/other) (Edition: Final Report). <https://rosap.nhtl.bts.gov/view/dot/68864>
- Steyn, H., Griffin, A., Rodegerdts, L. A., & United States. Department of Transportation. Federal Highway Administration. Office of Safety. (2015, September). Accelerating Roundabout Implementation in the United States - Volume IV of VII: Review of Fatal and Severe Injury Crashes at Roundabouts (tech. rep. No. FHWA-SA-15-072).
- Sussman, N. (2022). Turning around america's traffic crisis one roundabout at a time: Why more roundabouts means safer roads and fewer automobile deaths. *Wash. UL Rev.*, 100, 1241.
- Velden, M. v. d., D'Enza, A. I., & Palumbo, F. (2017). Cluster Correspondence Analysis. *Psychometrika*, 82(1), 158–185.