RESEARCH ARTICLE

Neural-Network-Optimized Vehicle Classification Using Clustered Image and Fiber-Sensor Datasets

PATRIK KAMENCAY[1], **(Member, IEEE), MIROSLAV MARKOVIC**[1], **JOZEF DUBOVAN**[1],
MILAN DADO[1], **(Senior Member, IEEE), AND DANIEL BENEDIKOVIC**[1,2], **(Member, IEEE)**

[1]Department of Multimedia and Information-Communication Technologies, University of Žilina, 01026 Žilina, Slovakia
[2]University Science Park, University of Žilina, 01026 Žilina, Slovakia

Corresponding author: Daniel Benedikovic (daniel.benedikovic@uniza.sk)

ABSTRACT Internet of Things (IoT) becomes indispensable for transport and automotive industry to advance functions in on-road traffic monitoring. Indeed, smart management tools and machine learning concepts are inevitable in vehicle categorization systems. However, to date, existing systems for vehicle classification are exclusively based on singular technological platforms only. This not only limits their long-term use and future scaling, but also sets restrictions to obtain high classification accuracies with modern machine learning tools feed with diversified big volume data. In this work, we design a novel convolutional neural network (CNN) that substantially improves the on-road vehicle classification. In particular, we experimentally harness, to the best of our knowledge for the first time, two different datasets from separated technological platforms based on close-circuit television (CCTV) and fiber Bragg grating (FBG) sensors, respectively. The hybrid CNN classification system, with individual CCTV and FBG datasets, substantially improves detection levels, reaching in-class accuracy up to 90% - 97%. Moreover, this classification concept includes an intrinsic back-up verification with respect to each platform compensating the shortcomings of individual technologies. Our demonstration can make key advances towards near-unity accuracy in vehicle classifications for IoT systems, capitalizing on cost-effective and well-established platforms.

INDEX TERMS Internet of Things, vehicle classification, convolutional neural networks, close-circuit television, fiber Bragg gratings.

I. INTRODUCTION

The Internet of Things (IoT) is seen as universal solution to merge diverse technologies into a conceptual network. IoT allows different devices to be mutually sensed with respect to their surroundings and then communicate with an instant response [1], [2], [3]. The IoT is particularly appealing for data-intensive applications [4]. ''Big data'' applications, which range from military, medicine, and smart buildings to intelligent transport and automotive industry, are good examples. In recent years, the IoT becomes indispensable for transport and automotive industries to advance activities such as vehicle classification, instant on-road control, or traffic management, among others [1], [2], [3], [4], [5], [6].

The ever-rising number of vehicles on roads calls upon smart solutions with dynamic monitoring and novel platforms. Present traffic capacities are unable to satisfy the infrastructural needs set by smart cities. Over the years, many IoT platforms were proposed and demonstrated to address these problems. Modern platforms can enhance traffic monitoring, management of road resources, and vehicle classification. So far, image recognition method remained the first choice for such platforms, considering the quality-cost trade-off [7], [8], [9]. Although these platforms are still widely used, they face a range of challenges. This includes limitation by weather conditions, slow data collection and inefficient processing. Nowadays, sensing-to-

The associate editor coordinating the review of this manuscript and approving it for publication was Muguang Wang.

communication platforms take a massive leap. They can be found as radio-founded networks [10], [11], [12], [13], virtual clouds and storages [14], [15], and a variety of ground-based concepts [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36].

Ground-based solutions can provide neat monitoring, management, and car classification. They are typically based on inductive and magnetic loops or may benefit from many sensing methods [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36]. The sensing maps a measurable physical quantity into virtual data, creating a basis for computation, analytics, and adoption of artificial intelligence (AI) and machine learning (ML) algorithms. For a sensing part, anisotropic magneto-resistance [20], piezoelectricity [21], [22], or hydro-electricity [23] are typically exploited. Alternative classes for on-road traffic control rely on embedded strain gauge [24], vibration [25], or fiber-optics [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36] sensors.

