sustainability

MDPI

Article

Evaluating Road Hazard Maintenance Efficiency Using Citizen Science Data to Improve Road Safety

Jinguk Kim [1], Woohoon Jeon [1] and Seoungbum Kim [2,*]

[1] Department of Highway & Transportation Research, Korea Institute of Civil Engineering and Building Technology, 283, Goyang-daero, Ilsanseo-gu, Goyang-si 10223, Republic of Korea; jingukkim@kict.re.kr (J.K.); cwhoon@kict.re.kr (W.J.)

[2] Department of Urban Engineering, Engineering Research Institute, Gyeongsang National University, 501, Jinju-daero, Jinju-si 52828, Republic of Korea

* Correspondence: kimsb@gnu.ac.kr; Tel.: +82-55-772-1778

Abstract: Accidents caused by road hazards can be prevented through regular inspections by road management agencies. To this end, traffic agencies allocate substantial budgets and workforces to maintain the performance of roads. Additionally, traffic agencies require comprehensive data such as the classifications and sizes of road hazards. However, collecting spatial–temporal data on various road hazards is challenging, and evaluating it comprehensively with respect to work efficiency and budget allocation is difficult due to stakeholder interests across agencies. This study proposes a process of evaluating operational efficiency in terms of maintaining roads and preventing hazards by analyzing citizen scientist-based data. First, we collected data from drivers through a mobile application and applied text mining techniques to classify each complaint into several types of road hazard maintenance. Second, we developed an indicator to measure operational efficiency using the processed data and evaluated each regional agency per each type of maintenance. The results of this study provide evidence that specific types of road hazards occur prominently under specific agencies. In addition, the time required to provide maintenance for identical road hazards can vary among agencies. These results suggest that the maintenance budget for the entire national highway may need to be distributed differently based upon regional characteristics.

Keywords: citizen science; road hazard; maintenance efficiency; road safety; text mining

check for updates

Citation: Kim, J.; Jeon, W.; Kim, S. Evaluating Road Hazard Maintenance Efficiency Using Citizen Science Data to Improve Road Safety. *Sustainability* **2023**, *15*, 15074. https://doi.org/10.3390/su152015074

Academic Editor: Rosolino Vaiana

Received: 24 August 2023
Revised: 25 September 2023
Accepted: 27 September 2023
Published: 19 October 2023

1. Introduction

The length of roadways across the world is consistently increasing as the global economy grows. For example, the highway network in the US has increased by 10,000 miles every year since 1990 (from OECD statistics). However, despite the quantitative increase in roads, traffic accidents also increased significantly up until 2000. To reduce traffic accidents, governments and transportation authorities have been implementing several systematic and ongoing road safety projects, such as improving accident-prone areas and renovating dangerous roads.

Road accidents are caused by vehicle, driver, and road infrastructure malfunctions [1]. In terms of vehicles, the development of new technologies such as autonomous vehicles is expected to greatly contribute to the prevention of traffic accidents [2]. Although many studies on human factors are currently underway, practically applying these findings is expected to require substantial time and costs. Accidents caused by hazardous road environments can be prevented through regular inspections by road management agencies. However, with the continuous extension of roads, their continuous deterioration, and the shortage of road management personnel, substantial progress is unlikely. To prevent accidents through efficient road management, it is particularly important to identify and address road hazards in real time. To this end, traffic agencies allocate substantial budgets and increase the workforce each year to maintain or improve the performance of roads. In addition, traffic agencies require comprehensive foundational data such as the classifications and sizes of road hazards in each region and the time required to process each hazard.

Collecting spatial–temporal data on road hazards is challenging, and even if it is available, evaluating the data for budget allocation is difficult due to stakeholder interests across agencies. This study proposes a process of evaluating operational efficiency in terms of maintaining roads and preventing hazards by analyzing citizen scientist-based data. First, we collected complaints reported by drivers through a mobile application across an entire national highway in South Korea. Because the data were logged in text format, this study applied text mining techniques to classify each complaint into several types of road hazard maintenance. In addition, the data included the time required to clear each type of road hazard (hereafter called the processing time). Second, based on the information, we developed an indicator to measure operational efficiency using the processing time and evaluated each regional agency from the calculated indicator per each type of maintenance.

The remainder of this paper is organized as follows: Section 2 reviews previous studies related to this research. Section 3 presents the text mining methodology. Section 4 introduces the analyses of the area and data. Then, the empirical results obtained after applying the text mining technique are presented in Section 5. Section 6 presents the discussion and conclusions.

2. Literature Review

Road agencies have an obligation to ensure the safety of the roads under their jurisdiction. Thus, monitoring when and where the maintenance of road infrastructure is needed is a routine task for road agencies and should be part of their operational processes to reduce road safety risks [3]. The first step is to collect road hazard data, which can be achieved in various ways. As information and communications technology (ICT) evolves, numerous studies seek to recognize potential road hazards through various sensors. On the contrary, other studies still rely on manual data collection via either volunteers or citizen scientists. This study reviews recent research in terms of data collection methods for not only the most representative road hazards, such as potholes and roadkill, but also for other minor issues.

Road networks experience significant impacts due to potholes and roadkill on the roads, leading to accidents and fatal injuries [4]. In particular, it is important to detect and repair hazardous factors in a timely and quickly manner to minimize adverse impacts on traffic. Recent studies have used the power of artificial intelligence (AI) to monitor the conditions of road pavements. Machine learning and deep learning methods have been employed to assess the condition of road surfaces through either field trials or case studies.

Briefly reviewing the literature using machine learning, ref. [5] developed a supervised machine learning model that uses image data to detect and classify nine types of crack image data. To develop the model, they utilized a data augmentation technique and achieved an accuracy of 85%. Ref. [6] developed a cloud-assisted road condition-monitoring system that is capable of applying monitoring in real time and can classify road conditions with an accuracy of 88% accuracy. Ref. [7] employed SVM to provide real-time warnings for bumps and potholes, which is also capable of providing instructions to drivers to suddenly accelerate and brake. For road anomalies, the classification accuracy in is approximately 80%. Ref. [8] proposed a new approach for comprehensive pavement condition indicators. The authors argued that the model improves the accuracy when fewer data are available. However, it is expected that considerable effort will be required to collect relevant data in new regions.

