RiskSens: A Multi-view Learning Approach to Identifying Risky Traffic Locations in Intelligent Transportation Systems Using Social and Remote Sensing

Yang Zhang, Yiwen Lu, Daniel Zhang, Lanyu Shang, Dong Wang
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
yzhang42@nd.edu, ylu9@nd.edu, yzhang40@nd.edu, lshang@nd.edu, dwang5@nd.edu

***Abstract*—With the ever-increasing number of road traffic accidents worldwide, the road traffic safety has become a critical problem in intelligent transportation systems. A key step towards improving the road traffic safety is to identify the locations where severe traffic accidents happen with a high probability so the precautions can be applied effectively. We refer to this problem as *risky traffic location identification*. While previous efforts have been made to address similar problems, two important limitations exist: i) *data availability*: many cities (especially in developing countries) do not maintain a *publicly accessible* database for the traffic accident records in a city, which makes it difficult to accurately estimate the accidents in the city; ii) *location accuracy*: many self-reported traffic accidents (e.g., social media posts from common citizens) are not associated with the exact GPS locations due to the privacy concerns. To address these limitations, this paper develops the RiskSens, a multi-view learning approach to identifying the risky traffic locations in a city by jointly exploring the social and remote sensing data. We evaluate RiskSens using a real world dataset from New York. The evaluation results show that RiskSens significantly outperforms the state-of-the-art baselines in identifying risky traffic locations in a city.**

***Keywords*-Risky Traffic Location Identification, Social Sensing, Remote Sensing, Multi-view Learning, Intelligent Transportation Systems.**

I. Introduction

Social sensing has emerged as a new sensing paradigm where human sensors collectively report measurements about the physical world [1]. Examples include monitoring the traffic conditions using mobile crowdsensing, reporting free parking lots using geotagging, and obtaining real-time situation awareness in disaster response using online social media. Remote sensing, on the other hand, is a conventional sensing paradigm that captures abundant visual features of objects on earth surface by means of aircraft or satellite-based sensors [2]. While social sensing provides data collection opportunities at an unprecedented scale, the collected data are often sparse and noisy due to the complexities and limitations of human sensors [1]. In contrast, remote sensing data (e.g., images) are often comprehensive and fine-grained but require an intensive manual labeling process to be useful [3]. In this paper, we develop a multi-view learning framework to address the risky traffic location identification problem in intelligent transportation systems by exploring the collective power of social and remote sensing.

Recent progress has been made to address the traffic related problems in intelligent transportation systems from data mining, machine learning, and geographical information system communities [4]–[7]. Examples of those solutions include road safety mapping using deep learning [4], hazardous road inference using a kernel density based estimation [5], traffic flow prediction using linear regression [6], and metro passenger flow analysis using visual fusion [7]. The primary data sources of these solutions fall into two categories: i) open data published by the authorities (e.g., traffic accident reports from the police departments, CCTV camera and smart card data from the transportation department), and ii) self-reported data from common citizens (e.g., posts related to traffic accidents from online social media, reports from mobile crowdsensing apps like Waze). However, two important limitations exist in current solutions, which are elaborated below.

Data Availability. In Intelligent Transportation Systems, the availability of traffic accident data is essential to obtain the accurate risky location identification results [5], [8]. However, such traffic accident data is not always available. For example, many cities (especially in underdeveloped areas) do not maintain a *publicly accessible* database for the traffic accident records due to the lack of necessary resource, infrastructure, and transparency [4], or the privacy and legal concerns [9]. In fact, fewer than 1% municipals in US have open data portals for accessing the traffic accident records [1]. In addition, less than 3% US cities install road traffic cameras and traffic monitoring devices are prohibited by 10 states in US [2]. We also note that the large-scale mobile crowdsensing apps (e.g., Waze) collect a rich set of traffic information (e.g., traffic congestion and accidents) from their participants. However, such data is often owned by private companies and not accessible to the public due to the legal and privacy restrictions [10]. Therefore, a general and robust risky traffic location identification scheme that only uses the

[1]https://www.forbes.com/
[2]http://www.iihs.org/iihs

publicly available information is highly desirable.

Location Accuracy. The accurate locations of traffic accidents are also critical to solve the risky traffic location identification problem [11]. However, unlike the open data published by the authorities, the locations associated with self-reported data from common individuals (e.g., posts on social media) are often coarse-grained and inaccurate due to the privacy concerns [12]. For instance, Figure 1(a) shows a tweet reporting a traffic accident. We observe that the tweet only describes the accident happened in "Queens" rather than provides the precise GPS location of the accident. Therefore, it remains a critical challenge to obtain the accurate risky traffic locations by leveraging the self-reported data from common citizens.

(a) Social Sensing (b) Remote Sensing

Figure 1. A Risky Traffic Location Identified by both Social and Remote Sensing Paradigms

To address the above limitations, this paper develops the RiskSens, a multi-view learning approach to identifying the risky traffic locations by jointly exploring both social and remote sensing data. To address the *data availability* limitation, RiskSens leverages social media data (Figure 1(a)) and satellite imagery data (Figure 1(b)) that are publicly available (e.g., tweets collected from Twitter API [3] and satellite images collected from Google Map API [4]). We further develop two customized feature extraction components in RiskSens to effectively extract semantic and visual features that are related with traffic accidents from the heterogeneous data collected from the two disparate sensing paradigms. To address the *location accuracy* limitation, we develop a novel multi-view learning scheme to accurately identify the risky traffic locations by fusing the extracted features (i.e., semantic and visual features) under a principled framework.

To the best of our knowledge, the RiskSens is the first solution that jointly explores the social and remote sensing data to address the risky traffic location identification problem in intelligent transportation systems. The RiskSens is also a general framework that can be applied to other smart city applications (e.g., disaster response, crime alert, environment monitoring) that rely on the publicly available data from social and remote sensing paradigms. We evaluate the RiskSens on a real-world traffic dataset from New York City. The results show that our scheme significantly outperforms the state-of-the-art baselines in various application scenarios.

