Unsupervised feature extraction and clustering of aerial images for understanding hazardous road segments

John Francis (✉ jfrancis@turing.ac.uk)  
The Alan Turing Institute

Jonathan Bright  
The Alan Turing Institute

Saba Esnaashari  
The Alan Turing Institute

Younma Hashem  
The Alan Turing Institute

Deborah Morgan  
The Alan Turing Institute

Vincent J. Straub  
The Alan Turing Institute

Article

Keywords:

Posted Date: March 16th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2678278/v1

License: ☛ ☛ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on July 5th, 2023. See the published version at https://doi.org/10.1038/s41598-023-38100-1.
Abstract

Satellite and aerial image data are becoming more widely available, and analysis techniques based on supervised learning are advancing their use in a wide variety of remote sensing contexts. However, supervised learning requires training datasets which are not always available or easy to construct. In this respect, unsupervised machine learning techniques present important advantages. This work presents a novel pipeline to demonstrate how available aerial imagery can be used to better the provision of services related to the built environment, using the case study of road traffic collisions (RTCs) across three cities in the UK. In this paper, we show how aerial imagery can be leveraged to extract latent features of the built environment from the purely visual representation of top-down images. Through the clustering of hazardous road segments with these latent image features, this work demonstrates how aerial images and machine learning can provide a data-driven aid for road safety experts to enhance their nuanced understanding of how and where different types of RTCs occur.

Introduction

The field of remote sensing, which is concerned with detecting and monitoring the physical characteristics of an area from a distance, has grown substantially with the proliferation of freely accessible high-quality aerial and satellite imagery. Remote sensing as a domain within computer vision has recently led to advances in agriculture [1], climate change [2], transportation [3], and disaster response [4]. With governments such as the UK aiming to harness the power of machine learning (ML) to deliver “first-class public services” [5], there is a need for academic research that demonstrates how new data and technologies within remote sensing can help allow for the improved provision of public services. This project takes on the important safety issue of road traffic collisions (RTCs) as a case study for exploring how old challenges can be tackled from fresh perspectives through the use of modern data sources, specifically, aerial imagery, and ML methodologies.

Globally, approximately 1.3 million people die each year because of RTCs, and a majority of these deaths fall among vulnerable road users such as pedestrians and cyclists [6]. In the UK, the country we focus on here as a case study due to the government’s prioritisation of road safety and availability of aerial imagery, fatal or serious injuries occur on public roads every 16 minutes [7]. In 2020, the United Nations General Assembly resolved to halve the number of global deaths and injuries from RTCs by 2030, noting that the “overwhelming majority” of these cases are preventable [8]. To achieve this ambitious goal, new technologies and data will be required to enhance road safety experts’ implementation of RTC interventions.

The study of RTCs has a long history with research dating back to the early days of motorisation in the first two decades of the 20th century. While early research focused on the characteristics that make up “accident-prone drivers”, today’s road safety research is more focused on how to best implement policies and interventions aimed at creating a safe road system that works for a variety of road users [9]. Although ML has been utilized in recent research concerned with predicting future RTCs or identifying
hazardous road locations [10,11], this project shows how aerial imagery can equip public officials, who already have some knowledge of which points within the road network are dangerous, with a new perspective. Specifically, we demonstrate how to leverage ML methods to cluster hazardous road segments solely on their built form which can help improve qualitative analysis of and intervention planning for reducing future RTCs. To do this, we utilize unsupervised machine learning techniques to extract latent features from aerial images and then leverage hierarchical cluster techniques to group hazardous road segments. We show how aerial image data can provide nuance to understanding how and where different types of RTCs occur, which can in turn allow for improved targeting of interventions designed at tackling this classical issue.

