Accident Analysis and Prevention 215 (2025) 108018

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Understanding e-scooter rider crash severity using a built environment typology: A two-stage clustering and random parameter model analysis

Amirhossein Abdi [a], Steve O'Hern [b,*]

[a] Institute of Transport Studies, Monash University, Australia
[b] Institute for Transport Studies, University of Leeds, UK

ARTICLE INFO

Keywords:
Electric scooter safety
Crash severity analysis
Built environment typology
Multistage modelling

ABSTRACT

E-scooters are an emerging transport mode that is transforming urban mobility; however, their proliferation has raised concerns about safety. This study combines UK e-scooter crash data with built environment characteristics from the crash locations. A two-stage framework was followed: first, a typology of built environments was developed using K-means++; second, crash severity within each cluster was analysed using a random parameter binary logit model. Four built environment clusters were identified: (1) car-centric and mixed-use zones, (2) commercial and industrial zones, (3) intersection-dense areas, and (4) residential and central areas. Collisions with motor vehicles, younger e-scooter riders, and higher speed limits were the most common risk factors across the clusters, with the first two clusters showing a higher impact of these factors on the likelihood of severe crashes. In the first and second clusters, riding on the carriageway significantly increased injury severity. In the second cluster, three collision types were significant, more than in other clusters where only side-impact collisions were significant. This indicates high e-scooter–motor vehicle friction in the second cluster. Among all collision types, head-on collisions increased the likelihood of severe outcomes more than others. In the third and fourth clusters, peak hours were associated with a lower likelihood of severe crashes, while this variable showed the opposite impact in the first cluster. The results highlight that consideration of the surrounding built environment is paramount when analysing e-scooter crash severity, as unique contributing factors were identified specific to each built environment type, along with varying magnitudes or directions of marginal effects.

1. Introduction

Given their simple designs, convenience, and affordability, micromobility vehicles have experienced a rapid increase in popularity, which, as a result, is reshaping urban mobility (O'Hern and Estgfaeller, 2020). Among the available options, electric scooters (e-scooters) have gained notable popularity, both privately owned and used through sharing schemes (Lee et al., 2021). Concurrently there has been growth in e-scooter-related crashes (Badeau et al., 2019; Brauner et al., 2022), and there is a need to understand the risk factors associated with e-scooter usage to facilitate safer uptake.

Studies have investigated the injury risk factors of e-scooter crashes using a variety of data sources, including emergency presentations and hospital admissions, crash reports, and other sources such as news reports, naturalistic riding data, and surveys. In the early stages of e-scooters emergence, hospital and emergency department reports were studied predominantly. Such studies have provided fruitful information on the nature of the severity of the injuries and some widespread influential factors, including demographics of injured riders, nighttime riding, alcohol consumption, and helmet use (Kleinertz et al., 2021; Lavoie-Gagne et al., 2021; Moftakhar et al., 2021; Stigson et al., 2021; Vernon et al., 2020). With the increased number of e-scooter involved incidents and frictions with vehicles on the roads, e-scooters have started to attract the attention of road safety analysts more than before. The results of mining crash records reflect somewhat similar findings to the hospital and emergency room analysis, but with some enhancements, including crash datasets from police reports that provide additional details on crash scene circumstances, vehicle interactions, and roadway and environmental conditions at crash time and location (Cicchino et al., 2021; Gao and Zhang, 2024; Longo et al., 2024; Shah et al., 2021). Other sources have facilitated elaboration on riders risk-taking and driving behaviours as well as e-scooter performance and manoeuvrability in different scenarios (Brauner et al., 2022; Ma et al., 2021; White et al., 2023).

* Corresponding author.
E-mail address: s.ohern@leeds.ac.uk (S. O'Hern).

https://doi.org/10.1016/j.aap.2025.108018
Received 27 November 2024; Received in revised form 13 March 2025; Accepted 22 March 2025
Available online 28 March 2025

In addition to the above-discussed factors, the built environment is shown to significantly shape crash risk. Research on micromobility safety has identified that road design, land use patterns, and infrastructure influence crash frequency and severity (Azimian and Jiao, 2022; Costa et al., 2024; Saha et al., 2020). Despite this, crash datasets often lack such detailed built environment variables, requiring the use of external data sources and data fusion techniques (Costa et al., 2024). While studies on cycling crashes have merged built environment data sources with crash reports (Bi et al., 2023; Heydari et al., 2022), research on e-scooters remains limited as Mehranfar and Jones (2024) emphasise that this gap limits the ability to develop targeted safety interventions for e-scooters and call for further research integrating built environment data.

To date, from a methodological perspective, research on e-scooter rider safety has primarily relied on generalized linear statistical models, such as discrete outcome models (Cicchino et al., 2021; Gao and Zhang, 2024; Gioldasis et al., 2021) and count data models (Azimian and Jiao, 2022; Heydari et al., 2022). These techniques, which incorporate fixed and random parameters, aim to identify significant crash risk factors based on crash reports and address potential underlying heterogeneity. However, the heterogeneity in crash data arising from the varying characteristics of e-scooter crash locations has not been explored and requires an additional level of analysis alongside crash modelling. This paper addresses this gap by retaining the highly interpretable random parameter discrete outcome models for crash severity analysis while introducing another layer of heterogeneity exploration for built environment factors. Using a two-stage hybrid approach, K-means++ is first applied as a clustering method to partition continuous features, a suitable choice given that all built environment factors in the dataset are continuous (Aggarwal and Reddy, 2018; Arthur and Vassilvitskii, 2007; Ikotun et al., 2023). Discrete outcome models specific to each identified group of crash locations are then applied in the next stage to analyse the varying effects of risk factors and possible heterogeneity among them (Mannering et al., 2016).

Considering this, the main objectives of this paper are: (1) to integrate e-scooter crash data with built environment characteristics around crash sites using OpenStreetMap; (2) to apply a two-stage framework that first identifies built environment clusters using an unsupervised machine learning technique; (3) to subsequently identify risk factors specifically influencing crash severity through interpretable random parameter discrete outcome models. Fig. 1 summarises the workflow of this research based on the described objectives.

2. Literature review

The growing role of e-scooters in urban transport has raised safety concerns, urging researchers to explore their underlying risk factor so that targeted safety strategies can be developed (Mehranfar and Jones, 2024). Micromobility safety research has identified crashes to be the result of complex interactions among numerous factors, including human factors, vehicle-related factors, and infrastructural and environmental conditions at the time and place of the crash (Costa et al., 2024; Gao and Zhang, 2024). Table 1 lists commonly reported e-scooter crash severity risk factors, based on a review of the literature. From a general viewpoint, current research on e-scooter safety can be reviewed under three categories: studies using emergency presentation and hospital admission data (Section 2.1), studies focused on crash datasets (Section 2.2), and research relying on surveys, naturalistic riding data, and observations (Section 2.3).

Because the field of research on e-scooter safety is in its infancy, most studies have relied on hospital admission and emergency data, as e-scooter crash records are scarce. The structure used in this section allows for a comprehensive examination of existing studies while recognising that over time, research based on crash datasets alone will become more frequent. Also, consistent with the focus of this research, we have reviewed works on micromobility safety that have fused the built environment with crash data, as well, in Section 2.2. Finally, Section 2.4 delineates the research gaps that this investigation seeks to address, drawing upon the analysed literature.

2.1. Emergency presentation and hospital admission data

Multiple studies from around the world have analysed e-scooter injuries using hospital data. For instance, Vernon et al. (2020) in Atlanta, Georgia; Kleinertz et al. (2021) in Hamburg, Germany; and Moftakhar et al. (2021) in Vienna, Austria, reported a significant rise in injuries during nighttime and weekends. Moftakhar et al. (2021) found that injuries peaked between 8:00p.m. and 1:59 a.m., often involving riders under the influence of alcohol. Notably, Vernon et al. (2020) identified that 32 % of injuries occurred during a nighttime ban between 9p.m. and 4 a.m. Kleinertz et al. (2021) compared temporal patterns of e-scooter and bicycle crashes and found that bicycle crashes were more commonly associated with commuting, while e-scooter crashes were increasingly associated with leisure, leading to an increase in injuries during weekends and nighttime.

When considering spatial factors, in Helsinki, Finland, Dibaj et al. (2024) found that injuries are more concentrated in the city centre, however, as service areas expanded crashes became more dispersed. In Washington, D.C., Cicchino et al. (2021), identified that injuries were more commonly sustained on sidewalks (58 %) as opposed to on the roads (23 %). However, injuries sustained on the road were on average more severe. Both Kleinertz et al. (2023) and Stigson et al. (2021) further highlighted poorly maintained and uneven road surfaces as contributing factors.

Demographic characteristics, like age and gender, along with helmet use were associated with injury severity. Numerous studies have found males constitute a higher proportion of injured e-scooter riders (Jones et al., 2023; Kleinertz et al., 2023; Lavoie-Gagne et al., 2021; Moftakhar et al., 2021; Stigson et al., 2021). Two studies identified specific age groups as being more prone to severe outcomes. Stigson et al. (2021) found that males aged 15–34 were more frequently involved in crashes, with their low rate of helmet-wearing contributing to the increased

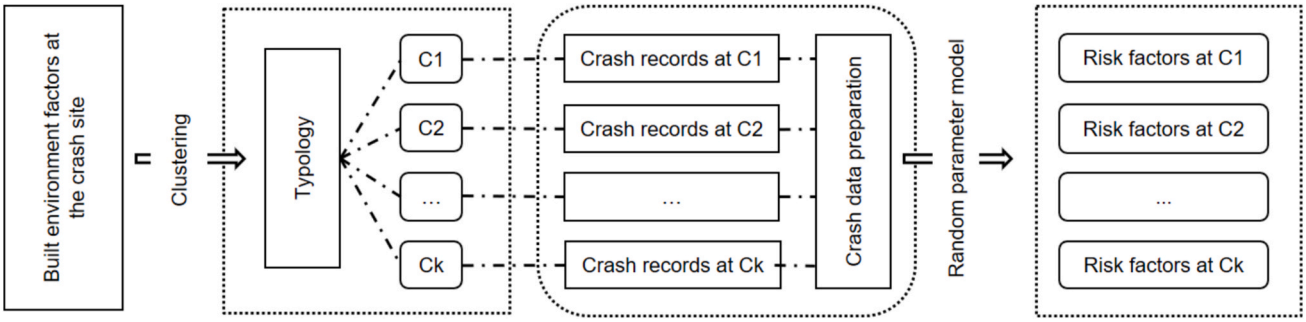

Fig. 1. Research design and workflow.

Table 1
Summary of e-scooter crash severity risk factors.

Risk Factor	Data Source	Finding(s)	Author(s)
Nighttime and Darkness	EPHA* dataset	Associated with increased injury severity	(Kleinertz et al., 2021; Lavoie-Gagne et al., 2021; Moftakhar et al., 2021; Vernon et al., 2020)
	Crash dataset		(Gao and Zhang, 2024; Longo et al., 2024)
	Other (News reports)		(Yang et al., 2020)
Alcohol Consumption	EPHA dataset	Associated with increased injury severity	(Jones et al., 2023; Kleinertz et al., 2023, 2021; Lavoie-Gagne et al., 2021; Moftakhar et al., 2021)
	Other (Public press portal)		(Brauner et al., 2022)
Gender	EPHA dataset	Male riders are more frequently involved in severe injury crashes	(Stigson et al., 2021; Vernon et al., 2020)
	Crash dataset		(Gao and Zhang, 2024)
	Other (News reports)		(Yang et al., 2020)
Age	EPHA dataset	Age is found to be a significant factor; however, the cohorts used by researchers vary, with some studies including younger cohorts and others including older adults as risk factors for injury	(Lavoie-Gagne et al., 2021; Moftakhar et al., 2021)
	Crash dataset		(Gao and Zhang, 2024)
Weekday vs. Weekend	EPHA dataset	Results differ among studies: Weekend crashes are reported to be more frequent and severe (Stigson et al., 2021; Vernon et al., 2020), while some studies find that weekday crashes tend to be more severe (Gao and Zhang, 2024)	(Stigson et al., 2021; Vernon et al., 2020)
	Crash dataset		(Gao and Zhang, 2024)
Helmet Use	EPHA dataset	Reduces the severity of head injuries, though usage is low across studies, and its impact on overall injury severity varies	(Jones et al., 2023; Kleinertz et al., 2023; Lavoie-Gagne et al., 2021; Moftakhar et al., 2021; Stigson et al., 2021; Vernon et al., 2020)
Pavement Condition	EPHA dataset	Poor pavement conditions generally increase the risk of injury; however, a study by Gao and Zhang (2024) found that poor pavement was associated with lower injury severity in collisions involving motor vehicles. Nonetheless, this relationship was not significant in single-vehicle e-scooter crashes	(Kleinertz et al., 2023)
	Crash dataset		(Gao and Zhang, 2024; Longo et al., 2024)
	Other (Naturalistic riding data)		(Ma et al., 2021)
Location of Incident	EPHA dataset	Injuries on roads (i.e., on the carriageway) tend to be more severe compared to off-road sites such as sidewalks	(Cicchino et al., 2021)
	Crash dataset		(Gao and Zhang, 2024)
Interaction with Motor Vehicles	Crash dataset	Associated with increased injury severity (Shah et al., 2021; Yang et al., 2020).	(Shah et al., 2021)
	Other (News reports)		(Yang et al., 2020)
Frequent Riders	EPHA dataset	Associated with more severe injuries	(Cicchino et al., 2021)
Handling Issues and Malfunctioning Scooters	EPHA dataset	Associated with increased risk and injury severity	(Kleinertz et al., 2023; Stigson et al., 2021)

*EPHA: Emergency presentation and hospital admission

severity of head injuries. In San Diego, California, Lavoie-Gagne et al. (2021) reported that individuals over 40 years and under the influence of alcohol or other substances had an increased risk of severe injuries. Additionally, when comparing e-scooter and bike users, Kleinertz et al. (2021), observed that e-scooter riders involved in crashes were younger and used helmets less frequently than cyclists.