Fiber-optic technologies are proved to be compelling for traffic surveillance as they afford a virtually unlimited capacity compared to electrical wires, while fiber sensors constitute a close-to-ideal choice for optical sensing. Co-existence between high-speed network and optical sensing platform is highly desired to keep both cost and complexity at low levels [36]. Fiber Bragg gratings (FBGs) emerge as an attractive optical platform to form single- and multi-point sensor arrays [26], [27], [35], [36], [37], [38], [39]. In FBGs, the reflected light adds constructively in the backward direction and creates a distinct narrowband drop in the transmission, while the rest of the light passes through the grating. The external perturbations induce a spectral shift compared to the initial position and the change can be clearly determined. This way, the FBG platform unleashes promises for effective vehicle categorization, especially in dense traffic situations. However, FBG systems have a range of obstacles, which includes instrumentation, installation, and calibration issues, while require improved data collection, vehicle categorization, and overall detection accuracy - IoT system parts that are mostly driven by modern processing techniques.

In this work, we propose and demonstrate a novel CNN architecture for enhanced vehicle classification. Here, we harness a clustered dataset, originating from a hybrid technological platform. Hybrid system combines data inputs from CCTV system and FBG sensors, respectively, a vital co-existence has remained unexplored to date. Clustered CNN-based classification improves vehicle detection accuracy and provides an intrinsic back-up verification with respect to each separate technology.

II. TESTING PLATFORMS AND DATA COLLECTION

In this work, the technological base for vehicle classification comprises two experimental platforms: an optical sensor network and a visual-based CCTV system. Indeed, each of them can be operated separately, but this brings a range of limitations. Measurements were performed over three weeks. The total number of vehicles passing through the platform was up to 2000 per day.
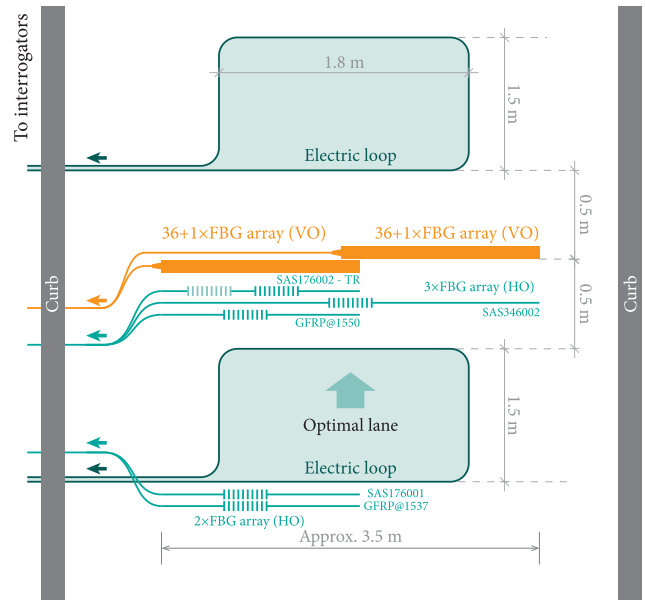

FIGURE 1. Schematic top view of the experimental platform based on fiber Bragg gratings.

A. FBG SENSOR NETWORK

The sensor network comprises a FBG array. This in-house testing platform, schematically shown in Fig. 1, is situated at the campus of the University of Zilina, Slovakia. The FBG sensor array is connected to two active interrogators. Interrogators collect the output data from the sensor array. The Sensing array includes active induction loops to help detection of passing vehicles. For this work, inductive loops are disconnected, and FBG sensor array is only used passively, i.e. without additional electricity, which reduces power consumption. From an installation point of view, FBG sensors were mounted invasively into the ground, i.e. directly into the existing road in the two-layer asphalt arrangement.

In the optical array, FBGs are positioned both vertically (parallel to the vehicle wheels) and horizontally (perpendicular to the vehicle wheels), which enable to set-up different experimental topologies. For our investigation, we are primarily focused on the spectra response of the FBG sensor array, for now, without the possibility to operate platform dynamically to detect the vehicle speed or exploit weigh-in-motion functions. The horizontally oriented (HO) fiber sensors are used at the beginning only to verify platform operation and working conditions. Vertical FBG array is divided into two 1.75 m long parts, located on both sides of the road, respecting the road center. The sensor array on each side comprises 36 vertically-oriented (VO) FBGs. The reference spacing between the individual optical sensors is 100 mm. FBG sensors are used to measure the

