

# Empowering Personal Navigation Technologies with Road Safety Awareness

Runsheng Xu  
Northwestern University  
Evanston, IL, United States  
RunshengXu2017@u.northwestern.edu

Shibo Zhang  
Northwestern University  
Evanston, IL, United States  
shibo.zhang@northwestern.edu

Peixi Xiong  
Northwestern University  
Evanston, IL, United States  
peixixiong2018@u.northwestern.edu

Allen Yilun Lin  
Northwestern University  
Evanston, IL, United States  
allen.lin@eecs.northwestern.edu

Brent Hecht  
Northwestern University  
Evanston, IL, United States  
bhecht@northwestern.edu

## ABSTRACT

Recent research has found that navigation systems usually assume that all roads are equally safe, directing drivers to dangerous routes, which led to catastrophic consequences. To address this problem, this paper aims to begin the process of adding road safety awareness to navigation systems. To do so, we first created a definition for road safety that navigation systems can easily understand by adapting well-established safety standards from transportation studies. Based on this road safety definition, we then developed a machine learning-based road safety classifier that predicts the safety level for road segments using a diverse feature set constructed only from publicly available geographic data. Evaluations in four different countries show that our road safety classifier achieves satisfactory performance. Finally, we discuss the factors to consider when extending our road safety classifier to other regions and potential new safety designs enabled by our road safety predictions.

## CCS CONCEPTS

• Information interfaces and presentation (e.g., HCI);

## KEYWORDS

Road Safety; Personal Navigation Technologies; GPS; Map Apps

## ACM Reference Format:

Runsheng Xu, Shibo Zhang, Peixi Xiong, Allen Yilun Lin, and Brent Hecht. 2018. Empowering Personal Navigation Technologies with Road Safety Awareness. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/1122445.1122456>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

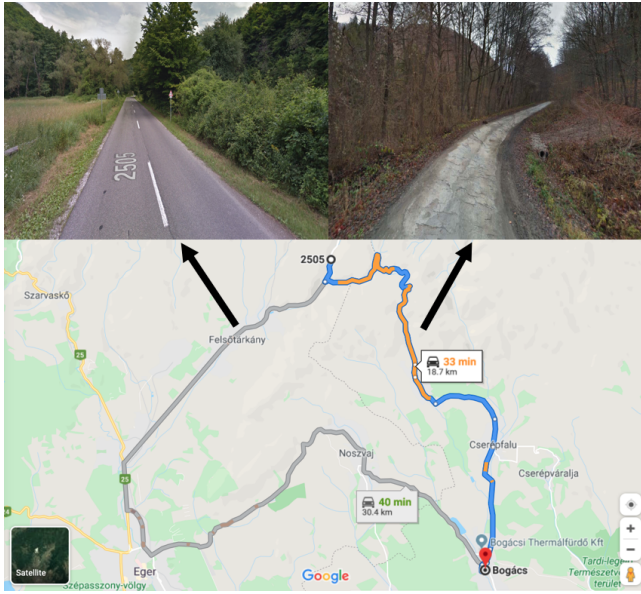
<https://doi.org/10.1145/1122445.1122456>

## 1 INTRODUCTION

While reliable in most cases, personal navigation technologies (e.g., Google Maps) occasionally route drivers to dangerous roads, resulting in catastrophic incidents [16]. Prior Geographic Human Computer Interaction research identified 47 catastrophic incidents in which navigation systems directed drivers to unpaved roads, narrow lanes, and roads with low clearance, causing drivers to get stranded, to crash with roadside objects, and to hit overpasses [42]. In these incidents, drivers bore financial losses, suffered severe injuries, and, in extreme cases, lost their lives.

One key reason behind these catastrophic incidents is that personal navigation systems usually lack road safety information and thus assume all roads are equally safe [42]. As such, navigation systems might generate routes that prioritize other parameters such as travel time over travel safety, resulting in short-yet-dangerous routes. Indeed, many short-yet-dangerous route recommendations that potentially lead to crashes can be found on Google Maps, the most popular personal navigation system used by over 50 million people every day [12]. For instance, Figure 1 compares two routes with the same origins and destinations generated by Google Maps in a rural area of Hungary. Without road safety information, Google Maps recommended a route that is faster but more dangerous (highlighted). The roads on this route are narrow and poorly surfaced, which, according to transportation research [15], are more dangerous for drivers. In comparison, the alternative route (gray) is slower but safer - it passes through roads that are wide and well-maintained.

To help navigation systems generate safer routes for drivers, we sought to begin the process of adding road safety awareness into personal navigation devices. We accomplished this goal in two stages. First, we developed a straightforward road safety definition that personal navigation systems can easily understand by adapting well-established safety standards from transportation studies. Specifically, we focused on the Star Ratings standard from International Road Assessment Programme (iRAP), which gives a single measure of the "built-in" safety for every 100-meter road segment (i.e., safety measured by a road design alone) [4]. Based on the Star Ratings standard, we created a straightforward binary definition for high-risk roads and safe roads, which navigation systems can smoothly incorporate.



**Figure 1: Without road safety information, Google Maps recommended a faster but more dangerous route (highlighted). By contrast, the alternative route (grey) is slower but safer. Images are from Google Maps and Google StreetView.**

In the second stage of our work, we utilized machine learning to produce large-scale predictions using our binary road safety definition. Although the road safety labels we defined can be directly converted from the Star Ratings scores, existing Star Ratings scores only cover a small percentage of roads, which is insufficient for navigation systems. To increase the coverage of road safety labels, we developed a classifier that automatically predicts high-risk roads and safe roads. Our classifier relied on a key insight: many road attributes that are highly relevant to road safety are with open access and can be used as features to predict the road safety labels. Leveraging open data from geographic peer-production projects and government transportation agencies as features, we built and tested our road safety classifier with straightforward machine learning algorithms in four different countries (Greece, Slovenia, Croatia, and the Netherlands). Our results show that our model achieved useful performance, consistently outperforming baseline approaches.