Alcohol consumption is another risk factor associated with e-scooter crashes. In Essex, UK, Jones et al. (2023) found that intoxicated riders had more severe injuries and required more comprehensive medical care post-crash. Alcohol was also mentioned in association with other risky behaviours; for instance, riders under the influence of alcohol demonstrated a lack of helmet use (Kleinertz et al., 2021). Further, the literature indicates that riding under the influence of alcohol most often happens at night and over weekends (Kleinertz et al., 2021; Moftakhar et al., 2021).

The increase in e-scooter injuries over time is a concern highlighted by Moftakhar et al. (2021) who documented a sharp rise in injuries from 13 cases in 2018 to 116 in 2019. This increase in cases represents the growing popularity of e-scooters as well as emerging safety concerns, leading some jurisdictions to implement bans and restrictions on e-scooters (Dibaj et al., 2024).

2.2. Crash dataset studies

E-scooters safety literature based on analysis of crash datasets tends to align with studies using hospital datasets, however, it also provides some further insights due to the factors present in crash records. Shah et al. (2021) analysed police-reported e-scooter crashes in Nashville, US. They found that e-scooter crashes were mainly concentrated in downtown areas. In addition, nighttime conditions, and younger riders were increasingly prone to crashes, consistent with the trends seen with emergency room admissions (e.g., Dibaj et al. (2024)). When considering crash types, Shah et al. (2021), identified cross traffic and crashes with right turning motor vehicles as common crash mechanisms.

Longo et al. (2024) examined e-scooter collisions in Bari, Italy, based on the application of a binary logit model for the outcome of non-injury and injury collision categories. Again, the results found that nighttime and road surface problems are associated with an increased likelihood of injury, consistent with the studies conducted by Kleinertz et al. (2023, 2021), Moftakhar et al. (2021), Stigson et al. (2021), and Vernon et al. (2020). The results also showed that, compared with single e-scooter crashes, e-scooter collisions with pedestrians have significantly higher odds of injury. Regarding the crash location, crashes frequently occurred on undivided road segments, though the severity was higher for divided

roads and was associated with their higher operating speeds.

Gao and Zhang (2024) studied UK e-scooter collisions within a two-year period. Random parameter models were developed for single e-scooter and two-vehicle e-scooter collisions. The application of the random parameter models improved earlier works, since it accounted for unobserved heterogeneity. Results of both single- and two- vehicle crashes identify nighttime, darkness, weekdays, males, and older riders as common risk factors. Also, adverse road surfaces increased the likelihood of severe outcomes in single vehicle crashes, while decreasing the severity for two vehicle crashes. Junctions and higher speed limit roads were also influential factors for the severity of e-scooter motor vehicle collisions.

In addition to the factors discussed above, the built environment plays an important role in the safety of e-scooter users; however, built environment characteristics are rarely recorded in crash reports (Chen and Shen, 2016). Road infrastructure is particularly important, for example, studies conducted by Morrison et al. (2019) and Gao and Zhang (2024) suggested that separated cycling lanes with protected infrastructure reduce the risks of crash involvement as well as severity levels for both cyclists and e-scooter riders.

Beyond road infrastructure, according to Chen and Shen (2016) and Branion-Calles et al. (2020), mixed land use and dense built environments are associated with increased crash rates. This is likely due to increased pedestrian and traffic flow through such areas. Similarly, Azimian and Jiao (2022) found that areas with increased activity, such as with more restaurants and educational centres experience more e-scooter crashes, while areas with higher incomes and better sidewalks were associated with fewer crashes. Similar results were observed by Heydari et al. (2022) who found that higher crash frequencies correlated with increased walking, cycling, higher crime rates, and larger Black, Asian, and Minority Ethnic populations, while areas with more greenspace saw fewer incidents.

2.3. Other works

In addition to crash records and emergency room visits, a range of other data sources and research methods have been used to enhance the understanding of e-scooter safety.

Yang et al. (2020) extracted news reports on 169 e-scooter crashes that took place in the United States over three years. Their results reflected a notable rise in reporting of crashes from 2018, with the highest proportion of news reports from California, Indiana, Texas, Florida, and Georgia. Articles predominantly involved male riders, particularly those aged 18–40 years, with incidents mainly reported on arterial roads and at intersections. Over half of these crashes involved severe injuries or death, particularly amongst children and teenagers, although this may reflect some reporting bias. Nighttime riding was increasingly reported for fatal crashes, and helmet non-use and alcohol were associated with higher injury severity cases. In a more recent content mining work, Brauner et al. (2022) extracted police reports on e-scooter crashes from a public German press portal. They found similar risk factors including nighttime riding, riding while intoxicated with alcohol, not wearing a helmet, as well as riding in pairs.

Several studies have used instrumented and naturalistic approaches to study e-scooter safety. Ma et al. (2021) describes instrumenting an e-scooter with a kinematic data acquisition system in Norfolk, Virginia. The findings from this study provided an environmental impact evaluation of riding behaviour and established that e-scooter riders are exposed to worse vibration events compared to cyclists, particularly on concrete pavements and sidewalks. Work by Ma et al. (2021) was limited to only one rider and performed under limited scenarios, which was addressed in a more recent work by White et al. (2023) at Virginia Tech, US. In their naturalistic riding study 3500 h of riding data was passively recorded over a six-month period. Analysis revealed infrastructural factors, including surface transitions and fixed element conflicts, caused a significant portion of all incidents. Aggressive riding and group riding increased risk, as did riding in conditions other than daylight. Additionally, dry road surface conditions were shown to raise risk, which was attributed to less cautious riding.

Gioldasis et al. (2021) conducted a survey among e-scooter riders in Paris, France, investigating risk-taking behaviours including riding under the influence of alcohol or drugs and using smartphone devices while riding. The results indicate that risky behaviours were more likely to be performed by younger and male riders, and riders undertaking longer trips. In another study concerning unsafe rider behaviour, Haworth et al. (2021) observed shared and privately owned e-scooters in Brisbane, Australia. The findings showed that e-scooters were mostly ridden on footpaths, which created frequent interactions with pedestrians (about 40 % of cases), however, actual conflicts which may turn into collisions were rare (less than 2 %). Amongst riders using shared e-scooters, half were performing an illegal behaviour, including not wearing a helmet, riding in pairs, and riding on the road. Observations of illegal behaviours were notably less amongst private owners (12.2 %).

2.4. Identified gaps in e-scooter safety research

Despite the growing popularity of e-scooters and the corresponding increase in hospital reported injuries, research on e-scooter safety remains limited. As demonstrated by the literature review, previous studies have primarily focused on post-crash reports from police and hospital datasets. Furthermore, these studies rarely address unobserved heterogeneity, with the exception of some notable studies like Gao and Zhang (2024). This paper builds on these efforts by developing random parameter models, enabling a more accurate assessment of the risk factors for e-scooters while considering built environment factors.

The built environment plays a crucial role in the safety of vulnerable road users. However, the existing literature on e-scooter crash severity has predominantly focused on temporal factors, demographics, and risky behaviours. A significant challenge in this area is that crash records often lack detailed contextual information, hindering analysis and leading to the underrepresentation of built environment factors in the literature (Costa et al., 2022). To address this challenge in this paper, we use data fusion techniques to combine crash data with open-source built environment data. Additionally, while our review of the literature highlights that some studies have examined frequency of e-scooter crashes from a built environment perspective (Azimian and Jiao, 2022; Heydari et al., 2022), limited analysis has been conducted in this respect for e-scooter crash severity. This study fills this gap by identifying built environment types and their associated risk factors for e-scooter crash injury severity.

A further limitation of previous studies examining the impact of built environment factors on road safety is the use of traditional generalized linear models (Azimian and Jiao, 2022; Heydari et al., 2022), treating environmental factors as contributors to crashes. In contrast, our research takes a different approach by exploring the types of built environments surrounding crash sites and then identifying the significant risk factors associated with each environmental context. To this aim, we applied a hybrid modelling approach, integrating an unsupervised machine learning technique, which identifies built environment types with a random parameter discrete outcome model. This approach allows us to identify significant factors associated with e-scooter crashes for each identified built environment cluster.

3. Data

3.1. Crash data

The study analysed e-scooter crash data from the UK between 2020 and 2023, obtained from the Department for Transport's STATS19 dataset (Department for Transport, 2024). Fig. 2 (a) shows the locations of e-scooter crashes across the UK, and Fig. 2 (b) illustrates the city-wise distribution of crashes categorized by severity level. The database

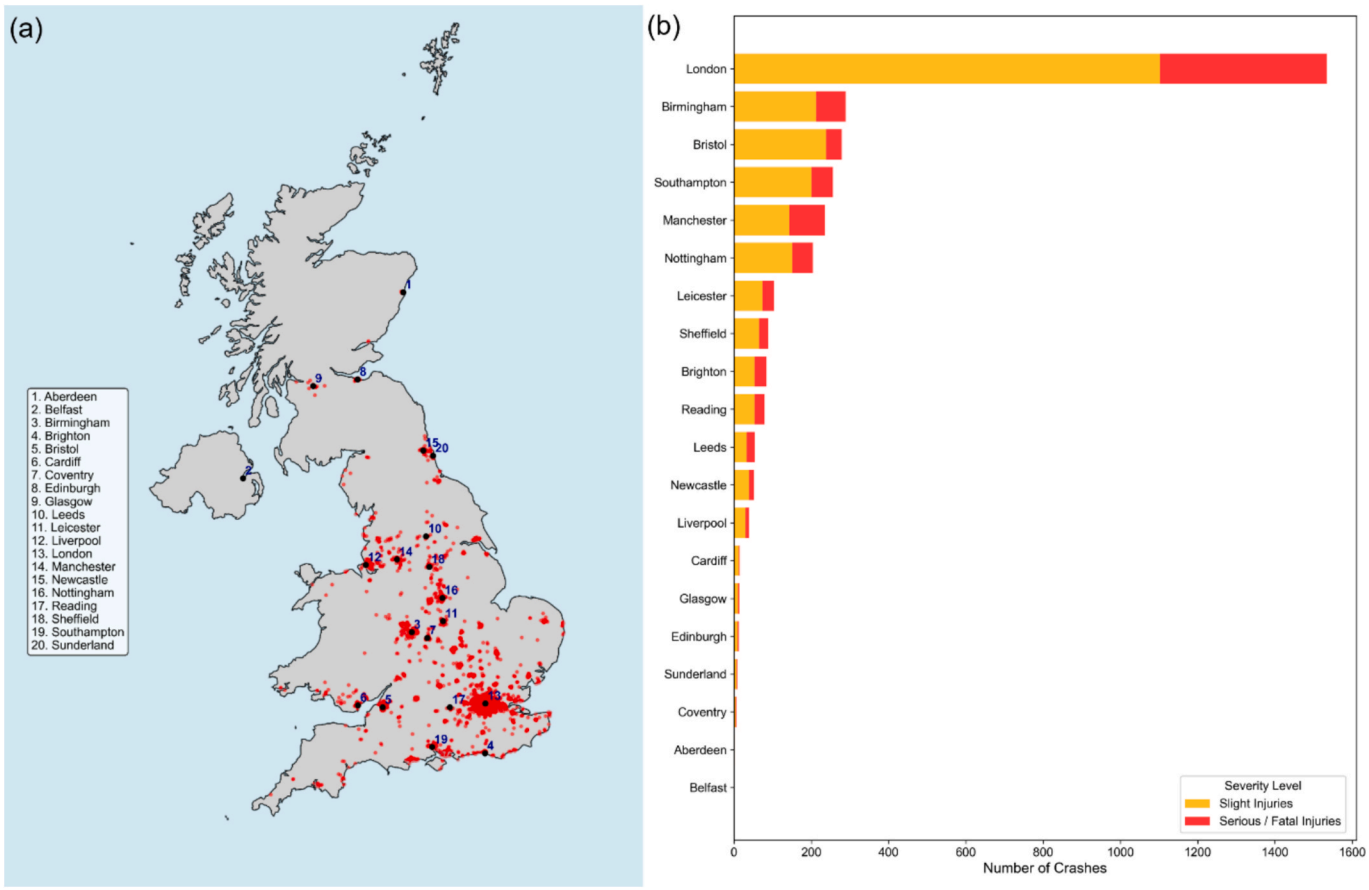

Fig. 2. (a) E-scooter crash locations across the UK (b) City-wise distribution of e-scooter crashes by severity level.

categorizes crash severity into three levels: slight, serious, and fatal crashes. A fatal crash is defined where a person (here, the e-scooter rider) dies within 30 days of the collision. Serious injury crashes involve a person who is hospitalised and requires medical treatment. Slight injuries do not require hospital treatment (Rella Riccardi et al., 2022).