There are similar efforts using deep learning methods. Ref. [9] regenerated image data from an RGB-D pavement surface dataset and developed deep convolutional neural networks for pothole detection, which is capable of extracting depth information when depth data are not available. Ref. [10] collected thermal imaging data and applied the CNN approach for pothole detection. Ref. [11] proposed a method that yields reliable pothole detection results under small sample conditions. After data augmentation, they tested the CNN fusion model and the detection accuracy improved up to 90%. Ref. [12] focused on identifying the severity and type of cracks at the same time using the Mobilenet-SSD approach. This study included the assumption that the severity of cracks is directly related to the area of said cracks. Ref. [13] modified a CNN to detect potholes in real time, where they removed some convolution layers and used different dilation rates. Ref. [14] focused on generating pseudo images for a training dataset by combining GAN with Poisson blending artificially. They improved the accuracy of pothole detection by 5% when the original image data were small. Ref. [15] introduced a new approach to obtain labeled training data. After training two mainstream deep learning frameworks (YOLO v2 and R-CNN), they evaluated them using a new dataset extracted via Google API. Ref. [16] developed a location-aware CNN to detect potholes. They argued that the model captures discriminative regions with potholes rather than the global context and as a result outperforms existing methods. Ref. [17] employed a crowd sensing-based deep learning approach to detect potholes. The model is capable of distinguishing potholes from destabilizations of vehicles due to speed bumps or driver behaviors. Ref. [18] employed a deep learning method to detect pothole areas and their depth using a mobile point cloud and images. Ref. [19] employed five different datasets and compared the performance of detecting potholes between 3D scene reconstruction methods and deep learning techniques. Ref. [20] employed the object detection technology of a CNN to identify five different pothole types. Ref. [21] proposed fully automated roadway safety assessment using a deep convolutional neural network. They used a street-level panorama image dataset, and the network is capable of estimate various road-level attributes.

Collisions between wildlife and vehicles pose a potential threat to both wildlife populations and road user safety. Data collection methods for wildlife–vehicle collisions (WVCs) include accident reports by the police [22]; historical data from hunters, citizen scientists, or volunteers [23–28]; sensor-driven data collection such as lidar [29] and smartphone [30]. The study by [4] is unique, given the fact that they utilized a YOLO v3 computer vision algorithm to detect two road hazards (potholes and roadkill) at the same time.

Road attributes such as traffic signs and trees may not be the direct cause of traffic accidents, but they are still crucial maintenance items to improve road safety. AI technology can also be used to evaluate the conditions of road sign integrity [31–34]. For example, ref. [34] used the deep learning method to develop an algorithm that evaluates road sign integrity and conditions. They validated their algorithm using Google images. There are also other efforts [35–39] that use deep learning techniques or video image processing to measure how far roadside objects (e.g., big trees, electric poles, and other roadside vegetation) are from the road boundary. See Table 1 for a summary of these previous studies.

Table 1. Summary of the data collection method for road hazard types.

Literature	Pavement Conditions	Roadkill	Road Attribute Maintenance
[4]	Image data	Image data	
[5]	Image data		
[6]	Smart phone		
[7]	Smart phone		
[8]	Field data		
[9]	Image data		
[10]	Image data		
[11]	Image data		
[12]	Image data		
[13]	Image data		
[14]	Image data		
[15]	Image data		
[16]	Image data		
[17]	Smart phone		
[18]	Image data		
[19]	Image data		

Table 1. *Cont.*

Literature	Pavement Conditions	Roadkill	Road Attribute Maintenance
[20]	Image data		
[21]			Image data
[22]		Accident report	
[23]		Historical data	
[24]		Historical data	
[25]		Historical data	
[26]		Historical data	
[27]		Historical data	
[28]		Historical data	
[29]		LIDAR data	
[30]		Smart phone	
[31]			Image data
[32]			Image data
[33]			Image data
[34]			Image data
[35]			Image data
[36]			Image data
[37]			Image data
[38]			Image data
[39]			Image data

In addition to the road hazards suggested in the review of previous studies, there are numerous others that can affect driver safety. Collecting data associated with various road hazards can be challenging because they are widely distributed in space and may not be timely. To this end, as noted above, research using artificial intelligence techniques has been actively conducted. However, there are many challenges in applying these techniques because of a large dataset for training, a class imbalance issue, and the need to retrain for a new site [40]. More importantly, automatically detecting, identifying, and classifying all road hazards is almost impossible.

This study sought to employ road hazard data to evaluate operational efficiency to clear given road hazards among road management agencies. Because this study dealt with all observable road hazards by users, we analyzed road hazard-related data in text from volunteers and citizen participations rather than in-vehicle sensor-based data.

Of course, road hazard detection using state-of-the-art technologies will continue to develop and is expected to be applied in practice someday. However, if a system to monitor citizen feedback for road hazard maintenance is in place, it has the advantage of being immediately applicable to a wide range of road networks. Another benefit is that such a system can monitor various hazards that may occur on the road. Conversely, data collection can be the biggest weakness in recognizing diverse and widely distributed road hazard factors using the latest technology.

3. Methodology

As a first step, this study needed to interpret road user complaints stored in text format and classify them into different types of road hazards. Thus, the Power Query feature of Excel was used for text mining. Power Query is a tool that imports data from external data sources and then cleanses, transforms, and restructures them as needed. It can also be used for powerful keyword analysis, such as extracting keywords from text-based sentences and sentence separation. To analyze the given text data, this study underwent five steps, as follows: First, we collected and extracted raw text-based data, and then used Power Query to extract one or more keywords from the sentences (step 1 through step 2). After merging all of the extracted keywords, we classified them for analytical purposes (step 3). Next, we searched the original text data for the presence of the classified keywords obtained from step 3 and classified the sentences accordingly (step 4). We applied this process to all sentences and derived analysis results according to the required objectives (step 5).

4. Analysis Area and Data

Highways in South Korea can be broadly divided into national highways managed by local governments and expressways managed by the Korea Expressway Corporation. The expressways are similar to turnpikes in the US and are generally better maintained than national highways. However, the combined total length of the national highways is 9155 km (refer to Table 2), which is more than twice the combined 4036 km length of expressways. Regarding national highways, complaints from users can be reported via a phone, but in reality, road users often do not identify the number of the traffic agency, and it may take over two days for the report to be transferred to the appropriate road agency. Thus, this study focused on national highways, where complaints are relatively frequent and the response times is expected to be somewhat long. Table 2 shows the regional and local agencies managing the national highways in South Korea, the local offices in each region, and the length of the roads they manage.