We continue below with an overview of related work, followed by a description of our road safety definition in the context of navigation systems. We then describe the construction and evaluation of our road safety classifier, which predicts the road safety labels according to our definition. Finally, we discuss how our road safety classifier may be extended to regions beyond the four countries studied in this paper, and how road safety awareness can enable new technology designs that significantly enhance the safety of drivers who use navigation systems.

## 2 RELATED WORK

### 2.1 Safety Risks Associated with Navigation Systems

Many studies in Geographic Human-Computer Interaction [31] have investigated the risks associated with personal navigation devices. While early work primarily focused on identifying (e.g., [13, 35, 40]) and addressing (e.g., [14, 53, 64]) risks related to distraction introduced by the navigation systems, recent studies have highlighted other types of risks, including the risks of recommending unsafe routes that are prone to accidents. Most notably, Lin et al. [42] analyzed 158 catastrophic incidents associated with personal navigation systems and found that 47, or about 1/3 of the events they collected, were attributed to dangerous routes recommended by the navigation systems. For example, Lin and colleagues found that personal navigation systems sometimes routed drivers through roads that are unpaved, narrow, or with low clearance, causing vehicles to get stranded, to crash with roadside objects, or to hit overpasses. Furthermore, Lin et al. also highlighted a set of road attributes, including road surface, lane width, and vertical clearance, that are critical for navigation systems to consider in order to generate safe routes for drivers.

Our paper has a twofold relationship with the work by Lin et al. First, our work directly addressed the risks associated with dangerous routes identified in their work by adding road safety awareness into navigation systems. Second, as we show later, the road attributes that Lin et al. recognized as critical to road safety informed how we defined and predicted road safety for navigation systems.

### 2.2 Computational Approaches to Road Safety Assessment

Computer science community has developed many computational approaches to automate the well established road safety assessment in transportation studies. One popular approach is the image-based approach, which applies computer vision techniques on road images to estimate road features needed for safety assessment [39, 43, 66] and/or to directly predict overall safety scores [62]. Most relevant to our work, Song et al. [62] applied Convolutional Neural Network and multi-task learning to automate the Star Ratings road assessment. By analyzing street-level panorama images, their system can estimate a variety of road safety features such as lane width, number of lanes, and road curvature and can predict the overall safety score defined in Star Ratings standard.

While our work are relevant to existing road safety assessments (especially to the Star Ratings standard), we focus on a different problem. Compared with previous work which focused on automating these well-established road assessment process, we used these road assessment to study road safety in the targeted context of navigation systems. In addition, instead of using images as the primary data source for prediction, our work explore a different data source - crowdsourced geographic data from OpenStreetMap, which can be complementary to image data for both tasks of automating road safety assessment and generating road safety data for navigation systems. We return to this point in the Discussion section.

### 3 DEFINING ROAD SAFETY FOR NAVIGATION SYSTEMS

In this section, we detail how we defined road safety for navigation purposes. We first describe how we selected an appropriate safety standard from transportation studies and then explain how we simplified this standard to create a straightforward definition that navigation systems can easily understand.

Road safety standards developed by transportation researchers can be broadly classified into two types. The first type of standard is based on crash statistics (e.g., [8, 25, 61]), which computes a road safety score as the ratio of the number of crashes to a certain measure of exposure such as the population in a region. The second type of standard is road attribute-based (e.g., [4, 21]) and computes the road safety score as a polynomial of many design attributes of the roads, including but not limited to width, surface, and vertical clearance.

We selected the most widely used road attribute-based standard - the Star Ratings standard [4] - as the basis for our road safety definition. Unlike the crash statistics-based standards which are influenced by factors irrelevant to road's physical safety such as local driving habit [37], road attributed-based standards, especially Star Ratings, measure only the safety that is "built-in" to the road [4]. Proposed and maintained by the International Road Assessment Programme (iRAP) <sup>1</sup>, the Star Ratings standard has been partially implemented in more than 100 countries around the world. The Star Ratings standard calculates a single score ranging from 1-star (most dangerous) to 5-star (safest) using a formula that involves an extensive list of safety-related road design attributes. The Star Ratings standard is particularly attractive to our purpose because its definition incorporates all road attributes that Lin et al. identified as the major contributors to GPS-related catastrophic incidents [42], such as width, surface, and vertical clearance of the roads.

However, Star Ratings standard can not be directly used by navigation systems because the five-star systems are difficult to interpret for navigation purposes. As a standard created for transportation studies, one of the main purposes of the Star Ratings is to offer detailed and accurate information to evaluate the cost of road improvement projects [3]. As such, while the five-star system in the Star Ratings is perfect for differentiating the amounts of money required for the road improvement projects, it does not offer interpretations about which roads are desirable for drivers to travel - a critical piece of information for navigation systems.

To address the lack of interpretability issues, we simplified the Star Ratings standard to create our road safety definition. Among the five levels in the Star Ratings system, the three-star level is a critical threshold. Most notably, three-star is considered as an economical criterion to ensure a desirable safety level by iRAP [4] and has been adopted by the World Bank and other government agencies as a requirement for the roads construction projects [49] around the world. In other words, roads with three stars and above are considered to have reached the minimum safety for traveling purposes.

Leveraging this critical threshold, we collapsed the five-star system in the Star Ratings into two classes and created the following binary definitions for road safety:

- **Safe roads:** road segments with three stars and above
- **High-risk roads:** road segments with one or two stars.

### 4 PREDICTING ROAD SAFETY

Using the above binary definition of road safety, it is trivial to convert the existing Star Ratings data into road safety labels. However, the resulting road safety labels have one significant limitation - they do not cover road segments without existing Star Rating scores. In fact, due to the high implementation cost of the Star Ratings standard which requires manually labeling road design attributes from images of the road, the existing Star Ratings scores are extremely sparse. For example, according to iRAP's official statistics [4], even in regions that have the best implementation such as Europe, only an average of 5.4% of road segments are rated with Star Ratings scores. As such, any binary road safety labels generated from these Star Ratings scores do not have sufficient coverage for personal navigation technologies.