In this study, the data quality was improved through a three-step method. First records with a large amount of missing information were filtered and removed; such records represented less than 5 % of the dataset. Next cases with missing values which could be replaced from other data sources, were addressed. For example, we used weather APIs, such as OpenWeatherMap (OpenWeatherMap, 2024) and Open-Meteo (Open-Meteo Team, 2024) to supplement cases with missing weather conditions, and a Python library called Astral (Kennedy, 2022) to impute diurnal times. Finally, missing data which could not be supplemented were treated using the KNN imputation technique. This approach had been previously applied by Abdi and O'Hern (2024) for imputation of crash datasets. This method computes the mean of the closest k neighbours for continuous variables, while for categorical variables, it assigns the most frequent value observed among the neighbours (Zhang, 2012). If no frequent value is clear, a value is chosen randomly. The final dataset contained 3,550 entries, including 2,352 minor injury crashes, 1,146 serious injury crashes, and 52 fatal crashes. Table 2 summarises the variables extracted from the crash dataset, cross-tabulated by injury severity level.

3.2. Measuring the built environment

As demonstrated in the review of the literature, the built environment is closely linked to e-scooter rider safety. However, where previous studies have included the built environment as regression variables, in this study we aim to translate the built environment variables by creating a typology of environments for e-scooter crashes. In this approach, first we selected the dimensions of the built environment to include in the development of the clusters. Next, we selected a buffer size for cases, which influences the range of built environment variables included in the analysis. In reviewing the literature, a range of buffer sizes have been previously applied. Costa et al. (2024) used a 25-meter radius because the variables of interest were primarily related to immediate surroundings, such as urban designs and nearby buildings. However, broader contextual factors like land use and road network factors have led researchers to use larger buffers, as seen by Abdi et al. (2022), Mathew et al. (2022), and Wu et al. (2023), who utilized radii from 200 to 3000 m. The nature of variables in this study aligns more closely with the works of Mathew et al. (2022), and Wu et al. (2023), as such we obtained data on the built environment within the radius of 250 m around crash sites and validated this choice using clustering performance measures, as discussed in the Results and Discussion Section. Using the geographic coordinates of crash locations, we extracted information from OpenStreetMap (OSM). Table 3 summarizes the descriptive statistics of the built environment variables used and provides their definitions.

4. Methodology

In this study, the innovation lies in applying clustering and random parameter modelling to an unexplored problem, where the built environment surrounding e-scooter crashes shapes its risk factors in terms of their significance and magnitude of effect on crash severity. This requires the identification of homogeneous clusters of crash sites that are similar in terms of their built environment characteristics. Latent class models, including Latent Class Analysis (LCA) and Latent Profile Analysis (LPA), are widely used for data clustering in the field of accident

Table 2
Descriptive statistics of crash factors cross-tabulated by severity levels.

Explanatory Variables	Slight Injury	Serious Injury	Fatal
	n (%)	n (%)	n (%)
Road Type			
roundabout	102 (4.34 %)	52 (4.53 %)	2 (3.85 %)
one-way street	151 (6.42 %)	79 (6.89 %)	4 (7.69 %)
carriageway	2039 (86.69 %)	1004 (87.61 %)	45 (86.54 %)
slip road	60 (2.55 %)	11 (0.96 %)	1 (1.92 %)
Speed Limit (mph)			
equal or less than 30	1186 (50.42 %)	405 (35.34 %)	15 (28.85 %)
31–49	962 (40.89 %)	602 (52.52 %)	28 (53.85 %)
equal or greater than 50	204 (8.67 %)	139 (12.13 %)	9 (17.31 %)
Road Surface Condition			
adverse condition (road surface in a poor condition, e.g., being wet, snowy, flooded, oily, muddy, or frosted)	206 (8.76 %)	166 (14.49 %)	16 (30.77 %)
normal condition	2146 (91.24 %)	980 (85.51 %)	36 (69.23 %)
Light Condition			
daylight	1602 (68.11 %)	705 (61.51 %)	29 (55.77 %)
darkness – lights lit	504 (21.43 %)	299 (26.09 %)	15 (28.85 %)
darkness – lights unlit	196 (8.34 %)	101 (8.81 %)	6 (11.54 %)
darkness – no lighting	50 (2.12 %)	41 (3.58 %)	2 (3.85 %)
Weather Condition			
clear	1499 (63.73 %)	500 (43.62 %)	17 (32.69 %)
cloudy	699 (29.71 %)	402 (35.08 %)	18 (34.62 %)
adverse (snow/rain/fog)	154 (6.55 %)	244 (21.30 %)	17 (32.69 %)
Crash Time			
late night/early morning (00:00–05:59)	151 (6.42 %)	101 (8.81 %)	8 (15.38 %)
morning commute hours (06:00–09:59)	502 (21.34 %)	298 (26.00 %)	15 (28.85 %)
midday (10:00–14:59)	698 (29.68 %)	299 (26.09 %)	10 (19.23 %)
afternoon commute hours (15:00–17:59)	594 (25.25 %)	202 (17.62 %)	9 (17.31 %)
evening (18:00–20:59)	303 (12.88 %)	142 (12.39 %)	6 (11.54 %)
night (21:00–23:59)	104 (4.42 %)	104 (9.07 %)	4 (7.69 %)
Day of Week			
weekday	1854 (78.83 %)	903 (78.80 %)	41 (78.85 %)
weekend	498 (21.17 %)	243 (21.20 %)	11 (21.15 %)
Season			
spring	608 (25.85 %)	298 (25.96 %)	14 (26.92 %)
summer	797 (33.89 %)	401 (34.98 %)	17 (32.69 %)
autumn	595 (25.30 %)	300 (26.17 %)	13 (25.00 %)
winter	352 (14.97 %)	147 (12.82 %)	8 (15.38 %)
E-scooter Rider Sex			
male	1701 (72.38 %)	798 (69.65 %)	35 (67.31 %)
female	651 (27.62 %)	348 (30.35 %)	17 (32.69 %)
E-scooter Rider Age			
less than 35	1234 (52.47 %)	626 (54.62 %)	29 (55.77 %)
35–45	768 (32.65 %)	367 (32.02 %)	18 (34.62 %)
greater than 45	350 (14.88 %)	153 (13.35 %)	5 (9.62 %)
Second Party Type			
e-scooter only crash	545 (23.17 %)	308 (26.88 %)	0 (0.00 %)
pedestrian	184 (7.82 %)	46 (4.01 %)	0 (0.00 %)
bicycle	94 (4.00 %)	21 (1.83 %)	0 (0.00 %)
motorcycle	103 (4.38 %)	49 (4.28 %)	8 (15.38 %)
car/heavy vehicle	1426 (60.63 %)	722 (63.00 %)	44 (84.62 %)
Manners of Collision with Vehicle			
head-on	282 (18.44 %)	103 (13.36 %)	10 (19.23 %)
side-impact with front of the e-scooter	338 (22.11 %)	155 (20.10 %)	7 (13.46 %)
side-impact with front of the 2nd party vehicle	292 (19.10 %)	149 (19.33 %)	11 (21.15 %)
sideswipe with opposite direction	126 (8.24 %)	68 (8.82 %)	11 (21.15 %)
sideswipe with same direction	219 (14.32 %)	61 (7.91 %)	9 (17.31 %)
rear-end with front end of the e-scooter	122 (7.98 %)	111 (14.40 %)	3 (5.77 %)
rear-end with front end of the 2nd party vehicle	150 (9.81 %)	124 (16.08 %)	1 (1.92 %)

research. According to Li et al. (2024), the primary focus of LCA is on categorical variables; conversely, LPA is applicable to continuous data yet faces some challenges, including the tendency to converge on local optima and complications in satisfying statistical assumptions, especially when the number of input variables expands. In light of the 19 continuous variables analysed in this study, the K-means++ clustering algorithm was selected due to its higher efficiency compared to traditional K-means, its assumption-free nature as a machine learning method, and its ability to deal with continuous data (Aggarwal and Reddy, 2018; Arthur and Vassilvitskii, 2007).

Despite these advantages, there is still a need to carefully address potential misleading results arising from the increased number of input factors and their varying scales. To handle this, specific steps were taken during data preparation and modelling. First, all variables were rescaled to a unified range. Second, the K-means++ method was preferred over traditional K-means since it inherently improves seed initialization and reduces the risk of falling into local optima in such a variable space (Arthur and Vassilvitskii, 2007). Third, suitable cluster validation measures, like the Silhouette Score and Davies-Bouldin Index, which focus on maximizing similarity within clusters and dissimilarity between clusters, were adopted to ensure the most homogenous groups of crash surroundings were identified.

Once identified, the study employed a highly interpretable statistical discrete outcome model with random parameters specific to each group of built environments and explored how influential factors' significance, magnitude, and direction of effect change among the built environment types. Additionally, given the low number of fatal crashes in the dataset, we integrated fatal and serious injury severity together to achieve more reliable estimates (Ye and Lord, 2014). Therefore, a binary logit model with random parameters is employed. The following subsections explain the mathematical relationships behind the methods used.

Table 3
Descriptive Statistics of Built Environment Variables.

Built Environment	Definition	Mean (SD)
Land Use		
Share of Residential Land Use	Residential land use percentage in the crash buffer zone	38.57 % (9.73 %)
Share of Commercial Land Use	Commercial land use percentage in the crash buffer zone	18.36 % (7.52 %)
Share of Industrial Land Use	Industrial land use percentage in the crash buffer zone	4.21 % (3.68 %)
Share of Recreational Land Use	Recreational land use percentage in the crash buffer zone	11.49 % (5.12 %)
Share of Retail Land Use	Retail land use percentage in the crash buffer zone	13.76 % (5.84 %)
Share of Educational Land Use	Educational land use percentage in the crash buffer zone	9.28 % (4.37 %)
Entropy Index*	$\frac{\sum_{i=1}^{n} p_i \log p_i}{\log n}$; p_i: percentage of land use i, n: total number of the land uses in the crash buffer zone	0.45 (0.17)
Infrastructure and Road Network Density		
Density of Intersections	Number of intersections within the crash buffer zone divided by its area (km^2)	32.36 (19.71)
Density of Bike-lanes	The summation of bike lane length (km) within the crash buffer zone divided by its area (km^2)	2.06 (1.37)
Density of Sidewalks	The summation of sidewalk length (km) within the crash buffer zone divided by its area (km^2)	3.98 (1.50)
Motorway Density	The summation of motorway road type length (km) within the crash buffer zone divided by its area (km^2)	0.47 (0.28)
Trunk Density	The summation of trunk road type length (km) within the crash buffer zone divided by its area (km^2)	1.06 (0.55)
Primary Density	The summation of primary road type length (km) within the crash buffer zone divided by its area (km^2)	1.79 (0.63)
Secondary Density	The summation of secondary road type length (km) within the crash buffer zone divided by its area (km^2)	2.07 (0.87)
Tertiary Density	The summation of tertiary road type length (km) within the crash buffer zone divided by its area (km^2)	2.34 (1.24)
Residential Street Density	The summation of residential street road type length (km) within the crash buffer zone divided by its area (km^2)	1.61 (0.76)
Accessibility and Amenities		
Distance to the City Centre	The distance between the crash location and the centre of the city using Haversine Formula (km) (GeeksforGeeks, 2022)	2.97 (1.37)
Bar/Cafe/Restaurant Density	Number of bar/cafe/restaurant within the crash buffer zone divided by its area (km^2)	4.57 (3.71)
Density of Transit (Train/Bus) Stations	Number of transit stations within the crash buffer zone divided by its area (km^2)	3.60 (2.08)

* Entropy ranges from 0 to 1, indicating land use diversity. Zero signifies a single use, while one reflects a balanced and varied distribution (Song et al., 2013).

