

## Article

# Leveraging Positive-Unlabeled Learning for Enhanced Black Spot Accident Identification on Greek Road Networks

Vasileios Sevetlidis <sup>1,2</sup>, George Pavlidis <sup>2,\*</sup>, Spyridon G. Mouroutsos <sup>3</sup> and Antonios Gasteratos <sup>1</sup>

<sup>1</sup> Department of Production and Management Engineering, Democritus University of Thrace, Vas. Sofias 12, GR-67100 Xanthi, Greece; vsevetli@pme.duth.gr (V.S.); agaster@pme.duth.gr (A.G.)

<sup>2</sup> Athena Research Center, University Campus at Kimmeria, GR-67100 Xanthi, Greece

<sup>3</sup> Department of Electrical and Computer Engineering, Democritus University of Thrace, University Campus at Kimmeria, GR-67100 Xanthi, Greece; sgmour@ee.duth.gr

\* Correspondence: gpavlid@athenarc.gr

**Abstract:** Identifying accidents in road black spots is crucial for improving road safety. Traditional methodologies, although insightful, often struggle with the complexities of imbalanced datasets, while machine learning (ML) techniques have shown promise, our previous work revealed that supervised learning (SL) methods face challenges in effectively distinguishing accidents that occur in black spots from those that do not. This paper introduces a novel approach that leverages positive-unlabeled (PU) learning, a technique we previously applied successfully in the domain of defect detection. The results of this work demonstrate a statistically significant improvement in key performance metrics, including accuracy, precision, recall, F1-score, and AUC, compared to SL methods. This study thus establishes PU learning as a more effective and robust approach for accident classification in black spots, particularly in scenarios with highly imbalanced datasets.

**Keywords:** black spot identification; imbalanced datasets; positive-unlabeled learning



**Citation:** Sevetlidis, V.; Pavlidis, G.; Mouroutsos, S.G.; Gasteratos A. Leveraging Positive-Unlabeled Learning for Enhanced Black Spot Accident Identification on Greek Road Networks. *Computers* **2024**, *13*, 49. <https://doi.org/10.3390/computers13020049>

Academic Editor: Kartik B. Ariyur

Received: 22 December 2023

Revised: 26 January 2024

Accepted: 6 February 2024

Published: 8 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Identifying accidents at black spots on road networks remains a critical task for enhancing road safety measures. These high-risk areas, characterized by a higher concentration of accidents, have traditionally been identified through various methodologies [1]. These range from statistical and Geographic Information Systems (GIS)-based analyses to accident reconstruction and road safety audits [2]. However, while these methods have provided valuable insights, they often come with limitations, such as strict assumptions or challenges in handling complex, imbalanced datasets.

In parallel, ML techniques have emerged as a promising avenue for tackling the problem of accident classification at black spots. Supervised learning (SL) methods have been explored, but their performance has often been found lacking, primarily due to the imbalanced nature of the dataset and the complexity of the problem space [3]. In a previous work [4], we introduced a novel dataset concerning accidents at black spots, called Black Spots in North Greece (BSNG), and provided a baseline using SL methods. The dataset highlighted the limitations of existing SL approaches in this context.

Weakly supervised learning (WSL) techniques have shown significant promise in various fields requiring defect or anomaly detection, particularly when labeled data are scarce or imbalanced [5]. These methods, including positive-unlabeled (PU) learning, have been successful in building robust classifiers that can predict the probability of a sample being positive, given the partially assigned labels, while requiring significantly fewer labeled data, as we have shown in a previous work [6].

Motivated by these observations, this paper aims to apply the concept of PU learning to the domain of accident classification at black spots. Given the nature of the BSNG dataset, where accidents at black spots are scarce and it is difficult to divine their pattern,

PU learning offers a promising avenue for improving model performance. This paper provides a comprehensive review of both traditional and ML methodologies, with a focus on the advantages of employing PU learning for accident classification at black spots.

### Contributions

This work makes several key contributions to the field of road safety analytics and ML. First, it introduces the application of positive-unlabeled (PU) learning [7] to the problem of accident classification at black spots, addressing the limitations of traditional SL methods in handling imbalanced datasets. Second, it provides a comprehensive comparative analysis between PU learning and existing SL methods, demonstrating statistically significant improvements ( $p < 0.01$ ) in performance metrics such as accuracy, precision, recall, F1-score, and AUC. Last but not least, the methodology presented herein serves as a blueprint for researchers and practitioners looking to apply weakly supervised learning techniques to similar problems in other domains, thereby broadening the scope and impact of this research.

## 2. Literature Review

### 2.1. Traditional Methods for Black Spot Identification

The primary goal of identifying black spots is to identify specific locations or road segments within a road network where accidents occur at a higher rate, with the ultimate objective of lowering both the frequency and severity of accidents in these areas. It is important to note that different countries may have distinct definitions of what constitutes a “black spot” in their road networks, as discussed in studies like [8–11]. However, a general definition for a “black spot” is a particular area or length of road that experiences a significant number of accidents or incidents, often resulting in serious injuries or worst, death. Typically, these places are recognized through the analysis of data from various sources, including police reports, records of traffic accidents, and other governmental databases.

Historically, conventional approaches to pinpointing black spots have heavily leaned on statistical analyses and GIS [12–15]. These techniques typically incorporate criteria that can vary based on factors such as the country in question, the type of road network, and the data available. To illustrate, certain countries may define a black spot by considering the total number of accidents within a designated timeframe, while others may emphasize the severity of these accidents as a defining factor. Table 1 provides a comparative overview of how different countries classify a “black spot”.

A breadth of literature underscores the diverse methodologies employed in the identification of black spots and the continuous advancements in this critical domain of road safety. These studies collectively underscore the multidimensional nature of black spot identification, illustrating the instrumental role of traditional methodologies like statistical analysis and GIS in advancing road safety and accident prevention.

**Table 1.** Different nations provide different definitions on what constitutes a black spot.

| Country/Area | Methodology                  | Sliding Window (m)        | Threshold                    | Severity Included | Time Frame (Years) |
|--------------|------------------------------|---------------------------|------------------------------|-------------------|--------------------|
| Denmark      | Poisson                      | variable length           | 4                            | No                | 5                  |
| Croatia      | Segment ranking              | 300                       | 12                           | Implicitly        | 3                  |
| Flanders     | Weighted method              | 100                       | 3                            | Yes               | 3                  |
| Hungary      | Accident indexing            | 100 (spot)/1000 (segment) | 4                            | No                | 3                  |
| Switzerland  | Accident indexing            | 100 (spot)/500 (segment)  | Statistical, critical values | Implicitly        | 2                  |
| Germany      | Weighted indexing            | Likelihood                | 4                            | No                | 5                  |
| Portugal     | Weighted method              | 200                       | 5                            | Yes               | 5                  |
| Norway       | Poisson, statistical testing | 100 (spot)/1000 (segment) | 4                            | Accident cost     | 5                  |
| Greece       | Absolute count               | 1000                      | 2                            | No                | N/A                |

Methods of statistical analysis, like regression and time-series analysis, are frequently employed to discern trends and connections between road accidents and variables such as the layout of roads, the amount of traffic, the behavior of drivers, and weather patterns.

These methods have a solid track record in the realm of road safety modeling, playing a crucial role in pinpointing hazardous locations and formulating preventive strategies [16]. Various statistical models have been applied, including Poisson regression [17,18], binomial regression [19], negative binomial regression [20], Poisson–lognormal regression [21], zero-inflated regression [22], generalized estimation equations [23], negative multinomial models [24], random effects models [20], and random parameter models [25]. Furthermore, numerous models have been developed to assess the severity of crashes, including binary logit, binary probit, Bayesian ordered probit, Bayesian hierarchical binomial logit, generalized ordered logit, log–linear model, multinomial logit, multivariate probit, ordered logit, and ordered probit models [26]. However, their application can be complex and requires specialized expertise, potentially introducing subjectivity into the analysis. These methods often demand extensive manual work and are constrained by data quality and availability. Finally, they lack adaptability to changing conditions, limiting real-world applicability.

GIS technologies are instrumental in plotting road accidents and pinpointing critical areas within the road network [27]. They reveal connections between spatial phenomena that may remain hidden when using non-spatial databases [28,29]. In recent decades, a plethora of studies has explored the application of GIS in traffic safety and accident analysis, with numerous organizations and researchers documenting its efficacy [30,31]. Such analyses encompass various techniques, including intersection analysis [32], segment analysis [33], cluster analysis [34], and density analysis modeling [34]. Notably, Lasisi et al. [35] homed in on the prediction of accidents at highway–rail grade crossings through a hybrid methodology integrating machine learning with GIS. The models they developed showcased impressive results, achieving a high accuracy rate of 98.9% alongside a Receiver Operating Characteristic (ROC) score of 0.9838, underlining the significant promise of merging machine learning and GIS for the purpose of accident forecasting.

However, GIS technology does come with its drawbacks. It is costly and demands specialized expertise and knowledge [36], potentially restricting its adoption by certain organizations. The data's quality is pivotal for the precision of the GIS analysis outcomes [37]. For instance, unreliable or incomplete data can lead to questionable analysis results. In addition, while GIS is proficient in processing spatial details, it may not offer extensive insight into temporal aspects that influence accidents, like driver behavior [38]. In essence, GIS technology furnishes a graphical depiction of black spot locations in road networks and supports the amalgamation of diverse datasets.