To scale up the coverage of our road safety labels, we proposed and evaluated a machine learning-based road safety classifier that automatically predicts our binary road safety labels for road segments. As such, the structure of this section follows the best practices of applied machine learning research (e.g., [41, 45, 63]). We first define the feature set used in our road safety classifier. We then summarize how we collected the ground truth datasets, including both the features and the road safety labels, and highlight the unique characteristics of these data. Finally, we detail the training and evaluation strategies and find that our classifier achieves useful performances that outperform the baseline approaches.

#### 4.1 Feature Definitions

Our road safety classifier is only as powerful as the features it leverages. Our choice of features is grounded in well-established theories from transportation studies that demonstrated these features' high relevance to road safety. In addition, to facilitate other researchers to replicate our results and leverage our approach, we purposefully only selected features that come from two publicly accessible sources - geographic peer-production projects and government transportation agencies.

**4.1.1 OpenStreetMap Features.** The most important features used in our road safety classifier are the features from OpenStreetMap (OSM). As known as the "Wikipedia of maps" [2], OpenStreetMap is a highly successful geographic peer-production project in which volunteers gather to produce free and accurate data for all geographic entities, including roads, for the entire world. Geographic data from OSM have been widely used in a variety of commercial-level map products such as Bing Maps and Facebook Live Maps [5].

The 12 OSM features we selected, shown in the top section of Table 1, can be broadly classified into two categories. The first category includes the OSM data that directly map to road attributes used in the Star Ratings standard. Because our road safety label is defined based on the Star Rating standard, we hypothesized that these OSM features would be highly relevant to our road safety labels. We manually checked the definitions of over 100 road attributes defined by OSM community [20] and identified eight such OSM

<sup>1</sup><https://www.irap.org/>

features such as maximum speed, road surface, and road lighting condition.

The second category of OSM features that our road safety classifier used characterizes the accessibility from the location of a road segment. Accessibility in transportation planning describes how easy it is to reach other areas from a given location [46]. We selected features about accessibility because prior research in transportation studies has established that places with lower accessibility are associated with poorer road safety [11, 27, 50]. In this paper, we focused on accessibility metrics that solely depend on roads' topological structure. Specifically, we used isochrones, which show the area a person can reach from a given location within a given period [51]. We computed the 15-minute isochrone from the midpoint of the road segment and included the area of the isochrone, the total population within the isochrone, and the reach factor (detailed in Table 1) as our features.

**4.1.2 Traffic Statistics Features.** In addition to features from OSM, we also included traffic statistics as features. We do so because prior research in transportation studies found that traffic statistics, such as the operating speed and traffic volume, are highly correlated to road safety [23]. The bottom section of Table 1 details the traffic statistics features, including annual average daily traffic, the mean operating speed of cars, the 85% operating speed of cars.

However, compared with the OSM features, traffic statistics features suffered from a major drawback. While OSM data around the world are accessible through a centralized repository maintained by the OSM community<sup>2</sup>, the accessibility of traffic statistics highly varies - local and national government agencies make own decision on whether to publish the data or not.

We took this ecological validity concern into account when designing our machine learning experiments. Specifically, we trained and evaluated two versions of our classifier - one with all features and the other with OSM features only - in order to gain a comprehensive understanding of the ecological validity of our model. As we show later, parsimonious models with only OSM features performed almost as good as the models trained on all features, highlighting that our model is robust in areas where traffic statistics are unavailable.

## 4.2 Ground Truth Datasets

**4.2.1 Building Ground Truth Datasets.** We prepared four ground truth datasets, each representing a sample of road segments from the following four countries - Croatia, Slovenia, Greece, and the Netherlands. As we show later, these four countries were selected because they represent countries with relatively different OSM feature qualities. By conducting experiments on ground truth datasets with varying feature qualities, we gain a deeper understanding of how well our road safety classifier will generalize to other regions.

To build these four ground truth datasets, we first generated binary road safety labels. To do so, we downloaded the existing Star Ratings data in these four countries, including both the coordinates of the road segments and the Star Rating scores of these road segments, from iRAP website [4]. We then converted these Star Rating scores to binary road safety labels.

For each road segment in our ground truth datasets, we obtained the features data from a variety of sources. To obtain the OSM feature, we queried the OSM-related APIs [6, 58] using the coordinates of the road segments. Concerning traffic statistics features, we obtained them from government websites and iRAP websites (which pre-compiles traffic data from the government.)

**4.2.2 Data Characteristics and Preprocessing.** Our ground truth datasets have certain characteristics that require additional preprocessing. First, the two classes for road safety are imbalanced. Table 2 shows that, consistently across all four datasets, only a small fraction of the road segments (8% - 17%) are in the high-risk class, presenting challenges for some machine learning algorithms [29]. As such, we applied a data re-sampling technique, a common countermeasure for handling class-imbalance problems in applied machine learning research (e.g., [7, 9, 54]). Specifically, we applied the commonly-used Synthetic Minority Over-sampling Technique (SMOTE) [17] on our training set, which over-samples minority class by creating "synthetic" samples that fall within the proximity of real samples.

Second, as shown in Table 3, OSM feature qualities vary among four ground truth datasets. Like many other peer-production projects such as Wikipedia, OSM does not stipulate mandatory attributes. Instead, OSM gives contributors the freedom to add as many, or as few attributes. Our analyses show that, while most of the OSM features we defined have high data completeness, a handful of features, especially features that represent road attributes, suffer from missing value problems. More importantly, analysis presented in Table 3 indicates that, for these OSM features, the specific rates of missing values are different for different countries we studied, a trend that is consistent with prior studies on OSM data quality [10, 26, 34]. For example, while the rates of missing values are lower in the Croatia (avg. 37%) dataset and the Slovenia dataset (avg. 42%), they are considerably higher for Greece dataset (avg. 51%) and the Netherlands dataset (avg. 49%). To partially address the varying degree of missing data, we applied the SOFT-IMPUTE algorithm [48], a data imputation technique that employs a soft-thresholded singular value decomposition to calculate the missing elements in the feature matrix iteratively.