4.1. Unsupervised Machine learning

In this research, the K-means++ algorithm, an enhanced version of K-means technique, is used. The advantage of using an unsupervised method is that there is no dependent variable, allowing for automatic searching for clusters in the dataset; this aligns with our objective of identifying built environment categories surrounding e-scooter crashes. To ensure each variable contributes proportionately to the clustering process, we first normalized the variables using the min–max technique. The mathematical formula for min–max normalization is as follows:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

In Eq. (1), x' is the scaled value, and x is the actual value. The K-means algorithm clusters the data by separating observations into K groups of equal variance by minimizing a criterion known as the inertia or within-cluster sum-of square (Aggarwal and Reddy, 2018). The mathematical expression of the criterion is as follows (Shutaywi and Kachouie, 2021):

$$D\left(\{C_k\}_{k=1}^{K}\right) = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||x_i - \mu_k||^2 \tag{2}$$

Eq. (2) represents the total within-cluster variance or distortion, where C_k denotes $k-th$ built environment cluster in the set of clusters $k = \{1, 2, \cdots, K\}$, x_i represents the $i-th$ individual data point (in this case, the built environment characteristics around the crash location), μ_k is the centroid (mean) of the $k-th$ cluster, and $||.||^2$ is the squared Euclidean distance. K-means++, proposed by Arthur and Vassilvitskii (2007), follows the same criterion as K-means but overcomes two of its shortcomings. Firstly, initiation in K-means++ is done by choosing well-separated centroids (μ_k) to possibly avoid falling into a local optimum. Secondly, compared with the previous algorithm, K-means++ requires fewer iterations to converge, resulting in reduced run time. In this paper, K-means++ clustering quality was evaluated using the Silhouette Score and the Davies-Bouldin (DB) Index, which are recommended measures for determining the appropriate number of clusters (Aggarwal and Reddy, 2018).

The Silhouette Score calculates the average similarity of data points within a cluster relative to their distance from data points in other clusters (Aggarwal and Reddy, 2018). The values range between -1 and $+1$. A higher value is preferred as it signifies a better fit within its own cluster and a weaker association with neighbouring clusters. Mathematically, the Silhouette Score for the $i-th$ individual data point (x_i) in cluster π_k is defined as follows (Shutaywi and Kachouie, 2021):

$$SilhouetteScore_i = \frac{b(x_i) - a(x_i)}{max(b(x_i), a(x_i))} \tag{3}$$

In Eq. (3), $a(x_i)$ is the average distance x_i to all other elements in the cluster π_k (within dissimilarity), and $b(x_i)$ represents the minimum average distance from the data point x_i to all points in clusters other than its own cluster (between dissimilarity), which can be expressed as:

$$b(x_i) = \min_{l \neq k}(d_l(x_i)) \tag{4}$$

where $d_l(x_i))$ denotes the average distance between the data point x_i and all points in the $l-th$ cluster, and the minimum is taken over all clusters l that are not equal to k.

The Davies-Bouldin (DB) Index measures the ratio of within-cluster dispersion to between-cluster separation, with a lower value indicating better clustering quality (Davies and Bouldin, 1979). The DB Index is defined as follows:

$$DB(C) = \frac{1}{K} \sum_{k=1}^{K} \max_{k \neq l} \left\{ \frac{\Delta(C_k) + \Delta(C_l)}{\delta(C_k C_l)} \right\} \tag{5}$$

$\Delta(C_k)$ is the intra-cluster distance for cluster k, representing the average distance between each point in the cluster and its centroid, thus measuring cluster compactness. Similarly, $\Delta(C_l)$ is the intra-cluster distance for cluster l. $\delta(C_k C_l)$ is the inter-cluster distance between the centroids of clusters k and l, measuring how well-separated the clusters are. In this study, the scikit-learn package, a Python package, is used for developing K-means++ clustering (Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel et al., 2011).

4.2. Random parameter binary logit model

Given the binary variable outcome for severity considered in this study, we employed a random parameter binary logit model to investi-

gate crash severity, its influential factors, and to account for underlying unobserved heterogeneity in the dataset. For this purpose, the response variable for the $i-th$ crash instance could be considered fatal or serious injury ($Y_i = 1$) or slight injury ($Y_i = 0$). Eq. (6) represents the general form of severity equation (Washington et al., 2020):

$$Y_i = \beta_n X_{ni} + \varepsilon_{ni} + \eta_{ni} \tag{6}$$

where β_n denotes the estimated parameter for $n-th$ variable. X_{ni} represents $n-th$ indicator variable for crash instance i. ε_{ni} denotes the error term and assumed to follow a Gumbel distribution. η_{ni} is the random parameter component with a mean of zero and a distribution that determines respective parameters. The probability function for crash severity is defined as Eq. (7) (Washington et al., 2020):

$$Pr = \int \frac{exp^{\beta_n X_{ni}}}{\sum_j exp^{\beta_n X_{ni}}} f(\beta|\varphi) d\beta \tag{7}$$

Here, $f(\beta|\varphi)$ is the density function of β, and φ is the vector of density parameters. the index of summation is j and is defined based on levels of severity. When β remains fixed, the equation represents a typical binary logit model. In this study, each estimated parameter was examined for randomness across observations to address unobserved heterogeneity. Variables with statistically significant mean and standard deviation values were treated as random across all crash instances. The inclusion of random parameters increases computational complexity in the numerical integration during maximum likelihood estimation (Train, 2001). Therefore, maximum simulated likelihood estimation with Halton draws is used to achieve accurate estimates (Halton, 1960; Train, 2001).

The overall fit of the model is evaluated using the Pseudo-R-squared value, ρ^2, as shown in Eq. (8). Higher values of ρ^2 indicate a better fit of the model to the data.

$$\rho^2 = 1 - LL(\beta)/LL(0) \tag{8}$$

Here, $LL(0)$ represents the log-likelihood of the null model, and $LL(\beta)$ denotes the log-likelihood of the converged model. The estimation of discrete outcome models in our work is conducted using the PandasBiogeme package in Python (Bierlaire, 2020).

4.3. Marginal effects

Eq. (9) illustrates how marginal effects are calculated for the indicator variables. The marginal effect for an indicator variable is described as the change in the probability of a particular severity outcome when the indicator variable transitions from 0 to 1 (Washington et al., 2020).

$$M_{x_{ijn}}^{P_{ij}} = P_{ij}\left[when x_{ijn} = 1\right] - P_{ij}\left[when x_{ijn} = 0\right] \tag{9}$$

P_{ij} represents the probability of severity outcome j for crash instance i, while x_{ijn} denotes the n –th indicator variable given severity level j and observation i. Eq. (9) determines the change in probability for each individual crash observation. To obtain the overall marginal effect, the individual probabilities can be aggregated.

5. Results and discussion

This section presents the outcomes of the analysis of e-scooter crashes using the two-stage clustering and discrete outcome model. First, the clustering results are presented, along with a discussion of the identified built environment factors within each cluster. Subsequently, crash severity models are estimated specific to each cluster to discuss how crash characteristics differ across different environments.

5.1. Built environment clustering

Using the variables in Table 3, we first validated the appropriate buffer size through various clustering scenarios. To do so, we implemented the clustering algorithm on the crash-built environment sets, capturing crash surroundings with varying sizes from 25 m to 3000 m. As shown in Fig. 3 (a), the 250 m buffer size produces the most homogeneous built environment clusters. Smaller buffer sizes result in more localised clustering of crash site environments as their buffers are likely to share fewer commonalities due to a lower chance of overlapping areas, as illustrated by the example of two crash points with 25 m buffers in Fig. 3 (b). Larger buffer sizes, on the other hand, create overly generalised clusters, decreasing dissimilarity between crash surroundings as the overlapping area of buffers increases for the crash points (Fig. 3 (c)). Both extremes negatively impact clustering performance. Smaller buffers gradually reduce the similarity between crash points within clusters, whereas larger buffers decrease the dissimilarity between clusters, which both ultimately leads to poorer overall performance measures (high Davies-Bouldin (DB) Index and a low Silhouette Score).

Having identified 250 m as the suitable buffer size for analysis, Fig. 4 illustrates the optimal number of clusters for this buffer based on the performance measures. K-means++ models with cluster counts ranging from 2 to 10 were evaluated. The Silhouette Score reached its maximum, and the DB Index its minimum, with four clusters. Additional optimal cluster counts were identified at 7 and 8 clusters, which performed well but were less well-separated than the four-cluster model due to lower performance metrics. Moreover, these extra clusters increased computational costs because they required additional discrete outcome models to be estimated, while considerably smaller sample sizes made it infeasible to achieve reliable estimates using 7 or 8 segments of crash data (Naghizadeh and Metaxas, 2020; Ye and Lord, 2014). Consequently, the four-cluster model was selected.

Fig. 5 exemplifies the spatial distribution of crash locations in London as the major area where e-scooter incidents occur, both before and after clustering the built environment of crashes. Table 4 also summarises the distribution of the crash indicator variables across four clusters.

Fig. 6 presents the results of the optimal model with four clusters, using radar plots to compare values across clusters and heat maps to display the variables in three categories: land use factors (Fig. 6 (a)), infrastructure and road network density factors (Fig. 6 (b)), and accessibility and amenities factors (Fig. 6 (c)). Plot values are normalized between 0 and 1 to facilitate comparisons. Given that urban spaces are shaped by a mix of environmental factors; some clusters share similarities in some aspects. The built environment clusters are categorised based on notable distinctions as follows: car-centric and mixed-use zones (C1), commercial and industrial zones (C2), intersection-dense areas (C3), and residential and central areas (C4). A brief description of each built environment cluster is provided below:

Car-centric and Mixed-use Zones (C1): The first cluster contains sites with balanced densities of land use types. This land use composition resulted in the highest entropy index (0.89), which indicates a diverse range of activities. The infrastructure supports mixed modes of transportation with moderate densities of intersections (0.62), bike lanes (0.47), and sidewalks (0.65). This cluster has the second highest motorway density (0.63) and the highest primary streets (0.86) density, supporting vehicular traffic. Residential street density is moderate (0.50), balancing residential and other uses. The cluster represents areas that are moderately distant from the city centre (0.68). It has moderate densities of bars, cafes, restaurants (0.57), and transit stations (0.58).

Commercial and Industrial Zones (C2): Cluster two has high commercial (0.83) and industrial activities (0.92), with the lowest residential (0.20), recreational (0.16), educational (0.15), and retail (0.41) land uses. The unbalanced land use results in the lowest entropy index (0.54). This cluster shows the lowest densities of intersections (0.53), bike lanes (0.36), and sidewalks (0.39). Compared to other clusters,

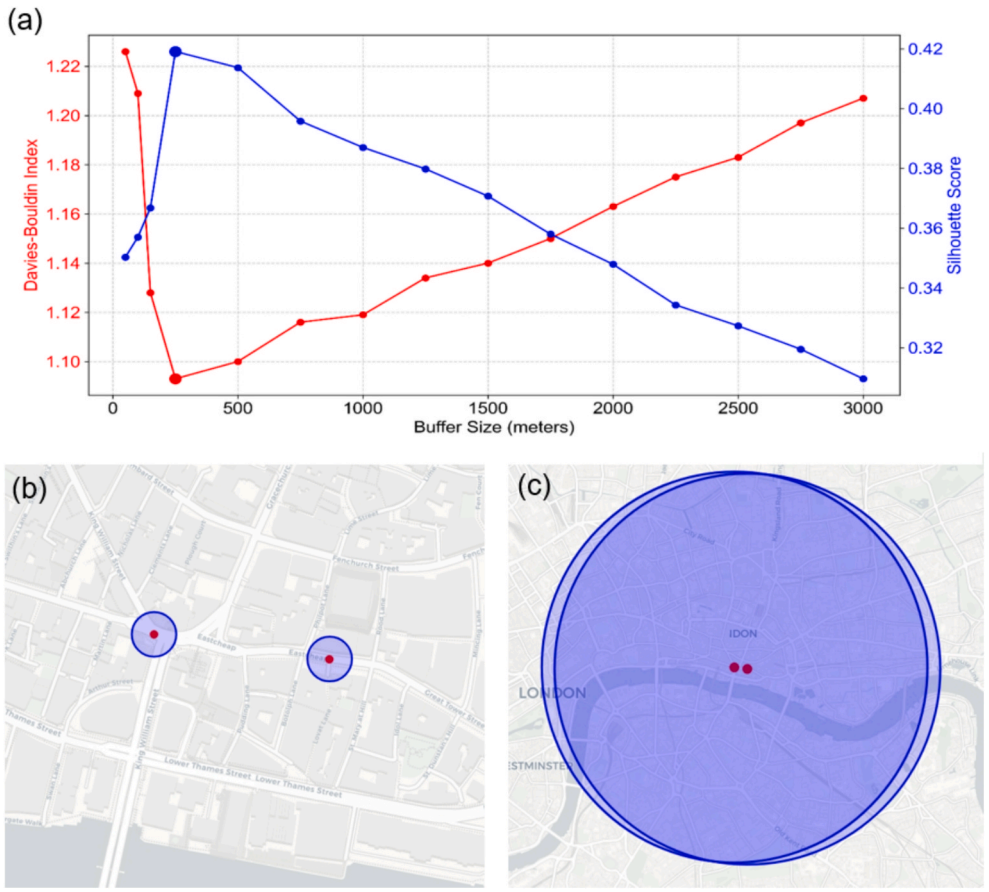

Fig. 3. (a) Selection of buffer size (b) Example of two crash points with a 25 m buffer (c) Example of two crash points with a 3000 m buffer.