Table 2. Five national highways maintained by regional and local traffic agencies in 2021.

Regional Office	Length (Meters)	Local Office	Length (Meters)
Office A	985,500	A1	466,800
		A2	518,700
Office B	1,265,584	B1	418,103
		B2	329,738
		B3	517,743
Office C	1,637,317	C1	404,954
		C2	326,439
		C3	525,315
		C4	380,609
Office D	2,854,451	D1	589,874
		D2	420,406
		D3	460,091
		D4	745,880
		D5	638,200
Office E	2,412,463	E1	768,415
		E2	425,395
		E3	604,943
		E4	613,710
	Total		9,155,315

To overcome the limitations of the conventional reporting system for national highways, the Department of Transportation in South Korea developed the Road Inconvenience Reporting System (RIRS), which allows communication with the appropriate road agency. The system has been collecting reports of complaints from any road using GPS technology since 28 March 2014. The RIRS provides a simple and convenient way for road users to report road hazards via a smartphone app, while also allowing road managers to receive location and image information, enabling accurate identification of issues and prompt response to them. Information collected through the RIRS app is stored as historical data, along with details such as report ID, registration time, complaint content, location, agency, processing status, and time taken to process the complaint, as shown in Table 3. This study analyzed a total of 17,738 complaints collected from the RIRS between 2014 and 2022, along with data on complaint processing times.

Table 3. Examples of road complaint reports.

ID	YYYYMMDD HH:MM	Complaint	Regional Office	Status	Processing Time (h)
8574	20140401 8:00	Stop sign replacement	A1	Completed	1
8584	20140401 9:48	Rock on the road	B2	Completed	6
8623	20140402 9:34	Roadkill	C1	Completed	55
8639	20140402 14:27	Uneven pavement	E1	Completed	166

5. Results

5.1. Classification of Road Hazard Types

To systematically analyze the complaints recorded in the RIRS, it was necessary to first extract representative keywords for each complaint. We converted 17,738 reported complaints into text format and used the Power Query feature in Excel to extract a total of 472 keywords. Next, we carefully reviewed these extracted keywords and defined eight representative road hazard types, as shown in Table 4. Complaints that were unclear or difficult to classify into a road hazard type were categorized as "unclassified". In addition, basic statistics such as the number of records and keywords and the percentage of each road hazard type are summarized in the table. Of the total 17,738 complaints, 94% were classified into one of the eight complaint types. Unclassified complaints accounted for only approximately 5%.

Table 4. Major keywords and basic statistics by each road hazard type.

Road Hazard Type	Major Keywords	Number of Records	Number of Keywords	%
Facilities	Facilities, Signage, Manholes, Guardrails, Structures, Reflectors, etc.	4797	164	27%
Roads	Roads, Accident, Sidewalk, Gradient, Lane Mark, Road Marking, etc.	3323	89	19%
Road Hazards	Obstructions, Soil, Falling Rocks, Tires, Plastic, Dirt, Boxes, Cleaning, etc.	3119	123	18%
Roadkill	Animal Carcasses, Deer, Boars, Animals, Cats, Badgers, Pigeons, etc.	1799	19	10%
Potholes	Potholes, Subsidence, Sinkholes, Ditches, Holes, Hollows, etc.	1649	13	9%
Illegal	Illegal, Advertisements, Parking, Placards, Vendors, Illegal Signs, etc.	945	17	5%
Drainage	Drainage, Drains, Canals, Rainwater, Dikes, Puddles, Flooding, etc.	811	28	5%
Weeding	Weeding, Weeds, Twigs, Bushes, Trees, Plants, Roadside Trees, etc.	279	19	2%
Unclassified	-	1016	NaN	5%
Total		17,738	472	100%

NaN: Not a Number.

5.2. Processing Time by Agency and Road Hazard Type

To understand what types of road hazard maintenance requests are received by each road agency, the number of complaint reports for each type was tabulated by regional and local office, as shown in Table 5.

In Table 5, Office D received 5423 (32%) of the total records. At the local office level, the location with the highest number of records was A2, receiving 2630 (15%). The regional office with the lowest records was Region C, with 1915 (11%), and at the local agency level, E3, with 194 (1.1%).

Focusing on the type of road hazards, "Road Signs" was the highest at 29%, while "Weeding" was the least common. There was a substantial variance in the number of complaint records across local offices. For example, in the case of "Roadkill", D2 received a relatively high proportion of complaints. On the contrary, it is noteworthy that many complaints were reported in E1 and B3 for "Potholes". These examples reveal that specific complaints are concentrated for specific agencies due to their environmental factors. As shown in Table 5, the road hazard types that showed the most prominent variation in the number of reported complaints by local office were "Roadkill", "Road Hazards", "Potholes", and "Illegal Ads". These results suggest that financial resources to resolve specific road hazards should be allocated based upon the circumstances of each agency.

Table 5. Number of complaint reports by agency and type.

Office		Roads	Roadkill	Drainage	Illegal Ads	Road Signs	Road Hazards	Weeding	Potholes	Sum 1	Sum 2
Office A	A1	424	64	92	20	281	64	4	126	1075	3705
	A2	538	120	172	175	632	656	71	266	2630	
Office B	B1	59	8	22	11	172	26	6	16	320	2102
	B2	261	25	44	3	211	128	25	145	842	
	B3	257	39	51	26	305	71	8	183	940	
Office C	C1	79	46	17	201	133	71	19	26	592	1915
	C2	131	32	42	33	253	65	5	51	612	
	C3	107	49	15	8	121	38	4	37	379	
	C4	60	84	11	3	90	28	3	53	332	
Office D	D1	230	44	23	75	494	115	10	34	1025	5423
	D2	145	590	50	43	289	895	19	121	2152	
	D3	210	66	34	29	303	151	14	119	926	
	D4	201	38	8	11	376	88	3	29	754	
	D5	99	97	18	15	126	161	10	40	566	
Office E	E1	143	206	60	30	312	106	12	284	1153	3566
	E2	154	107	82	20	436	244	30	54	1127	
	E3	30	14	10	2	76	30	6	26	194	
	E4	195	165	60	239	184	181	30	38	1092	
Column Sum		3323	1794	811	944	4794	3118	279	1648	16,711	

The time required to process road hazards may vary with the capabilities of each agency, and the time required to clear them directly impacts the satisfaction of road users. As shown in Table 3, the analysis data included not only the record of complaints, but also the time taken to resolve them. In this study, we further analyzed the processing time in Table 3 to evaluate and compare work efficiency across agencies. Table 6 summarizes the total processing time taken by each agency and road hazard type.