Aerial image data presents unique challenges in terms of extracting data and meaning. On the one hand, aerial images offer incredible, high-resolution data on the built environment. However, extracting meaning from image data is not as straightforward as working with prototypical tabular data. Image data requires linkages to existing knowledge (road networks in our case) as well as novel metrics of similarity in order to enable comparisons between images. To address these challenges, we present a pipeline that extracts meaning from aerial images with minimal supervision, a process which has been shown to have widespread utility [12], and focus on three case study locations within the UK, Cambridge, Gloucester, and Oxford, to demonstrate the application of the pipeline for road safety analysis.

The rest of this paper is structured as follows. We first review related work on the built environment and road safety, especially prior research that has sought to use aerial imagery and ML methods. Then, we describe the data and methods used in our paper. To apply our methods, we present the results of our analytic pipeline for the three cities we use as case studies: Cambridge, Gloucester, and Oxford. Finally, we discuss the results and conclude by outlining the limitations of our study and considering future directions for similar methodologies in other public settings.

**Literature Review**

Prior research on the built environment has found success by conceptualizing features at the macro- or neighbourhood-level and at the micro- or street-level [13]. At the neighbourhood-level, features such as urban density and street connectivity have been found to be related to outcomes such as walkability [14]. At the street-level, features such as street trees and street width have been associated with housing price [15] and air quality [16], respectively. This study extracts both macro and micro features of the built environment from aerial images to give a more holistic understanding of the larger geographic context around hazardous road segments.

Studies of the built environment have previously made efforts to quantify the many physical attributes of the urban environment. Access to imagery when analysing the built environment has been lauded for better enabling standardised assessments, in turn facilitating comparability across other image-based studies of the built-environment [17,18]. Ewing and Handy [19], for instance, found mixed success using human annotators to measure more than 100 characteristics of the built environment from video clips. Other researchers have used machine learning to try and computationally measure some of these
features from street view images [20]. Measuring features such as building facades, street condition, or even temporally dynamic features, such as the number of parked cars, is tricky, with various cultural or selection biases introduced no matter the method used for quantification. By relying solely on the latent visual features residing within images, this study attempts to bypass some of the issues that arise from manually deriving measures of the built environment.

Given the potential of ML to address global challenges [12], using ML methods alongside imagery to study the built environment has become popular in recent years to study crime incidents [21], urban morphology [22], as well as human perceptions of their urban environments [23]. In many ways, raw imagery provides a more unfiltered view of the built environment as opposed to the more common annotated datasets such as land use or street features in a tabular format.

One popular area of study that relies heavily on data about the built environment is road safety research. In their comprehensive review of RTC research, Gutierrez-Osorio & Pedraza [24] found that governmental data such as road infrastructure design features were among the most commonly used sources. Importantly, numerous features of the built environment have been found to contribute to RTCs. Looking at RTCs involving pedestrians, Guo et al [25] found that more densely connected road networks were related to higher incidence rates. Meanwhile, a qualitative study of fatal collisions in the UK found that excessive speed on road bends was a clear factor in many fatal accidents, particularly in low-lit areas during hours of darkness and on more rural roads [26].

More recently, deep learning methods have become common in road safety research, primarily as a means for making predictions with tabular data to better capture the spatial and temporal aspects of RTC data. Convolution neural networks [10], negative binomial models [27], and long short-term memory networks [28] have all been shown to predict RTCs with relatively high accuracy and claim to be more effective than classical regression techniques. Other methods such as decision trees [29] and extreme gradient boosting [30] have been used to explore the relative importance of various built environment attributes on RTC frequency and injury severity respectively. Closely related to this project, prior research by Zhang et al. [11] used graph neural networks to extract latent visual features from satellite imagery in order to identify hazardous traffic locations in conjunction with social media data. Their work showed that ML-derived image features out-performed conventional and deep learning models for traffic risk forecasting in New York. Additionally, the authors observed that “locations with similar accident rates tend to share similar visual features” [11, pg. 2]. While Zhang et al. [11] focused on mining historical data to make predictions about future hazardous locations, our work hopes to help officials better understand past RTCs to allow for improved interventions that can prevent future RTCs.