[3]https://developer.twitter.com/

[4]https://cloud.google.com/maps-platform/

II. Related Work

A. Social Sensing and Remote Sensing

Social sensing has emerged as a new sensing paradigm for big data applications where human sensors collectively report measurements about the physical world [12]. Social sensing has been widely used in intelligent transportation systems [13], edge computing systems [**?**], [14], urban infrastructure and environment monitoring [15], and disaster response and situation awareness [16]. Meanwhile, recent advances in remote sensing have made it a powerful tool to acquire high resolution satellite imagery that includes abundant visual features about the earth surface with affordable costs [2]. Examples of remote sensing applications include precision agriculture [17], traffic monitoring [18], and urban planning [19]. This paper focuses on a challenging risky traffic location identification problem in intelligent transportation systems by exploring the data from both social and remote sensing applications.

B. Risky Traffic Location Identification

Previous efforts have made good progress to address the traffic related problems in data mining, machine learning, and geographical information system communities [4]–[7]. For example, Najjar *et al.* developed a deep learning based approach to create road safety map using the traffic-accident reports from the police departments [4]. Bil *et al.* proposed a kernel density estimation based procedure to identify hazardous road locations using road network data and accidents data collected by government authorities [5]. He *et al.* designed a linear regression based approach to predicting the traffic flow using semantic features extracted from Twitter data [6]. Itoh *et al.* developed a data visualization based framework to analyze the metro passenger flow using both social media and smart card data collected in Tokyo [7]. Those approaches either i) require the publicly accessible database of the traffic accident records or ii) do not solve the risky traffic location identification problem that we focus on. In contrast, we develop a novel multi-view learning framework to explicitly address the *data availability* and *location accuracy* challenges in accurately identifying the risky traffic locations in a city.

C. Multi-View Learning

Our work is also related to multi-view learning techniques in machine learning which has been applied in areas like natural language processing, computer vision, robotics, computational medicine and sustainable computing [20]–[23]. For example, Zhou *et al.* developed a human-computer conversation system based on word sequence view and utterance sequence view that jointly model the natural language [20]. Su *et al.* proposed an approach to 3D shape recognition problem using multiple 2D views of an object [21]. Zeng *et al.* leveraged multiple views of a robot to estimate its 6D pose [22]. Valmarska *et al.* developed a

method for aggregating multiple views of symptoms of Parkinson's disease patients to support the decisions about patients' therapies [23]. To the best of our knowledge, the RiskSens is the first multi-view learning based approach to solving the risky traffic location identification problem in intelligent transportation systems by exploring both the social and remote sensing data.

III. Problem Definition

In this section, we formulate the risky traffic location identification problem in intelligent transportation systems. We first define the terms that will be used in the problem statement.

Definition 1: **Sensing Area**: We define a sensing area to be a target area (e.g., New York City) where the social and remote sensing data are collected to identify risky traffic locations.

Definition 2: **Sensing Cell** (SC): We divide the sensing area into disjoint sensing cells (e.g., $60\text{m} \times 60\text{m}$ squares as shown in Figure 2) where each cell represents a subarea of interest. In particular, we define X to be the number of cells in the sensing area and SC_x to be the x^{th} sensing cell in the sensing area $(x = 1, 2, \cdots, X)$.

Definition 3: **Social Sensing Data** (SD): We define the Social Sensing Data (SD) to be the self-reports about traffic accidents from common citizens on social media (e.g., tweets shown in Figure 1). In particular, we define $SD = \{SD^1, SD^2, ..., SD^X\}$ where SD^x represents the subset of SD that are associated with locations in sensing cell SC_x where $x = 1, 2, \cdots, X$.

Definition 4: **Remote Sensing Data** (RD): We define the Remote Sensing Data (RD) to be the satellite imagery data collected from the online map service (e.g., Google Map [5]). In particular, we define $RD = \{RD_1, RD_2, ..., RD_X\}$ where RD_x represents the satellite imagery of sensing cell SC_x where $x = 1, 2, \cdots, X$.

Figure 2 shows a group of satellite images collected from four different locations in New York City through Google Map service. The locations in Figure 2(a) have less than 2 accidents/year while the locations in Figure 2(b) have more than 30 accidents/year. We can clearly observe the visual similarity on both low-level (i.e., colors, textures) and high-level (i.e., objects, landscapes) features between the locations with similar accidents rates. Such visual similarity provides important clues for RiskSens to explore the remote sensing data in addressing the risky traffic location identification problem.

Definition 5: **Traffic Risk Level** ($\mathcal{R}$): We define the traffic risk level ($\mathcal{R}$) to indicate the traffic safety condition of a given location. For example, the locations in Figure 2(a) have *lower* traffic risk level than the locations in Figure 2(b). In particular, we define $\mathcal{R}_x$ to be the traffic risk level of cell SC_x $(i = 1, 2, \cdots, X)$. We further define $\overline{\mathcal{R}_x}$ and $\widehat{\mathcal{R}_x}$ as the *real* and *estimated* traffic risk level for cell SC_x respectively. Following a similar procedure in [4], we categorize the traffic risk level ($\mathcal{R}$) into k different categories (i.e., $\mathcal{R}_x \in \{1, 2, \cdots, k\}$ where k represents a severity level of traffic risk (e.g., high, medium, low)).

[5]https://cloud.google.com/maps-platform/

(a) Less than 2 Accidents/Year (b) More than 30 Accidents/Year

Figure 2. Google Map Satellite Imagery Examples in New York City.

The goal of the risky traffic location identification problem is to correctly estimate the real values of the traffic risk level of the sensing cells based on the collected social sensing data SD and remote sensing data RD. In particular, our goal is to derive the estimated $\widehat{\mathcal{R}_x}$ to be as close to the real value $\overline{\mathcal{R}_x}$ as possible in each sensing cell. Our problem is formally defined as:

$$\arg\max_{\widehat{\mathcal{R}_x}} \Pr(\widehat{\mathcal{R}_x} = \overline{\mathcal{R}_x} \mid SC, SD, RD) \quad \forall 1 \leq x \leq X \tag{1}$$

This above problem is challenging due to the unstructured and heterogeneous nature of the social and remote sensing data. In this paper, we develop the RiskSens scheme to jointly fuse the semantic and visual features extracted from social and remote sensing data using a multi-view learning based framework. The details of the RiskSens scheme are discussed in the next section.