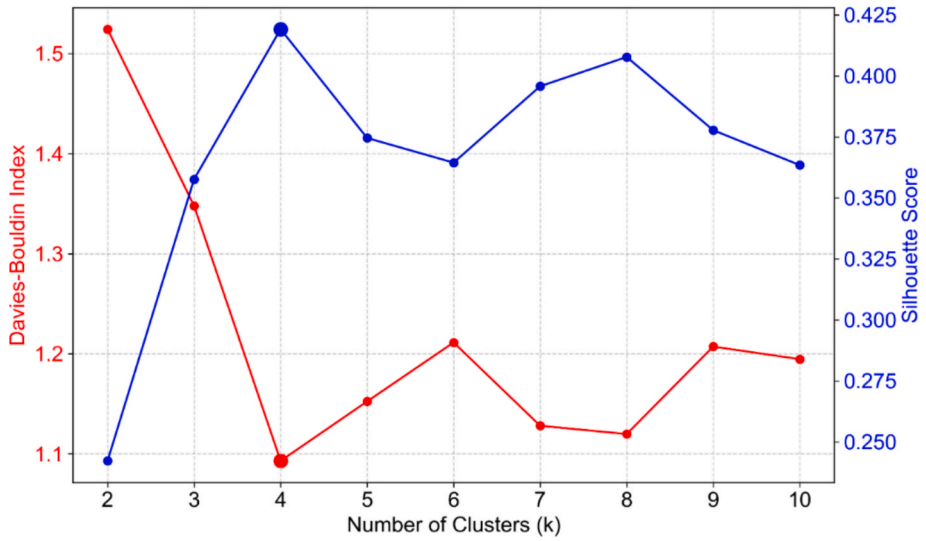

Fig. 4. Silhouette Score and DB Index across cluster numbers.

motorway density is the highest (0.74), while trunk (0.61), primary (0.69), and secondary street densities (0.58) are moderate to high. Residential street density is low (0.27). Similarly, there are low densities of bars, cafes, restaurants (0.35), and transit stations (0.42), with the sites having the furthest distance from the city centre (0.83).

Intersection-Dense Areas (C3): This cluster features low residential land use (0.31), higher commercial (0.61) and industrial (0.43) activities and moderate recreational (0.46) and educational (0.42) land uses. The land use mixture is reflected in the entropy index (0.66). These areas are distinguished by their high density of intersections (0.92). Bike lanes (0.68) and sidewalks (0.81) are higher, while motorway (0.25) and trunk density (0.36) are lower than in earlier clusters. The road network shows the highest density of primary (0.74) and secondary (0.81) roads, and tertiary street densities are also high (0.77). The residential street density is moderate (0.46), but higher than the first two clusters. The area is moderately distant from the city centre (0.54) with moderate to

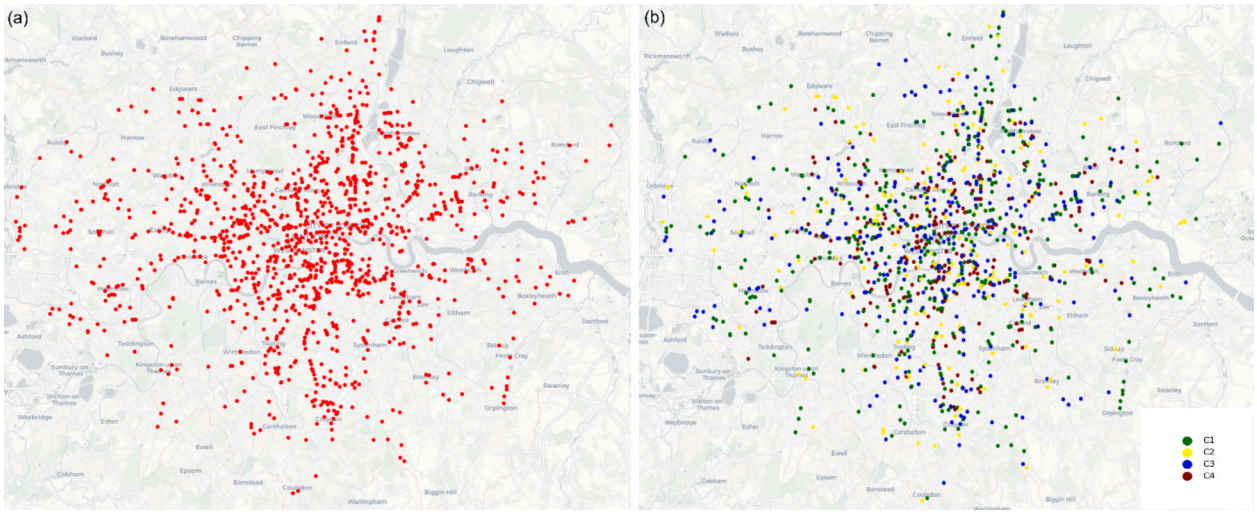

Fig. 5. Spatial distribution of e-scooter crash locations in London (a) before and (b) after clustering.

high densities of bars, cafes, restaurants (0.67), and transit stations (0.81).

Residential and Central Areas (C4): This cluster has the highest share of residential land use (0.91), with low shares of commercial (0.23) and industrial land use (0.17). Recreational land use is moderate (0.64); as are retail (0.58) and educational land use (0.52. The entropy index (0.78) indicates a relatively high diversity of land uses. High densities of intersections (0.81), bike lanes (0.76), and sidewalks (0.92), high scores for bars, cafes, restaurants (0.79), and transit stations (0.73) and low densities of motorways (0.11) and trunks (0.23) in this cluster suggest less focus on high-speed travel. However, the primary (0.53), secondary (0.71), and tertiary street densities (0.78) indicate a well-connected local network. The residential street density (0.81) reinforces this clusters residential focus. This cluster contains data points with substantially the lowest distance from the city centre (0.35).

5.2. E-scooter crash severity analysis

After the identification of built environment types, we estimated a crash severity model for each cluster. First, we applied a range of fixed parameter binary logit models for the identified clusters and then developed them toward a random parameter model to quantify the impact of crash characteristics on severity level. In this process, we tested normal, log-normal, triangular, and uniform distributions for random parameters. The normal distribution provided the best statistical fit for the random parameter density function, aligning with previous research findings (Gao and Zhang, 2024). Variables associated with t-statistics at or over a 90 % confidence level in a two-tailed *t*-test and for random parameters with significant standard deviations were kept for the final model specification.

As shown in Table 5, the comparison of model fit measures indicates slightly better performance of random parameter models over fixed parameter ones, which are presented as the optimal outputs of this research through Tables 6–9. Additionally, as presented in the last column of Table 5, to validate the soundness of homogeneities within clusters in a way that significantly enhances modelling performance, we estimated a model on the dataset containing the crashes from four identified clusters and compared it with four separate models using the Likelihood Ratio (LR) test. The calculated statistic follows a χ^2 distribution with degrees of freedom equal to the summation of the number of estimated parameters in all cluster-specific models minus the number of estimated parameters in the model that does not use clustering results. Results show statistically significant improvements when the data is segmented prior to severity analyses in both binary logit models with fixed and random parameters.

5.2.1. Car-centric and mixed-use zones

Table 6 shows that crashes on carriageways increase the probabilities of serious and fatal injury outcome by 0.261. Higher speeds are another dominant risk-increasing factor, increasing serious and fatal outcomes of riders by 0.252. The statistically significant standard deviation shows variance in how the factor influenced the outcome of severity. The findings are in line with Gao and Zhang (2024), who identified that e-scooter users are prone to severe crashes on carriageways and roads with higher speed limits, highlighting the need for e-scooters to operate in safe speed environments such as the measures proposed in Helsinki (Dibaj et al. 2024).

Adverse road conditions were found to significantly raise the likelihood of serious and fatal injuries 0.035. Similar to this finding, the role of surface condition in e-scooter safety research is noted by Ma et al. (2021), Stigson et al. (2021), and White et al. (2023). Adverse weather had a protective effect in this cluster. This might be a result of more cautious behaviour and decreasing speeds in adverse weather. Also, the discomfort of riding under such conditions might decrease the rate of usage. These findings align with Sun et al. (2024), who investigated negative relations between micro-mobility crash severity and adverse weather.

Temporal indicators, such as morning and afternoon peak hours and weekdays, were statistically significant in increasing the probability of severe crashes. Such results can be attributed to the mixed-use nature of crash sites in this cluster, which see high traffic volume and activity levels in these periods. Besides, the morning commute hours' indicator was found to be a significant random parameter representing variation of this factor on the crash severity.

Regarding riders' demographics, young riders were more prone to increased serious and fatal injuries than other age groups, with a statistically significant standard deviation unveiling heterogeneity in this group. Females had reduced risk of serious and fatal injuries by 0.068, aligning with findings by Gao and Zhang (2024).

Rider collisions with cars and heavy vehicles raise the likelihood of a severe crash by 0.310. Side-impact collisions between e-scooters and the front of the second-party vehicle increase the likelihood of severe outcome by 0.267. Across all models in Tables 6-9, side-impact collisions are the most common significant collision type. In this regard, Pérez-Zuriaga et al. (2023) previously identified side impacts as an important collision type for e-scooter crashes, often resulting in more severe injuries than for bicycle riders. As seen in the literature review, Morrison et al. (2019) and Gao and Zhang (2024) highlight the

Table 4
Summary of crash indicator variables by cluster.

Explanatory Variables	C1	C2	C3	C4
	n (%)	n (%)	n (%)	n (%)
Road Type				
roundabout	13 (1.05 %)	16 (2.62 %)	98 (9.93 %)	29 (4.09 %)
one-way street	21 (1.69 %)	10 (1.64 %)	126 (12.77 %)	77 (10.86 %)
carriageway	1205 (96.94 %)	583 (95.42 %)	756 (76.60 %)	544 (76.73 %)
slip road	4 (0.32 %)	2 (0.33 %)	7 (0.71 %)	59 (8.32 %)
Speed Limit (mph)				
equal or less than 30	701 (56.40 %)	346 (56.63 %)	363 (36.78 %)	196 (27.64 %)
31–49	340 (27.35 %)	171 (27.99 %)	583 (59.07 %)	498 (70.24 %)
equal or greater than 50	202 (16.25 %)	94 (15.38 %)	41 (4.15 %)	15 (2.12 %)
Road Surface Condition				
adverse condition (road surface in a poor condition, e.g., being wet, snowy, flooded, oily, muddy, or frosted)	201 (16.17 %)	31 (5.07 %)	111 (11.25 %)	45 (6.35 %)
normal condition	1042 (83.83 %)	580 (94.93 %)	876 (88.75 %)	664 (93.65 %)
Light Condition				
daylight	818 (65.81 %)	409 (66.94 %)	650 (65.86 %)	459 (64.74 %)
darkness – lights lit	341 (27.43 %)	94 (15.38 %)	179 (18.14 %)	204 (28.77 %)
darkness – lights unlit	65 (5.23 %)	83 (13.58 %)	121 (12.26 %)	34 (4.80 %)
darkness – no lighting	19 (1.53 %)	25 (4.09 %)	37 (3.75 %)	12 (1.69 %)
Weather Condition				
clear	674 (54.22 %)	377 (61.70 %)	553 (56.03 %)	412 (58.11 %)
cloudy	402 (32.34 %)	156 (25.53 %)	336 (34.04 %)	225 (31.73 %)
adverse (snow/rain/fog)	167 (13.44 %)	78 (12.77 %)	98 (9.93 %)	72 (10.16 %)
Crash Time				
late night/early morning (00:00–05:59)	80 (6.44 %)	59 (9.66 %)	77 (7.80 %)	44 (6.21 %)
morning commute hours (06:00–09:59)	291 (23.41 %)	102 (16.69 %)	214 (21.68 %)	208 (29.34 %)
midday (10:00–14:59)	356 (28.64 %)	204 (33.39 %)	292 (29.58 %)	155 (21.86 %)
afternoon commute hours (15:00–17:59)	283 (22.77 %)	105 (17.18 %)	206 (20.87 %)	211 (29.76 %)
evening (18:00–20:59)	156 (12.55 %)	103 (16.86 %)	137 (13.88 %)	55 (7.76 %)
night (21:00–23:59)	77 (6.19 %)	38 (6.22 %)	61 (6.18 %)	36 (5.08 %)
Day of Week				
weekday	1021 (82.14 %)	524 (85.76 %)	732 (74.16 %)	521 (73.48 %)
weekend	222 (17.86 %)	87 (14.24 %)	255 (25.84 %)	188 (26.52 %)
Season				
spring	336 (27.03 %)	150 (24.55 %)	231 (23.40 %)	203 (28.63 %)
summer	412 (33.15 %)	208 (34.04 %)	353 (35.76 %)	242 (34.13 %)
autumn	315 (25.34 %)	166 (27.17 %)	251 (25.43 %)	176 (24.82 %)
winter	180 (14.48 %)	87 (14.24 %)	152 (15.40 %)	88 (12.41 %)
E-scooter Rider Sex				
male	848 (68.22 %)	420 (68.74 %)	710 (71.94 %)	556 (78.42 %)
female	395 (31.78 %)	191 (31.26 %)	277 (28.06 %)	153 (21.58 %)
E-scooter Rider Age				
less than 35	687 (55.27 %)	342 (55.97 %)	515 (52.18 %)	345 (48.66 %)
35–45	421 (33.87 %)	208 (34.04 %)	304 (30.80 %)	220 (31.03 %)
greater than 45	135 (10.86 %)	61 (9.98 %)	168 (17.02 %)	144 (20.31 %)
Second Party Type				
e-scooter only crash	296 (23.81 %)	114 (18.66 %)	237 (24.01 %)	206 (29.06 %)
pedestrian	75 (6.03 %)	33 (5.40 %)	64 (6.48 %)	58 (8.18 %)
bicycle	33 (2.65 %)	19 (3.11 %)	34 (3.44 %)	29 (4.09 %)
motorcycle	61 (4.91 %)	30 (4.91 %)	44 (4.46 %)	25 (3.53 %)
car/heavy vehicle	778 (62.59 %)	415 (67.92 %)	608 (61.60 %)	391 (55.15 %)
Manners of Collision with Vehicle				
head-on	111 (13.23 %)	142 (31.91 %)	83 (12.77 %)	59 (14.11 %)
side-impact with front of the e-scooter	126 (15.02 %)	54 (12.13 %)	187 (28.77 %)	133 (31.82 %)
side-impact with front of the 2nd party vehicle	259 (30.87 %)	11 (2.47 %)	109 (16.77 %)	73 (17.46 %)
sideswipe with opposite direction	75 (8.94 %)	37 (8.31 %)	55 (8.46 %)	38 (9.09 %)
sideswipe with same direction	101 (12.04 %)	63 (14.16 %)	89 (13.69 %)	36 (8.61 %)
rear-end with front end of the e-scooter	88 (10.49 %)	41 (9.21 %)	63 (9.69 %)	44 (10.53 %)
rear-end with front end of the 2nd party vehicle	79 (9.42 %)	97 (21.80 %)	64 (9.85 %)	35 (8.37 %)

importance of separating infrastructure that is sharable by e-scooters and bicycles as a potential approach to reducing motor vehicle crashes.