Identifying black spots presents challenges, primarily due to the variability in accident numbers compared to regular road segments, influenced by location, road design, and traffic volume [39]. Thus, a multidisciplinary approach, considering various data sources and analytical techniques, is essential.

Recognizing these challenges and limitations, there is growing interest in employing ML techniques, particularly deep learning, to enhance black spot identification [40]. However, the application of ML in this field remains relatively unexplored, primarily due to limited large datasets, necessitating further research to bridge this gap.

## 2.2. Machine Learning in Black Spot Identification

ML techniques, particularly deep learning, are increasingly being used in transportation research for black spot identification [41,42]. These data-driven methods offer adaptability to new data and changing conditions, overcoming limitations of traditional methods, which often involve strict assumptions and manual labor and struggle with complex datasets [43]. Studies like Theofilatos et al.'s [44] and Fan et al.'s [45] demonstrate the use of advanced ML and deep neural networks to predict road accidents with considerable accuracy. These studies highlight the potential of deep learning in analyzing traffic accident data, considering various factors including road conditions and weather.

Mbarek et al. [46] developed a model using the extreme learning machine algorithm, ordinal regression, and XGBoost, which accurately identified black spots on rural roads in Morocco with an accuracy of 98.6%. According to the study, the significant factors

contributing to accidents included pavement width, road curve type, and position. A different venue was explored in [47], which retrieved data from social networks, particularly Twitter, to create supervised classification models. These models classify tweets about the occurrence of road accidents at black spots and include the construction of mobile applications to notify drivers about accidents in real-time. A data-driven machine learning solution for screening accident black spots on road networks was proposed in [48]. The solution utilizes features of the road network and nearby locations associated with accidents to predict black spots accurately. Similarly, another data-driven study utilized machine learning and supervised learning methods to investigate the causes and severity of road accidents at black spots on road networks. This approach aids in understanding the contributing factors and potentially mitigating their impact [49]. The spatial distribution of road traffic accidents and the identification of factors associated with these accidents using a decision tree classification approach was performed in [50]. This machine learning method was instrumental in identifying accident hot spots during peak and off-peak hours. SVM, random forest, and a multi-layer perceptron neural network were used in [51] to classify road accident hot spots on the Brazilian federal road network. The neural network model was notably effective, achieving the highest accuracy of 83% in predicting severe or non-severe accident risks.

Other recent studies have underscored the potential of advanced machine learning approaches for enhancing the identification and analysis of black spots on road networks. A hybrid machine learning approach to analyze road accidents in black spots, utilizing various algorithms to classify accidents based on their consequences and identifying the best suitable model for each zone was proposed in [52]. The effectiveness of the random forest model as the most suitable algorithm for predicting crash severity levels, marking a significant step in accident severity analysis, was highlighted in [53]. The ability of decision tree, LightGBM, and XGBoost to provide deeper insights into accident classification was investigated in [54]. The study mainly focused on the causes and severity of road accidents. The impact of traffic management factors on the causes and severity of road accidents at black spots, employing machine learning methods for a comprehensive analysis, was analyzed in [55]. An algorithm named META-DES-RF was proposed to predict injury severity, showcasing the potential of machine learning in classifying and understanding the severity of road accidents [56]. A fuzzy algorithm was employed to classify road traffic accident data, identifying key factors related to accident severity with an accuracy of 85.94%, marking a significant advancement in accident analysis [57]. Finally, a combination of random forest and convolutional neural network models was proposed in [58] to identify significant factors strongly correlated with accident severity.

However, the effectiveness of these ML methods can be hampered by challenges such as imbalanced datasets, noise, and missing data [59,60]. Traditional statistical models, limited by rigid assumptions, face difficulties in handling these issues. To overcome these challenges, ensemble methods like random forest and AdaBoost have been employed, but they too have limitations in managing the dynamic nature of black spot identification [3].

Deep learning advancements, like deep neural networks, have been applied to predict accidents in real-time using comprehensive datasets. However, while these methods have shown higher accuracies compared to traditional models, they still depend heavily on the nature of the dataset. The variability in data type, size, and the specific context of the road network are critical factors influencing model accuracy.

Acknowledging these limitations, there is a clear need for more robust and reliable ML approaches in this field. This paper proposes a shift from traditional SL to an outlier detection framework. It introduces the application of positive-unlabeled learning to address these challenges, offering a novel and promising direction for enhancing black spot identification methods.

### 2.3. Weakly Supervised Learning in Defect Detection

In numerous sectors, ranging from production to healthcare, the task of detecting defects is paramount [61]. Promptly spotting flaws helps avert issues related to quality, leading to cost reductions by circumventing the need for costly revisions or product recalls. As Industry 4.0 progresses, there is a growing reliance on automated systems for inspection to ensure defects are identified instantly, guaranteeing that only the finest quality products reach the consumers [62]. Nevertheless, the acquisition of extensive labeled datasets to train defect detection models that are precise poses a substantial hurdle [63]. The process of manually labeling is not only tedious and costly [64–66] but inaccuracies in labeling can also severely impact the performance of the resulting model [67,68].

Within this framework, weakly supervised learning (WSL) has surfaced as an appealing substitute. In contrast to conventional SL that necessitates extensive annotated data, weakly SL allows models to extract knowledge from merely a portion of the labeled data or even from data that is not labeled at all [69]. This methodology has garnered considerable interest for its ability to diminish dependence on manual labeling and enhance the learning process's efficacy.

Several approaches are categorized under weakly supervised learning (WSL), such as multi-instance learning [70], co-training [71], and positive-unlabeled learning [72]. These techniques are designed to construct models with a minimal set of labeled data or even in the absence of any labeled data. For instance, multi-instance learning involves a collection of instances where each is identified as positive, negative, or unlabeled. Co-training involves the parallel training of multiple classifiers, each utilizing distinct perspectives of the data, to enhance the collective accuracy of the models. Positive-unlabeled learning is centered around developing a binary classification model by utilizing datasets in which only the positive instances are labeled.

Recent research has underscored the efficacy of weakly supervised learning (WSL) in the domain of defect detection [73]. For example, techniques centered around surface segmentation via CycleGAN, trained with image-level labels, have surpassed the performance of fully supervised methods in industrial datasets [74]. Additionally, a different study amalgamated a modest quantity of labeled data with a vast pool of unlabeled data, resulting in a model whose accuracy rivaled that of its fully supervised counterpart [6]. These findings accentuate the promise held by weakly supervised learning in defect detection endeavors and spotlight the overall effectiveness of various methodologies.

Specifically, the integration of weakly supervised learning in defect detection, particularly in road safety and accident classification, has seen notable advancements. Chatterjee et al. [75] proposed a machine learning approach utilizing front-view images for crack and defect detection on road surfaces. This method efficiently managed various road surface conditions and types of cracks, pinpointing the defective regions in the images. A variety of machine learning techniques including support vector machines (SVM), k-nearest neighbors (kNN), and multi-layer perceptron (MLP) models and their performance in detecting cracks and potholes in road images was explored in [76]. The use of a convolutional neural network (CNN) for automatically detecting and classifying road surface images was discussed in [77]; the proposed model achieved high accuracy for crack detection and categorization into 10 distinct classes. Machine learning techniques for the categorization of road surface conditions through smartphone sensors to effectively classify smooth road, potholes, and deep transverse cracks was proposed in [78]. Lastly, a novel self-supervised learning method based on masked image modeling was proposed by Zhang et al. [79] for driver distraction behavior detection, achieving an impressive accuracy of 99.60%, nearly matching the performance of advanced supervised learning methods. These studies illustrate the diverse approaches and significant potential of weakly supervised and machine learning techniques in defect detection and road safety enhancement, offering innovative solutions to longstanding challenges in the field.

In conclusion, weakly supervised learning (WSL) emerges as a potent strategy in situations where labeled data is limited. This method employs a modest set of positively

labeled data, a feature encoder, an anomaly detection technique, and a binary classifier to yield notable outcomes. As such, it stands as a significant substitute for conventional SL approaches, particularly in practical settings where acquiring extensive labeled data is problematic.

#### 2.4. Bridging the Gap

The identification of black spots on road networks and defect detection in various industries represent two distinct but related challenges that both aim to enhance safety and quality through the early identification of high-risk areas or defects. However, while both domains have seen significant advancements through traditional methods and ML techniques, there exists a gap in the literature that explores the intersection of these two fields. This gap is particularly evident when considering the common challenges they face, such as the handling of imbalanced datasets, the need for robust classifiers, and the limitations of fully SL methods.

Given these observations, there is a compelling need for research that bridges the gap between these two domains. The common challenges they face, such as imbalanced datasets and the need for robust classifiers, make a strong case for the exploration of weakly SL methods like PU learning in the context of black spot identification. Such an interdisciplinary approach could offer innovative solutions that leverage the strengths of both domains, thereby enhancing the effectiveness of black spot identification methods and potentially saving lives.

This paper aims to fill the research gap by integrating advanced analytical techniques from both fields. Thus, the application of PU learning to the domain of black spot identification is proposed, offering a novel approach that promises to address the limitations of existing methods in both domains.