## 4.3 Training and Evaluation

We trained our road safety classifier with a variety of machine learning algorithms, including both ensemble models and conventional models. With respect to ensemble models, we used an XGBoost model [19], a gradient boosting tree framework that has been shown to be effective in a variety of applied machine learning tasks (e.g. [18, 32, 57]). In addition to the ensemble model, we also employed a number of traditional classifiers, including Logistic Regression, Random Forest, and Decision Tree. Our experiments revealed that XGBoost consistently outperformed other machine learning algorithms. As such, we report only the classification accuracies for XGBoost in the Results section.

To build training and testing sets, we can not use standard k-fold cross-validation due to the particular nature of our dataset. Road segments, like many other types of spatial data, are adjacent to each other. As such, training and test sets generated by standard k-fold cross-validation will be correlated, violating the independent

<sup>2</sup>[https://wiki.openstreetmap.org/wiki/Downloading\\_data](https://wiki.openstreetmap.org/wiki/Downloading_data)

**Table 1: Features for our road safety classifier**

| Source                      | Feature Name       | Definition   |
|-----------------------------|--------------------|--|
| OSM Features                | highway            | one-hot encoding indicating the type of the road   |
|                             | oneway             | a binary value of whether this road segment is one way                                       |
|                             | maxspeed           | an integer for the maximum speed of this segment   |
|                             | surface            | one-hot encoding of the surface material   |
|                             | smoothness         | an integer for road surface quality level  |
|                             | lit                | a binary value of whether street lights equipped   |
|                             | bridge             | a binary value of whether there is a bridge above or below this road segment                 |
|                             | lanes              | an integer for the number of lanes   |
|                             | nodes              | an integer for the number of nodes nearby the road segment                                   |
|                             | iso area           | a float value of the area of isochrone   |
|                             | reach factor       | a float value that measures the complexity of the road network around this road segment [24] |
|                             | totpop             | an integer for the population covered by the isochrone                                       |
| Traffic Statistics Features | direction          | the driving direction of this segment  |
|                             | aadt               | an integer for the annual average daily traffic of the road segment                          |
|                             | mean speed of cars | an integer for vehicle mean operating speed  |
|                             | 85% speed of cars  | an integer for the vehicle 85th percentile   |

**Table 2: The summary statistics for road safety class labels in four ground truth datasets. Our statistics show that two road safety classes are highly imbalanced. (HR: Croatia; GR: Greece; SL: Slovenia; NL: The Netherlands)**

| Class     | HR     | GR     | SL     | NL     |
|-----------|--------|--------|--------|--------|
| High-risk | 11.71% | 8.39%  | 16.31% | 8.00%  |
| Safe      | 88.29% | 91.61% | 83.69% | 92.00% |
| Total     | 6320   | 41582  | 22873  | 55985  |

**Table 3: The missing values rates for OSM features. Features that are not shown in the table do not have missing value problems. Our statistics show that the missing values rates differ among different countries. (HR: Croatia; GR: Greece; SL: Slovenia; NL: The Netherlands)**

| Feature    | HR     | GR     | SL     | NL    |
|------------|--------|--------|--------|-------|
| oneway     | 9.2%   | 23.3%  | 12.03% | 19.7% |
| maxspeed   | 33.2%  | 62.7%  | 41.0%  | 51.4% |
| surface    | 35.7%  | 43.1%  | 44.9%  | 50.3% |
| smoothness | 65.3%  | 78.5%  | 68.2%  | 73.5% |
| lit        | 41.3%  | 47.2%  | 45.8%  | 47.6% |
| average    | 36.94% | 50.96% | 42.38% | 48.5% |

identical distribution (i.i.d) assumption fundamental to many statistical machine learning algorithms [60]. As such, we adopted a strategy from Song et al. [62] to maximize the independence between the samples in the training set and the samples in the test set. Specifically, after building the initial training and test sets with the standard 10-fold cross-validation, we swapped a handful of samples between training sets and test sets to ensure that samples in the test set are at least 500 meters away from the samples in the training set.

For evaluation, we report the average precision, recall, and F1 score using 10-fold cross-validation with the above spatially-aware

strategy. We selected precision, recall, and F1 score as performance metrics because these metrics are more informative for evaluating performance on imbalanced data [56]. In addition, we report the performance for high-risk road class and safe road class separately to highlight the different extent of performance improvements for each class. Because this paper is the first to define and to attempt to predict road safety labels in the context of navigation systems, we cannot compare our classifier’s performance with prior work. This is a relatively common scenario when applying machine learning to newly defined problems (e.g., [30, 44, 52]). As such, we followed the best practice to compare the performance of our classifier with that of the straightforward baseline approaches [30, 44, 52]. In this paper, we leveraged the most powerful baseline approach, which randomly generates predictions by respecting the class distributions in ground truth dataset.

## 4.4 Results

Results in Table 4 show our model’s ability to predict safety labels for road segments in our four country-defined ground truth datasets. Our results highlight three high-level trends. First, across the board, our road safety classifier consistently outperformed the baseline approach on all metrics. In particular, the performance increase is most significant for high-risk road class - the minority class. For example, on the Netherlands dataset, our road safety classifier that used all features can increase the F1 score on the high-risk road class from 0.07 to 0.62, a nine-fold increase. Similarly, considerable performance increase (at least five-fold increase) can be found in the results for high-risk road class in other datasets. Even for the safe road class, which already has very high baseline accuracies, our road safety classifier still consistently make smaller improvements. For example, on the Greece dataset and the Netherlands dataset on which our classifier can achieve over 0.9 baseline F1 scores for safe road class, our classifier trained on all features still further increases the F1 scores by about 0.05. Overall, these results demonstrate the clear benefits of applying our road safety classifier - for areas without existing road safety labels, applying our road

safety classifiers offers predictions that are consistently better than baseline approaches. On the Croatia dataset, our classifier even reaches near-perfect predictions.