An area of major concern for road safety research is omitted variable bias. Often, many factors affect the likelihood of a RTC which do not have detailed data available for analysts. Statistical methods that are meant to account for this unobserved heterogeneity often struggle with increased complexity, complicating interpretability and transferability [31]. While it is extremely difficult and costly to collect
data on some of these unknown factors, such as the mental state of drivers, raw aerial image features allow for a more complete view of aspects of the built environment that contribute to each incident.

One less studied area is the use of unsupervised ML methods to extract latent image features from the built environment. In one study, Singleton et al. measure local spatial structures from the latent image features derived from Sentinel 2 satellite data, creating a new measure of geodemographic classification [32]. Another study extracted the latent features from street view imagery to allow for a more interpretable method of predicting street quality and street network attributes [33]. Further work described the areas around leisure and retail amenities using latent image features extracted from storefront images [34]. One common thread among these works is the use of dimension reduction techniques such as k-means or principal component analysis (PCA) to allow for an easier interpretation of the latent image features.

Clustering algorithms are among the most used analytic tools among RTC researchers [24]. Clustering algorithms calculate the similarity across features using a distance function to partition objects into clusters. Unlike supervised ML tasks such as classification, clustering does not require data points to have a predefined target category to train the algorithm on. One limitation of commonly used clustering algorithms is the possibility that there can be a multitude of valid possible solutions. Additionally, just because clusters can be formed, does not mean that the clusters themselves are meaningful. This study utilizes agglomerative clustering with Euclidean linkage distance, which is a bottom-up hierarchical clustering technique that tends to form higher quality clusters although it can be expensive computationally [35]. Another benefit of agglomerative clustering is that the method is computationally robust, so the same result is achieved every time.

Taken together, prior work has shown that combining aerial imagery with ML has the potential to help address global challenges relating to the built environment, specifically, RTCs. While urban planners have long studied the physical characteristics that influence traffic patterns and RTCs, harnessing aerial imagery and ML can enable governments to make more accurate forecasts and devise policies informed by city-level data. Importantly, rolling out such technologies also creates new dilemmas, especially if the focus is on ML solely as a predictive tool. For example, what should a city planner do if presented with a statistical probability of 60%—or even 97%—that a road segment will be the site of a future RTC? Should the government invest more resources in interventions to make that location safer for pedestrians, or divert them away? Here, we do not undertake such a normative stance or offer a predictive analysis, instead, we seek to show the potential of using ML to help road safety experts understand their cities better by deriving useful insights about the built environment from aerial imagery and historical RTC data. As such, our results demonstrate the potential of unsupervised ML techniques in extracting aerial image features that can effectively cluster hazardous road segments.

**Data**

In this section, we will describe the data used in our paper. Our data are taken from three sources: aerial imagery from the EDINA Aerial Digimap Service; street network data from OpenStreetMap; and RTC data from the UK Department for Transport (DfT). These three data sources are described in turn below.
Aerial Image Data. Vertical aerial image RGB data for this project were retrieved from the EDINA Aerial Digimap Service at 25cm resolution [36]. In total 540 orthorectified 1km by 1km images were acquired across three mid-sized UK towns. Cambridge, Gloucester, and Oxford were chosen as they are comparable in terms of population and land area; this, coupled with their proximity to one another and shared morphological heritage [37] allows us to reasonably assume that their built environments will contain similar features. Image data varies in availability from year to year, the full details of which can be found in Supplementary Table S1. Multiple years of images were acquired to bolster the number of available training samples, however for final analyses only a single year’s worth of images from each locale are included.

Street Network Data. Street network data for this project was retrieved from OpenStreetMap [38] for a single time point in June 2020. We observed minimal change in these street networks over the time period of the acquired aerial image data, so a central time point was selected. Only roads which are accessible to cars were included in this analysis, so the network was filtered to remove cycleways, footpaths, and other types of non-vehicular roads.