### 3. Methodology

To tackle this issue, this work employs positive-unlabeled (PU) learning, a specialized framework for handling imbalanced or partially labeled datasets [80,81]. In PU learning, the approach starts with a set of positively labeled examples (high-risk areas) and a larger set of unlabeled examples. The initial step involves creating an “anti-class” from the unlabeled data. This counter-class serves as a negative proxy class, distinct from the known positive class. The rationale behind this is to provide a contrasting set that helps the model focus on the characteristics that differentiate the positive class.

#### 3.1. Transforming Supervised Learning into Outlier Detection

In SL, the objective is to learn a function  $f$  that maps an input  $x$  to an output  $y$ , based on training examples  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ . Here,  $x$  represents data features, and  $y$  is the target label. The goal is to learn a function  $f$  that maps inputs to their correct labels, denoted as:

$$f : X \rightarrow Y$$

The function  $f$  is learned during training, where the algorithm minimizes the error between predicted values  $f(x)$  and actual values  $y$  over the training examples.

In contrast, outlier detection identifies data points significantly different from the majority of the data. It is often considered an unsupervised task, as it does not rely on labeled data. Given a dataset  $D = \{x_1, x_2, \dots, x_n\}$ , an outlier detection algorithm aims to identify a subset  $O \subseteq D$  such that each  $x \in O$  is an outlier. The determination of whether a data point is an outlier depends on a measure of distance or deviation from the dataset's central tendency or distribution.

The mathematical formulation of outlier detection varies depending on the method used. For instance, in a distance-based approach, a data point  $x_i$  is considered an outlier if:

$$|\{x_j : d(x_i, x_j) > \theta, j \neq i\}| > \phi$$

where  $d(x_i, x_j)$  is the distance measure between points  $x_i$  and  $x_j$ ,  $\theta$  is a threshold distance, and  $\phi$  is a threshold number of points.

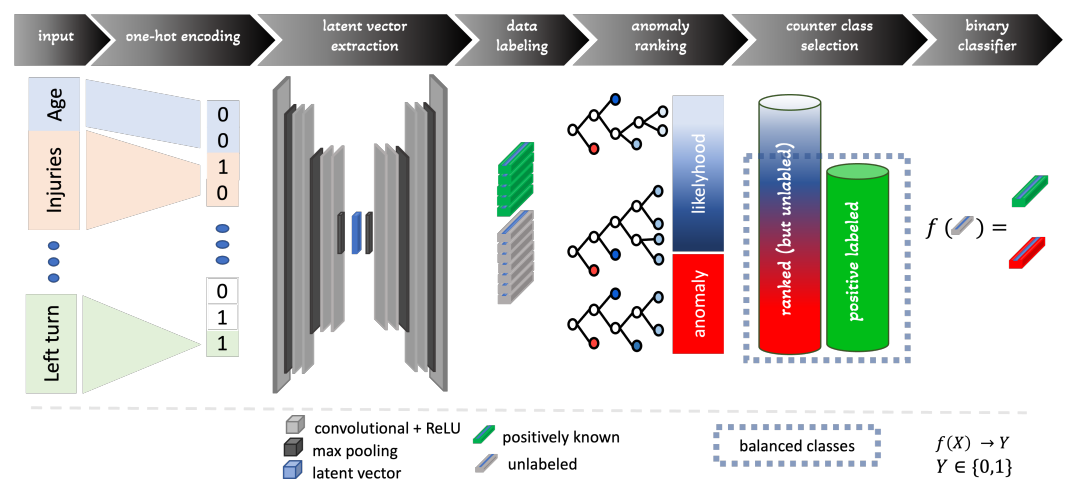
In PU learning, instead of having both positive and negative labeled examples, only positive examples are available and all the rest are unlabeled data, which could have as members only the samples of one class or the other. The dataset appears as  $\{(x_1^+, y_1^+), \dots, (x_p^+, y_p^+)\}$  for positive samples and  $\{x_{p+1}^u, \dots, x_n^u\}$  for unlabeled samples, where  $y^+$  indicates a positive label and  $u$  denotes unlabeled. In this context, positive examples are considered ‘normal’ data, and the goal is to identify outliers (negative examples) in the unlabeled dataset.

A commonly adopted positive-unlabeled learning approach involves a process with two main steps. First, the method identifies reliable negatives from the unlabeled set. This identification can be achieved using various techniques, including clustering, applying distance metrics, or employing a classifier initially trained only on the positive examples and then applied to the unlabeled set. The second step involves training a classifier using both the original positive examples and these newly identified negative examples.

Subsequently, this classifier can be applied to a new set of unlabeled data and instances classified as negative can be considered as outliers relative to the positive class.

### 3.2. Proposed Pipeline

In this work, the process of black spot identification commences with one-hot encoding to effectively represent the categorical data. An autoencoder is utilized to extract the latent space from the dataset. Following this, the data are partitioned into two distinct subsets: “known black spots” and unlabeled data (a mix of accidents situated either at a black spot or at a regular location). The isolation forest method is trained on the positively known class, namely the black spots, enabling the detection of anomalies (non-black spot accident areas) within the unlabeled data [82]. Subsequently, the predictions are ranked in ascending order, with a focus on identifying the least to the most anomalous samples. From this ranked list, a portion equivalent to the population of the known data is retrieved from the topmost anomalous samples. Finally, a classification algorithm is trained on these two “balanced” classes, culminating in the comprehensive methodology employed in this study for black spot identification. The proposed pipeline is shown graphically in Figure 1.



**Figure 1.** A step-by-step graphical presentation of the proposed pipeline.

#### Categorical Encoding

Categorical variables are transformed using one-hot encoding. This involves converting each category into a new categorical variable and assigning a binary value of 1 or 0. For example, a variable RoadType with categories Urban, Rural, and Highway would be transformed into three new variables: IsUrban, IsRural, and IsHighway (see Table 2).

This transformation allows ML algorithms to process the categorical data effectively, thereby improving the model’s performance.

**Table 2.** Description of road types

| RoadType | IsUrban | IsRural | IsHighway |
|----------|---------|---------|-----------|
| Urban    | 1       | 0       | 0         |
| Rural    | 0       | 1       | 0         |
| Highway  | 0       | 0       | 1         |

### 3.3. Self-Supervised Deep Learning Model

The second pivotal step in the proposed feature extraction pipeline is the application of a self-supervised deep learning model. This model aims to reduce the dimensionality of the feature space while capturing the most salient characteristics of the data. Usually this is achieved with the employment of a bottleneck architecture that consists of an encoder and a decoder, forming an autoencoder structure.

The encoder is a neural network that takes the high-dimensional input features  $x$  and maps them to a lower-dimensional latent vector  $z$ . The mapping function  $f_{\text{encoder}}$  is defined as:

$$z = f_{\text{encoder}}(x; \theta_{\text{encoder}})$$

where  $\theta_{\text{encoder}}$  are the parameters of the encoder.

The decoder is a separate neural network that aims to reconstruct the original input features from the latent vector  $z$ . The mapping function  $f_{\text{decoder}}$  is defined as:

$$\hat{x} = f_{\text{decoder}}(z; \theta_{\text{decoder}})$$

where  $\hat{x}$  is the reconstructed input and  $\theta_{\text{decoder}}$  are the parameters of the decoder.

The objective of the self-supervised deep learning architecture is to minimize the reconstruction error between the original features  $x$  and the reconstructed features  $\hat{x}$ . The loss function  $\mathcal{L}$  is defined as:

$$\mathcal{L}(x, \hat{x}) = \|x - \hat{x}\|^2$$

The minimization of this loss leads the architecture to learn to capture the most important characteristics of the data in the latent vector  $z$ , which is then used for subsequent classification tasks.

### 3.4. Utilization of Latent Vectors

The final step of the proposed feature extraction pipeline involves the utilization of the latent vectors generated by the self-supervised deep learning architecture. These vectors serve as the input for the subsequent classification model and encapsulate the most salient features of the data.

The latent vectors  $z$  are used as the input features for a binary classification model. The model aims to distinguish between black spots and non-black spots based on these high-dimensional vectors. Formally, the classification function  $f_{\text{classifier}}$  is defined as:

$$y_{\text{pred}} = f_{\text{classifier}}(z; \theta_{\text{classifier}})$$

where  $y_{\text{pred}}$  is the predicted label, and  $\theta_{\text{classifier}}$  are the parameters of the classifier.

The objective function for the classifier is to minimize the binary cross-entropy loss between the predicted labels  $y_{\text{pred}}$  and the true labels  $y_{\text{true}}$ . The loss function  $\mathcal{L}_{\text{classifier}}$  is defined as:

$$\mathcal{L}_{\text{classifier}} = -\left(y_{\text{true}} \log(y_{\text{pred}}) + (1 - y_{\text{true}}) \log(1 - y_{\text{pred}})\right)$$

The primary benefit of employing latent vectors is their capacity to provide a condensed representation of data. This attribute is especially advantageous for diminishing the computational complexity associated with the classification task. In traditional ML models, high-dimensional data often require significant computational resources, both in terms of memory and processing power. Reducing the dimensionality of the feature space without losing essential information possibly enables the latent vectors to run more



complex algorithms faster and more efficiently [83], thereby expediting the research and development process.