The second trend in Table 4 is that although our classifier consistently outperformed the baseline, we notice that our classifier’s absolute accuracies, especially for the high-risk road class, vary among different ground truth datasets. For example, our classifier achieves high accuracies with over 0.8 F1 scores for high-risk road class on the Croatia dataset and on the Slovenia dataset. In comparison, F1 scores for high-risk road class on the Greece dataset and the Netherlands dataset are moderate ( $\sim 0.55$ ). One possible explanation is the different OSM feature qualities in these four countries - recall Table 3 shows that the Greece dataset and the Netherlands dataset have high rates of missing values ( $\sim 50\%$ ) for some OSM features, while the Croatia dataset and the Slovenia dataset have lower rates of missing values ( $\sim 40\%$ ). Although our classifier’s performance on the lower-end still outperformed the baseline significantly, we suspected that the performance could further decrease when applied to regions that have even higher missing values rates. We return to this point in the discussion section.

Table 4 shows that the parsimonious models using only OSM features perform almost as well as the models using all features. Our results show only a negligible decrease in the F1 score for most of the experiments. The fact that parsimonious models are able to perform equally well demonstrates the robustness of our approach - even in areas where local government does not provide traffic statistics, the performance of our road safety classifier will not significantly decrease.

**Table 4: Our road safety classifier’s performance on the Croatia dataset. The performance of the baseline approach is in the bracket. (HR: Croatia; GR: Greece; SL: Slovenia; NL: The Netherlands)**

|    | Class     | Features | Precision   | Recall      | F1-Score    |
|----|-----------|----------|-------------|-------------|-------------|
| HR | High-risk | All      | 0.95 (0.13) | 0.94 (0.12) | 0.94 (0.12) |
|    |           | OSM      | 0.93 (0.13) | 0.94 (0.12) | 0.93 (0.12) |
|    | Safe      | All      | 0.99 (0.88) | 0.99 (0.89) | 0.99 (0.88) |
|    |           | OSM      | 0.97 (0.88) | 0.98 (0.89) | 0.98 (0.88) |
| GR | High-risk | All      | 0.58 (0.08) | 0.45 (0.08) | 0.51 (0.08) |
|    |           | OSM      | 0.57 (0.08) | 0.44 (0.08) | 0.50 (0.08) |
|    | Safe      | All      | 0.95 (0.92) | 0.97 (0.92) | 0.97 (0.92) |
|    |           | OSM      | 0.93 (0.92) | 0.96 (0.92) | 0.96 (0.92) |
| SL | High-risk | All      | 0.79 (0.16) | 0.83 (0.16) | 0.81 (0.16) |
|    |           | OSM      | 0.78 (0.16) | 0.81 (0.16) | 0.80 (0.16) |
|    | Safe      | All      | 0.95 (0.83) | 0.95 (0.83) | 0.95 (0.83) |
|    |           | OSM      | 0.94 (0.83) | 0.95 (0.83) | 0.94 (0.83) |
| NL | High-risk | All      | 0.74 (0.07) | 0.53 (0.08) | 0.62 (0.07) |
|    |           | OSM      | 0.71 (0.07) | 0.44 (0.08) | 0.54 (0.07) |
|    | Safe      | All      | 0.96 (0.92) | 0.97 (0.92) | 0.97 (0.92) |
|    |           | OSM      | 0.93 (0.92) | 0.96 (0.92) | 0.94 (0.92) |

**Table 5: The percentage of unsafe routes among randomly sampled routes in Croatia. The criteria for unsafe routes specify the minimum percentage of high-risk road segments for routes to be considered unsafe routes. For example, ">10%" means that routes contain at least 10% of high-risk road segments are unsafe routes.**

| Area  | Option   | Criteria for Unsafe Routes |       |       |       |
|-------|----------|----------------------------|-------|-------|-------|
|       |          | >10%                       | >30%  | >50%  | >70%  |
| Urban | Shortest | 90.4%                      | 67.4% | 30.0% | 7.6%  |
| Urban | Fastest  | 90.6%                      | 66.5% | 32.3% | 8.7%  |
| Rura  | Shortest | 98.8%                      | 85.6% | 52.9% | 25.2% |
| Rural | Fastest  | 99.1%                      | 85.4% | 52.3% | 26.1% |

#### 4.5 Estimating Users’ Exposures to Unsafe Routes

While prior studies (e.g., [42]) have discovered the risks associated with unsafe routes given by navigation systems, the prevalence of the unsafe routes remains unknown. With our road safety classifier, we could, for the first time, quantify how often users of navigation systems will be exposed to unsafe routing results.

To do so, we conduct a small case study in Croatia, where our road safety classifier has the best performance. To account for the potential differences between urban and rural areas, we selected two regions - Croatia’s capital Zagreb (roughly 900 km<sup>2</sup>) and a equally sized rural region in Croatia. We then generated large number of routes for these two areas. Specifically, we followed the best practices in routes evaluation studies (e.g., [28, 36, 67]) which first randomly sample origin-destination pairs and then generate routes between these origin-destination pairs. In our case, we sampled 10,000 origin-destination pairs for each region and used Mapzen Valhalla [1], a popular open-sourced routing service, to generate the shortest routes and the fastest routes between these origin-destination pairs. Finally, we applied our road safety prediction model on the road segments included these 20,000 routes to quantify the unsafe routes.

Table 5 shows the estimated percentages of unsafe routes by type of regions and type of route options. In addition, we tested different criteria for defining a unsafe route by varying the minimum percentage of high-risk road segments the route contains. The results in Table 5 suggest that users of navigation systems have high exposure to potentially unsafe routing results. Specifically, if routes with at least 10% of its road segments are high-risk road segment are considered unsafe routes, over 90% of routing results given by navigation systems are potentially dangerous in Croatia, regardless of the type of regions and the type of route options. This number might look startling but is actually reasonable given the fact that 11.71% of road segments in Croatia are high-risk (see Table 2.) Even if we use a lenient criteria for unsafe routes - routes with at least 70% of high-risk road segments are unsafe, experiment results still show that as much as a quarter routing results are potentially dangerous in rural area of Croatia. While driving on these potentially dangerous route will not necessarily result in accidents, users’ high exposure to potentially unsafe routes still highlight the prevalence of the risks.