deformation caused by pressure, once the vehicle passes through the platform. Double-sided sensor topology helps to create a back-up solution for situations, when the optical fiber is damaged, or fiber connections are interrupted. Output data from the on-road sensor cluster are transmitted through single-mode optical fibers (SMF-28) to two four-channel interrogators. The interrogators cover a spectral range from 1485 nm to 1610 nm. Interrogators are situated in the local data center, 700 m away from the experimental platform. On a side of interrogators, there are standard opto-electrical input/output (I/O) interfaces. This includes: 4 optical ports, two USB ports, 1 Ethernet and HDMI port and a serial interface RS485. Interrogators use a 12 V DC industrial adapter as input power. An RJ-45 Ethernet port allows connection to a wired 10/100/1000 Mbps local area network. Optical fiber ports are used to connect the interrogator to the sensor array, in particular, each I/O fiber port connects 19 FBGs.
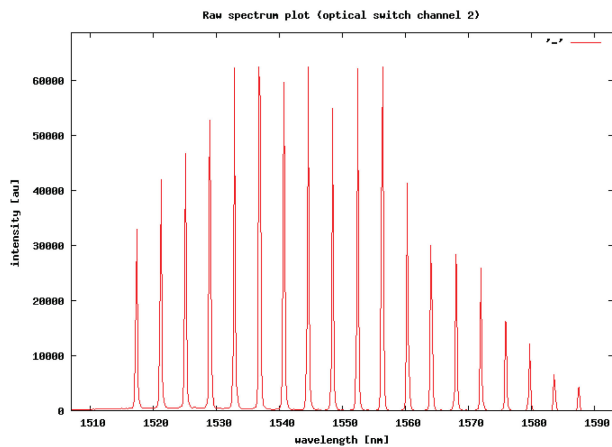

FIGURE 2. The spectral output of the interrogator's channel. The output was retrieved from the fiber Bragg grating array.
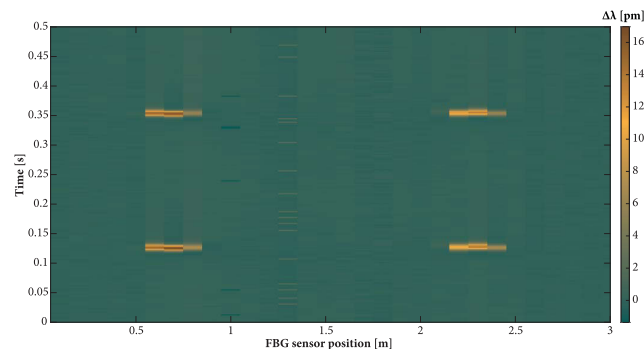

FIGURE 3. Spectral outputs from the sensor array using vertical sensor arrangement. Two-dimensional pressure map of the vehicle passing the platform as a function of time and FBG sensor position. Here, the vehicle detected on the map has a wheelbase of 2570 mm.

The output spectrum from the sensing array at one of interrogators channel is shown in Fig. 2. Interrogators process the data offline from the sensor network offline, with a recording frequency 500 samples per second. Examples of

spectral outputs from FBG sensors for vertically positioned configurations are shown in Fig. 3. Here, we can observe a detail of vehicle passage with an optimal trajectory over a sensor array, having an axle wheelbase of 2570 mm. Measured values correspond to the wavelength shift of the reflected optical signal. This way, we create a two-dimensional pressure map of the vehicle as a function of the time and FBG sensor position. Sensor data were collected from 2000 tests using 36 vertical FBG sensors. The dataset comprises 500 samples per second × 4 s time slot per single vehicle.

The time period is sufficiently long to cover the majority of vehicles that passed in and out of the platform, with a restricted area speed of 30 km/h. Moreover, correct detection for front and back axles is always obtained at different positions. In turn, this helps to reduce the FBG dataset down to 600 correct measurements × 36 vertical FBGs per one passing vehicle. Furthermore, from the dataset analysis, we found out that a maximum of 5 sensors from the whole array is primarily responsible for the spectral outputs, which lowers the final dataset to 600 tests × 5 vertical sensors. The spectral shifts initiated by the remaining sensors (compared to the 5 dominant) are close to zero, and thus their contribution can be neglected. Last, but not least, measured data are normalized to toned image values (shaded gray colors), having a range from 0 and 1. Output data are stored in Tagged Image Files Format (TIFF). Output TIFF data, shown in Fig. 4, are then used to train and test CNNs and to categorize particular vehicles.
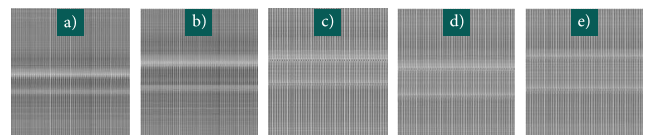

FIGURE 4. TIFF outputs obtained from a fiber Bragg grating array for: a) small car, b) personal car, c) SUV, d) Minivan, and e) VAN/Trucks.