The second benefit lies in the latent vectors' proficiency in encapsulating the most prominent attributes of the data. This is crucial for improving the model's performance metrics such as accuracy, precision, and recall. In the context of black spot identification, where the dataset may contain a wide range of variables from different domains, capturing the most important characteristics is essential for building a robust model. The latent vectors, generated through a self-supervised deep learning architecture, encapsulate these critical features, thereby enhancing the model's ability to generalize well to new, unseen data.

Lastly, the utilization of latent vectors provides a unified feature space that is amenable to various ML algorithms. This offers a level of flexibility that is often lacking when using raw or pre-processed data. Researchers and practitioners can experiment with different types of classification models, from decision trees and random forests to neural networks, without the need to re-engineer the feature extraction process. This adaptability not only speeds up the iterative process of model selection but also opens the door to leveraging more advanced ML techniques as they become available.

### 3.5. Anomaly Ranking with Isolation Forest

Anomaly ranking is a crucial step in the proposed pipeline for black spot identification, and it is achieved through the use of the isolation forest algorithm. This method is particularly adept at identifying outliers or anomalies within a dataset.

The isolation forest algorithm is based on the principle of isolating anomalies, rather than profiling normal data points. The fundamental concept posits that anomalies are 'sparse and distinct', rendering them more prone to isolation.

In mathematical terms, the isolation forest algorithm creates multiple decision trees, or 'isolation trees', to isolate each data point. For each tree, a random subset of features is selected, and a random split value between the minimum and maximum values of these features is chosen. This process is repeated recursively until each data point is isolated, i.e., it falls into its own path in the tree.

The key parameter in this algorithm is the path length, which is the number of splits required to isolate a sample. Anomalies, being few and different, tend to have shorter path lengths in these trees, as they are easier to isolate.

The anomaly score is computed as follows:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

where  $s(x, n)$  is the anomaly score of the sample  $x$ ,  $E(h(x))$  is the average path length of  $x$  over all the trees in the forest,  $n$  is the number of external nodes, and  $c(n)$  is a normalization factor.

The rationale behind using isolation forest in anomaly ranking is due to method's boundary creation, which effectively separate normal from anomalous unlabeled samples. Unlike distance-based or density-based methods, the isolation forest performs well in high-dimensional spaces, making it suitable for complex datasets like those involved in black spot identification. The random partitioning followed by this method and the multiple training instances provide a form of ensemble learning, making the method adaptable to various data distributions and less prone to overfitting to a specific feature or pattern. Finally, isolation forest is highly scalable with respect to the number of samples, making it suitable for large datasets.

### 3.6. Class Balancing through Counter Example Generation

The class balancing step in the proposed methodology involves creating a new class of counter examples, which are equal in population to the positively known class. These counter examples are derived from the initially unlabeled samples that are identified as most anomalous by the isolation forest algorithm.

The generation of counter examples is an important step in preparing the dataset for training a binary classifier. In this process, the samples from the unlabeled dataset that are ranked as most anomalous by the isolation forest are selected. This selection is based on the anomaly scores, with a higher score indicating a greater likelihood of the sample being a counter example. The number of samples selected as counter examples is made equal to the number of samples in the positively known class.

### 3.7. Training of the Binary Classifier

With the generation of counter examples, the dataset now comprises two balanced classes: the positively known and the counter examples. These classes are then used to train a binary classifier. The balanced nature of the dataset is crucial in this context, as it prevents the classifier from being biased towards the majority class, which is a common issue in imbalanced datasets. This approach ensures that the classifier is trained on a representative and unbiased dataset, enhancing its ability to generalize and accurately classify new, unseen data.

## 4. Experiments

### 4.1. Dataset and Preprocessing

The dataset employed in this study is the Black Spot Dataset of North Greece (BSNG), which was meticulously compiled from a variety of sources [4]. These sources include police reports, construction agencies, and academic experts. The dataset provides a comprehensive view of road accidents and safety conditions in North Greece.

The data were organized into a structured format using spreadsheets, where each row represents a record and each column represents an attribute or feature. The dataset has 1810 samples of traffic accident audits from which 310 are black spots and 1500 are regular accidents. All samples were described initially by 35 features. During pre-processing highly correlated variables were either merged into a single one or the redundancy was discarded, finally, using the following:

- Accident location;
- Incident and road environment details (month, week of year, number of deaths, serious injuries, minor injuries, total number of injuries, number of vehicles involved, road surface type, atmospheric conditions, road surface conditions, road marking, lane marking, road width, road narrowness, turn sequence, road gradient, straightness, right turn, left turn, boundary line marking left and right, accident severity, type of first collision);
- Driver information (gender and age);
- Vehicle information (type, age, and mechanical inspection status).

Duplicate values were identified and promptly removed. A small percentage of the data records exhibited missing values. To maintain the integrity of the dataset, interpolation was applied between these records and their closest neighbors. In cases where too many features were missing, the records were discarded.

Special attention was given to anonymizing the data records. Personally identifiable information (PII) was excised, and data points at the individual level were grouped together to obscure details specific to individuals. This approach significantly reduces the possibility of associating a data record in the BSNG with any particular individual involved in an accident.

During the data preprocessing phase, numerical values were scaled, and features that were not numerical were encoded. Qualitative attributes were categorized with specific labels, whereas quantitative attributes underwent normalization. Each step of the transformation process was meticulously documented to ensure that the procedures could be replicated. The variables underwent a transformation to labeled and one-hot encodings, rendering the data compatible with various machine learning (ML) algorithms.

#### 4.2. Experimental Setup

To assess the effectiveness of the proposed feature extraction pipeline and classification model, a series of experiments were carried out (All experiments were executed on a machine equipped with an Intel Core i9 processor, 64 GB RAM, and an NVIDIA GeForce RTX 3080 GPU). To establish a common ground for all experiments, each experimental iteration used the same randomly shuffled training and test splits, with an 80–20 ratio, respectively. In all experiments, hyperparameters were optimized using 5-fold cross-validation, with the aim of maximizing the F1 score, as in [4]. Initially, the dataset was used in its original state, without any transformations or augmentations applied (dataset A). This served as the baseline for the BSNG dataset, highlighting the inherent challenges.

The next set of experiments involved applying encoding procedures to categorize the data and ensure uniform representation (dataset B). This step aimed to present the classifiers with data that had a consistent appearance.

Subsequently, a third round of experiments was conducted using a neural network to encode the BSNG samples and employing MixUp augmentation, as suggested in [4]. For Mixup, 11 additional samples were generated for each of 6000 randomly selected pairs (with replacement) using a beta distribution  $\beta(0.2, 0.2)$ , comprising 67,448 training samples. This procedure aimed to increase the dataset's population while maintaining its statistical profile, making it suitable for models that need larger datasets to converge (dataset C).

Finally, the dataset was utilized within the framework of PU learning, without any augmentation techniques applied, using the initial sample count (dataset PU); 263 black spot samples were used for establishing the positively known class. This allowed for a direct comparison of the effectiveness of the PU learning approach in extreme cases, such as black spot detection, with traditional SL models.

The proposed self-supervised method used for the latent vector extraction was an autoencoder. The autoencoder's architecture had a bottleneck layout with the encoder consisting of three layers with ReLU activations and nodes of size (256, 64, 32). The decoder has the reverse layer order. The optimizer used was Adam with a learning rate of  $10^{-4}$ . The isolation forest was trained on 80% of the known black spots. The ensemble consisted of 250 trees and contamination set to 0.05.

In terms of the classifiers, a thresholded Poisson regressor was trained with  $\alpha = 0.7$  and tolerance  $10^{-4}$ ; a Gaussian Process with the squared exponential kernel  $k(x_n, x_m) = e^{-\frac{\|x_n - x_m\|^2}{2l^2}}$  was trained and the rest of the parameters were optimized during fitting; the value  $k = 3$  was set for the k-nearest neighbors algorithm; the decision tree used the Gini impurity as the splitting criterion and its depth left to be optimized during fitting; an ensemble of 500 trees were trained for the random forest algorithm and the Gini impurity was used here as well; the extra randomized trees had an ensemble size of 1000 trees, due to the method's low memory requirements and fast training; and the MLP was an architecture of three fully connected layers with (32, 24, 6) nodes, ReLU activation, a learning rate of  $10^{-4}$ , the Adam solver, and 100 training epochs. The classifier parameters were determined using a grid search, a method that we also employed in our previous study [4].

## 5. Results

This study presents a detailed comparison of machine learning performances across different datasets and learning frameworks, specifically standard supervised learning (SL) and positive-unlabeled (PU) learning. In SL, three datasets (A, B, and C) were examined.

Dataset A comprised unaltered BSNG samples. The accuracy varied between 68% and 80.7%, with precision and recall of black spots ranging lower. Random forest achieved the highest accuracy (80.71%), outperforming other methods, but all methods struggled with precision and recall. Dataset B, with an encoding procedure, showed improved accuracy (28% to 83%) and better precision and recall of black spots. SVM led in performance with an accuracy of 82.92% and precision of 50.01%. Dataset C, incorporating encoding and augmentation, presented enhanced results. Extra randomized trees excelled with an accuracy of 82.36% and precision of 45.5%.

In the PU learning context, accuracy ranged from 69.8% to 87.84%, with random forest having an accuracy of 87.84% and precision of 49.45%.