A second trend in Table 5 is that while route options have little influence over the prevalence of unsafe routes, urbanness/ruralness do affect the percentage of unsafe routes. Specifically, using navigation systems in rural areas of Croatia will be exposed to 8% to 20% more unsafe routes than doing so in urban areas. While the higher percentage in rural areas can be partially attributed to the fact that randomly generated routes might be biased towards more local roads over highways (which draws most of the traffic in the rural area), our results still highlight the higher risks of driving on rural roads which is consistent with crash statistics from National Highway Traffic Safety Administration [65].

## 5 DISCUSSION

Our results demonstrate that using straightforward machine learning algorithms and publicly available geographic data as features, we can successfully predict road safety labels with accuracies that are significantly higher than the baseline approach. In the following, we first discuss the factors to consider when extending our road safety classifier to regions beyond the areas studied in this paper. With an eye towards future work, we then discuss various technology designs through which our road safety predictions could help personal navigation technologies reduce the risks of dangerous routes.

### 5.1 Extending Our Approach to Other Regions

Our results show that, while the performance of our classifier consistently beat the baseline, its performance in a region might be affected by the OSM feature quality in that region.

As such, when applying our road classifier to other regions, one important factor to consider is the quality of OSM features in this region. Researchers have found that OSM data quality is usually better in urban/developed areas than in rural/undeveloped areas [34, 47, 68]. As such, our road safety classifier might have higher performances for more developed regions such as Europe and have lower performances for less developed regions, such as South Asia and Africa. Similarly, urban areas, which have higher OSM qualities, might observe higher performance than the rural areas. As such, future work that extends our classifier to other regions should examine the specific OSM data quality, especially the rates of the missing value, before applying our approach. For areas where OSM data quality and coverage are low, one possible mitigation is to combine the image-based approaches for road safety (e.g., [55, 62] with our approach to compensate for data quality issue.

### 5.2 Integrating Road Safety Predictions into Navigation Systems

Our work can produce large-scale reliable road safety predictions which navigation systems can directly leverage. These road safety predictions present many opportunities for designing new algorithms and user interactions for protecting drivers' safety. Below, we briefly mention the opportunities for new algorithm designs and then focus our attention on new user interactions that ensure driver safety.

For algorithm innovation, one clear opportunity is building new routing algorithms that prioritize road safety. As a starting point,

one may simply adopt the existing algorithms designed for generating safe routes that pass through low-crime areas (e.g., [22, 38, 59]). For example, the crime-index score in the cost functions of these safe routing algorithms can be replaced by an aggregated road risk score of the candidate route. Alternatively, instead of adding waypoints in low-crime areas, these safe routing algorithms can add waypoints that detour through a safe road.

However, even with new routing algorithms that prioritize road safety, drivers might still unavoidably travel on high-risk roads given their time constraints or their choices of destinations. In these scenarios, road safety awareness-enabled user interactions serve as the critical last layer of protection. These essential user interactions can happen in two phases - during route recommendation and while driving. During route recommendations, navigation systems can put users in control of the trade-off between road safety and other parameters. To begin with, as a preference setting, navigation systems can explicitly ask for the maximum distance or portion of high-risk roads that users are willing to tolerate within a route and incorporate this constraint when running the routing algorithms. In addition, navigation systems can explicitly communicate the trade-off between time and route safety among alternative routes through routes overview. For example, in addition to informing users of the distance and time of the candidate routes, navigation systems can highlight the distance or the portion of high-risk road segments through textual or visual communications.

During the trip, when passing through high-risk roads, drivers must focus their entire attention on the road [33]. As such, researchers have called for navigation systems that can timely alter drivers for abnormal road environments (e.g., [16]). With our road safety predictions, navigation systems will know precisely where the high-risk road segments are. Given this knowledge, navigation systems can generate audio and visual alerts when drivers are approaching the dangerous road segments to call drivers' attention to the road. With cars becoming increasingly automated and drivers delegating more responsibility to the technology, navigation systems with road safety awareness will be even more vital to ensure user safety.

## 6 CONCLUSION

A dangerous assumption adopted by existing navigation systems is that all roads are equally safe. Navigation systems with this wrong assumption have routed drivers to dangerous routes, threatening driver safety. In this paper, we sought to address this risk by adding road safety awareness to the navigation systems. Specifically, we defined road safety in the context of the navigation systems and proposed a road safety classifier to automatically predict the road safety for many roads using out-of-the-box machine learning algorithms and diverse features from public geographic data. We demonstrate we can predict the road safety labels with accuracies that are significantly higher than baseline approaches. Finally, we discussed the factors to consider when extending our road safety classifier to other regions and the novel safety designs on navigation systems enabled by our road safety predictions.

## REFERENCES

- [1] [n.d.]. Valhalla. <https://github.com/valhalla/valhalla>.