TABLE 1. Training and testing dataset for CNN in case of different vehicle categories.

Class	Description	Dataset		
		X	Y	Z
1	Small cars	1346	940	406
2	Personal cars	2148	500	648
3	SUV	790	550	240
4	Minivan	690	480	210
5	Van/Trucks	250	170	80

X: Total number of data; Y: Data used for training; and Z: Data used for testing.

The system for car identification encompasses the body type of the vehicle, the number of axles, wheelbase as well

as the number of vehicles driving in and out of the platform. Table 1 sums up car categories and corresponding datasets. In total, there are 5 specific car categories (see Tab. 1). Corresponding parameters (specific car characteristics) are listed in Tab. 2. Specific characteristics for 5 vehicle categories are also depicted in Fig. 5 and explained below [37], [38]:

TABLE 2. Special characteristics for given vehicle class [37], [38].

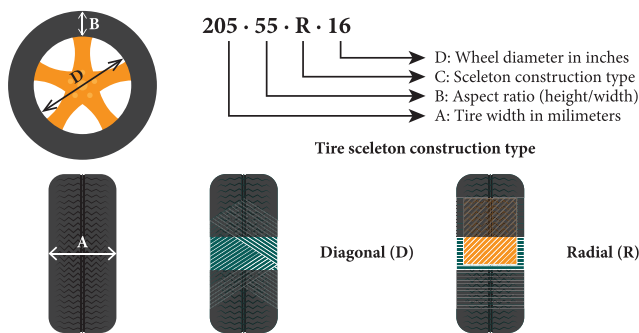| Class	Description	Characteristics
1 & 2	Small cars & Personal cars	205/55R 16 195/65R 15 165/70R 14
3	SUV	215/65R 16 215/70R 16 235/55R 18
4	Minivan	205/65R 16 215/65R 16 225/65R 16
5	Van/Trucks	225/50R 18 235/45R 18 245/45R 18 255/45R 18

FIGURE 5. Example of special characteristics for a given vehicle class (small cars).

- A: The tire width is marked by the first three numbers (from the side to side) in millimeters.
- B: The next two values determine the aspect ratio (the tire height as a percentage of the width). In our case, the aspect ratio is 55. This means that the sidewall of the tire is 55 percent higher than the tire width.
- C: The construction type is represented by a single letter which describes the type of the internal construction of the tire (R is for radial tires and D is for tires built with diagonal plies).
- D: The last number describes the diameter of the wheel in inches.

B. CCTV SYSTEM
CCTV system represents a conventional image-based recognition technique. Here, visual recognition was carried out with an industrial camera situated in a close proximity to the experimental FBG platform (of about 100 m). Data were retrieved according to the local GDPR rules. CCTV system operates with digitalized data from a recorded video

stream. This video stream is then transformed into static images. Once the vehicles pass through the platform, image characteristics are changed. The camera has a resolution of 1920 × 1080 pixels, with the speed of 30 frames per second. Camera was supplied by Power over Ethernet (PoE) interface and the video stream was recorded on a computer using Open Broadcaster Software Studio. For vehicle classification, only images of passing cars were selected from the entire video stream, and then only points of interest were captured from those images, i.e. that is the passing vehicle itself. For this, images are shrunk down to a square resolution with 800 × 800 pixels and 1:1 aspect ratio. Images with shadows, blurs or images affected by weather conditions were excluded from further processing. By repeating this process, we retrieved final images of vehicles that are assigned to the specific car categories. Different vehicles from the CCTV system are shown in Fig. 6. In turn, this creates a final image dataset for neural network training and testing.
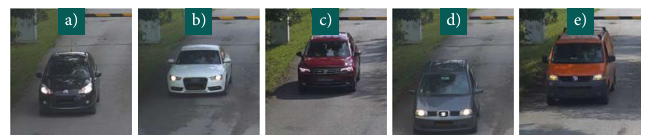

FIGURE 6. TIFF outputs obtained from a CCTV system: a) small car, b) personal car, c) SUV, d) Minivan, and e) VAN/Trucks.