Moreover, a rigorous comparative analysis of the performance of two machine learning frameworks, namely SL and PU learning was carried out. To make a statistical comparison between the frameworks, the best results from the two frameworks were used in the context of the 5-fold cross-validation experiment. F1-score score, a critical metric for classification tasks, was chosen as the primary performance measure. Our findings reveal a noteworthy distinction in F1-score performance between the two methods. Using the proposed framework, in particular, the random forest algorithm, produced an F1-score of 49.03 with a narrow standard deviation of 0.67, while the best method of the traditional SL, namely MLP, lagged significantly behind, achieving an F1-score of 37.79 with a larger standard deviation of 2.77. To assess the statistical significance of this performance gap, a paired *t*-test was conducted, revealing a substantial *t*-statistic of 8.8 ( $p < 0.01$ ). This *p*-value of 0.00046 indicates that the observed difference in F1 scores is highly significant, reaffirming the suitability of the proposed PU learning framework over the SL in a statistically robust manner.

In summary, our experiments demonstrated the importance of selecting the appropriate machine learning method and dataset based on the specific characteristics of the problem at hand. For instance, when dealing with a large dataset, a deep neural network might be the most appropriate choice. However, for PU learning, random forest appears to be a robust option, especially given the small dataset size and its class-imbalanced nature.

Table 3 provides a comprehensive overview of the performance metrics for each dataset and method. These results offer valuable insights for researchers and practitioners in the field of PU learning.

**Table 3.** Comparison of performances between the proposed positive-unlabeled-learning-based framework and the SL framework with various data augmentation approaches found in the literature. (SL Datasets: A = original, B = one-hot encoded, C = encoded and augmented as proposed in [4]; positive-unlabeled learning dataset: PU).

| Dataset | Method           | Acc (std) |        | Prec (std) |        | Rec (std) |        | F1 (std) |        | AUC (std) |        |
|---------|------------------|-----------|--------|------------|--------|-----------|--------|----------|--------|-----------|--------|
| A       | Poisson          | 74.93     | (3.20) | 19.14      | (1.67) | 14.51     | (2.08) | 16.51    | (2.54) | 50.94     | (4.12) |
|         | Gaussian Process | 69.14     | (2.81) | 15.27      | (3.02) | 17.74     | (3.22) | 16.41    | (3.14) | 48.73     | (2.67) |
|         | kNN              | 68.31     | (2.85) | 14.66      | (2.33) | 27.74     | (3.18) | 16.05    | (2.63) | 48.23     | (2.98) |
|         | SVM              | 68.04     | (3.12) | 14.28      | (1.98) | 16.10     | (0.89) | 28.00    | (1.06) | 49.80     | (2.77) |
|         | Decision Tree    | 76.03     | (2.72) | 30.15      | (2.09) | 30.64     | (2.78) | 30.41    | (2.63) | 58.01     | (3.12) |
|         | Random Forest    | 80.71     | (1.94) | 40.01      | (2.48) | 25.81     | (2.09) | 31.37    | (2.63) | 58.91     | (1.98) |
|         | Xtra Trees       | 77.96     | (2.47) | 33.33      | (2.73) | 29.03     | (2.28) | 31.03    | (2.67) | 58.53     | (2.42) |
|         | MLP              | 79.61     | (2.47) | 25.00      | (1.83) | 10.67     | (0.75) | 13.96    | (1.01) | 51.84     | (2.11) |
| B       | Poisson          | 70.24     | (0.83) | 14.62      | (1.27) | 14.51     | (0.92) | 14.28    | (1.04) | 48.12     | (1.11) |
|         | Gaussian Process | 79.33     | (1.01) | 21.73      | (0.84) | 18.10     | (0.06) | 19.54    | (1.02) | 51.04     | (0.92) |
|         | kNN              | 71.62     | (1.14) | 16.39      | (0.79) | 16.12     | (0.98) | 16.26    | (1.01) | 49.59     | (0.89) |
|         | SVM              | 82.92     | (0.98) | 50.01      | (1.33) | 30.64     | (0.96) | 28.02    | (0.91) | 61.16     | (1.22) |
|         | Decision Tree    | 73.55     | (1.02) | 20.68      | (0.91) | 19.35     | (0.83) | 19.99    | (0.95) | 52.03     | (0.99) |
|         | Random Forest    | 80.16     | (0.94) | 38.63      | (1.12) | 27.41     | (0.79) | 32.07    | (0.87) | 59.22     | (1.09) |
|         | Xtra Trees       | 81.26     | (1.05) | 43.24      | (0.98) | 25.81     | (0.86) | 32.32    | (1.01) | 59.41     | (1.11) |
|         | MLP              | 28.65     | (0.78) | 18.32      | (0.89) | 91.93     | (1.34) | 30.56    | (0.93) | 53.77     | (1.01) |
| C       | Poisson          | 36.63     | (2.50) | 13.79      | (1.08) | 51.61     | (3.20) | 21.76    | (2.11) | 44.49     | (2.58) |
|         | Gaussian Process | 66.94     | (2.30) | 20.40      | (1.59) | 32.25     | (2.98) | 25.00    | (2.29) | 53.27     | (3.13) |
|         | kNN              | 63.25     | (2.02) | 14.85      | (1.56) | 24.19     | (2.26) | 18.42    | (1.90) | 47.81     | (2.60) |
|         | SVM              | 81.81     | (3.10) | 43.75      | (2.63) | 22.25     | (1.89) | 29.78    | (2.29) | 58.31     | (3.70) |
|         | Decision Tree    | 69.42     | (2.21) | 21.83      | (1.82) | 30.64     | (2.92) | 25.52    | (2.08) | 54.02     | (2.73) |
|         | Random Forest    | 79.33     | (2.74) | 37.73      | (2.31) | 32.25     | (2.25) | 34.78    | (2.42) | 60.64     | (3.14) |
|         | Xtra Trees       | 82.36     | (2.83) | 45.45      | (2.57) | 16.12     | (1.71) | 24.44    | (2.03) | 56.07     | (2.79) |
|         | MLP              | 78.23     | (2.62) | 36.92      | (2.63) | 38.79     | (2.54) | 37.77    | (2.49) | 62.54     | (3.21) |

Table 3. Cont.

| Dataset | Method           | Acc (std) |        | Prec (std) |        | Rec (std) |        | F1 (std) |        | AUC (std) |        |
|---------|------------------|-----------|--------|------------|--------|-----------|--------|----------|--------|-----------|--------|
| PU      | Poisson          | 75.14     | (0.22) | 40.03      | (0.53) | 41.67     | (0.36) | 40.82    | (0.79) | 64.45     | (0.68) |
|         | Gaussian Process | 74.07     | (0.31) | 30.06      | (0.78) | 32.50     | (0.52) | 31.20    | (0.19) | 56.03     | (0.88) |
|         | kNN              | 79.66     | (0.66) | 22.60      | (0.35) | 27.79     | (0.17) | 24.93    | (0.51) | 54.72     | (0.69) |
|         | SVM              | 81.82     | (0.82) | 43.75      | (0.33) | 22.23     | (0.62) | 29.80    | (0.30) | 64.31     | (0.98) |
|         | Decision Tree    | 69.80     | (0.31) | 42.90      | (0.46) | 39.31     | (0.61) | 41.02    | (0.14) | 62.68     | (0.12) |
|         | Random Forest    | 87.84     | (0.24) | 49.45      | (0.43) | 48.61     | (0.28) | 49.03    | (0.67) | 69.20     | (0.39) |
|         | Xtra Trees       | 83.58     | (0.52) | 49.03      | (0.56) | 47.44     | (0.71) | 48.22    | (0.54) | 68.81     | (0.51) |
|         | MLP              | 79.50     | (0.76) | 46.39      | (0.83) | 47.44     | (0.23) | 46.63    | (0.13) | 63.19     | (0.71) |

### 5.1. Comparative Analysis

For the purpose of comparative analysis of black spot identification methods, several approaches have been investigated in order to put the results of this study into perspective. Xu et al. [84] utilized vehicle kinetic parameters obtained from road experiments to address the three-class problem of identifying safe, low-risk, and black spot segments. Their model achieved an impressive accuracy rate of 76.34%. Conversely, Tanprasert et al. [85] employed street-view image data as input for a deep neural architecture and attained a 69.91% accuracy in distinguishing safe from black spot segments, with a notable 75.86% accuracy when specifically querying the black spot class. Fan et al. [45] took a different approach, gathering data from crash accident reports and applying SVM to achieve remarkable precision and recall rates of 88%, along with an F1-score of 88%. However, when focusing solely on the black spot class, their model yielded a slightly lower accuracy of 62%. It is worth noting that the studies by Xu et al. and Fan et al. [45,84] provided comprehensive evaluations of accuracy, precision, and recall, while Tanprasert et al. [85] concentrated solely on accuracy. Additionally, all the aforementioned works faced challenges related to imbalanced datasets, with Xu et al. and Fan et al. [45,84] favoring the non-black spot class, and Tanprasert et al. [85] favoring the black spot class. These findings offer valuable insights into the diverse methods employed for black spot identification in road safety research.

### 5.2. Discussion

The BSNG dataset presents a considerable obstacle for classification algorithms, as demonstrated by the results depicted in Table 3. Given the dataset's imbalanced composition, evaluating performance solely on the basis of correct prediction rates is unsuitable. This section delves into the outcomes derived from diverse ML algorithms in the framework of binary classification endeavors. The metrics assessed encompass accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC).