- [2] 2012. OpenStreetMap: 'It's the Wikipedia of maps'. <https://www.theguardian.com/theobserver/2012/feb/18/openstreetmap-world-map-radicals>
- [3] 2017. Here's how iRAP can help you save lives. [https://www.irap.org/how-we-can-help/?et\\_open\\_tab=et\\_pb\\_tab\\_1#mytabs1](https://www.irap.org/how-we-can-help/?et_open_tab=et_pb_tab_1#mytabs1)
- [4] 2017. International Road Assessment Programme. <https://www.irap.org/>
- [5] 2019. Major OpenStreetMap Consumers. [https://wiki.openstreetmap.org/wiki/Major\\_OpenStreetMap\\_Consumers](https://wiki.openstreetmap.org/wiki/Major_OpenStreetMap_Consumers)
- [6] 2020. Overpass Turbo. <https://overpass-turbo.eu>
- [7] N. Japkowicz A. Nickerson and E. Milios. 2001. Using unsupervised learning to guide re-sampling in imbalanced data sets. In *In Proceedings of the Eighth International Workshop on AI and Statistics*. 261–265.
- [8] Syeed Adnan Raheel Shah; Tom Brijs; Naveed Ahmad. 2017. Road Safety Risk Evaluation Using GIS-Based Data Envelopment Analysis—Artificial Neural Networks Approach. In *Appl. Sci.*
- [9] Nathalie Japkowicz Andrew Estabrooks, Taeho Jo. 2004. A Multiple Resampling Method for Learning from Imbalanced Data Sets. In *Computational Intelligence* 20. 18–36.
- [10] V Antoniou and A Skopeliti. 2015. Measures and indicators of VGI quality: An overview. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences* 2 (2015), 345.
- [11] Antonio Antunes, Alvaro Seco, and Nuno Pinto. 2003. An accessibility–maximization approach to road network planning. *Computer-Aided Civil and Infrastructure Engineering* 18, 3 (2003), 224–240.
- [12] Marie C. Baca. 2019. Google Maps now shows speed traps, potentially raising the ire of law enforcement. <https://www.washingtonpost.com/technology/2019/10/22/google-maps-now-helps-iphone-users-avoid-speed-traps/>
- [13] Kenneth Majlund Bach, Mads Gregers Jäger, Mikael B Skov, and Nils Gram Thomassen. 2009. Interacting with in-vehicle systems: understanding, measuring, and evaluating attention. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*. British Computer Society, 453–462.
- [14] Karlin Bark, Cuong Tran, Kikuo Fujimura, and Victor Ng-Thow-Hing. 2014. Personal Navi: Benefits of an augmented reality navigational aid using a see-thru 3D volumetric HUD. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 1–8.
- [15] Michael Beenstock and Dalit Gafni. 2000. Globalization in road safety: explaining the downward trend in road accident rates in a single country (Israel). *Accident Analysis & Prevention* 32, 1 (2000), 71–84.
- [16] Stephen M Casner, Edwin L Hutchins, and Don Norman. 2016. The challenges of partially automated driving. *Commun. ACM* 59, 5 (2016), 70–77.
- [17] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Int. Res.* 16, 1 (June 2002), 321–357. <http://dl.acm.org/citation.cfm?id=1622407.1622416>
- [18] Hongming Chen, Ola Engkvist, Yinhai Wang, Marcus Olivecrona, and Thomas Blaschke. 2018. The rise of deep learning in drug discovery. *Drug discovery today* 23, 6 (2018), 1241–1250.
- [19] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 785–794.
- [20] OpenStreetMap contributors. [n.d.]. Map Features - OpenStreetMap Wiki. [https://wiki.openstreetmap.org/wiki/Map\\_Features](https://wiki.openstreetmap.org/wiki/Map_Features). (Accessed on 01/18/2019).
- [21] Michael A. Fields Eric R. Green Douglas W. Harwood, Reginald R. Souleyrette. 2014. *Comparison of Countermeasure Selection Methods for Use in Road Safety Management*. Technical Report. Transportation Research Board, Washington, DC.
- [22] Kaiqin Fu, Yen-Cheng Lu, and Chang-Tien Lu. 2014. Treads: A safe route recommender using social media mining and text summarization. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 557–560.
- [23] Thomas F Golob and Wilfred W Recker. 2003. Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. *Journal of transportation engineering* 129, 4 (2003), 342–353.
- [24] GIScience Research Group. 2020. Openrouteservice. <https://github.com/GIScience/openrouteservice>.
- [25] Ghazwan Al Haji. 2005. Towards a Road Safety Development Index (RSDI): Development of an International Index to Measure Road Safety Performance. *Linköping University Electronic Press: Linköping, Sweden* (2005).
- [26] Mordechai Haklay, Sofia Basiouka, Vyron Antoniou, and Aamer Ather. 2010. How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The cartographic journal* 47, 4 (2010), 315–322.
- [27] Susan L Handy and Kelly J Clifton. 2001. Evaluating neighborhood accessibility: Possibilities and practicalities. *Journal of transportation and statistics* 4, 2/3 (2001), 67–78.
- [28] Shazia Haque, Lars Kulik, and Alexander Klippel. 2006. Algorithms for reliable navigation and wayfinding. In *International Conference on Spatial Cognition*. Springer, 308–326.
- [29] Haibo He and Edwardo A Garcia. 2009. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering* 21, 9 (2009), 1263–1284.
- [30] Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 237–246.
- [31] Brent Hecht, Johannes Schöning, Thomas Erickson, and Reid Priedhorsky. 2011. Geographic human-computer interaction. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. 447–450.
- [32] Tomislav Hengl, Jorge Mendes de Jesus, Gerard BM Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangguan, Marvin N Wright, Xiaoyuan Geng, Bernhard Bauer-Marschallinger, et al. 2017. Soil-Grids250m: Global gridded soil information based on machine learning. *PLoS one* 12, 2 (2017).
- [33] Tim Horberry, Janet Anderson, Michael A Regan, Thomas J Triggs, and John Brown. 2006. Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention* 38, 1 (2006), 185–191.
- [34] Toby Jia-Jun Li Andrew Hall Aaron Halfaker Johannes Schöning Brent Hecht Isaac Johnson, Allen Yilun Lin. 2016. Not at Home on the Range: Peer Production and the Urban/Rural Divide. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI, 13–25.
- [35] Brit Susan Jensen, Mikael B Skov, and Nissanthan Thiruvachandran. 2010. Studying driver attention and behaviour for the three configurations of GPS navigation in real traffic driving. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1271–1280.
- [36] Isaac Johnson, Jessica Henderson, Caitlin Perry, Johannes Schöning, and Brent Hecht. 2017. Beautiful... but at What Cost?: An Examination of Externalities in Geographic Vehicle Routing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 15.
- [37] Vanya Jones, Juhée Cho, Jacqueline Milani, and Andrea Gielen. 2011. Driving Habits and Risk Exposure in Older Drivers: Lessons Learned From the Implementation of a Self-Regulation Curriculum. *Traffic injury prevention* 12 (2011), 468–74.
- [38] Jaewoo Kim, Meeyoung Cha, and Thomas Sandholm. 2014. Socroutes: safe routes based on tweet sentiments. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 179–182.
- [39] Ch Koch and I Brilakis. 2011. Improving pothole recognition through vision tracking for automated pavement assessment. In *Proc. of EGICE Workshop*. 1–8.
- [40] Andrew L. Kun, Tim Paek, Željko Medenica, Nemanja Memarović, and Oskar Palinko. 2009. Glancing at personal navigation devices can affect driving: experimental results and design implications. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 129–136.
- [41] H. V Ribeiro L. G. A. Alves and F. A. Rodrigues. 2018. Crime prediction through urban metrics and statistical learning. In *Physica A*, Vol. 505. 435–443.
- [42] Allen Yilun Lin, Kate Kuehl, Johannes Schöning, and Brent Hecht. 2017. Understanding Death by GPS: A Systematic Study of Catastrophic Incidents Associated with Personal Navigation Technologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1154–1166.
- [43] Jin Lin and Yayu Liu. 2010. Potholes detection based on SVM in the pavement distress image. In *2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science*. IEEE, 544–547.
- [44] Yilun Lin, Bowen Yu, Andrew Hall, and Brent Hecht. 2017. Problematising and addressing the article-as-concept assumption in wikipedia. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2052–2067.
- [45] Dandan Lyu Xiaozhe Liu Hanyu Long Ling Chen, Yifang Ding. 2017. Deep Multi-Task Learning Based Urban Air Quality Index Modelling. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*.
- [46] Todd Litman. 2003. Measuring transportation: traffic, mobility and accessibility. *Institute of Transportation Engineers. ITE Journal* 73, 10 (2003), 28.
- [47] Giovanni Quattrone Mashhadi, Afra and Licia Capra. 2013. Putting ubiquitous crowd-sourcing into context. In *Proceedings of the 2013 conference on Computer supported cooperative work*.
- [48] Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. 2010. Spectral Regularization Algorithms for Learning Large Incomplete Matrices. 11 (03 2010), 2287–2322.
- [49] Congress of the United States. 2015. Letters to World Bank. (2015).
- [50] Jerry Olsson. 2009. Improved road accessibility and indirect development effects: evidence from rural Philippines. *Journal of Transport Geography* 17, 6 (2009), 476–483.
- [51] David O'Sullivan, Alastair Morrison, and John Shearer. 2000. Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach. *International Journal of Geographical Information Science* 14, 1 (2000), 85–104.
- [52] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume*