C. NEURAL NETWORK STRUCTURE AND INPUT DATASETS
CNN comprises a three-dimensional architecture built upon a convolutional layer, MaxPooling, and fully connected layers. The general schematic of CNN is shown in Fig. 7.
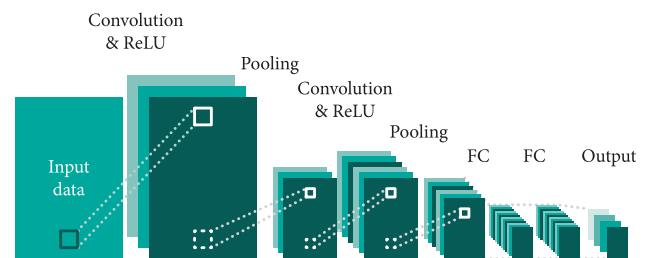

FIGURE 7. Schematics of the convolutional neural network.

CNN's are conventionally leveraged to recognize two-dimensional (2-D) image patterns from the pixels of the input data with minimal pre-processing. CNN classifies an input image into the following four categories: supine, left side, prone, and right side.

Main operations performed by CNN rely on: (i) convolutional layer to extract various features from the input image. For a reliable process, these layers (and thus operations they perform) are repeated several times; (ii) the non-linearity (ReLU) layer serves as an activation function for data processing; (iii) MaxPooling or sub-sampling layer that searches for the largest element in the feature map and forms a bridge (inter-layer) between initial convolutional layer and ending fully connected (FC) layer; and (iv) fully connected

layer (classification). This way, CNN can learn more and more complex symptoms. In the first layers, the network encodes low-level image features (such as edge detectors and simple color transitions). In the subsequent layers, features for shapes (such as semicircle or multicolor gradients) are described. The last CNN layers comprise features responsible for individual image objects or complex image shapes.

TABLE 3. Image sizes for conventional open-access CNNs.

CNN	Size [pixels]
AlexNet	227 × 227
GoogleNet	224 × 224
ResNet-50	224 × 224
ResNet-101	224 × 224

III. RESULTS AND DISCUSSION

For given datasets, training and testing is realized with re-sized image inputs to match the size of open-access CNNs (*AlexNet*, *GoogleNet*, *ResNet-50*, and *ResNet-101*), as defined in Tab. 3. Those images pass through the CNN convolutional layers, with filters that extract local image features.

The CNN activation function determines values of the outputs for individual neurons based on their internal potential. The internal potential is a by-product between weights and the input. Outputs from individual windows are combined to create a new down-scaled feature map. Once this process applies to all new maps, an additional set of feature maps are created, which then form inputs to the next convolutional layer, where the whole process is repeated. The output of the convolutional layer is flattened via fully connected layer. Finally, the Softmax output layer for image classification is used. This way, we determine and link specific input image with an appropriate vehicle class.

FIGURE 8. Example of the confusion matrix. (a) General scheme and (b) ResNet-50.

The confusion matrix, schematically shown in Fig. 8a, summarizes the result of the classifier [39], [40]. Rows of the confusion matrix are indexed by output variable classes (corresponding to the reality), while columns are sorted according to classes predicted by the model (estimation/prediction). Thus, each column within the confusion matrix represents predicted classes. The individual rows in the confusion matrix represent the current (correctly assigned) classes. Success parameters, specifically *P*, *R*, and *F1-score*, were calculated from the obtained values. The parameter of precision (*P*) is the ratio between true positive (*TP*) and the sum of positive data (true positive (*TP*) and false positive (*FP*)). On the other hand, the parameter of recall (*R*) is the ratio between true positive (*TP*) and the sum of data from the actual class (true positive (*TP*) and false negative (*FN*)). The *F1-score* denotes a combination between precision and recall, i.e. weighted average value of precision and recall. The resulting accuracy is obtained as the ratio between the sum of correctly predicted samples and the sums of all samples. In particular, Fig. 8b shows the actual numbers of the confusion matrix, here for conventional open-access CNN - *ResNet-50*. This matrix summarizes the result of the classification. The rows of the confusion matrix represent the actual class (small cars, personal cars, SUV-s, minivans and the VAN/Trucks).