IV. Solution

In this section, we present the RiskSens scheme to address the risky traffic location identification problem formulated in the previous section. We first present an overview of the RiskSens scheme and then discuss its components in detail.

A. Overview of the RiskSense scheme

The RiskSens scheme consists three components: i) Social Sensing Feature Extraction (SSFE), ii) Remote Sensing Feature Extraction (RSFE), and iii) Multi-View Information Fusion (MVIF). First, the SSFE component extracts semantic features (i.e., location descriptions, severity level of accidents) from online social media data. Second, the RSFE component extracts various visual features from satellite imagery that are related to the road traffic accidents. Finally, the MVIF component fuses the heterogeneous features extracted from both SSFE and RSFE components to estimate the traffic risk level of each sensing cell. The overall architecture of the RiskSens scheme is shown in Figure 3.

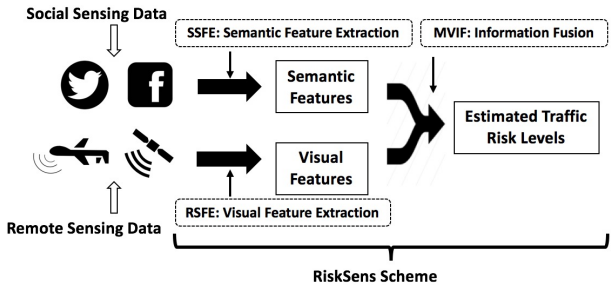


Figure 3. Overview of RiskSens Scheme

B. Social Sensing Feature Extraction (SSFE)

In this subsection, we describe the SSFE component that extracts the semantic features (e.g., locations descriptions, severity level of accidents) that are tightly related to the traffic risk level of a cell from noisy and unstructured online social media data. In particular, we define the Social Sensing Feature (SSF) as follows.

Definition 6: **Social Sensing Feature (SSF)**: We define the Social Sensing Feature (SSF) to include: i) *location description* and ii) *severity level* of an accident reported in the social media data. In particular, for each social media post s from the social sensing data SD, we extract the following semantic features:

$$\{(\boldsymbol{\mu}_s, \sigma_s), c_s\} = \mathcal{P}(s) \quad \forall s \in SD \tag{2}$$

where $(\boldsymbol{\mu}_s, \sigma_s)$ represents the extracted location description of an accident. In particular, $\boldsymbol{\mu}_s \in \mathbb{R}^2$ is the expectation of accident location given the social media post s and $\sigma_s \in (0, +\infty)$ is uncertainty of accident location given s. $c_s \in [0, 1]$ measures the severity of a reported accident given s. $\mathcal{P}$ is the function that maps raw social media data (e.g., tweets) to $\{(\boldsymbol{\mu}_s, \sigma_s), c_s\}$ using a set of application specific regular expressions [24]. For example, the tweet "minor accident on I-278 near Bronx River Parkway" will map to a relatively small c value (because "minor" indicates the accident is not severe), a $\boldsymbol{\mu}$ value representing the coordinate of the intersection between I-278 and Bronx River Parkway, and a relatively large σ value (because "near" is a vague and uncertain description of location).

We observe that the traffic risk level of a sensing cell depends on both the number of accidents in the cell and the severity level of the accidents. Therefore, we aggregate the extracted semantic features into a unified index to estimate the traffic risk level using the social sensing data. In particular, we define the Aggregated Semantic Feature (ASF) of the sensing cell SC_x as follows.

Definition 7: **Aggregated Semantic Feature (ASF)**: the expectation of the traffic risk within a cell using the social sensing data. It is calculated as follows:

$$ASF_x = \sum_{s \in SD^x} c_s \iint_{SC_x} p_s(q_\lambda, q_\phi) \mathrm{d}q_\lambda \mathrm{d}q_\phi \tag{3}$$

where SD^x is the set of social sensing reports associated with the locations in cell SC_x. s is a social media post from SD^x. q_λ, q_ϕ are coordinates in a 2D plane mapped from longitude and latitude, and $p_s(q_\lambda, q_\phi)$ is the distribution of the location of the accident reported by s (i.e., $\mathcal{N}(\boldsymbol{\mu}_s, \sigma_s^2 I)$). Intuitively, this model reflects two key observations: i) different social media posts indicate traffic accidents with different levels of confidence (e.g., deadly vehicle crashes v.s. minor collisions) and should not be treated equally; ii) social media users are usually not accurate in terms of reporting the accident locations (e.g., they often report landmark buildings or road names near an accident instead of its actual GPS location).

C. Remote Sensing Feature Extraction (RSFE)

In this subsection, we describe the Remote Sensing Feature Extraction (RSFE) component that extracts various visual features from satellite imagery that are related to the traffic risk level of a cell. In particular, we define Remote Sensing Feature (RSF) as follows.

Definition 8: **Remote Sensing Feature (RSF)**: a set of visual feature vectors of sensing cell SC_x:

$$RSF_x = \{\gamma(RD_x) \mid \gamma \in \Gamma\}, \quad \forall 1 \leq x \leq X \tag{4}$$

where Γ is a set of visual feature extraction functions, each of which maps a raw satellite image to a feature vector. The input to RSFE is the raw satellite imagery RD_x for cell SC_x. The output of RSFE is RSF_x. Intuitively, RSF_x is generated by both low-level (e.g., color, texture) and high-level (e.g., object, scene) feature extraction functions in Γ. For example, we can use low-level feature extractors like color histogram [25] to capture the dominant color of a satellite image (e.g., cells dominated by green tend to have low traffic risk as shown in Figure 2(a)) and local binary pattern (LBP) [26] to capture texture patterns in a satellite image (e.g., dense zebra crossings is an indicator of high traffic risk as shown in Figure 2(b)). In addition, we can use ImageNet [3] and deep convolutional neural network tools that extract higher-level features (e.g., scenes like traffic congestion). The visual features extract by the RSFE component are summarized in Table I.