Road Traffic Collision Data. RTC data is based on the point-level incidents provided by the DfT road accidents and safety statistics publication for the years 2017–2020 and compiled with the stats19 R package which enables access to the UK’s official RTC database [39]. Importantly, the RTC data used in this article only includes collisions involving personal injury on public roads which have been reported to police within 30 days. In the UK, data on vehicular damage-only RTCs do not generate a police report and incident counts are not published. While this paper focuses only on RTCs involving personal injury, we believe these incidents are the most critical to mitigate and have the potential to benefit most from this analysis. Besides the location of the RTC, descriptive variables are also recorded for each incident including the type of road user involved in the incident. Full details on the RTCs included can be found in Supplementary Table S2.

Methodology

Figure 1 shows the methodological pipeline employed for this analysis. First, road networks and aerial images were combined into one dataset of evenly spaced road segment points, which were further linked to RTC data. Then, a convolutional autoencoder (CAE), PCA, and hierarchical clustering are used to extract image features from the data before grouping road segments for analysis and interpretation. We describe each of these steps in turn in the sections below.

Data Combination. Starting with the street network data, points were sampled evenly along the street network every 50 metres. Two square buffers, one with a radius of 25 metres and one with a radius of 100 metres, were created for each point. We expect that having two buffers will allow for better feature extraction overall, as the 25-metre buffer should capture micro features about the road itself, while the 100-metre buffer should capture macro features about the local street network and the urban form. The small buffer size was chosen to capture the full width of the dataset’s widest road sample, while the large
buffer size was chosen as it has been used previously as a search radius for traffic accident research [40]. Each of the square buffers was matched to an aerial image and clipped to create two sets of image patches for feature extraction. The 100 metre buffers were then resampled from a 25cm resolution to a one metre resolution so both sets of images would have a size of 200x200 pixels. This process generated 107,305 images for the small buffer, and 105,441 images for the large buffer. These numbers differ as a small number of the large buffers did not fall entirely within the acquired aerial images. While all images were used to train the CAEs, only points with both a small and large buffer were included in the final feature extraction.

For each point and associated images, RTCs were assigned if they occurred within 50 meters of each point. This process allows for a single RTC point to be assigned to multiple road segments if it occurred somewhere in the middle. Additionally, RTC data was aggregated for the years 2017–2020 so each image was assigned four years of RTCs to help account for some of the natural variance that can cause the number of RTCs in a particular location to fluctuate from year to year. Figure 2 maps out the location of each of the 1,575 road segments which contained at least one RTC in each of the three cities under study. For final analyses, this study focuses only on 334 hazardous road segments, also shown in Fig. 2, which are defined here as points that contained at least three RTCs over the study period, similar to definitions used throughout Europe [41].

It is worth noting that, due to the dense road networks in parts of the study cities, road segments that were sampled every 50m along the street network could end up significantly closer than 50m in straight line distance (especially when they occur around intersecting roads). While these segments contained relevant information to warrant inclusion in the model training and feature extraction steps, to avoid double counting similarly located segments in the clustering process, a further reduction of points was performed to ensure no segments remained within a 50m buffer radius of any other point. A breakdown of the number of segments at each step in the analytic pipeline can be seen in Table 1.

### Table 1
**Data samples used throughout the analytic pipeline.**

<table>
<thead>
<tr>
<th>Locality</th>
<th>Image year selected for final analyses</th>
<th># of segments used for feature extraction</th>
<th># of segments used in training</th>
<th># of segments after point reduction</th>
<th># of hazardous road segments (&gt;3 RTCs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloucester</td>
<td>2018</td>
<td>20,736</td>
<td>42,076</td>
<td>3,029</td>
<td>67</td>
</tr>
<tr>
<td>Cambridge</td>
<td>2020</td>
<td>15,676</td>
<td>32,022</td>
<td>2,776</td>
<td>129</td>
</tr>
<tr>
<td>Oxford</td>
<td>2019</td>
<td>17,142</td>
<td>33,207</td>
<td>3,154</td>
<td>138</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>53,554</td>
<td>107,305</td>
<td>8,959</td>
<td>334</td>
</tr>
</tbody>
</table>