In Dataset A (original data), decision tree and random forest seem to perform well across most metrics, with high accuracy and F1-score. This indicates that these methods handle the original data effectively. When using one-hot encoded data (Dataset B), support vector machines (SVM) and random forest show strong performance. SVM particularly excels in precision and AUC, indicating its effectiveness in binary classification with one-hot encoded features. Dataset C, which includes encoded and augmented data, presents different performance trends. Random forest and Xtra trees maintain their strong performance, while support vector machines (RBF SVM) stands out in terms of precision and AUC. Finally, The PU dataset differs from the others in that it involves positive and unlabeled samples. In this context, random forest consistently achieves high performance across multiple metrics, including precision, recall, and AUC.

The situation wherein a binary classifier exhibits high accuracy but a low F1-score within an imbalanced dataset implies a noteworthy performance characteristic. However, this apparent success is nuanced when the dataset exhibits a substantial class imbalance. The low F1-score, which harmonizes precision and recall, reflects the model's inability to effectively manage the imbalanced nature of the data. Specifically, it signifies the model's challenge in accurately classifying instances from the minority class. This performance

discrepancy between accuracy and F1-score underscores the model's inclination to favor the majority class in its predictions.

Moreover, it is frequently observed that a binary classifier attains elevated accuracy while exhibiting a diminished F1-score. This phenomenon typically arises from the classifier's strong bias toward the majority class, resulting in a substantial count of true negatives (TN) and true positives (TP), which collectively bolster the accuracy metric. Nevertheless, this pronounced emphasis on the majority class often leads to sub-optimal recall for the minority class, thereby causing a notable reduction in the F1-score.

Poisson and Gaussian process models generally yielded moderate results, with PU learning showing promise for addressing imbalanced datasets. The k-nearest neighbor (kNN) displayed mixed performance, with PU learning consistently excelling in precision and recall, making it a valuable choice for handling imbalanced data. Decision tree demonstrated balanced performance, while extra randomized trees exhibited competitive accuracy. Conversely, random forest showcased a balanced trade-off between precision and recall, with PU learning emerging as a robust option. Multilayered perceptron displayed variable results, emphasizing the importance of selecting the appropriate method based on dataset characteristics and specific task requirements.

From the perspective of the ML approaches, traditional SL often exhibits moderate to good accuracy but struggles with imbalanced datasets, resulting in lower recall rates. One-hot encoded SL, while improving precision, can suffer from lower recall, particularly when positive instances are sparse. In contrast, encoded and augmented methods are designed to address imbalanced datasets, consistently achieving improved recall while maintaining competitive precision. PU learning emerges as a robust approach, consistently delivering high precision, recall, and F1-score, making it a compelling choice, especially for tasks involving imbalanced or partially labeled data.

Based on the evidence presented in the performance metrics of the ML methods, there is a compelling argument to suggest that positive-unlabeled (PU) learning might be more suitable than traditional SL for the specific classification tasks evaluated in the study. The results of Table 3 clearly demonstrate that PU learning outperformed traditional SL methods in several key metrics. Across multiple ML algorithms, PU learning consistently showcased higher precision, recall, and F1-score. The aforementioned are reflected also in are presented also in Figure 2, which each sub-figure is a ML method (from top-left to bottom-right: Poisson, Gaussian Process, K-NN, SVM, Decision Tree, Extra Randomised Trees, Random Forest and MLP) and the columns within a sub-figure are classification performance metrics as described before. This superiority in performance can be attributed to PU learning's ability to handle imbalanced datasets effectively, a common challenge in real-world applications.

Traditional SL methods struggled with imbalanced data, often showing a trade-off between precision and recall. In contrast, PU learning demonstrated a more balanced performance, excelling in precision and recall simultaneously. This characteristic is of paramount importance in scenarios where correctly identifying positive instances while minimizing false positives is critical.

Furthermore, the PU approach stands in stark contrast to traditional supervised learning (SL), which adheres to the dogma that 'the more data, the better the model'. Notably, our experiments reveal that PU learning's performance is on par with that of SL, despite utilizing only a fraction of the available data, as we have previously demonstrated in our research [6].

In summary, the choice of the learning framework and method greatly impacts the performance of the model. In PU learning, ensemble methods like random forest and Xtra trees show promise. However, the choice should be made considering the specific dataset characteristics and problem requirements.



**Figure 2.** Performance comparison of eight classification methods used in different learning frameworks.

## 6. Conclusions

The shift from SL to PU learning represents a paradigm change in model training. Treating a portion of the data as unlabeled, PU learning accounts for the uncertainty inherent in real-world data, which is often partially labeled. This change in approach framework allowed the models to focus on identifying true positive instances without being constrained by the limitations of traditional SL methods.

In conclusion, the evidence presented in this study strongly supports the argument that PU learning is more suitable than SL for the given classification tasks. The observed improvements in precision, recall, and F1-score demonstrate the potential benefits of adopting a PU learning approach, particularly when dealing with imbalanced datasets.

Changing the approach framework from SL to PU not only yielded better results but also highlights the adaptability of ML techniques to real-world data challenges, making PU learning a valuable tool for various applications where accurate positive instance identification is paramount.

Future research could build upon the findings of this study to further enhance the performance of the weakly supervised framework presented. A promising avenue involves developing tailored augmentation techniques, facilitating the use of deep neural networks (DNNs). Given the current size of the BSNG dataset, which limits the applicability of DNNs, such augmentation strategies could prove pivotal in overcoming these constraints and unlocking new potentials in the domain of black spot identification.

**Author Contributions:** Conceptualization, V.S.; methodology, V.S. and G.P.; software, V.S.; validation, G.P., S.G.M. and A.G.; formal analysis, V.S.; investigation, V.S., G.P., S.G.M. and A.G.; resources, G.P.; data curation, V.S. and G.P.; writing—original draft preparation, V.S.; writing—review and editing, G.P. and A.G.; visualization, V.S., G.P., S.G.M. and A.G.; supervision, A.G.; project administration, G.P. and A.G.; funding acquisition, G.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were used in this study. These data can be found here: <https://github.com/iokarama/BSNG-dataset> (accessed on 1 December 2023).

**Acknowledgments:** The authors extend their sincere gratitude to the reviewers whose insightful comments and constructive critiques significantly contributed to enhancing the quality and clarity of this manuscript. To improve the readability of the English language in narrative sections throughout the paper we have used chatGPT (<https://chat.openai.com>).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|      |                                     |
|------|-------------------------------------|
| ML   | Machine Learning                    |
| SL   | Supervised Learning                 |
| PU   | Positive-Unlabeled                  |
| BSNG | Black Spots of North Greece         |
| GIS  | Geographical Information Systems    |
| AUC  | Area Under the Curve                |
| ROC  | Receiver Operational Characteristic |
| SVM  | Support Vector Machine              |
| RBF  | Radial Basis Function               |
| DNN  | Deep Neural Network                 |
| ReLU | Rectified Linear Unit               |

## References

1. Debrabant, B.; Halekoh, U.; Bonat, W.H.; Hansen, D.L.; Hjelmberg, J.; Lauritsen, J. Identifying traffic accident black spots with Poisson–Tweedie models. *Accid. Anal. Prev.* **2018**, *111*, 147–154. [\[CrossRef\]](#)
2. Elvik, R. *State-of-the-Art Approaches to Road Accident Black Spot Management and Safety Analysis of Road Networks*; Transportøkonomisk Institutt: Oslo, Norway, 2007.
3. Tiwari, M.; Nagar, P.; Arya, G.; Chauhan, S.S. Road Accident Analysis Using ML Classification Algorithms and Plotting Black Spot Areas on Map. In Proceedings of the International Conference on Micro-Electronics and Telecommunication Engineering, Ghaziabad, India, 24–25 September 2021; pp. 685–701.
4. Karamanlis, I.; Kokkalis, A.; Profillidis, V.; Botzoris, G.; Kiourt, C.; Sevetlidis, V.; Pavlidis, G. Deep Learning-Based Black Spot Identification on Greek Road Networks. *Data* **2023**, *8*, 110. [\[CrossRef\]](#)
5. Božič, J.; Tabernik, D.; Skočaj, D. Mixed supervision for surface-defect detection: From weakly to fully supervised learning. *Comput. Ind.* **2021**, *129*, 103459. [\[CrossRef\]](#)