Table 6. Complaint processing time by office and complaint type.

Office		Roads	Roadkill	Drainage	Illegal Ads	Road Signs	Road Hazards	Weeding	Potholes	Sum 1	Sum 2
Office A	A1	35 K	6 K	6 K	1 K	19 K	7 K	1 K	12 K	87 K	526 K
	A2	86 K	12 K	22 K	34 K	94 K	124 K	13 K	53 K	439 K	
Office B	B1	7 K	1 K	3 K	1 K	29 K	3 K	1 K	19 K	62 K	178 K
	B2	17 K	2 K	4 K	1 K	12 K	8 K	1 K	9 K	54 K	
	B3	13 K	2 K	2 K	6 K	30 K	4 K	1 K	4 K	62 K	
Office C	C1	11 K	1 K	3 K	9 K	19 K	4 K	1 K	2 K	50 K	280 K
	C2	9 K	1 K	2 K	4 K	17 K	4 K	1 K	3 K	41 K	
	C3	15 K	12 K	1 K	1 K	23 K	2 K	1 K	1 K	56 K	
	C4	106 K	3 K	1 K	1 K	11 K	4 K	1 K	7 K	133 K	

Table 6. *Cont.*

Office		Roads	Roadkill	Drainage	Illegal Ads	Road Signs	Road Hazards	Weeding	Potholes	Sum 1	Sum 2
Office D	D1	9 K	2 K	1 K	2 K	41 K	2 K	1 K	1 K	58 K	445 K
	D2	24 K	59 K	1 K	1 K	24 K	59 K	1 K	10 K	179 K	
	D3	11 K	3 K	2 K	1 K	115 K	6 K	1 K	7 K	146 K	
	D4	7 K	1 K	K	1 K	12 K	3 K	1 K	1 K	24 K	
	D5	6 K	12 K	1 K	1 K	8 K	7 K	1 K	3 K	38 K	
Office E	E1	2 K	3 K	6 K	1 K	12 K	1 K	1 K	8 K	32 K	534 K
	E2	42 K	25 K	18 K	21 K	137 K	72 K	6 K	13 K	334 K	
	E3	3 K	1 K	1 K	1 K	6 K	2 K	1 K	4 K	18 K	
	E4	107 K	4 K	3 K	6 K	7 K	21 K	1 K	2 K	150 K	
Column Sum		511 K	150 K	77 K	87 K	616 K	333 K	29 K	160 K	1962 K	

In addition, Figure 1 presents a box plot of the processing time by local office per event (i.e., Table 6 divided by Table 5) for a given road hazard type. In each box plot, the top and bottom edge of a box show the first and third quartiles, while the horizontal line within the box shows the median value of the processing time for a given type. Above and below each box, the "T"-shaped whiskers extend to the furthest point within 150% of the interquartile range to bound the range of the data. Following standard conventions, all of the points outside of this range are considered to be outliers and are indicated with plus symbols.

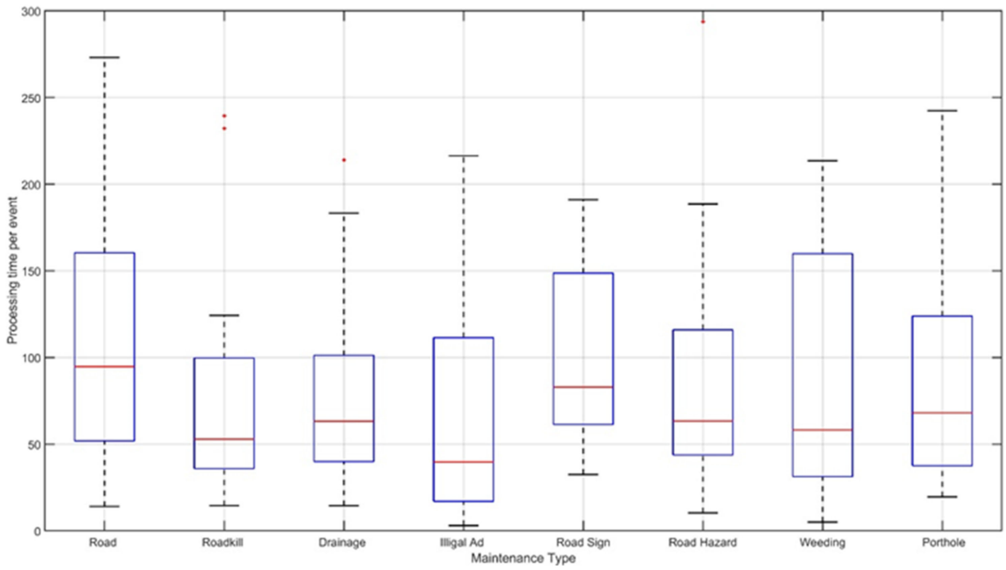

Figure 1. Box plot of the processing time by local office per event for a given road hazard type.

Comparing the median values of the processing time among road hazard types, "Roads" maintenance took the longest time, while "Illegal Ads" the shortest. The difference in processing time between them was 55 h, with the others distributed within the gap. Overall, the processing time seemed quite different across road hazard types. Investigating the variance of processing across local office for a given road hazard type, "Roads" maintenance type showed the highest variance, which was 165,344 h^2. The lowest variance came from "Drainage", 3113 h^2. It turns out that the minimum processing time per event for "Drainage" came from D4 and maximum from E2, and their difference was over 190 h. From Figure 1, it can be concluded that the time required to process the same type of complaint can vary by local agency.

5.3. Development and Evaluation of An Operational Efficiency Indicator

Thus far, we have investigated road complaint report data to define road hazard types and analyzed the number of reports and processing times by road hazard type and agency. We found that the operational efficiency of each office differs depending on the type of road hazard. However, deriving these basic statistics to confirm operational efficiency at each office can be slightly complex and cumbersome, as illustrated above. Therefore, we aimed to develop an indicator that can quantitatively calculate operational efficiency, and then evaluate the maintenance performance of road agencies based on the indicator.