**Feature Extraction.** Previous works have utilized CAEs to extract latent features from images [32,33]. By creating a convolutional network that can recreate an image, the feature space in the middle of the CAE
forms a set of latent features that should contain a good representation of the data in a one-dimensional space. For this project, two identical CAEs were trained separately on the small and large image buffers respectively. The full model architecture can be found in Supplementary Figure S1. Each of the CAEs included 518,691 parameters split across an encoder and decoder culminating in a sigmoid activation. Both models were trained for 50 epochs using an Adam optimizer.

The input tensor to each of these CAEs is a 3x200x200 RGB image. Through the encoding portion of the model, whereby the latent structure of the images is learned, each image is compressed into a set of 4,608 features. The size of this feature space was chosen to avoid over-compressing the data such that features could not be learned by the decoder algorithm, while also not creating so many features that would result in a sparse vector where features are hyper specialized to a small set of images.

While over 100,000 images were used to train each of the models, for the final feature extraction, only one timepoint from each locality was used to extract a final set of features. Timepoints were selected based on their temporal proximity to each other and their coverage of the study area which varied year over year. By including latent features from each of these two CAEs in the final clustering analysis, we expect that a set of built environment variables will be constructed that represent a holistic view of the area around each road segment, capturing both street-level and neighbourhood-level features.

Dimension Reduction and Clustering. After extracting the latent features from the two CAEs, we next ran a PCA separately on each set of latent features to aid in dimension reduction. The PCA forms a representation of the latent feature space with a higher degree of interpretability than the features generated by the CAEs, while also concentrating a larger percentage of the variability from the image features into a smaller number of principal components [42].

Agglomerative clustering was then performed on the combined sets of principal components to create a set of clusters from hazardous road segments. Importantly, the principal components computed from the latent image features likely contains common features that are shared by most images. To ensure a clustering analysis which could differentiate hazardous road segments, we only included a small subset of the principal components which displayed the highest correlation to at least one of the RTC variables. The number of highest correlated principal components selected varied by locale (Cambridge N = 946; Gloucester N = 73; Oxford N = 154). The final number of components used was determined iteratively by running the agglomerative clustering analysis on different sets of components and analysing the silhouette scores from 2–15 clusters to determine which sets of components with the highest correlation to RTCs also identified the most well-formed clusters.

Dendrograms in conjunction with metrics such as the silhouette scores were used to select the final number of clusters for each location. Dendrograms are a pictorial display of the hierarchical process of points being merged into successive clusters, with the lines connecting clusters representing the distance between each set of points [43]. Silhouette scores on the other hand are one measure of how tightly clusters are internally alongside how well separated clusters are from each other [44]. Final clusters were chosen primarily due to their jump in linkage distance between merged groups alongside visual
inspection of cluster members, while silhouette scores primarily helped avoid the selection of particularly ill-formed clusters. This clustering analysis was done separately for Cambridge, Gloucester, and Oxford to allow some variance based on the unique combination of features that make up each city.

After clustering, descriptive analysis and visual inspection of the clusters was performed to determine the utility of this process. Clusters were examined to find potential outliers or interesting patterns on the various RTC variables. One cluster from each locale was then chosen to focus on and interrogate for potential usefulness. Shared aspects of each cluster's road segments were identified which appear to be particularly associated with common RTC types within the cluster.

**Results And Discussion**

**Clustering Hazardous Road Segments.** Agglomerative clustering for Cambridge, Gloucester, and Oxford generated three separate dendrograms shown in Fig. 3. The dotted horizontal line in each plot signifies the cut point for choosing the number of clusters, whereby the number of intersecting vertical lines represents the number of final clusters. For Cambridge and Gloucester, three clusters were determined to be most appropriate, while for Oxford four clusters of hazardous road segments were deemed most appropriate for this analysis. There is no one size fits all approach to selecting the number of clusters, exemplified by the fact that not each locale was determined to contain the same number of clusters.