A. CONVENTIONAL NEURAL NETWORKS

First, we trained and tested conventional open-access CNNs (*1 - AlexNet; 2 - GoogLeNet; 3 - ResNet-50 and 4 - ResNet-101*) using FBG dataset only. These CNNs have input sizes as specified in Tab. 3. Results obtained from open-access CNNs are provided in Tab. 4 and Tab. 5, respectively. From the obtained results, we can conclude that conventional CNNs (*AlexNet, GoogleNet, ResNet-50 and ResNet-101*) are better suited for pure image data recognition. For this reason, the afore-mentioned standard CNNs, in all cases fed by sensor data only, obtained results that are less precise compared to the CNN architectures based on CCTV-based image inputs [27].

TABLE 4. Results of vehicle classification (Confusion matrix) using FBG-based dataset through conventional CNNs: a) AlexNet, b) GoogleNet, c) ResNet-50 and d) ResNet-101.

Confusion matrix (AlexNet)						Confusion matrix (GoogleNet)				
	1	2	3	4	5	1	2	3	4	5
1	**205**	115	53	30	3	**294**	82	17	11	2
2	137	**324**	99	78	10	165	**310**	99	66	8
3	26	66	**112**	31	5	16	45	**153**	23	3
4	10	30	53	**107**	10	10	12	24	**159**	5
5	3	10	10	12	**45**	4	3	6	6	**61**

Confusion matrix (ResNet-50)						Confusion matrix (ResNet-101)				
	1	2	3	4	5	1	2	3	4	5
1	**289**	84	16	14	3	**297**	80	15	12	2
2	179	**327**	96	38	8	168	**351**	90	35	4
3	12	33	**181**	12	2	10	30	**189**	9	2
4	8	11	20	**165**	6	7	9	19	**172**	3
5	2	2	3	5	**68**	2	2	3	4	**69**
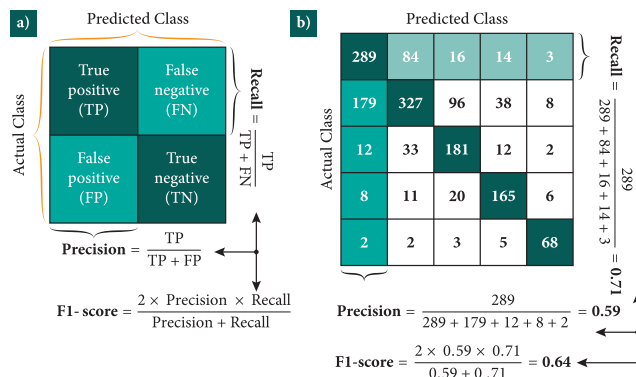
To monitor the generalization performance and select the optimal study model, the dataset was divided into a training

TABLE 5. Results of vehicle classification using FBG-based dataset through conventional CNNs: a) AlexNet, b) GoogleNet, c) ResNet-50 and d) ResNet-101.

	Classification results (AlexNet)				Classification results (GoogleNet)		
	Precision	Recall	F1-score		Precision	Recall	F1-score
1	0.54	0.50	0.52	1	0.6	0.72	0.65
2	0.60	0.50	0.55	2	0.69	0.48	0.57
3	0.34	0.47	0.39	3	0.51	0.64	0.57
4	0.41	0.51	0.45	4	0.60	0.76	0.67
5	0.62	0.56	0.59	5	0.77	0.76	0.76
	Classification results (ResNet-50)				Classification results (ResNet-101)		
	Precision	Recall	F1-score		Precision	Recall	F1-score
1	0.59	0.71	0.64	1	0.61	0.73	0.66
2	0.72	0.50	0.59	2	0.74	0.54	0.62
3	0.57	0.75	0.65	3	0.59	0.79	0.68
4	0.71	0.79	0.75	4	0.74	0.82	0.78
5	0.78	0.85	0.81	5	0.86	0.86	0.86

set with a data amount of 70% and a validation set, having a 30% of data amount. The performance of the proposed CNN with a training set was continually improved, while the CNN performance with a validation set reached a saturation point, after which the network starts to overfit training data, and then the network learning algorithm was terminated. The implementation of the CNN models relies on the TensorFlow framework. Moreover, from the results summarized in Tab. 4, we can observe that *GoogleNet* and *ResNet-50* performed better than *AlexNet*. The obtained precision levels are 34% - 62% for *AlexNet*, 51% - 77% for *GoogleNet*, 59% - 78% for *ResNet-50*, 61% - 86% for *ResNet-101*. From a precision point of view (classification results), by comparing the same vehicle class for different types of standard CNNs, we can clearly see this trend. These results arise from the fact that the *GoogleNet* network uses combinations of inception modules, and each of them encompasses Max-Pooling, convolutions at different scales, and concatenation operations.