5.2.2. Commercial and industrial zones

As shown in Table 7, like the first cluster, higher speed limits

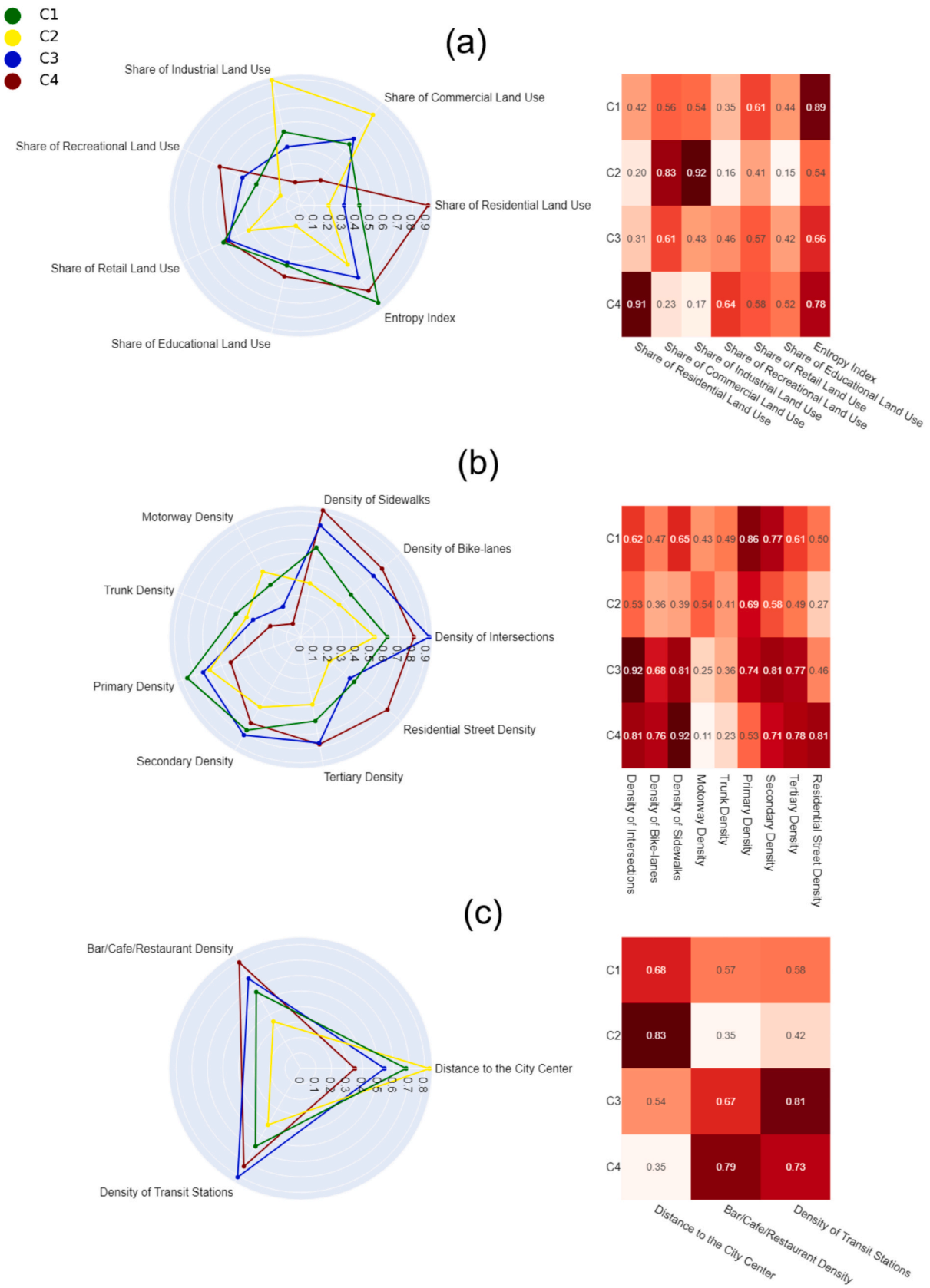

Fig. 6. Radar plots and heat maps comparing cluster values for (a) land use, (b) infrastructure and road network density, and (c) accessibility and amenities at crash locations.

Table 5
Model comparison results.

Model	ρ^2				$LR^* = -2\left[LL\left(\beta^{withoutclustering}\right) - \sum_{k=1}^{4} LL\left(\beta^{cluster(k)}\right)\right]$
	C1	C2	C3	C4	
Fixed parameter binary logit	0.188	0.189	0.193	0.165	$(LR = 51.17) > (\chi^2_{0.99,27} = 46.96)$
Random parameter binary logit	0.207	0.214	0.211	0.171	$(LR = 65.15) > (\chi^2_{0.99,33} = 54.78)$

*Note: The log-likelihood values of the cluster-specific models use the same variables and specifications as the model without clustering to ensure that differences in $LL(\beta)$ values arise solely from the clustering of the data.

Table 6
Estimated parameters and marginal effects for e-scooter crashes (built environment type: car-centric and mixed-use zones).

Variable	Coefficient	t-statistics	Marginal Effects (Serious Injury/ Fatal)
Constant	−1.64***	−6.78	
Road Type: carriageway	1.72***	3.60	0.261
Speed Limit: greater than 50 mph (standard deviation of parameter distribution)	2.21** (4.87**)	2.29 (2.11)	0.252
Road Surface Condition: adverse	0.32*	1.82	0.035
Weather Condition: adverse	−0.78*	−1.81	−0.063
Crash Time: morning commute hours (standard deviation of parameter distribution)	1.08*** (1.83*)	3.29 (1.71)	0.046
Crash Time: afternoon commute hours	0.56***	2.94	0.061
Day of Week: weekday	0.49***	2.63	0.032
Rider Age: less than 35 (standard deviation of parameter distribution)	2.33*** (4.11*)	2.73 (1.88)	0.275
Rider Sex: female	−1.12*	−1.70	−0.068
Second Party Type: car/heavy vehicle	1.72***	4.37	0.310
Manner of Collision with Vehicle: side-impact with front of the 2nd party vehicle	1.26***	4.90	0.267
Model Statistics			
Number of observations	1243		
$LL(0)$	−1134.672		
$LL(\beta)$	−899.875		
ρ^2	0.207		

***, ** and * represent 99%, 95%, and 90% confidence level, respectively.

Table 7
Estimated parameters and marginal effects for e-scooter crashes (built environment type: commercial and industrial zones).

Variable	Coefficient	t-statistics	Marginal Effects (Serious Injury/ Fatal)
Constant	−1.38***	−4.65	
Road Type: carriageway	1.46***	3.57	0.228
Speed Limit: greater than 50 mph (standard deviation of parameter distribution)	2.47*** (4.64**)	2.37 (2.11)	0.405
Light Condition: darkness lights unlit/ no lighting (standard deviation of parameter distribution)	2.05** (3.23**)	2.01 (2.40)	0.335
Weather Condition: adverse	−0.51*	−1.71	−0.026
Day of Week: weekday	0.26**	2.32	0.021
Rider Age: less than 35 (standard deviation of parameter distribution)	2.17**	2.29	0.243
Rider Sex: female	−1.92**	−2.20	−0.347
Second Party Type: car/heavy vehicle	3.45***	3.14	0.419
Manner of Collision with Vehicle: head-on	1.62***	3.85	0.293
Manner of Collision with Vehicle: sideswipe with same direction	1.55**	2.31	0.271
Manner of Collision with Vehicle: rear-end with front end of the 2nd party vehicle	0.66**	2.02	0.038
Model Statistics			
Number of observations	611		
$LL(0)$	−1346.891		
$LL(\beta)$	−1059.238		
ρ^2	0.214		

***, ** and * represent 99%, 95%, and 90% confidence level, respectively.

significantly increased the probability of serious and fatal injuries. Another common significant risk factor is the indicator variable representing collisions with cars and trucks, raising the likelihood of serious and fatal outcomes by 0.419. Despite the same direction of effect for the mentioned variables in clusters 1 and 2, the marginal effects are noticeably more pronounced here. This may be due to the increased presence of high-speed roads and low density of intersections, which could indicate higher vehicle speeds. While there have been some efforts to lower speeds in the UK, with campaigns for 20mph zones, more work is needed to create safer road environments for vulnerable road users (Ekmekci et al., 2024).

Other common significant factors between the first and second clusters include the indicator variable for carriageways, which increases the severe crash risk by 0.228, the inclement weather indicator being associated with a lower likelihood of severe crashes, the weekday indicator variable being associated with a 0.021 increase in the likelihood of fatal and serious injury crashes, and female riders and younger cohorts showing a lower and higher likelihood of sustaining serious or fatal injuries, respectively.

With respect to significant factors specific to the crash sites in this cluster, nighttime lighting, either riding in unlit or poorly lit conditions posed a significant risk. As a random parameter, it greatly increases the probability of serious and fatal injuries by 0.335. Similarly, Shah et al. (2021) emphasized that inadequate lighting on suburban roads is a high-risk factor for riders, and the provision of further street lighting could help to reduce this issue.

Three types of collisions were identified as risk factors: head-on, side-swipes, and rear-end collisions, with head-on collisions showing the highest rise in fatal and serious injury levels, increasing the likelihood by 0.293, followed by side-swipes at 0.271. Another point worth noting is that compared to other clusters, this cluster has the highest number of significant collision types, which is possibly due to the dominance of vehicular traffic facilities, and further stresses the need for separate infrastructure in areas with high volumes of motor vehicle traffic.

Table 8
Estimated parameters and marginal effects for e-scooter crashes (built environment type: intersection-dense areas).

Variable	Coefficient	t-statistics	Marginal Effects (Serious Injury/ Fatal)
Constant	−1.15***	−3.71	
Speed Limit: 30–50 mph (standard deviation of parameter distribution)	1.12*** (4.43**)	2.78 (2.15)	0.162
Road Surface Condition: adverse	0.51**	1.92	0.030
Light Condition: darkness lights unlit/ no lighting (standard deviation of parameter distribution)	1.70* (2.53*)	1.84 (1.68)	0.241
Crash Time: morning commute hours	−0.83**	−2.23	−0.039
Crash Time: afternoon commute hours	−0.59*	−1.76	−0.025
Rider Age: less than 35 (standard deviation of parameter distribution)	1.05** (2.18*)	2.41 (1.90)	0.153
Second Party Type: car/heavy vehicle	0.95**	2.26	0.137
Manner of Collision with Vehicle: side-impact with front of the e-scooter	1.18***	3.81	0.182
Manner of Collision with Vehicle: side-impact with front of the 2nd party vehicle	2.19***	3.10	0.342
Model Statistics			
Number of observations	987		
$LL(0)$	−987.534		
$LL(\beta)$	−779.125		
ρ^2	0.211		

***, ** and * represent 99%, 95%, and 90% confidence level, respectively.

Table 9
Estimated parameters and marginal effects for e-scooter crashes (built environment type: residential and central areas).