10. Association for Computational Linguistics, 79–86.
- [53] Vanessa Doebler Peissner, Matthias and Florian Metze. 2011. Can voice interaction help reducing the level of distraction and prevent accidents. (2011).
- [54] Vasile Palade Rukshan Batuwita. 2010. Efficient resampling methods for training support vector machines with imbalanced datasets. In *The 2010 International Joint Conference on Neural Networks*.
- [55] Arpan Sainju and Zhe Jiang. 2019. Mapping road safety features from streetview imagery: A deep learning approach. *arXiv preprint arXiv:1907.12647* (2019).
- [56] Takaya Saito and Marc Rehmsmeier. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one* 10, 3 (2015).
- [57] Marwin HS Segler, Thierry Kogej, Christian Tyrchan, and Mark P Waller. 2018. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS central science* 4, 1 (2018), 120–131.
- [58] Openroute Service. 2018. . <https://openrouteservice.org/>
- [59] Sumit Shah, Fenye Bao, Chang-Tien Lu, and Ing-Ray Chen. 2011. Crowdsafe: crowd sourcing of crime incidents and safe routing on mobile devices. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 521–524.
- [60] Steven Feiner Shekhar, Shashi and Walid G. Aref. 2015. From GPS and virtual globes to spatial computing-2020.. In *GeoInformatica*, Vol. 19. ACM, 799–832.
- [61] E.; Brijis T.; Wets G.; Vanhoof K. Shen, Y.; Hermans. 2012. Road safety risk evaluation and target setting using data envelopment analysis and its extensions. In *Accid. Anal. Prev.*
- [62] Weilian Song, Scott Workman, Armin Hadzic, Xu Zhang, Eric Green, Mei Chen, Reginald Souleyrette, and Nathan Jacobs. 2018. FARSA: Fully Automated Roadway Safety Assessment. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 521–529.
- [63] Jessica A Pater Munmun De Choudhury Stevie Chancellor, Yannis Kalantidis and David A Shamma. 2017. Multimodal Classification of Moderated Online Pro-Eating Disorder Content. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3213–3226.
- [64] Gioconda Tarqui, Luis A Castro, and Jesus Favela. 2013. Reducing drivers' distractions in phone-based navigation assistants using landmarks. In *Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction*. Springer, 342–349.
- [65] Caitlin N Webb. 2020. *Geospatial Summary of Crash Fatalities*. Technical Report.
- [66] Lei Zhang, Fan Yang, Yimin Daniel Zhang, and Ying Julie Zhu. 2016. Road crack detection using deep convolutional neural network. In *2016 IEEE international conference on image processing (ICIP)*. IEEE, 3708–3712.
- [67] Dennis Zielstra and Hartwig H Hochmair. 2012. Using free and proprietary data to compare shortest-path lengths for effective pedestrian routing in street networks. *Transportation Research Record* 2299, 1 (2012), 41–47.
- [68] Dennis Zielstra and Alexander Zipf. 2010. A comparative study of proprietary geodata and volunteered geographic information for Germany. In *13th AGILE International Conference on Geographic Information Science*.