**Descriptive Analysis of Clusters.** After identifying the final number of clusters in each location, descriptive analyses examined the various RTC variables to see if there were differences across clusters. Figure 4 shows selected RTC variables from each locale highlighting some of the most interesting RTC features. The goal of this descriptive analysis was to identify one cluster of interest from each locale to examine in detail based on variation we observed among the different RTC variables. Looking at the breakdown of RTCs by type within each cluster of Gloucester, one cluster combines road segments with a higher average rate of RTCs involving motorcycles than either of the other two clusters. In Cambridge, another cluster stands out as having a high average number of RTCs involving cyclists. Alternatively, in Oxford RTCs involving pedestrians were observed at a higher rate in one cluster than any of the other clusters. These three clusters will be examined qualitatively through visual inspection and mapping to explore the type of insights that can be gained from this method.

**Visual Interpretation of Clusters.** Figure 5 shows a visual inspection of five randomly chosen road segments from the chosen clusters. Each road segment is shown with their two paired buffers. Policymakers from these cities will be best positioned to interpret the clusters and determine what interventions might be appropriate, however there are some high-level conclusions that can be made from a simple visual inspection. For instance, the Cambridge cluster's road segments all highlight long straight roadways with merging lanes. This cluster also has a high rate of serious RTCs, likely suggesting higher speeds are involved. The danger of these higher speeds could be further enhanced by the curved nature of these intersections. Interestingly, Cambridgeshire has recently begun fast-tracking lower speed limits in
selected zones throughout the county to combat this very issue [45], highlighting one possible intervention which could be used to help reduce RTCs in this cluster.

In Gloucester, roundabouts appear in nearly all the images of the selected cluster. Given the high rate of RTCs involving motorcycles within this cluster, it could be that motorcycles have particular difficulties with roundabouts in Gloucester, perhaps emphasizing the need for clarity in their right of way assignments. This echoes Gloucestershire's recent Local Transport Plan which called for interventions of increased education programmes aimed at motorcyclists [46].

Looking at the chosen Oxford cluster, these road segments contain the highest rate of RTCs involving pedestrians and appear to be in denser urban areas with T road segments. This could indicate places where pedestrians need to cross the street, but don't currently have a straightforward means to do so. Targeting this cluster will be essential for Oxford as creating more pedestrian-friendly roads is a key component of Oxfordshire's Vision Zero plan to eliminate fatalities from RTCs by 2050 [47].

The right side of Fig. 5 shows the geographical distribution of these selected clusters. In each locale there is occasional grouping of points along thoroughfares and around large junctions. Some geographical clustering is expected given the use of aerial images, especially since the large buffers may overlap in places. Despite this, for each location hazardous road segments with similar built forms are identified across large distances in disparate parts of each city. This highlights the difficulties faced by policymakers in tackling an issue that spans across diverse communities.

**Harnessing the Built Environment.** Initial inspection of the feature extraction process and the created clusters indicates that the methodology presented in this paper can identify meaningful clusters of hazardous road segments in three UK cities. In each location, a group of hazardous road segments was reviewed which share similar image-derived features. Because these road segments share common characteristics pulled solely from their built form, when it is found that these road segments also share similar challenges, such as a propensity for RTCs involving pedestrians, policymakers should have more confidence that these locations can be targeted by a comparable set of interventions. Looking at the whole set of hazardous road segments, there is not a strong reason for public officials to believe that just because similar RTC incidents frequently occur at multiple locations, that the cause of those RTCs and therefore planned interventions should also be similar in those locations. By utilizing aerial image features, this work suggests that the similarities found in the built environment in different parts of these cities can be leveraged to create better targeted interventions at reducing future RTCs.