D. Multi-view Information Fusion (MVIF)

In this subsection, we describe the Multi-View Information Fusion (MVIF) component that estimates the traffic risk level of a cell by fusing the semantic and visual features extracted from both SSFE and RSFE. In MVIF, we first design a set of *views* where each view makes its own estimation of the traffic risk level of a cell by leveraging a subset of features extracted by the SSFE and RSFE. We then develop an enhanced Maximum Likelihood Estimation (MLE) framework to judiciously fuse the estimation made by different views into a unified estimation of the traffic risk level of a cell. Finally, we propose a Recursive View

Table I
FEATURE DESCRIPTORS OF SENSING CELLS

Class	Name		Dimensions
Low-level Visual Features	Color Histogram [25]		768
	GIST [27]		960
	LBP [26]		126
	GLCM [28]		7
High-Level Visual Features	ImageNet-Pretrained CNN	AlexNet [3]	4096
		VGG19 [29]	4096
		ResNet152 [30]	2048
		Inception-v3 [31]	2048
		DenseNet201 [32]	1920

Elimination (RVE) algorithm to improve the estimation accuracy and reduce the complexity of our model.

1) Multiple Views Generation:

Definition 9: **Feature Set** F: For the sensing cell SC_x, we define the feature set F_x to be a set that contains the aggregated semantic feature (ASF defined in 7) extracted from the SSFE and the remote sensing feature set (RSF defined in 8) extracted from the RSFE as follows:

$$F_x = ASF_x \cup RSF_x \qquad \forall 1 \leq x \leq X \tag{5}$$

where X is the number of sensing cells.

Definition 10: **View** φ: We define the view φ as follows:

$$r_x^y = \varphi^y(F_x), \text{where } r_x^y \in \{0, 1, 2, \cdots, k\}, \quad \forall 1 \leq x \leq X, \forall 1 \leq y \leq Y \tag{6}$$

where φ^y is a view that can make its own estimation of the traffic risk level of a cell by leveraging a subset of features extracted by the SSFE and RSFE. For example, we can create a view by applying the k-nearest neighbor classification algorithm to ASF to estimate traffic risk level of a cell from social sensing data by specifying k to be the number of traffic risk levels defined in Definition 5. We use $\Phi = \{\varphi^y\}_{y=1}^{Y}$ to denote the set of all views in our model where Y is the total number of views. r_x^y is the estimated traffic risk level from view φ^y at sensing cell SC_x. k represents the number of traffic risk levels of a cell. We add 0 to $\{1, 2, \cdots, k\}$ to represent the case where a view does not have sufficient confidence on its estimation of the traffic risk level in a cell (we refer to it as "abstain" case) [33]. We save the output r_x^y of the Y views on X sensing cells into an estimation matrix $\boldsymbol{M} = \left[r_x^y\right]_{X \times Y}$.

We propose three categories of views by leveraging different combinations of features in F_x:

- *Social-only views:* only ASF feature are used for the estimation from a view. In particular, we apply a set of semantic feature based classification algorithms (e.g., k-nearest neighbor) to estimate the traffic risk level of a cell.
- *Remote-only views:* only RSF features are used for the estimation from a view. In particular, we apply a set of visual feature based classification algorithms (e.g. sparse annotation propagation [34]) to estimate the traffic risk level of a cell.
- *Social-remote views:* both ASF and RSF features are used for estimation from a view. In particular, we apply a set of mix feature (i.e., both semantic and visual features) classification algorithms (e.g., pseudo-labeling [35]) to estimate the traffic risk level of a cell.

2) Mutli-View Integration:

Definition 11: **View Coverage** β: We define the View Coverage β^y as follows:

$$\beta^y = \frac{1}{X} \sum_{x=1}^{X} \mathbb{1}_{\{r_x^y \neq 0\}} \tag{7}$$

where $\mathbb{1}$ is the indicator function and β^y is the view coverage of the view φ^y, which is defined as the the probability that φ^y makes a non-zero (i.e., not "abstain") estimation r_x^y for a randomly selected cell in SC. In addition, we define the coverage threshold β_C where only a view with the coverage larger than β_C will be considered in the integrated estimation results. This threshold controls the trade-off between the noisiness and sparsity of the estimation matrix $\boldsymbol{M}$.

Given the above definitions, the next step is to combine the individual estimation on the traffic risk level of a cell from each view into a unified estimation to maximize the overall estimation accuracy. In particular, we formulate this problem as an enhanced maximum likelihood estimation problem [33]. We define the log-likelihood function of estimation matrix $\boldsymbol{M}$ and traffic risk level $\boldsymbol{\mathcal{R}}$ as follows:

$$\begin{aligned} l(\boldsymbol{\omega}, \boldsymbol{A}) &= \log p(\boldsymbol{M}, \boldsymbol{\mathcal{R}}; \boldsymbol{\alpha}, \boldsymbol{\omega}) \\ &= \log p(\boldsymbol{M} \mid \boldsymbol{\mathcal{R}}; \boldsymbol{A}) + \log p(\boldsymbol{\mathcal{R}}; \boldsymbol{\omega}) \\ &= \sum_{x=1}^{X} \sum_{i=1}^{k} \mathbb{1}_{\{\mathcal{R}_x = i\}} \omega_i \sum_{y=1}^{Y} \log \Bigg((1 - \beta^y) \mathbb{1}_{\{r_x^y = 0\}} \\ &\quad + \sum_{j=1}^{k} \beta^y \alpha_{ij}^y \mathbb{1}_{\{r_x^y = j\}} \Bigg) \end{aligned} \tag{8}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots, \omega_k)^T$ are a set of prior probabilities that represent the odds of a randomly selected cell with a certain traffic risk level. $\boldsymbol{A} = (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \cdots, \boldsymbol{\alpha}^Y)$ are a set of view confusion matrices where $\boldsymbol{\alpha}^y\, (1 \leq y \leq Y)$ is the confusion matrix of view φ^y (i.e., $\alpha_{ij}^y = \Pr(r_x^y = j \mid \mathcal{R}_x = i), \forall 1 \leq i, j \leq k$). We estimate the values of $\boldsymbol{\omega}$ and $\boldsymbol{A}$ (i.e., $\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{A}}$) using the stochastic gradient descent approach [36]. Using the $\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{A}}$, we can further estimate