6. Sevetlidis, V.; Pavlidis, G.; Balaska, V.; Psomoulis, A.; Mouroutsos, S.; Gasteratos, A. Defect detection using weakly supervised learning. In Proceedings of the 2023 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings, Copenhagen, Denmark, 17–19 October 2023.
7. Bekker, J.; Davis, J. Learning from positive and unlabeled data: A survey. *Mach. Learn.* **2020**, *109*, 719–760. [[CrossRef](#)]
8. Elvik, R. Evaluations of road accident blackspot treatment: A case of the iron law of evaluation studies? *Accid. Anal. Prev.* **1997**, *29*, 191–199. [[CrossRef](#)] [[PubMed](#)]
9. Alsop, J.; Langley, J. Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accid. Anal. Prev.* **2001**, *33*, 353–359. [[CrossRef](#)] [[PubMed](#)]
10. Newstead, S.V.; Corben, B.F. *Evaluation of the 1992–1996 Transport Accident Commission Funded Accident Black Spot Treatment Program in Victoria*; Monash University Press: Clayton, VIC, Australia, 2001; Number 182.
11. Robinson, D. Changes in head injury with the New Zealand bicycle helmet law. *Accid. Anal. Prev.* **2001**, *33*, 687–691. [[CrossRef](#)] [[PubMed](#)]
12. Oppe, S. *Detection and Analysis of Black Spots with Even Small Accident Figures*; Institute for Road Safety Research SWOV: Amsterdam, The Netherlands, 1982.
13. Dereli, M.A.; Erdogan, S. A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. *Transp. Res. Part A Policy Pract.* **2017**, *103*, 106–117. [[CrossRef](#)]
14. Budzyński, M.; Kustra, W.; Okraszewska, R.; Jamroz, K.; Pyrchla, J. The use of GIS tools for road infrastructure safety management. *E3S Web Conf.* **2018**, *26*, 00009. [[CrossRef](#)]
15. Chang, K.T. *Introduction to Geographic Information Systems*; Mcgraw-Hill Boston: Boston, MA, USA, 2008; Volume 4.
16. Silva, P.B.; Andrade, M.; Ferreira, S. Machine learning applied to road safety modeling: A systematic literature review. *J. Traffic Transp. Eng. (Engl. Ed.)* **2020**, *7*, 775–790. [[CrossRef](#)]
17. Miaou, S.P.; Hu, P.S.; Wright, T.; Rathi, A.K.; Davis, S.C. Relationship between truck accidents and highway geometric design: A Poisson regression approach. *Transp. Res. Rec.* **1994**, *26*, 471–482.
18. Sagamiko, T.; Mbare, N. Modelling road traffic accidents counts in Tanzania: A poisson regression approach. *Tanzan. J. Sci.* **2021**, *47*, 308–314.
19. Abdel-Aty, M.A.; Radwan, A.E. Modeling traffic accident occurrence and involvement. *Accid. Anal. Prev.* **2000**, *32*, 633–642. [[CrossRef](#)] [[PubMed](#)]
20. Chin, H.C.; Quddus, M.A. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accid. Anal. Prev.* **2003**, *35*, 253–259. [[CrossRef](#)] [[PubMed](#)]
21. Ma, J.; Kockelman, K.M.; Damien, P. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accid. Anal. Prev.* **2008**, *40*, 964–975. [[CrossRef](#)] [[PubMed](#)]
22. Lord, D.; Washington, S.P.; Ivan, J.N. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accid. Anal. Prev.* **2005**, *37*, 35–46. [[CrossRef](#)] [[PubMed](#)]
23. Lord, D.; Persaud, B.N. Accident prediction models with and without trend: Application of the generalized estimating equations procedure. *Transp. Res. Rec.* **2000**, *1717*, 102–108. [[CrossRef](#)]
24. Caliendo, C.; Guida, M.; Parisi, A. A crash-prediction model for multilane roads. *Accid. Anal. Prev.* **2007**, *39*, 657–670. [[CrossRef](#)] [[PubMed](#)]
25. Lord, D.; Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 291–305. [[CrossRef](#)]
26. Savolainen, P.T.; Mannering, F.L.; Lord, D.; Quddus, M.A. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accid. Anal. Prev.* **2011**, *43*, 1666–1676. [[CrossRef](#)]
27. Szénási, S.; Jankó, D. A method to identify black spot candidates in built-up areas. *J. Transp. Saf. Secur.* **2017**, *9*, 20–44. [[CrossRef](#)]
28. Aghajani, M.A.; Dezfoulian, R.S.; Arjroody, A.R.; Rezaei, M. Applying GIS to identify the spatial and temporal patterns of road accidents using spatial statistics (case study: Ilam Province, Iran). *Transp. Res. Procedia* **2017**, *25*, 2126–2138. [[CrossRef](#)]
29. Erdogan, S.; İlçi, V.; Soysal, O.M.; Kormaz, A. A model suggestion for the determination of the traffic accident hotspots on the Turkish highway road network: A pilot study. *Bol. De Ciências Geodésicas* **2015**, *21*, 169–188. [[CrossRef](#)]
30. Zhu, H.; Zhou, Y.; Chen, Y. Identification of potential traffic accident hot spots based on accident data and GIS. *MATEC Web Conf.* **2020**, *325*, 01005. [[CrossRef](#)]
31. Turki, Z.; Ghédira, A.; Ouni, F.; Kahloul, A. Spatio-temporal analysis of road traffic accidents in Tunisia. In Proceedings of the 2022 14th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA), El Jadida, Morocco, 25–27 May 2022; pp. 1–7.
32. Al-Omari, A.; Shatnawi, N.; Khedaywi, T.; Miqdady, T. Prediction of traffic accidents hot spots using fuzzy logic and GIS. *Appl. Geomat.* **2020**, *12*, 149–161. [[CrossRef](#)]
33. Gundogdu, I.B. Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya. *Saf. Sci.* **2010**, *48*, 763–769. [[CrossRef](#)]
34. Shafabakhsh, G.A.; Famili, A.; Bahadori, M.S. GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. *J. Traffic Transp. Eng. (Engl. Ed.)* **2017**, *4*, 290–299. [[CrossRef](#)]
35. Lasisi, A.; Li, P.; Chen, J. Hybrid Machine Learning and Geographic Information Systems Approach—A Case for Grade Crossing Crash Data Analysis. *Adv. Data Sci. Adapt. Anal.* **2020**, *12*, 2050003. [[CrossRef](#)]

36. Azmi, N.N.; Sarif, A.S. The Development of a GIS database for blackspot area in Federal Route 24. *Prog. Eng. Appl. Technol.* **2023**, *4*, 949–955.
37. Thakare, K.; Shete, B.; Bijwe, A. A Review on the Study of Different Black Spot Identification Methods. *Int. Res. J. Eng. Technol.* **2021**, *9*, 1758–1763.
38. Chen, H. Black spot determination of traffic accident locations and its spatial association characteristic analysis based on GIS. *J. Geogr. Inf. Syst.* **2012**, *4*, 608–617. [[CrossRef](#)]
39. Goodchild, M.F.; Steyaert, L.T.; Parks, B.O.; Johnston, C.; Maidment, D.; Crane, M.; Glendinning, S. *GIS and Environmental Modeling: Progress and Research Issues*; John Wiley and Sons Inc.: Canada, 1996.
40. Iqbal, A.; Rehman, Z.U.; Ali, S.; Ullah, K.; Ghani, U. Road traffic accident analysis and identification of black spot locations on highway. *Civ. Eng. J.* **2020**, *6*, 2448–2456. [[CrossRef](#)]
41. Karamanlis, I.; Kokkalis, A.; Profillidis, V.; Botzoris, G.; Galanis, A. Identifying Road Accident Black Spots using Classical and Modern Approaches. *WSEAS Trans. Syst.* **2023**, *22*, 556–565. [[CrossRef](#)]
42. Karamanlis, I.; Nikiforiadis, A.; Botzoris, G.; Kokkalis, A.; Basbas, S. Towards Sustainable Transportation: The Role of Black Spot Analysis in Improving Road Safety. *Sustainability* **2023**, *15*, 14478. [[CrossRef](#)]
43. Fiorentini, N.; Losa, M. Handling imbalanced data in road crash severity prediction by machine learning algorithms. *Infrastructures* **2020**, *5*, 61. [[CrossRef](#)]
44. Theofilatos, A.; Chen, C.; Antoniou, C. Comparing machine learning and deep learning methods for real-time crash prediction. *Transp. Res. Rec.* **2019**, *2673*, 169–178. [[CrossRef](#)]
45. Fan, Z.; Liu, C.; Cai, D.; Yue, S. Research on black spot identification of safety in urban traffic accidents based on machine learning method. *Saf. Sci.* **2019**, *118*, 607–616. [[CrossRef](#)]
46. Mbarek, A.; Jiber, M.; Yahyaouy, A.; Sabri, A. Black spots identification on rural roads based on extreme learning machine. *Int. J. Electr. Comput. Eng.* **2023**, *13*, 3149–3160. [[CrossRef](#)]
47. Vasconcelos, S.P.; de Souza Baptista, C.; de Figueirêdo, H.F. Using a Social Network for Road Accidents Detection, Geolocation and Notification—A Machine Learning Approach. In Proceedings of the Fifteenth International Conference on Advanced Geographic Information Systems, Applications, and Services, Venice, Italy, 24–28 April 2023.
48. Kwok-Fai Lui, A.; Chan, Y.H.; Lo, K.H.; Cheng, W.T.; Cheung, H.T. Predictive Screening of Accident Black Spots based on Deep Neural Models of Road Networks and Facilities: A Case Study based on a District in Hong Kong. In Proceedings of the 2021 5th International Conference on Computer Science and Artificial Intelligence, Beijing, China, 4–6 December 2021; pp. 422–428.
49. Paul, A.K.; Boni, P.K.; Islam, M.Z. A Data-Driven Study to Investigate the Causes of Severity of Road Accidents. In Proceedings of the 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Virtual, 3–5 October 2022; pp. 1–7.
50. Abdullah, P.; Sipos, T. Exploring the Factors Influencing Traffic Accidents: An Analysis of Black Spots and Decision Tree for Injury Severity. *Period. Polytech. Transp. Eng.* **2024**, *52*, 33–39. [[CrossRef](#)]
51. Amorim, B.d.S.P.; Firmino, A.A.; Baptista, C.d.S.; Júnior, G.B.; Paiva, A.C.d.; Júnior, F.E.d.A. A Machine Learning Approach for Classifying Road Accident Hotspots. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 227. [[CrossRef](#)]
52. Sobhana, M.; Rohith, V.K.; Avinash, T.; Malathi, N. A Hybrid Machine Learning Approach for Performing Predictive Analytics on Road Accidents. In Proceedings of the 2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 21–23 December 2022; pp. 1–6.
53. Al-Mistarehi, B.; Alomari, A.H.; Imam, R.; Mashaqba, M. Using machine learning models to forecast severity level of traffic crashes by R Studio and ArcGIS. *Front. Built Environ.* **2022**, *8*, 860805. [[CrossRef](#)]
54. Megnidio-Tchoukouegno, M.; Adedeji, J.A. Machine learning for road traffic accident improvement and environmental resource management in the transportation sector. *Sustainability* **2023**, *15*, 2014. [[CrossRef](#)]
55. Wang, Y.; Zhai, H.; Cao, X.; Geng, X. Cause Analysis and Accident Classification of Road Traffic Accidents Based on Complex Networks. *Appl. Sci.* **2023**, *13*, 12963. [[CrossRef](#)]
56. Khattak, A.; Almujibah, H.; Elamary, A.; Matara, C.M. Interpretable Dynamic Ensemble Selection Approach for the Prediction of Road Traffic Injury Severity: A Case Study of Pakistan's National Highway N-5. *Sustainability* **2022**, *14*, 12340. [[CrossRef](#)]
57. Kumeda, B.; Zhang, F.; Zhou, F.; Hussain, S.; Almasri, A.; Assefa, M. Classification of road traffic accident data using machine learning algorithms. In Proceedings of the 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN), Chongqing, China, 12–15 June 2019; pp. 682–687.
58. Manzoor, M.; Umer, M.; Sadiq, S.; Ishaq, A.; Ullah, S.; Madni, H.A.; Bisogni, C. RFCNN: Traffic accident severity prediction based on decision level fusion of machine and deep learning model. *IEEE Access* **2021**, *9*, 128359–128371. [[CrossRef](#)]
59. Kaur, G.; Kaur, H. Black Spot and Accidental Attributes Identification on State Highways and Ordinary District Roads Using Data Mining Techniques. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*, 2312.
60. Balakrishnan, S.; Karuppanagounder, K. Accident Blackspot Ranking: An Alternative Approach in the Presence of Limited Data. In *Recent Advances in Transportation Systems Engineering and Management: Select Proceedings of CTSEM 2021*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 721–735.
61. Sevtilidis, V.; Giuffrida, M.V.; Tsafaris, S.A. Whole image synthesis using a deep encoder-decoder network. In Proceedings of the Simulation and Synthesis in Medical Imaging: First International Workshop, SASHIMI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, 21 October 2016; pp. 127–137.