Variable	Coefficient	t-statistics	Marginal Effects (Serious Injury/ Fatal)
Constant	−1.02***	−4.51	
Speed Limit: 30–50 mph (standard deviation of parameter distribution)	1.14***	3.20	0.164
Crash Time: morning commute hours	−1.14**	−2.25	−0.118
Crash Time: afternoon commute hours	−0.95**	−2.06	−0.109
Day of Week: weekend (standard deviation of parameter distribution)	1.19*** (3.14**)	2.83 (2.44)	0.182
Rider Age: less than 35 (standard deviation of parameter distribution)	1.33*** (2.35***)	3.54 (3.22)	0.194
Second Party Type: car/heavy vehicle	0.85*	1.72	0.139
Manner of Collision with Vehicle: side-impact with front of the e-scooter	1.10 ***	3.28	0.175
Manner of Collision with Vehicle: side-impact with front of the 2nd party vehicle	2.49***	4.29	0.361
Model Statistics			
Number of observations	709		
$LL(0)$	−1256.782		
$LL(\beta)$	−1041.325		
ρ^2	0.171		

***, ** and * represent 99%, 95%, and 90% confidence level, respectively.

5.2.3. Intersection-Dense areas

As presented in Table 8, side-impact collisions, including those involving the front of either an e-scooter or a motor vehicle, contribute highly to injury severity, with side-impacts involving the front of a motor vehicle having a higher probability of severe injuries. Lighting condition is also an important random parameter in this cluster, which shows that under unlit and poorly lit conditions, the probability of serious and fatal injuries increases by 0.241.

Speed limit is a significant random parameter with 30–50 mph roads, increasing the probability of serious and fatal outcomes by 0.162. Compared to severity analysis discussed for previous clusters, here, the speed limits with lower bounds are significant, and the marginal effects are relatively lower. This could be explained by the high density of intersections and low hierarchy roads compared to the first and second clusters, representing slower vehicular traffic.

Temporal indicators, namely morning and afternoon commute hours, were found to be significant, with marginal effects of −0.039 and −0.025, showing a decreasing effect on severe outcomes, in contrast to their direction of effect in the first cluster. This might be explained by the fact that locations with many intersections may experience more congestion and thus have lower vehicle speed making high-speed collisions between e-scooters and motor vehicles less probable, resulting in lower likelihood for high injury severity.

Finally, and consistent with earlier clusters, younger riders tend to have more severe outcomes. As shown by the statistically significant random parameter that increased the likelihood of serious and fatal injuries by 0.153. Again, in line with other clusters, collisions with cars and heavy vehicles are significant and increase the probability of serious and fatal injuries by 0.137.

5.2.4. Residential and central areas

As presented in Table 9, side-impact collisions were significant factors for this cluster. The calculated marginal effects show that side-impacts with the front of the vehicle and the front of the e-scooter increase severe crash likelihood by 0.361 and 0.175, respectively.

In a similar pattern to the previous cluster, morning and afternoon peak hours reduce the likelihood of severe crashes by 0.118, and 0.109, respectively. This effect can be attributed to notable characteristics of this built environment type, such as high land use entropy, and the density of intersections, which can hinder fast vehicular flow during rush hours. Also, similar to the previous cluster, the 30–50 mph roads were shown to significantly increase severe outcomes by 0.164. The similarity in the significance of this variable between the third and fourth clusters, while the first two had shown greater than 50 mph roads as significant factors, can be rooted in their differences in built environment characteristics, as the first two were shown to include more fast-speed facilities on average.

Another temporal factor contributing to this cluster is the weekend indicator, which exhibits a significant standard deviation and marginal effects of 0.182 for serious and fatal outcomes. Similar findings on severe crashes happening over the weekends due to leisure activities and alcohol consumption are also reported in studies by Vernon et al. (2020) and Stigson et al. (2021).

Similar to the previous models, collisions with cars and heavy vehicles are significant and increase the probability of serious and fatal injuries by 0.139. The younger rider indicator has also shown a significant random parameter, raising the probabilities of serious and fatal outcomes by 0.194.

5.3. Comparative analysis of crash risk factors and interactions across and within clusters

This section provides a comparative analysis of crash risk factors and their relationships, both within and between the identified built environment clusters. By exploring these patterns in this manner, this part seeks to identify relationships between the severity of crashes and

contributing factors within and across the clusters. At the same time, it may also help to further justify the validity of the grouping of the crash environments performed prior to the analysis of crashes.

Across all clusters, certain indicators, such as speed limits, collisions with motor vehicles, and rider age, were reportedly identified as significant and consistently showed an increasing effect on crash severity. Assuming a built environment typology of crash locations as a system, such factors can act as shared principles within this system, demonstrating consistent impacts. That being said, despite showing the same direction of effect on severity levels, the magnitude of the coefficients is not uniform, which indicates the role of the built environment in causing differences in this aspect. For instance, in areas with a higher density of high-speed facilities (as seen in the first two clusters), speeds greater than 50 mph increase severity to a greater magnitude than in other areas. In contrast, in intersection-dense areas, speeds between 30–50 mph act as a significant risk factor. This demonstrates that, while speed is a shared risk factor increasing severity, the built environment can adjust its magnitude depending on the context in which crashes occur. A similar interpretation applies to sociodemographic factors. As seen in this study, in car-centric locations (with a lower density of bike lanes) in the first cluster, the risk of severe injuries among younger riders escalates compared to other contexts, where marginal effects are lower. This, again, provides an example of how the built environment can shape the impact of risk factors.

In some cases, variables can even show more dissimilar patterns across built environments, losing their significance in some locations while retaining their role in others. For instance, the carriageway indicator variable is shown to be a significant factor, increasing the likelihood of severe crashes in the first two clusters while being insignificant in the others. This discrepancy may be explained by the clustering findings, which show that data points in the first two clusters have a higher density of vehicular facilities, particularly high-speed ones, and simultaneously a lower density of bike lanes. These infrastructural characteristics make e-scooter–motor vehicle frictions on carriageways more likely and severe than in areas that are more micromobility-friendly. Due to such infrastructural characteristics, various collision types like head-on, rear-end, and sideswipe crashes are shown to be significant in the second cluster, indicating a higher likelihood of vehicular interactions. In contrast, in the other clusters, interactions between motor vehicles and micromobility devices are primarily limited to side impacts, which are a frequent collision type at junctions where riders and drivers often meet even in the presence of divided micromobility lanes (as seen in the third cluster with a high density of junctions).

Another situation identified in our analysis, is when crash factors change their direction of effect on crash severity in different clusters. Morning and afternoon rush hours decreased the likelihood of serious and fatal crashes in the third and fourth clusters, while revealing an increasing effect in the first cluster. The answer to such a pattern lies in the characteristics of the environments. Areas with a higher density of intersections (third and fourth clusters) likely experience lower speeds and higher congestion during rush hours compared to the first cluster.

Fig. 7 gives a summary of the recognised patterns. These three possible situations and inferred patterns, within and across clusters, demonstrate how significantly the crash environment affects risk factors and emphasis the need for considering built environment types when addressing e-scooter crashes.

6. Conclusion and limitations

This paper identifies influential factors of e-scooter crash severity over different built environment types by integrating an unsupervised machine learning approach with random parameter binary logit models. The results are that car-centric areas and commercial/industrial areas have high risk of severe injury in higher speed zone environments and in collisions with motor vehicles. Central residential areas and areas with high densities of intersections show the same findings but with a lower magnitude of marginal effects compared to previous clusters. Moreover, the severe outcomes increase for the young riders, with car-centric and mixed-use locations found to be the most dangerous areas based on marginal effects.

Peak hours lower the likelihood of severe injuries in clusters representing central areas as well as intersection-dense zones. However, the converse effect was seen for the cluster containing car-centric and mixed-use areas. Considering manner of collision, side-impacts are the most common significant collision type across identified clusters; however, head-on collisions, which were shown to be significant in the commercial and industrial cluster, have the highest likelihood of serious and fatal outcomes.

Several limitations apply to this study. It is based on UK e-scooter crash data, which may limit the generalisability of the findings to settings outside of the UK and beyond areas where e-scooters currently operate. Furthermore, e-scooters have limited use in the UK, and except for some ongoing trials, they are not currently legal to operate on roads or footpaths. This may influence the characteristics of crashes, and the factors identified in this research may change if e-scooter legislation is relaxed and an increase in usage occurs.

Regarding the built environment, it is acknowledged that only select built environment factors have been considered in this study, which could be extended to include more variables. More importantly, augmenting crash data with other sources could further enhance the analysis, for example imagery data and computer vision could be utilised to

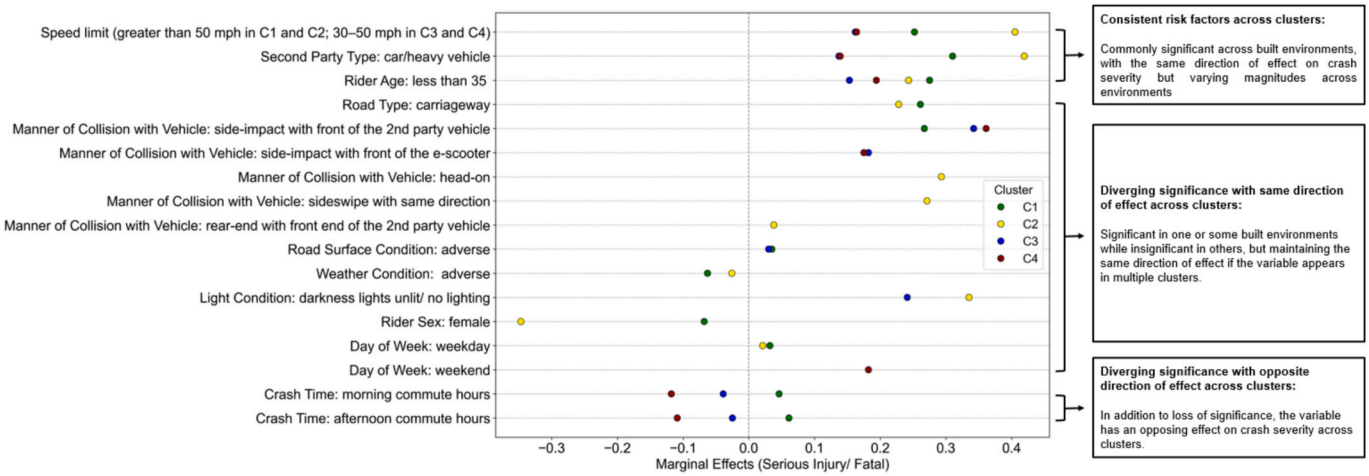

Fig. 7. Summary of crash risk factor comparisons.

extract further built environment characteristics at crash sites. Nonetheless, an increased number of factors could lead to high collinearity between potential factors and cause computational complexity in terms of clustering a high dimensional feature space. In this context, the application of dimensionality reduction techniques is proposed. Additionally, this study does not address the modifiable aerial unit problem, which may introduce potential sensitivity in crash severity analyses related to the buffer area size used for extracting built environment factors. Future studies could systematically investigate the impact of buffer size on e-scooter safety.

The results of crash severity analysis in this research are dependent on the clustering outcome in the previous stage. Therefore, the choice of clustering algorithm and the number of clusters can considerably affect the outcomes. In this respect, K-means++, despite its popularity as a K-means-based technique, has some limitations, such as assuming clusters have a spherical shape in the feature space and being incapable of identifying noise data, as noted by Behara et al. (2021). This requires future work in the application of advanced techniques. Such algorithms, for instance, are Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). These methods were not applied in this study due to sample size limitations, as they reduce the sample size used in the crash severity models by leaving some crashes unlabelled as noise, which requires larger samples for their application.

Finally, this research does not cover the longitudinal aspects of e-scooter safety. A future research avenue could consider temporal variations in the built environment and e-scooter safety, as well as the linkage between usage patterns and crashes. The latter requires usage data provided by city managers or operators to be analysed in parallel with crash trends. As micromobility travel modes continue to evolve, it is necessary to keep examining their effects on urban traffic safety.

The raw dataset used in this study is based on data from the Department for Transport (DfT) (Department for Transport, 2024). The cleaned and processed dataset is available upon request by contacting the first author.

CRediT authorship contribution statement

Amirhossein Abdi: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Steve O'Hern:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abdi, A., O'Hern, S., 2024. Investigating the severity of single-vehicle truck crashes under different crash types using mixed logit models. J Safety Res 88, 344–353. https://doi.org/10.1016/j.jsr.2023.12.001.

Abdi, A., Seyedabrishami, S., Llorca, C., Moreno, A.T., 2022. Exploring the effects of stationary camera spots on inferences drawn from real-time crash severity models. Sci Rep 12 (1), 20321. https://doi.org/10.1038/s41598-022-24102-y.

Aggarwal C.C. Reddy C.K. Data clustering 2018 Chapman and Hall/CRC 10.1201/9781315373515.

Arthur, D., Vassilvitskii, S., 2007. k-means++: the advantages of careful seeding. ACM-SIAM Symposium on Discrete Algorithm. 1027–1035.

Azimian, A., Jiao, J., 2022. Modeling factors contributing to dockless e-scooter injury accidents in Austin. Texas. Traffic Inj Prev 23 (2), 107–111. https://doi.org/10.1080/15389588.2022.2030057.

Badeau, A., Carman, C., Newman, M., Steenblik, J., Carlson, M., Madsen, T., 2019. Emergency department visits for electric scooter-related injuries after introduction of an urban rental program. Am J Emerg Med 37 (8), 1531–1533. https://doi.org/10.1016/j.ajem.2019.05.003.