the traffic risk level of a cell as follows:

$$\begin{aligned}\widehat{\mathcal{R}_x} &= \underset{i\in\{1,2,\cdots,k\}}{\arg\max}\ \Pr(\mathcal{R}_x = i \mid \boldsymbol{M}_x) \\ &= \underset{i\in\{1,2,\cdots,k\}}{\arg\max}\ \Bigg\{ \log\hat{\omega}_i + \sum_{y=1}^{Y} \log\Bigg((1-\beta^y)\mathbb{1}_{\{r_x^y=0\}} \\ &\quad + \sum_{j=1}^{k} \beta^y \hat{\alpha}_{ij}^y \mathbb{1}_{\{r_x^y=j\}}\Bigg)\Bigg\}, \quad \forall 1 \le x \le X \end{aligned} \tag{9}$$

The intuition of above equation is to estimate the accuracy of each view and judiciously combine the estimations from different views by assigning higher weights to more accurate views to maximize the overall estimation accuracy.

3) Recursive View Elimination: To obtain a more accurate and efficient model, we develop a Recursive View Elimination (RVE) algorithm to recursively remove the views with low estimation accuracy. In particular, we eliminate the view with the lowest estimated accuracy from the current set of views in each iteration of the algorithm and repeat the mutliview integration process on the remaining views until the union of the view coverage reaches a predefined threshold Θ[6]. We summarize MVID scheme in Algorithm 1. The input of this algorithm is the semantic and visual features extracted by the SSFE and RSFE components and the output is the estimated traffic risk levels of the cells.

Algorithm 1 MVIF Scheme with Recursive View Elimination (RVE)

1: Extract features of sensing cells: $\mathcal{F} \leftarrow \{F_1, F_2, \cdots, F_X\}$.
2: Initialize the view set to all available views: $\Phi \leftarrow \{\varphi^1, \varphi^2, \cdots, \varphi^Y\}$.
3: **while** True **do**
4: Calculate the overall coverage: $\beta^{\Phi} \leftarrow \frac{1}{X}\sum_{x=1}^{X} \mathbb{1}_{\{\sum_{\varphi^y\in\Phi} r_x^y \neq 0\}}$.
5: **if** $\beta^{\Phi} < \Theta$ **then**
6: **return** $\widehat{\boldsymbol{\mathcal{R}}}$
7: **end if**
8: Generate view estimation matrix $\boldsymbol{M}$ based on $\mathcal{F}$ and Φ.
9: Get estimations of parameters $\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{A}}$ based on Equation (8).
10: Update estimations of traffic risk level: $\widehat{\boldsymbol{\mathcal{R}}} \leftarrow (\widehat{\mathcal{R}_1}, \widehat{\mathcal{R}_2}, \cdots, \widehat{\mathcal{R}_X})^T$ based on Equation (9).
11: Calculate the estimated accuracy: $\hat{\alpha}^y \leftarrow \sum_{i=1}^{k} \hat{\omega}_i \frac{\hat{\alpha}_{ii}^y}{\sum_{j=1}^{k}\hat{\alpha}_{ij}^y}$ for each $\varphi \in \Phi$
12: select $\varphi^* \leftarrow \underset{\varphi\in\Phi}{\arg\min}\, \hat{\alpha}^{\varphi}$.
13: $\Phi \leftarrow \Phi \backslash \{\varphi^*\}$.
14: **end while**

V. Evaluation

In this section, we evaluate the performance of the RiskSens scheme using the real world traffic datasets collected from New York City. We compare the performance of RiskSens with state-of-the-art baselines. The evaluation results show that RiskSens significantly outperforms the baselines in terms of identification accuracy.

A. Dataset

Twitter NYC Traffic Report Dataset: In our evaluation, we use a dataset collected from Twitter [7] as our social sensing data source. This dataset consists of 239,734 traffic-related tweets that are posted in New York City over the time span from Jan.1st, 2016 to Jun.30th, 2018.

Google Map NYC Satellite Imagery Dataset: We use a dataset collected from Google Map [8] as our remote sensing data source. This dataset consists of 214,951 satellite images from New York City that are captured around June, 2018. Each image is of 640×640 resolution and represents a cell with a size of $60\mathrm{m} \times 60\mathrm{m}$.

NYPD Motor Vehicle Accident Report Dataset: We use public data provided by New York City Police Department (NYPD) [9] as the ground-truth data, upon which we evaluate our scheme and baseline methods. This dataset consists of 568,051 reports of traffic accidents that happened in New York City between Jan.1st, 2016 and Jun.30th, 2018. Each report contains the time and location of an accident as well as a set of numerical features indicating the severity of the accident. Examples of such features include the number of persons injured and killed, the number of pedestrians injured and killed, the number of cyclists injured and killed, and the number of motorists injured and killed. To obtain the ground truth traffic risk level ($\overline{\mathcal{R}_x}$, defined in Definition 5), we categorize all sensing cells in NYC into three different categories (i.e., "low", "neutral" and "high") based on the NYPD dataset by following the clustering procedure proposed in [4]. Please note that such detailed accident reports are not generally available in every city and we only use them for the purpose of ground truth in our evaluation.

B. Baseline Algorithms

We choose various risky traffic location identification schemes as the baselines. Except for random-based scheme (abbreviated as "RAND"), which randomly guesses the traffic risk levels, each of the baseline schemes leverages a subset of the sensing data listed in *Data Source* and integrates the information from the sensing data using an algorithm listed in *Integration Algorithm*.