The operational efficiency for a given complaint can be quantified by the processing time per road hazard type. To derive an operational efficiency indicator, we first calculated the Z-score to align the units of report count per kilometer and the processing time per event of all agencies. Second, we visualized the Z-score of the processing time per event, corresponding to the Z-score of the report count per kilometer for each agency on a two-dimensional plane, as shown in Figure 2. If the Z-score of a local office is on the positive diagonal, it can be considered that the agency has average operational efficiency. Conversely, if the Z-score is in the region above the diagonal, the efficiency is low; when it is in the region below the diagonal, the efficiency is high.

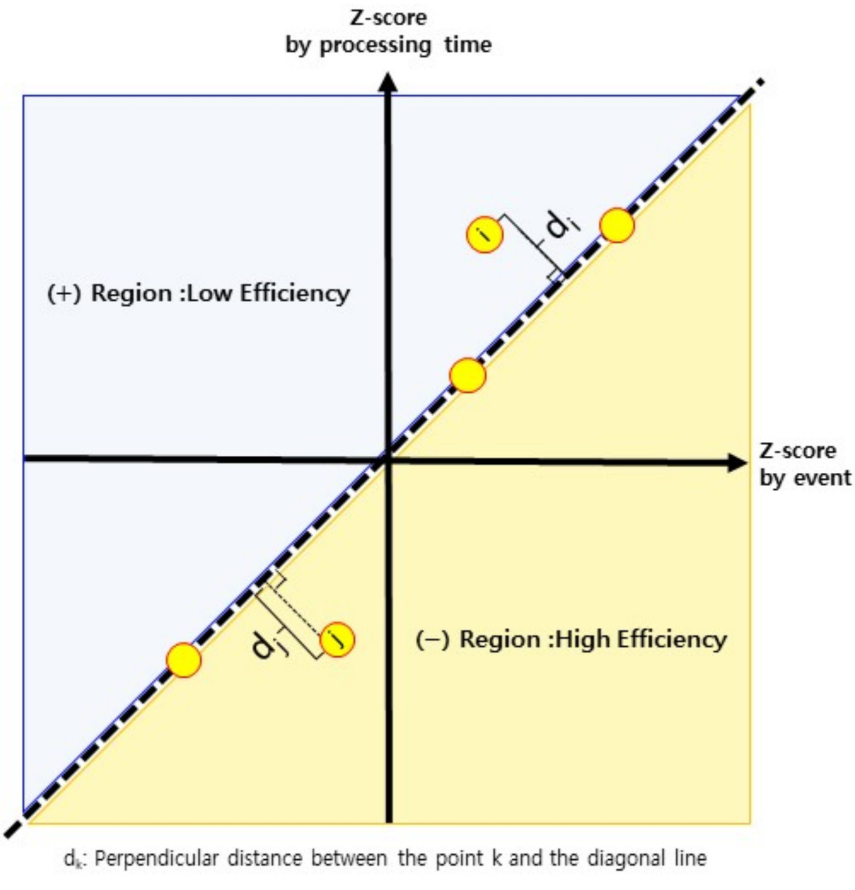

Figure 2. Operational efficiency indicator using the Z-score.

Figure 3 shows the scatter plots of Z-scores for report counts per kilometer and processing times per event for eight types of road hazards ((a) through (h)) and a scatter plot for all types (i). In each figure, "o" represents the Z-score of the processing time, corresponding to the Z-score of the complaint record count for a specific local office; this was used as an indicator of operational efficiency by calculating the orthogonal distance to the positive diagonal.