In addition, *ResNet-50* and *ResNet-101* are a type of CNN with a depth of 50 layers and 101 layers, respectively. The *ResNet-50* and *ResNet-101* models replace each two-layer residual block with a three layer bottleneck block. This then uses 1 × 1 convolutions to reduce and subsequently restore the channel depth, allowing for a reduced computational load when calculating the 3 × 3 convolution. Those specific features of both CNNs are particularly advantageous, facilitating more accurate categorization. On the other hand, the *AlexNet* is formed by 5 convolutional layers, followed by 3 fully connected layers, and finally added a 1000-way Softmax. This then corresponds to the probability of 1000 categories (in our case we used five vehicle classes as the output). Typically, this network includes repetition of few convolutional layers and each one is followed by max MaxPoolings and few dense layers. However, there was no standard for the filter sizes to be used, which is a significant shortcoming of this CNN.
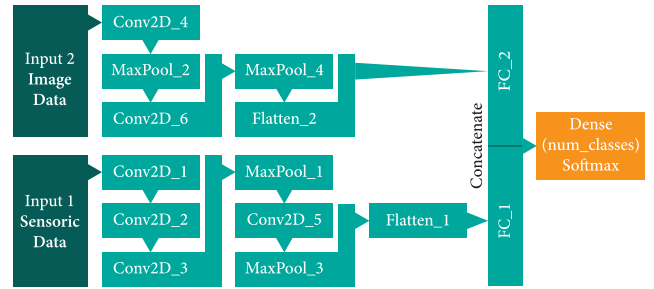

FIGURE 9. Proposed CNN with two independently operated CNN branches. Individual CNN branches are based on pure sensor and pure image datasets.

TABLE 6. Description of individual layers of the hybrid CNN.

Layer (type)	Output Shape	Activation
input_1 (Inputlayer)	(None, 600, 5, 1)	-
input_2 (Inputlayer)	(None, 32, 32, 3)	ReLU
conv2d_1 (Conv2D_1)	(None, 300, 4, 32)	ReLU
conv2d_2 (Conv2D_2)	(None, 100, 4, 32)	ReLU
conv2d_3 (Conv2D_3)	(None, 100, 4, 32)	ReLU
conv2d_4 (Conv2D_4)	(None, 29, 29, 16)	ReLU
max_pooling2d_1 (MaxPooling_1)	(None, 2, 2, 32)	ReLU
max_pooling2d_2 (MaxPooling_2)	(None, 14, 14, 32)	ReLU
conv2d_5 (Conv2D_5)	(None, 50, 2, 64)	ReLU
conv2d_6 (Conv2D_6)	(None, 14, 14, 32)	ReLU
max_pooling2d_3 (MaxPooling_3)	(None, 2, 1, 16)	ReLU
max_pooling2d_4 (MaxPooling_4)	(None, 5, 5, 16)	ReLU
flatten_1 (Flatten_1)	(None, 32)	ReLU
flatten_2 (Flatten_2)	(None, 400)	ReLU
concatenate_1 (Concatenate)	(None, 432)	ReLU
dense_1 (Dense)	(None, 5)	Softmax

B. HYBRID IMAGE- AND SENSOR-BASED NEURAL NETWORK

The proposed CNN, schematically shown in Fig. 9 and described in Tab. 6, comprises two parts: sensor- and image-based branches, respectively. This is a hybrid CNN structure with two independent inputs (sensor- and image-based data). The first network (sensor-based) is fed with 600-d inputs, while the second one (image-based) accepts the 32-d inputs. Both networks operate separately with respect to each other and then they are merged into the single CNN that performs the final classification.

The individual layers of the hybrid CNN are described below.

1) SENSOR DATA BRANCH
- The sensor data was used as input. These input data were resized.
- The second to fourth blocks of the sensor part were 2D CNN layers, which have 32 feature maps with 3 × 3 kernel dimension. The Rectifier linear unit (ReLU) was used as an activation function.