1) Data Source:

- **Social-Only**: Social-only scheme (abbreviated as "S") estimates the traffic risk level of a cell based on the social sensing data alone [37]. Social-only views described in subsection IV-D are used in the estimation.
- **Remote-Only**: Remote-only scheme (abbreviated as "R") estimates the traffic safety level of a cell based

[6]In practice, Θ is usually set to be a value close to 1 (e.g., $\Theta = 0.99$) to make sure most sensing cells are covered by at least one view. For cells that are not covered by any view, we make estimations based on $\hat{\boldsymbol{\omega}}$.

[7]https://developer.twitter.com/

[8]https://cloud.google.com/maps-platform/

[9]https://www1.nyc.gov/site/nypd/services/vehicles-property/reports.page

on remote sensing data alone [4]. Remote-only views described in subsection IV-D are used in the estimation.

- **Social-Remote**: Social-Remote scheme (abbreviated as "SR") estimates the traffic risk level of a cell based on both social sensing and remote sensing data, using pseudo-labeling technique [35]. Social-remote views described in subsection IV-D are used in the estimation.

2) Integration Algorithm:

- **Best Single**: Best single algorithm (abbreviated as "BS") selects an individual view that has the highest estimation confidence on training data from the set of views as the representative one to make the final estimation.
- **Majority Voting**: Majority voting [38] algorithm (abbreviated as "MV") picks the mode of estimations made by views for each cell.
- **Maximum Likelihood Estimation**: Maximum likelihood estimation [33] algorithm (abbreviated as "MLE") integrates estimations made by views based on Equations (8) to (9).

The combinations of the data source and integration algorithm discussed above comprise all baselines. For example, "R-MV" refers to the baseline that leverages remote sensing data and makes the estimation using the majority voting.

C. Evaluation Metrics

In our evaluation, we define the following metrics to evaluate the performance of all compared schemes.

Accuracy (a): We examine the overall accuracy of the estimation over three classes.

$$a = \frac{1}{X}\sum_{x=1}^{X} \mathbb{1}_{\widehat{\mathcal{R}_x}=\overline{\mathcal{R}_x}} \tag{10}$$

where X is the number of sensing cells and $\mathbb{1}$ is the indicator function.

Quadratic Weighted Kappa (κ-Score): We want to measure the degree of disagreement between the estimation and the ground truth. For example, the mis-classification of the risk level from "high" to "safe" is more severe than the mis-classification from "neutral" to "safe", and they should be assigned different weights. In particular, we calculate weighted Cohen's kappa score [39] over the three classes, with quadratic distances as the weight.

F1-Score: We pay special attention to the "high" class, which represents the identified high risk traffic locations. We calculate F1-Score of this class in our evaluation.

D. Evaluation Results

In the first set of experiments, we evaluate the performance of all schemes by changing the number of studied sensing cells from 2000 to 6000. The results are presented in Table II. We observe that the RiskSens scheme outperforms all of the baselines in all evaluation metrics as the number of sensing cells changes. In terms of classification accuracy, the performance gain achieved by RiskSens compared to the best-performing baseline on $X = 2000, X = 4000, X = 6000$ are $4.3\%, 7.8\%, 8.1\%$, respectively. Such performance gains of the RiskSens scheme are achieved by efficiently fusing heterogeneous information into a holistic view of the traffic risk and by judiciously selecting views (defined in Definition 10) with the highest estimated accuracy. Additionally, we observe that the performance gains of RiskSens scheme compared to baselines increase as the number of sensing cells increases. This demonstrates the consistent performance of the RiskSens scheme over different sample sizes.

In the second set of experiments, we evaluate the performance of all schemes in a more refined granularity of the sensing area. In particular, we study three of the major boroughs of New York City, namely Manhattan, Queens and Brooklyn. These sensing areas have different population density, road conditions and rate of traffic accidents. The results are presented in Table III. We observe that the RiskSens scheme consistently outperforms all the baselines (e.g., the accuracy gains over the best-performing baseline are $4.8\%, 8.0\%, 3.7\%$, respectively). The fluctuation in the estimation performances as the sensing area changes may be accredited to the variance in landscape of different sensing areas. For example, in Manhattan, some sensing cells are obstructed by tall buildings, making remote sensing data less reflective of the real traffic conditions. The traffic risk level of this kind of areas are more difficult to estimate. We observe that the RiskSens scheme maintains a significant advantage over baseline methods in these "difficult" areas. This is because RiskSens efficiently fuses information from multiple views, and one view may help suppressing the noise present in another view.

In the third set of experiments, we study the robustness of the RiskSens scheme by tuning the parameters of our model. One key parameter in our model is the coverage threshold β_C (defined in Definition 11) for each individual view. This parameter controls the trade-off between the noisiness and sparsity of the estimation matrix (defined in Definition 10). The results are shown in Figure 4 and Figure 5. We observe that the performance of the RiskSens scheme is stable against the variations of the empirical parameter β_C in a broad range (from 0.25 to 0.5) when we vary the number of the sensing cells and when we change the sensing area.

Finally, we analyze the relative importance of view categories defined in Definition 10 in Section IV. The results are presented in Table IV. We observe that both remote-only views and social-only views make significant contributions to the RiskSens scheme (i.e., removing either remote-only views or social-only views would result in about 10% drop in accuracy). This shows that the RiskSens scheme effectively leverages both social-sensing and remote-sensing data. In addition, we observe that social-remote views also

Table II
IDENTIFICATION ACCURACY FOR ALL SCHEMES ON DIFFERENT NUMBERS OF SENSING CELLS

		2000 Sensing Cells			4000 Sensing Cells			6000 Sensing Cells		
Category	Algorithm	Accuracy	κ-Score	F1-Score	Accuracy	κ-Score	F1-Score	Accuracy	κ-Score	F1-Score
Random-Based	RAND	0.349	0.031	0.349	0.342	0.006	0.334	0.332	0.002	0.341
Social-Only	S-BS	0.416	0.296	0.479	0.433	0.318	0.489	0.428	0.316	0.490
	S-MV	0.438	0.257	0.362	0.442	0.273	0.377	0.436	0.268	0.348
	S-MLE	0.488	0.388	0.514	0.490	0.397	0.519	0.489	0.394	0.513
Remote-Only	R-BS	0.469	0.193	0.471	0.384	0.075	0.455	0.407	0.118	0.339
	R-MV	0.501	0.319	0.431	0.482	0.291	0.419	0.491	0.308	0.447
	R-MLE	0.506	0.333	0.459	0.488	0.310	0.461	0.499	0.319	0.485
Social-Remote	SR-BS	0.412	0.205	0.419	0.418	0.224	0.442	0.419	0.231	0.447
	SR-MV	0.365	0.100	0.156	0.397	0.169	0.220	0.395	0.166	0.217
	SR-MLE	0.368	0.105	0.163	0.400	0.175	0.232	0.398	0.171	0.228
Our Scheme	**RiskSens**	**0.549**	**0.453**	**0.524**	**0.568**	**0.464**	**0.590**	**0.580**	**0.507**	**0.591**