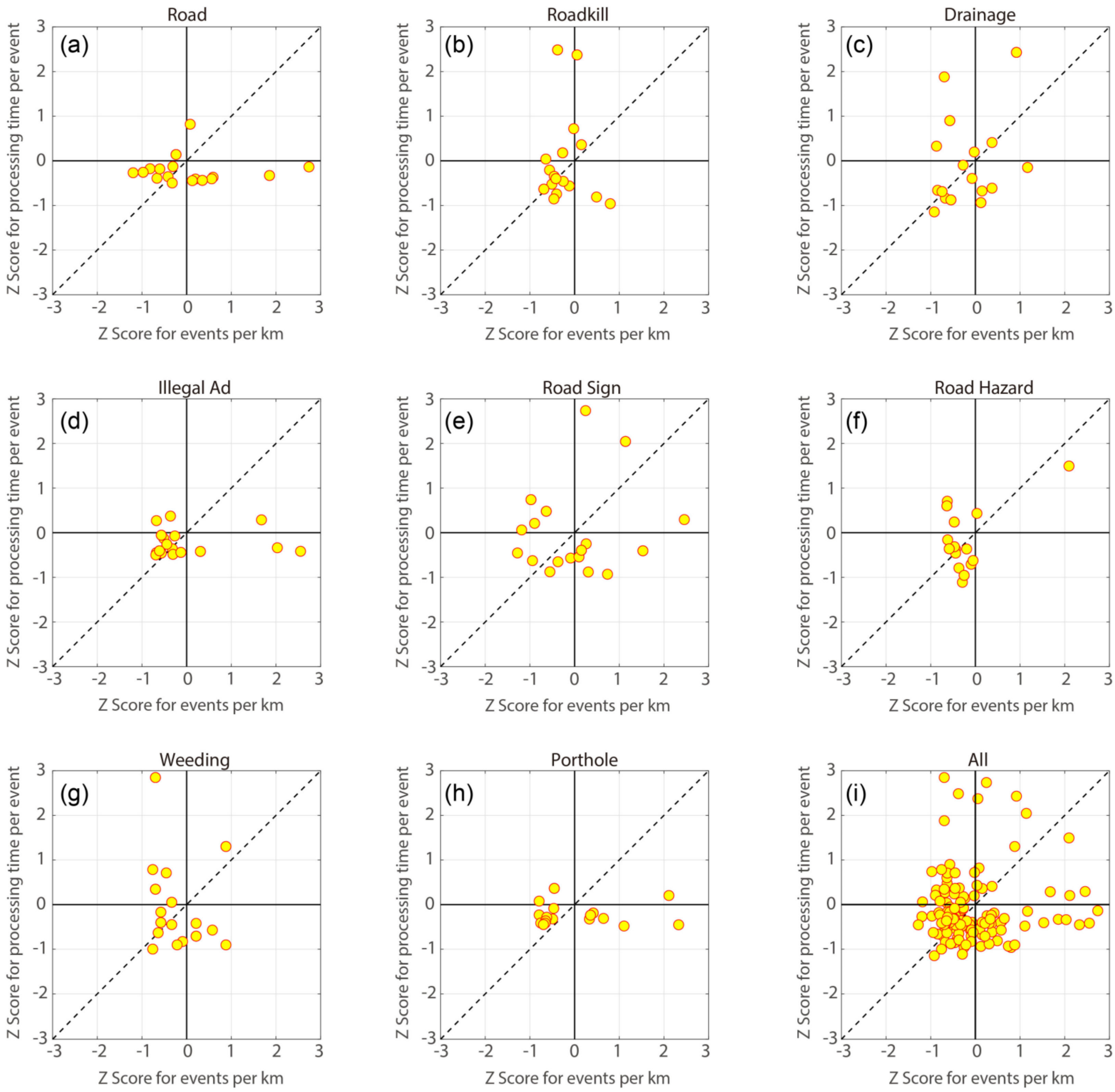

Figure 3. Z-score for processing time per event vs. Z-score for the number of events per kilometer: (**a**) Roads, (**b**) Roadkill, (**c**) Drainage, (**d**) Illegal Activity, (**e**) Facilities, (**f**) Road Hazards, (**g**) Weeding, (**h**) Potholes, and (**i**) All.

In Figure 3, most points are located near the diagonal line for each road hazard type. Particularly, for "Roadkill", the points are distributed along the diagonal line, implying the lack of a large difference in operational efficiency for this type across local offices. The results of calculating the operational efficiency indicator based on Figure 3 are shown in Table 7. In Table 7, a positive operational efficiency indicator value means that the Z-score coordinate in Figure 3 is located in the upper section of the positive diagonal, which can be interpreted as low operational efficiency. In Table 7, Office B is ranked the lowest, whereas Office C is ranked the highest. Notably, even though office C had the lowest number of complaint records (See Table 6), the majority of operational efficiency indicator values for the different types of road hazards showed positive values, indicating low operational efficacy compared to the other regional offices. Regarding local agencies, C4, with the lowest operational efficiency, showed positive indicator values for all types. The second least efficient local office was E2, with particularly low efficiency for "Potholes". The third least efficient office was D2, with a longer processing time for "Roads" compared to the other types. The office with the highest operational efficiency was E1, which showed negative values for all types.

Table 7. Results of the operational efficiency indicator calculations.

Office		Roads	Roadkill	Drainage	Illegal Ads	Road Signs	Road Hazards	Weeding	Potholes	Sum 1	Sum 2
Office A	A1	0.0	0.0	−0.2	0.1	−0.3	0.1	0.0	0.1	−0.2	−7.2
	A2	0.1	−2.4	−0.7	−0.2	−0.4	−2.4	−0.7	−0.4	−7.1	
Office B	B1	−1.5	0.3	−0.9	0.1	−0.5	0.5	0.7	−0.4	−1.7	−8.0
	B2	−0.4	−0.3	−0.2	0.1	−1.2	−0.3	−0.2	0.2	−2.2	
	B3	0.5	−0.9	−0.7	−2.1	−0.2	0.3	−1.3	0.2	−4.2	
Office C	C1	−0.7	0.3	0.2	0.2	−0.2	−0.1	−0.8	−0.7	−1.9	5.9
	C2	0.5	−0.2	1.8	−1.7	0.8	0.0	−0.4	0.4	1.1	
	C3	−0.6	0.0	−0.2	−0.5	−1.4	−0.5	0.3	0.2	−2.7	
	C4	0.3	1.6	1.1	3.0	0.6	1.9	0.3	0.6	9.5	
Office D	D1	−0.7	0.1	−0.6	0.5	−0.4	0.1	0.8	−1.1	−1.3	5.2
	D2	3.4	−0.3	0.1	0.7	0.9	0.9	1.1	0.3	7.1	
	D3	0.2	0.5	−0.1	0.3	0.2	−0.4	−0.1	0.2	0.9	
	D4	−0.1	−1.2	0.0	−0.1	−0.8	−0.6	−0.5	−2.0	−5.3	
	D5	0.7	0.5	0.8	0.1	0.6	0.3	0.3	0.6	3.9	
Office E	E1	−2.0	0.1	−1.6	−1.0	−1.5	−0.4	−1.7	−1.3	−9.5	4.1
	E2	0.5	0.0	1.0	0.4	0.8	0.9	0.1	3.4	7.1	
	E3	0.3	2.0	0.0	0.1	1.2	0.2	2.5	0.2	6.6	
	E4	−0.4	−0.1	0.1	0.0	1.8	−0.4	−0.5	−0.5	−0.1	

6. Conclusions

To prevent accidents through efficient road management, it is important to identify and address road hazards in real time. Hence, traffic agencies allocate substantial budget and personnel each year to maintain or improve the performance of the roads under their jurisdiction. However, the budget for road management is limited, and traffic agencies are distributed across regions; therefore, it is essential to determine the appropriate budget size and regional allocation for which comprehensive foundational data are required, including the classification and size of road hazards in each region and the time required to process each hazard. In this study, we proposed a text mining-based methodology to acquire such foundational data for allocating road management assets efficiently. We employed text-based complaint records reported by volunteers and citizen participation data collected using a mobile-based RIRS application. Taking advantage of the text mining technique, we defined road hazard types to be cleared for all complaint records. The analysis of road hazard types and complaint records for each road management agency revealed that specific types of road hazards (i.e., "Roadkill", "Road Hazards", "Potholes", and "Illegal Ads") occurred prominently under specific agencies. After extracting the processing time from the analysis of the data, we examined the operational efficiency of road management agencies through road hazard type. The results showed that the time required to process identical road hazard types can vary among agencies. These results suggest that the control tower overseeing the entire national highway may need to distribute its budget and support by region to resolve specific road hazards. Additionally, we developed an indicator that easily evaluates the operational efficiency of each management agency by combining the processing time and complaint record counts for each type.