62. Pavlidis, G.; Mouroutsos, S.; Sevetlidis, V. Efficient colour sorting of Chios mastiha. In Proceedings of the 2014 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings, Santorini, Greece, 14–17 October 2014; pp. 386–391.
63. Kritsis, K.; Kiourt, C.; Stamouli, S.; Sevetlidis, V.; Solomou, A.; Karetos, G.; Katsouros, V.; Pavlidis, G. GRASP-125: A Dataset for Greek Vascular Plant Recognition in Natural Environment. *Sustainability* **2021**, *13*, 11865. [[CrossRef](#)]
64. Sevetlidis, V.; Pavlidis, G.; Arampatzakis, V.; Kiourt, C.; Mouroutsos, S.G.; Gasteratos, A. Web acquired image datasets need curation: An exemplar pipeline evaluated on Greek food images. In Proceedings of the 2021 IEEE International Conference on Imaging Systems and Techniques (IST), Kaohsiung, Taiwan, 24–26 August 2021; pp. 1–6.
65. Sevetlidis, V.; Pavlidis, G.; Mouroutsos, S.; Gasteratos, A. Tackling dataset bias with an automated collection of real-world samples. *IEEE Access* **2022**, *10*, 126832–126844. [[CrossRef](#)]
66. Pavlidis, G.; Solomou, A.; Stamouli, S.; Papavassiliou, V.; Kritsis, K.; Kiourt, C.; Sevetlidis, V.; Karetos, G.; Trigas, P.; Kougioumoutzis, K.; et al. Sustainable ecotourism through cutting-edge technologies. *Sustainability* **2022**, *14*, 800. [[CrossRef](#)]
67. Sevetlidis, V.; Pavlidis, G. Effective Raman spectra identification with tree-based methods. *J. Cult. Herit.* **2019**, *37*, 121–128. [[CrossRef](#)]
68. Sevetlidis, V.; Pavlidis, G. Hierarchical Classification For Improved Compound Identification In Raman Spectroscopy. In Proceedings of the 3rd Computer Applications and Quantitative Methods in Archaeology (CAA-GR) Conference, Limassol, Cyprus, 18–20 June 2018; p. 133.
69. Zhou, Z.H. A brief introduction to weakly supervised learning. *Natl. Sci. Rev.* **2018**, *5*, 44–53. [[CrossRef](#)]
70. Foulds, J.; Frank, E. A review of multi-instance learning assumptions. *Knowl. Eng. Rev.* **2010**, *25*, 1–25. [[CrossRef](#)]
71. Wang, W.; Zhou, Z.H. A New Analysis of Co-Training. *ICML* **2010**, *2*, 3.
72. Letouzey, F.; Denis, F.; Gilleron, R. Learning from positive and unlabeled examples. In Proceedings of the International Conference on Algorithmic Learning Theory, Sydney, NSW, Australia, 11–13 December 2000; pp. 71–85.
73. Zhang, D.; Han, J.; Cheng, G.; Yang, M.H. Weakly supervised object localization and detection: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5866–5885. [[CrossRef](#)]
74. Xu, L.; Lv, S.; Deng, Y.; Li, X. A weakly supervised surface defect detection based on convolutional neural network. *IEEE Access* **2020**, *8*, 42285–42296. [[CrossRef](#)]
75. Chatterjee, S.; Saeedfar, P.; Tofangchi, S.; Kolbe, L.M. Intelligent Road Maintenance: A Machine Learning Approach for surface Defect Detection. In Proceedings of the Twenty-Sixth European Conference on Information Systems (ECIS2018), Portsmouth, UK, 23–28 June 2018; p. 194.
76. Fernandes, A.M.d.R.; Cassaniga, M.J.; Passos, B.T.; Comunello, E.; Stefenon, S.F.; Leithardt, V.R.Q. Detection and classification of cracks and potholes in road images using texture descriptors. *J. Intell. Fuzzy Syst.* **2023**, *44*, 10255–10274. [[CrossRef](#)]
77. Boucetta, Z.; Fazziki, A.E.; Adnani, M.E. A Deep-Learning-Based Road Deterioration Notification and Road Condition Monitoring Framework. *Int. J. Intell. Eng. Syst.* **2021**, *14*, 503–515. [[CrossRef](#)]
78. Basavaraju, A.; Du, J.; Zhou, F.; Ji, J. A machine learning approach to road surface anomaly assessment using smartphone sensors. *Ieee Sens. J.* **2019**, *20*, 2635–2647. [[CrossRef](#)]
79. Zhang, Y.; Li, T.; Li, C.; Zhou, X. A Novel Driver Distraction Behavior Detection Method Based on Self-Supervised Learning with Masked Image Modeling. *IEEE Internet Things J.* **2023**, *11*, 6056–6071. [[CrossRef](#)]
80. Xiao, Y.; Liu, B.; Yin, J.; Cao, L.; Zhang, C.; Hao, Z. Similarity-based approach for positive and unlabeled learning. In Proceedings of the IJCAI Proceedings-International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, 16–22 July 2011; Volume 22, p. 1577.
81. Nam, J.; Kim, S. Clami: Defect prediction on unlabeled datasets (t). In Proceedings of the 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE), Lincoln, NE, USA, 9–13 November 2015; pp. 452–463.
82. Hariri, S.; Kind, M.C.; Brunner, R.J. Extended isolation forest. *IEEE Trans. Knowl. Data Eng.* **2019**, *33*, 1479–1489. [[CrossRef](#)]
83. Tsintotas, K.A.; Sevetlidis, V.; Papapetros, I.T.; Balaska, V.; Psomoulis, A.; Gasteratos, A. BK tree indexing for active vision-based loop-closure detection in autonomous navigation. In Proceedings of the 2022 30th Mediterranean Conference on Control and Automation (MED), Athens, Greece, 28 June–1 July 2022; pp. 532–537.
84. Xu, Y.; Zhang, C.; He, J.; Liu, Z.; Chen, Y.; Zhang, H. Comparisons on methods for identifying accident black spots using vehicle kinetic parameters collected from road experiments. *J. Traffic Transp. Eng. (Engl. Ed.)* **2023**, *10*, 659–674. [[CrossRef](#)]
85. Tanprasert, T.; Siripanpornchana, C.; Surasvadi, N.; Thajchayapong, S. Recognizing traffic black spots from street view images using environment-aware image processing and neural network. *IEEE Access* **2020**, *8*, 121469–121478. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.