Table III
IDENTIFICATION ACCURACY FOR ALL SCHEMES ON DIFFERENT AREAS

		Manhattan			Queens			Brooklyn		
Category	Algorithm	Accuracy	κ-Score	F1-Score	Accuracy	κ-Score	F1-Score	Accuracy	κ-Score	F1-Score
Random-Based	RAND	0.333	0.027	0.349	0.329	0.002	0.315	0.328	0.006	0.319
Social-Only	S-BS	0.464	0.363	0.379	0.521	0.406	0.450	0.488	0.386	0.530
	S-MV	0.455	0.549	0.119	0.471	0.559	0.239	0.493	0.506	0.182
	S-MLE	0.472	0.549	0.119	0.472	0.559	0.239	0.544	0.482	0.461
Remote-Only	R-BS	0.489	0.297	0.555	0.507	0.375	0.653	0.516	0.256	0.568
	R-MV	0.502	0.553	0.416	0.549	0.599	0.560	0.494	0.529	0.421
	R-MLE	0.517	0.554	0.510	0.548	0.566	0.632	0.500	0.494	0.479
Social-Remote	SR-BS	0.460	0.375	0.549	0.451	0.486	0.608	0.462	0.295	0.586
	SR-MV	0.359	0.523	0.206	0.409	0.544	0.192	0.381	0.540	0.215
	SR-MLE	0.360	0.525	0.207	0.412	0.542	0.200	0.382	0.539	0.213
Our Scheme	**RiskSens**	**0.565**	**0.564**	**0.580**	**0.629**	**0.584**	**0.687**	**0.581**	**0.556**	**0.600**

make positive contributions to the RiskSense scheme. This demonstrates that RiskSens goes beyond simply combining the social-based and remote-based views and explores that latent correlation between the two.

Table IV
VIEW IMPORTANCE FOR THE RISKSENS SCHEME

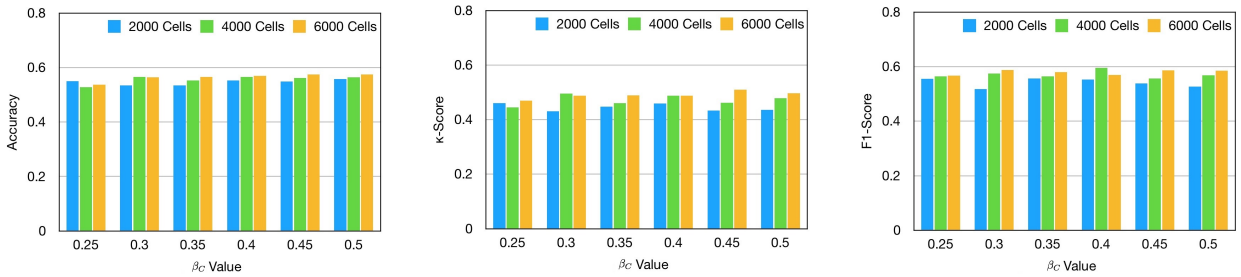
View Category	Accuracy Gain
Social-Only	9.3%
Remote-Only	10.4%
Social-Remote	1.1%

VI. CONCLUSION

This paper develops a RiskSens scheme to solve the risky traffic identification problem in intelligent transportation systems. The RiskSens scheme addresses two fundamental challenges that have not been fully addressed by current solutions, namely *data availability* and *identification granularity*. In particular, RiskSens explicitly fuses the heterogeneous features extracted from social and remote sensing data using a multi-view learning approach. The evaluation results on the real-world dataset from New York City demonstrate that the RiskSens scheme achieves significant performance gains compared to the state-of-the-art baselines under various experiment settings. The results of this paper are significant because they lay out an analytical foundation to address the risky traffic location identification problem by exploring the data collected from social and remote sensing applications.

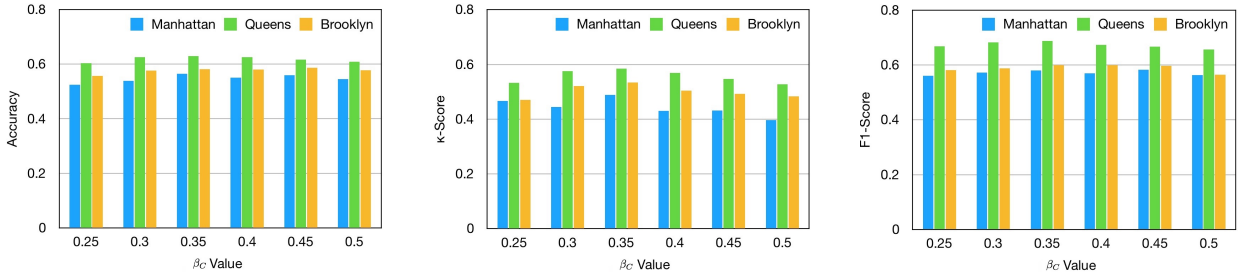
ACKNOWLEDGEMENT

This research is supported in part by the National Science Foundation under Grant No. CNS-1831669, CBET-1637251, CNS-1566465, and IIS-1447795, Army Research Office under Grant W911NF-17-1-0409, Google 2017 Faculty Research Award. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.