Behara, K.N.S., Bhaskar, A., Chung, E., 2021. A DBSCAN-based framework to mine travel patterns from origin-destination matrices: Proof-of-concept on proxy static OD from Brisbane. Transp Res Part C Emerg Technol 131, 103370. https://doi.org/10.1016/j.trc.2021.103370.

Bi, H., Li, A., Zhu, H., Ye, Z., 2023. Bicycle safety outside the crosswalks: Investigating cyclists' risky street-crossing behavior and its relationship with built environment. J Transp Geogr 108, 103551. https://doi.org/10.1016/j.jtrangeo.2023.103551.

Bierlaire, M., 2020. A short introduction to PandasBiogeme.

Branion-Calles, M., Götschi, T., Nelson, T., Anaya-Boig, E., Avila-Palencia, I., Castro, A., Cole-Hunter, T., de Nazelle, A., Dons, E., Gaupp-Berghausen, M., Gerike, R., Int Panis, L., Kahlmeier, S., Nieuwenhuijsen, M., Rojas-Rueda, D., Winters, M., 2020. Cyclist crash rates and risk factors in a prospective cohort in seven European cities. Accid Anal Prev 141, 105540. https://doi.org/10.1016/j.aap.2020.105540.

Brauner, T., Heumann, M., Kraschewski, T., Prahlow, O., Rehse, J., Kiehne, C., Breitner, M.H., 2022. Web content mining analysis of e-scooter crash causes and implications in Germany. Accid Anal Prev 178, 106833. https://doi.org/10.1016/j.aap.2022.106833.

Chen, P., Shen, Q., 2016. Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. Accid Anal Prev 86, 239–246. https://doi.org/10.1016/j.aap.2015.11.002.

Cicchino, J.B., Kulie, P.E., McCarthy, M.L., 2021. Severity of e-scooter rider injuries associated with trip characteristics. J Safety Res 76, 256–261. https://doi.org/10.1016/j.jsr.2020.12.016.

Costa, M., Lima Azevedo, C., Siebert, F.W., Marques, M., Moura, F., 2024. Unraveling the relation between cycling accidents and built environment typologies: Capturing spatial heterogeneity through a latent class discrete outcome model. Accid Anal Prev 200, 107533. https://doi.org/10.1016/j.aap.2024.107533.

Costa, M., Marques, M., Roque, C., Moura, F., 2022. CYCLANDS: Cycling geo-located accidents, their details and severities. Sci Data 9 (1), 237. https://doi.org/10.1038/s41597-022-01333-2.

Davies, D.L., Bouldin, D.W., 1979. A Cluster Separation Measure. IEEE Trans Pattern Anal Mach Intell PAMI-1 2, 224–227. https://doi.org/10.1109/TPAMI.1979.4766909.

Department for Transport, 2024. Road safety data [WWW Document]. Data.gov.uk. URL https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data.

Dibaj, S., Vosough, S., Kazemzadeh, K., O'Hern, S., Mladenović, M., 2024. Investigating e-scooter injury and severity before and after the restriction policies in Helsinki, Finland. Paper presented at Transportation Research Board Annual Meeting, Washington, District of Columbia, United States.

Ekmekci, M., Dadashzadeh, N., Woods, L., 2024. Assessing the impact of low-speed limit zones' policy implications on cyclist safety: Evidence from the UK. Transp Policy (oxf) 152, 29–39. https://doi.org/10.1016/j.tranpol.2024.04.014.

Gao, D., Zhang, X., 2024. Injury severity analysis of single-vehicle and two-vehicle crashes with electric scooters: A random parameters approach with heterogeneity in means and variances. Accid Anal Prev 195, 107408. https://doi.org/10.1016/j.aap.2023.107408.

GeeksforGeeks, 2022. Haversine formula to find distance between two points on a sphere. [WWW Document]. www.geeksforgeeks.org. URL https://www.geeksforgeeks.org/haversine-formula-to-find-distance-between-two-points-on-a-sphere/.

Gioldasis, C., Christoforou, Z., Seidowsky, R., 2021. Risk-taking behaviors of e-scooter users: A survey in Paris. Accid Anal Prev 163, 106427. https://doi.org/10.1016/j.aap.2021.106427.

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer Math (heidelb) 2 (1), 84–90. https://doi.org/10.1007/BF01386213.

Haworth, N., Schramm, A., Twisk, D., 2021. Comparing the risky behaviours of shared and private e-scooter and bicycle riders in downtown Brisbane. Australia. Accid Anal Prev 152, 105981. https://doi.org/10.1016/j.aap.2021.105981.

Heydari, S., Forrest, M., Preston, J., 2022. Investigating the association between neighbourhood characteristics and e-scooter safety. Sustain Cities Soc 83, 103982. https://doi.org/10.1016/j.scs.2022.103982.

Ikotun, A.M., Ezugwu, A.E., Abualigah, L., Abuhaija, B., Heming, J., 2023. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Inf Sci (n y) 622, 178–210. https://doi.org/10.1016/j.ins.2022.11.139.

Jones, K., Parkin, J., Rathod, N., Bhatt, V., 2023. Oral and maxillofacial injuries associated with e-scooter use at Broomfield Hospital: a cohort study of 24 months of data since e-scooter legalisation in the UK. Br Dent J. https://doi.org/10.1038/s41415-023-5506-5.

S. Kennedy Astral: Calculations for the position of the sun and moon [WWW Document] 2022 https://pypi.org/project/astral/.

Kleinertz, H., Ntalos, D., Hennes, F., Nüchtern, J.V., Frosch, K.-H., Thiesen, D.M., 2021. Accident Mechanisms and Injury Patterns in E-Scooter Users. Dtsch Arztebl Int. https://doi.org/10.3238/arztebl.m2021.0019.

Kleinertz, H., Volk, A., Dalos, D., Rutkowski, R., Frosch, K.-H., Thiesen, D.M., 2023. Risk factors and injury patterns of e-scooter associated injuries in Germany. Sci Rep 13 (1), 706. https://doi.org/10.1038/s41598-022-25448-z.

Lavoie-Gagne, O., Siow, M., Harkin, W., Flores, A.R., Girard, P.J., Schwartz, A.K., Kent, W.T., 2021. Characterization of electric scooter injuries over 27 months at an

urban level 1 trauma center. Am J Emerg Med 45, 129–136. https://doi.org/10.1016/j.ajem.2021.02.019.
Lee, K.-J., Yun, C.H., Yun, M.H., 2021. Contextual risk factors in the use of electric kick scooters: An episode sampling inquiry. Saf Sci 139, 105233. https://doi.org/10.1016/j.ssci.2021.105233.
Li, J., Li, C., Zhao, X., 2024. Optimizing crash risk models for freeway segments: A focus on the heterogeneous effects of road geometric design features, traffic operation status, and crash units. Accid Anal Prev 205, 107665. https://doi.org/10.1016/j.aap.2024.107665.
Longo, P., Berloco, N., Coropulis, S., Intini, P., Ranieri, V., 2024. Analysis of E-Scooter Crashes in the City of Bari. Infrastructures (basel) 9 (3), 63. https://doi.org/10.3390/infrastructures9030063.
Ma, Q., Yang, H., Mayhue, A., Sun, Y., Huang, Z., Ma, Y., 2021. E-Scooter safety: The riding risk analysis based on mobile sensing data. Accid Anal Prev 151, 105954. https://doi.org/10.1016/j.aap.2020.105954.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal Methods Accid Res 11, 1–16. https://doi.org/10.1016/j.amar.2016.04.001.
Mathew, S., Pulugurtha, S.S., Duvvuri, S., 2022. Exploring the effect of road network, demographic, and land use characteristics on teen crash frequency using geographically weighted negative binomial regression. Accid Anal Prev 168, 106615. https://doi.org/10.1016/j.aap.2022.106615.
Mehranfar, V., Jones, C., 2024. Exploring implications and current practices in e-scooter safety: A systematic review. Transp Res Part F Traffic Psychol Behav 107, 321–382. https://doi.org/10.1016/j.trf.2024.09.004.
Moftakhar, T., Wanzel, M., Vojcsik, A., Kralinger, F., Mousavi, M., Hajdu, S., Aldrian, S., Starlinger, J., 2021. Incidence and severity of electric scooter related injuries after introduction of an urban rental programme in Vienna: a retrospective multicentre study. Arch Orthop Trauma Surg 141 (7), 1207–1213. https://doi.org/10.1007/s00402-020-03589-y.
Morrison, C.N., Thompson, J., Kondo, M.C., Beck, B., 2019. On-road bicycle lane types, roadway characteristics, and risks for bicycle crashes. Accid Anal Prev 123, 123–131. https://doi.org/10.1016/j.aap.2018.11.017.
Naghizadeh, A., Metaxas, D.N., 2020. Condensed Silhouette: An Optimized Filtering Process for Cluster Selection in K-Means. Procedia Comput Sci 176, 205–214. https://doi.org/10.1016/j.procs.2020.08.022.
O'Hern, S., Estgfaeller, N., 2020. A Scientometric Review of Powered Micromobility. Sustainability 12 (22), 9505. https://doi.org/10.3390/su12229505.
Open-Meteo Team Open-Meteo API: Free Weather Forecast and Historical Data API [WWW Document] 2024 https://open-meteo.com/.
OpenWeatherMap, 2024. OpenWeatherMap API [WWW Document]. URL https://openweathermap.org/api.
Pérez-Zuriaga, A.M., Dols, J., Nespereira, M., García, A., Sajurjo-de-No, A., 2023. Analysis of the consequences of car to micromobility user side impact crashes. J Safety Res 87, 168–175. https://doi.org/10.1016/j.jsr.2023.09.014.
Rella Riccardi, M., Mauriello, F., Sarkar, S., Galante, F., Scarano, A., Montella, A., 2022. Parametric and Non-Parametric Analyses for Pedestrian Crash Severity Prediction in Great Britain. Sustainability 14 (6), 3188. https://doi.org/10.3390/su14063188.
Saha, D., Dumbaugh, E., Merlin, L.A., 2020. A conceptual framework to understand the role of built environment on traffic safety. J Safety Res 75, 41–50. https://doi.org/10.1016/j.jsr.2020.07.004.
Shah, N.R., Aryal, S., Wen, Y., Cherry, C.R., 2021. Comparison of motor vehicle-involved e-scooter and bicycle crashes using standardized crash typology. J Safety Res 77, 217–228. https://doi.org/10.1016/j.jsr.2021.03.005.
Shutaywi, M., Kachouie, N.N., 2021. Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering. Entropy 23 (6), 759. https://doi.org/10.3390/e23060759.
Song, Y., Merlin, L., Rodriguez, D., 2013. Comparing measures of urban land use mix. Comput Environ Urban Syst 42, 1–13. https://doi.org/10.1016/j.compenvurbsys.2013.08.001.
Stigson, H., Malakuti, I., Klingegård, M., 2021. Electric scooters accidents: Analyses of two Swedish accident data sets. Accid Anal Prev 163, 106466. https://doi.org/10.1016/j.aap.2021.106466.
Sun, Y., Lu, J., Ren, G., Ma, J., 2024. Exploring the heterogeneities of factors affecting e-bicyclist severities in delivery e-scooter crashes. Journal of Transportation Safety & Security 16 (4), 443–466. https://doi.org/10.1080/19439962.2023.2230544.
Train, K.E., 2001. Discrete Choice Methods with Simulation. Cambridge University Press. https://doi.org/10.1017/CBO9780511805271.
Vernon, N., Maddu, K., Hanna, T.N., Chahine, A., Leonard, C.E., Johnson, J.-O., 2020. Emergency department visits resulting from electric scooter use in a major southeast metropolitan area. Emerg Radiol 27 (5), 469–475. https://doi.org/10.1007/s10140-020-01783-4.
Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC. https://doi.org/10.1201/9780429244018.
White, E., Guo, F., Han, S., Mollenhauer, M., Broaddus, A., Sweeney, T., Robinson, S., Novotny, A., Buehler, R., 2023. What factors contribute to e-scooter crashes: A first look using a naturalistic riding approach. J Safety Res 85, 182–191. https://doi.org/10.1016/j.jsr.2023.02.002.
Wu, Y., Hu, X., Ji, X., Wu, K., 2023. Exploring associations between built environment and crash risk of children in school commuting. Accid Anal Prev 193, 107287. https://doi.org/10.1016/j.aap.2023.107287.
Yang, H., Ma, Q., Wang, Z., Cai, Q., Xie, K., Yang, D., 2020. Safety of micro-mobility: Analysis of E-Scooter crashes by mining news reports. Accid Anal Prev 143, 105608. https://doi.org/10.1016/j.aap.2020.105608.
Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. Anal Methods Accid Res 1, 72–85. https://doi.org/10.1016/j.amar.2013.03.001.
Zhang, S., 2012. Nearest neighbor selection for iteratively kNN imputation. Journal of Systems and Software 85 (11), 2541–2552. https://doi.org/10.1016/j.jss.2012.05.073.