**Minimal Supervision.** This methodology provides a straightforward pipeline with minimal supervision for road safety experts to harness the power of ML through modern forms of data for the improved provision of public roads. Notably, by designing a pipeline that utilizes unsupervised ML techniques but is intentionally not fully unsupervised, experts can be kept in the analytic loop to leverage their domain knowledge for enhanced qualitative analysis of hazardous road segments. Much research utilizing ML in road safety research is focused around using the newest methods to make better predictions about where RTCs will happen in the future or identifying where RTC hot spots are. However, simple descriptive
analyses of governmental datasets can inform policymakers exactly where the most problematic road segments are. Much of the difficulty around the provision of increasingly complex public services is figuring out how to leverage expert knowledge to allocate a limited set of public funds in the most effective manner. The methodology provided in this paper lays out a data driven process which enables road safety experts to employ data sources and technical tools which may be underutilized within the public sector.

Limitations. Although this paper presents some promising findings, there are some important limitations to consider. Most notably, the relatively simple visual feature extraction method used in this paper is likely insufficient to capture all the RTC-related features contained within aerial images given the complexity and scope of the potential feature space. More advanced ML architectures, such as the addition of attention mechanisms [11], may be able to better extract RTC-related features. Additionally, cluster analysis itself has some limitations. Methods which are used to determine the ideal number of clusters, such as the silhouette score, are not always reliable, expert knowledge is often required to make meaningful interpretations of groupings, and results can be difficult to replicate, especially with alternative clustering algorithms that make different assumptions about what constitutes a cluster [48]. Furthermore, refinement will be needed in coordination with public sector collaborators to ensure that this and similar tools can be integrated effectively into current public sector decision-making streams.

Conclusions And Future Research Directions

This paper presented a pipeline to extract features from aerial images to allow for the clustering of hazardous road segments in three UK cities. The clusters examined in this paper appeared well formed upon visual inspection, allowing an interpretable framework from which to assess the possible interventions that may be appropriate for reducing certain types of RTCs which are particularly prevalent in these groups. This research was designed to demonstrate how aerial imagery and ML methods can be leveraged to aid decisions that should reduce RTCs on public roads and improve governmental decision making. Building off this work, it is easy to see how similar methodologies could utilize aerial imagery to better understand other factors related to the built environment such as crime, energy use, pollution, wildlife management, and food access, to name but a few.

Declarations

Data Availability

The aerial imagery used in this study are available from the EDINA Aerial Digimap Service to members of subscribing higher and further education institutions in the UK. All other data sources are publicly available via OpenStreeMap and the UK’s Department for Transport. The code used in this study can be found at: https://github.com/ai-for-public-services/sat-img-demonstrator.

Author contributions
J.F. developed the presented idea, performed the analysis, and wrote the initial draft. J.B. supervised this work. All authors discussed the results and contributed to the final manuscript.

**Funding**

This work was supported by Towards Turing 2.0 under the EPSRC Grant EP/W037211/1 & The Alan Turing Institute.

**Competing interests**

The authors declare no competing interests.

**References**


**Figures**
Figure 1

Aerial Image Road Traffic Collision Methodological Pipeline. Plots and images shown here were generated for illustrative purposes only.
Figure 2

Locations of RTC and hazardous RTC road segments across the three UK towns.
Figure 3

Dendrograms used to select the final number of clusters for the three UK towns. Linkage distance should not be compared across the different towns as it depends on the number of variables included in the clustering, which differed by locale.
**Figure 4**

Dot plots of selected RTC variables by cluster in each of the three UK towns. A Z-score of 0 denotes a mean level of that type of RTC for the specified locale, Z-scores should not be compared across locales.
Figure 5

Visual interpretation of the selected cluster for each of the three UK towns. The maps on the right only shown the location of hazardous road segments from the selected cluster, with the randomly chosen images highlighted in red.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• supplementaryinfo.docx