(a) Accuracy (b) κ-Score (c) F1-Score

Figure 4. Robustness of RiskSens Scheme on Different Numbers of Sensing Cells



(a) Accuracy (b) κ-Score (c) F1-Score

Figure 5. Robustness of RiskSens Scheme on Different Areas

REFERENCES

[1] D. Wang, T. Abdelzaher, and L. Kaplan, *Social sensing: building reliable systems on unreliable data*. Morgan Kaufmann, 2015.

[2] J. Dash and B. O. Ogutu, "Recent advances in space-borne optical remote sensing systems for monitoring global terrestrial ecosystems," *Progress in Physical Geography*, vol. 40, no. 2, pp. 322–351, 2016.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[4] A. Najjar, S. Kaneko, and Y. Miyanaga, "Combining satellite imagery and open data to map road safety." in *AAAI*, 2017, pp. 4524–4530.

[5] M. Bíl, R. Andrášik, and Z. Janoška, "Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation," *Accident Analysis & Prevention*, vol. 55, pp. 265–273, 2013.

[6] J. He, W. Shen, P. Divakaruni, L. Wynter, and R. Lawrence, "Improving traffic prediction with tweet semantics."

[7] M. Itoh, D. Yokoyama, M. Toyoda, Y. Tomita, S. Kawamura, and M. Kitsuregawa, "Visual fusion of mega-city big data: an application to traffic and tweets data analysis of metro passengers," in *Big Data (Big Data), 2014 IEEE International Conference on*. IEEE, 2014, pp. 431–440.

[8] Q. Chen, X. Song, H. Yamada, and R. Shibasaki, "Learning deep representation from big and heterogeneous data for traffic accident inference." in *AAAI*, 2016.

[9] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Goldberg, "Respectful cameras: Detecting visual markers in real-time to address privacy concerns," in *Protecting Privacy in Video Surveillance*. Springer, 2009, pp. 65–89.

[10] Y. Gu, Z. S. Qian, and F. Chen, "From twitter to detector: Real-time traffic incident detection using social media data," *Transportation research part C: emerging technologies*, vol. 67, pp. 321–342, 2016.

[11] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2013, pp. 344–353.

[12] D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji, and L. Kaplan, "The age of social sensing," *arXiv preprint arXiv:1801.09116*, 2018.

[13] S. Ilarri, O. Wolfson, and T. Delot, "Collaborative sensing for urban transportation." *IEEE Data Eng. Bull.*, vol. 37, no. 4, pp. 3–14, 2014.

[14] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[15] Y.-C. Yu, "A mobile social networking service for urban community disaster response," in *Semantic Computing (ICSC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 503–508.

[16] D. Y. Zhang, C. Zheng, D. Wang, D. Thain, X. Mu, G. Madey, and C. Huang, "Towards scalable and dynamic social sensing using a distributed computing framework," in *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*. IEEE, 2017, pp. 966–976.

[17] D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps," *Biosystems engineering*, vol. 114, no. 4, pp. 358–371, 2013.

[18] P. Reinartz, M. Lachaise, E. Schmeer, T. Krauss, and H. Runge, "Traffic monitoring with serial images from airborne cameras," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 61, no. 3-4, pp. 149–158, 2006.

[19] X.-L. Chen, H.-M. Zhao, P.-X. Li, and Z.-Y. Yin, "Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes," *Remote sensing of environment*, vol. 104, no. 2, pp. 133–146, 2006.

[20] X. Zhou, D. Dong, H. Wu, S. Zhao, D. Yu, H. Tian, X. Liu, and R. Yan, "Multi-view response selection for human-computer conversation," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 372–381.

[21] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 945–953.

[22] A. Zeng, K.-T. Yu, S. Song, D. Suo, E. Walker, A. Rodriguez, and J. Xiao, "Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1386–1383.

[23] A. Valmarska, D. Miljkovic, N. Lavrač, and M. Robnik-Šikonja, "Towards multi-view approach to parkinsonâĂŹs disease quality of life data analysis," in *5th International Workshop on New Frontiers in Mining Complex Patterns at ECML-PKDD2016*, 2016.

[24] J. Wang and F. H. Lochovsky, "Data extraction and label assignment for web databases," in *Proceedings of the 12th international conference on World Wide Web*. ACM, 2003, pp. 187–196.

[25] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, "Efficient color histogram indexing for quadratic form distance functions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 17, no. 7, 1995.

[26] M. Pietikäinen, "Local binary patterns," *Scholarpedia*, vol. 5, no. 3, p. 9775, 2010.

[27] I. Sikirić, K. Brkić, and S. Šegvić, "Classifying traffic scenes using the gist image descriptor," *arXiv preprint arXiv:1310.0316*, 2013.

[28] X. Zhang, J. Cui, W. Wang, and C. Lin, "A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm," *Sensors*, vol. 17, no. 7, p. 1474, 2017.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[32] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks." in *CVPR*, vol. 1, no. 2, 2017, p. 3.

[33] A. J. Ratner, C. M. De Sa, S. Wu, D. Selsam, and C. Ré, "Data programming: Creating large training sets, quickly," in *Advances in neural information processing systems*, 2016, pp. 3567–3575.

[34] A. Jung, A. O. Hero III, A. Mara, and S. Jahromi, "Semi-supervised learning via sparse label propagation," *arXiv preprint arXiv:1612.01414*, 2016.

[35] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on Challenges in Representation Learning, ICML*, vol. 3, 2013, p. 2.

[36] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[37] G. Anastasi, M. Antonelli, A. Bechini, S. Brienza, E. D'Andrea, D. De Guglielmo, P. Ducange, B. Lazzerini, F. Marcelloni, and A. Segatori, "Urban and social sensing for sustainable mobility in smart cities," in *Sustainable Internet and ICT for Sustainability (SustainIT), 2013*. IEEE, 2013.

[38] L. Lam and S. Suen, "Application of majority voting to pattern recognition: an analysis of its behavior and performance," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 27, no. 5, pp. 553–568, 1997.

[39] J. Cohen, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit." *Psychological bulletin*, vol. 70, no. 4, p. 213, 1968.