We expect this research will help transportation authorities in road maintenance data acquisition and budget allocation. As road and traffic environments change, road maintenance workload and frequency also change over time. Therefore, it is very important to understand the workload for each type of maintenance in order to provide sustainable and consistent road services. Moreover, monitoring the maintenance status of roads is essential for the analysis of accident risk areas. This study proposed a framework to quantify the amount of road hazard maintenance. Through the simple method, various road maintenance workloads can be identified, and it is also easy to tally the work time required for each maintenance. Second, the budget for road maintenance is limited so that transportation authorities need to establish an appropriate budget allocation strategy. A possible MOE (measure of effectiveness) to determine which local office takes more budget would be work efficiency. The indicator proposed in this study could be a proxy value to measure work efficiency. Thus, transportation authorities can utilize the indicator to evaluate local offices in terms of work efficiency and determine priorities for resource allocation.

Generally, it is quite complicated to extract the necessary information from text data, a typical form of unstructured data, compared to structured data recorded numerically. In this study, we used Excel's Power Query feature to extract keywords from a vast amount of text data and classify complaint types that can represent the content of the complaint written by the user. Consequently, we classified 95% of over 17,000 complaints into eight road hazard types using the data mining methodology. However, there are some limitations in utilizing text mining in this study. First, a review of researchers is necessary in some steps for extracting keywords from complaint records, which can extend the time required for keyword extraction if the historical data volume is vast. Furthermore, approximately 5% of the complaints were unclassified as a result of data mining; therefore, future studies must also consider the reduction of unclassified complaints. Finally, the techniques proposed in this study are a basic approach that works on limited text forms. Various techniques associated with text clustering, text summarization, and information extraction should be applied to obtain more sophisticated results in future studies.

Author Contributions: Conceptualization, S.K. and J.K.; methodology and investigation, S.K. and W.J.; software and visualization, S.K.; validation, J.K.; project administration, W.J.; writing and editing, S.K. and J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research data can be provided upon request to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Borsos, A.; Birth, S.; Vollpracht, H.-J. The role of human factors in road design. In Proceedings of the 6th IEEE International Conference on Cognitive Infocommunications IEEE, Gyor, Hungary, 19–21 October 2015; pp. 363–367. [CrossRef]
2. Kang, J.G.; Lee, S.G. Traffic Accident Prediction Model by Freeway Geometric Types. *Korean Soc. Transp.* **2002**, *20*, 163–175.
3. Borghetti, F.; Marchionni, G.; De Bianchi, M.; Barabino, B.; Bonera, M.; Caballini, C. A new methodology for accidents analysis: The case of the State Road 36 in Italy. *Int. J. Transp. Dev. Integr.* **2021**, *5*, 278–290. [CrossRef]
4. Sharma, N.; Garg, R.D. Advanced transportation safety using real-time GIS-based alarming system for animal-prone zones and pothole areas. *J. Transp. Eng. Part A Syst.* **2023**, *149*, 04023003. [CrossRef]
5. Tran, V.P.; Tran, T.S.; Lee, H.J.; Kim, K.D.; Baek, J.; Nguyen, T.T. One stage detector (RetinaNet)-based crack detection for asphalt pavements considering pavement distresses and surface objects. *J. Civ. Struct. Health Monit.* **2021**, *11*, 205–222. [CrossRef]
6. Ameddah, M.A.; Das, B.; Almhana, J. Cloud-assisted real-time road condition monitoring system for vehicles. In Proceedings of the 2018 IEEE Global Communications Conference, Abu Dhabi, United Arab Emirates, 9–13 December 2018; pp. 1–6. [CrossRef]
7. Bose, B.; Dutta, J.; Ghosh, S.; Pramanick, P.; Roy, S. Detection of driving patterns and road anomalies. In Proceedings of the 2018 3rd International Conference on Internet of Things: Smart Innovation and Usages IEEE, Bhimtal, India, 23–24 February 2018; pp. 1–7. [CrossRef]
8. Marcelino, P.; Lurdes Antunes, M.D.; Fortunato, E. Comprehensive performance indicators for road pavement condition assessment. *Struct. Infrastruct. Eng.* **2018**, *14*, 1433–1445. [CrossRef]
9. Huang, Y.T.; Jahanshahi, M.R.; Shen, F.; Mondal, T.G. Deep Learning–Based Autonomous Road Condition Assessment Leveraging Inexpensive RGB and Depth Sensors and Heterogeneous Data Fusion: Pothole Detection and Quantification. *J. Transp. Eng. Part B Pavements* **2023**, *149*, 04023010. [CrossRef]
10. Bhatia, Y.; Rai, R.; Gupta, V.; Aggarwal, N.; Akula, A. Convolutional neural networks based potholes detection using thermal imaging. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 578–588. [CrossRef]
11. Pei, L.; Shi, L.; Sun, Z.; Li, W.; Gao, Y.; Chen, Y. Detecting potholes in asphalt pavement under small-sample conditions based on improved faster region-based convolution neural networks. *Can. J. Civ. Eng.* **2022**, *49*, 265–273. [CrossRef]

12. Ha, J.; Park, K.; Kim, M. A development of road crack detection system using deep learning-based segmentation and object detection. *J. Soc. E-Bus. Stud.* **2022**, *26*, 93–106. [CrossRef]
13. Ahmed, K.R. Smart pothole detection using deep learning based on dilated convolution. *Sensors* **2021**, *21*, 8406. [CrossRef]
14. Maeda, H.; Kashiyama, T.; Sekimoto, Y.; Seto, T.; Omata, H. Generative adversarial network for road damage detection. *Comput. -Aided Civ. Infrastruct. Eng.* **2021**, *36*, 47–60. [CrossRef]
15. Majidifard, H.; Jin, P.; Adu-Gyamfi, Y.; Buttlar, W.G. Pavement image datasets: A new benchmark dataset to classify and densify pavement distresses. *Transp. Res. Rec.* **2020**, *2674*, 328–339. [CrossRef]
16. Chen, H.; Yao, M.; Gu, Q. Pothole detection using location-aware convolutional neural networks. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 899–911. [CrossRef]
17. Varona, B.; Monteserin, A.; Teyseyre, A. A deep learning approach to automatic road surface monitoring and pothole detection. *Pers. Ubiquitous Comput.* **2020**, *24*, 519–534. [CrossRef]
18. Wu, H.; Yao, L.; Xu, Z.; Li, Y.; Ao, X.; Chen, Q.; Li, Z.; Meng, B. Road pothole extraction and safety evaluation by integration of point cloud and images derived from mobile mapping sensors. *Adv. Eng. Inform.* **2019**, *42*, 100936. [CrossRef]
19. Dhiman, A.; Klette, R. Pothole detection using computer vision and learning. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3536–3550. [CrossRef]
20. Li, J.; Zhao, X.; Li, H. Method for detecting road pavement damage based on deep learning. In Proceedings of the Health Monitoring of Structural and Biological Systems XIII, Denver, CO, USA, 1 April 2019; Volume 10972, pp. 517–526. [CrossRef]
21. Song, W.; Workman, S.; Hadzic, A.; Zhang, X.; Green, E.; Chen, M.; Souleyrette, R.; Jacobs, N. Farsa: Fully automated roadway safety assessment. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision, Lake Tahoe, NV, USA, 12–15 March 2018; pp. 521–529. [CrossRef]
22. Grilo, C.; Coimbra, M.R.; Cerqueira, R.C.; Barbosa, P.; Dornas, R.A.; Gonçalves, L.O.; Teixeira, F.Z.; Coelho, I.P.; Schmidt, B.R.; Pacheco, D.L.; et al. Brazil Road-Kill: A Dataset of Wildlife Terrestrial Vertebrate Road-Kills. *Bull. Ecol. Soc. Am.* **2019**, *100*, e01449. [CrossRef]
23. Medrano-Vizcaíno, P.; Brito-Zapata, D.; Rueda-Vera, A.; Jarrín-V, P.; García-Carrasco, J.M.; Medina, D.; Aguilar, J.; Acosta-Buenaño, N.; González-Suárez, M. First national assessment of wildlife mortality in Ecuador: An effort from citizens and academia to collect roadkill data at country scale. *Ecol. Evol.* **2023**, *13*, e9916. [CrossRef] [PubMed]
24. Mayer, M.; Nielsen, J.C.; Elmeros, M.; Sunde, P. Understanding spatio-temporal patterns of deer-vehicle collisions to improve roadkill mitigation. *J. Environ. Manag.* **2021**, *295*, 113148. [CrossRef] [PubMed]
25. Tiedeman, K.; Hijmans, R.J.; Mandel, A.; Waetjen, D.P.; Shilling, F. The quality and contribution of volunteer collected animal vehicle collision data in ecological research. *Ecol. Indic.* **2019**, *106*, 105431. [CrossRef]
26. Kreling, S.E.; Gaynor, K.M.; Coon, C.A. Roadkill distribution at the wildland-urban interface. *J. Wildl. Manag.* **2019**, *83*, 1427–1436. [CrossRef]
27. Heigl, F.; Stretz, C.R.; Steiner, W.; Suppan, F.; Bauer, T.; Laaha, G.; Zaller, J.G. Comparing road-kill datasets from hunters and citizen scientists in a landscape context. *Remote Sens.* **2016**, *8*, 832. [CrossRef]
28. Vercayie, D.; Herremans, M. Citizen science and smartphones take roadkill monitoring to the next level. *Nat. Conserv.* **2015**, *11*, 29–40. [CrossRef]
29. Chen, J.; Xu, H.; Wu, J.; Yue, R.; Yuan, C.; Wang, L. Deer crossing road detection with roadside LiDAR sensor. *IEEE Access* **2019**, *7*, 65944–65954. [CrossRef]
30. Olson, D.D.; Bissonette, J.A.; Cramer, P.C.; Green, A.D.; Davis, S.T.; Jackson, P.J.; Coster, D.C. Monitoring wildlife-vehicle collisions in the information age: How smartphones can improve data collection. *PLoS ONE* **2014**, *9*, e98613. [CrossRef]
31. Ng, Z.B.; Lim, K.M.; Lee, C.P. Traffic Sign Recognition with Convolutional Neural Network. In Proceedings of the 2021 9th International Conference on Information and Communication Technology IEEE, Yogyakarta, Indonesia, 3–5 August 2021; pp. 48–53. [CrossRef]
32. Farag, W. Traffic signs classification by deep learning for advanced driving assistance systems. *Intell. Decis. Technol.* **2019**, *13*, 305–314. [CrossRef]
33. Roxas, E.A.; Acilo, J.N.; Vicerra, R.R.P.; Dadios, E.P.; Bandala, A.A. Vision-based traffic sign compliance evaluation using convolutional neural network. In Proceedings of the 2018 IEEE International Conference on Applied System Invention, Chiba, Japan, 13–17 April 2018; pp. 120–123. [CrossRef]
34. Acilo, J.P.N.; Cruz, A.G.S.D.; Kaw, M.K.L.; Mabanta, M.D.; Pineda, V.G.G.; Roxas, E.A. Traffic sign integrity analysis using deep learning. In Proceedings of the 2018 IEEE 14th International Colloquium on Signal Processing and Its Applications, Penang, Malaysia, 9–10 March 2018; pp. 107–112. [CrossRef]
35. Sanjeewani, T.G.P.; Verma, B. Learning and analysis of AusRAP attributes from digital video recording for road safety. In Proceedings of the 2019 International Conference on Image and Vision Computing New Zealand IEEE, Dunedin, New Zealand, 2–4 December 2019; pp. 1–6. [CrossRef]
36. Jan, Z.; Verma, B.; Affum, J.; Atabak, S.; Moir, L. A convolutional neural network based deep learning technique for identifying road attributes. In Proceedings of the 2018 International Conference on Image and Vision Computing New Zealand IEEE, Auckland, New Zealand, 19–21 November 2018; pp. 1–6. [CrossRef]
37. Verma, B.; Zhang, L.; Stockwell, D. *Roadside Video Data Analysis: Deep Learning*; Springer: Singapore, 2017. [CrossRef]
38. Zhang, L.; Verma, B.; Stockwell, D. Spatial contextual superpixel model for natural roadside vegetation classification. *Pattern Recognit.* **2016**, *60*, 444–457. [CrossRef]

39. Chowdhury, S.; Verma, B.; Tom, M.; Zhang, M. Pixel characteristics based feature extraction approach for roadside object detection. In Proceedings of the 2015 International Joint Conference on Neural Networks IEEE, Killarney, Ireland, 12–17 July 2015; pp. 1–8. [CrossRef]
40. Sanjeewani, P.; Verma, B. Single class detection-based deep learning approach for identification of road safety attributes. *Neural Comput. Appl.* **2022**, *33*, 9691–9702. [CrossRef]