- In the next block, the MaxPooling layer with kernel dimension 2 × 2 was used. The Dropout layer with a probability set of 0.25 was used.
- In the sixth block, the 2D CNN was used with the same parameters as in the second to fourth blocks. However, the number of feature maps was doubled to 64.
- MaxPooling and Dropout layer have the same parameters as the fifth block.

2) IMAGE DATA BRANCH

- The image data was used as input. These input data were resized.
- The second block of the image part was the 2D CNN layer, which has 16 feature maps with 3 × 3 kernel dimension. The Rectifier linear unit (ReLU) was used as an activation function.
- In the next block, the MaxPooling layer was used (kernel dimension 2 × 2. The Dropout layer with a probability set of 0.25 was used.
- The fourth block, the 2D CNN was used with the same parameters as in the second block. However, the number of feature maps was doubled to 32.
- The MaxPooling layer and Dropout have the same parameters as the third block.

3) HYBRID SENSOR AND IMAGE DATASETS

- In our case, we tack on a fully connected layer with fifth neurons. Our final model using the inputs of both branches (sensor and image data) was defined.
- The connection of the final layers in the hybrid network is based on the output of both the sensor and image branches. The final output of the hybrid network was classified into 5 different classes of cars.

Features are extracted from individual sources of information by building appropriate network models, preferably models that are most suitable for given data types. Feature extraction from one source is independent with respect to the other one. Once all the features essential for prediction are extracted from both datasets, they are combined into the single-shared representation. In the next step, information is merged from two modalities to perform a final prediction. The information coming from different modalities is characterized by varying predictive power and noise topology. In our case, we take a weighted combination of the sub-networks so that each input modality can have a learned contribution towards the final output, i.e. towards a resulting prediction.

The hybrid CNN model comprises sensor and image parts. They are described below.

4) SENSOR PART

1. model = input_1()
2. layers.conv2d_1(kernel_size=3, activation='relu', input_shape=(n_features))
3. layers.conv2d_2(kernel_size=3, activation='relu')
4. layers.conv2d_3(kernel_size=3, activation='relu')
5. layers.max_pooling2d_1(pool_size=2)
6. layers.conv2d_5(kernel_size=3, activation='relu')
7. layers.max_pooling2d_3(pool_size=2)
8. layers.Flatten_1()
9. layers.Dropout(0.5))
10. layers.Dense(activation='relu')
11. layers.Dense(n_outputs, activation='softmax')

5) IMAGE PART

1. model = input_2()
2. layers.conv2d_4(kernel_size=3, activation='relu', input_shape=(n_features))
3. layers.max_pooling2d_2(pool_size=2)
4. layers.conv2d_6(kernel_size=3, activation='relu')
5. layers.max_pooling2d_4(pool_size=2)
6. layers.Flatten_2()
7. layers.Dropout(0.5))
8. layers.Dense(activation='relu')
9. layers.Dense(n_outputs, activation='softmax')

It is worth to note that hyper-parameters play a crucial role to control the training process and to obtain the optimal performance. The setting of hyper-parameters may affect the results of the training process. For this reason, it is critically important to train the CNN with an appropriate set of hyperparameters, including learning rate, batch size, weight factor initialization and optimizer. Details for optimal hyper-parameters are provided in Tab. 7, while the hyper-parameters description is provided below:

TABLE 7. Description of CNN hyperparameters.

Hyperparameters	Optimal value
Kernel size	[3]
Pooling layers	['MAX'], [2]
Activation function	ReLu
Dropout size	[0.5]
Learning rate	[0.001]
Optimizer	['ADAM']
Batch size	[128]
Number of epochs	[35]

- Adam optimizer: This is an extended version of the stochastic gradient descent commonly used in deep learning applications.
- Number of epochs: This can be set as an integer value ranging from one to infinity. To increase the number of epochs is advantageous, if the dataset contains a big data count.
- Early stopping: The training process stops when the validation performance deteriorates for 10 consecutive epochs. This helps to avoid poor performance of the neural network with non-training data, while learning well learning on training data.
- Training function: This is the overarching algorithm used to train the neural network and to associate certain inputs with specific outputs.
- Minimum gradient: This refers to the minimum magnitude of the gradient descent required for the training of the neural network to terminate.

- Activation function to introduce non-linearity into the models was used. In our case, the Rectified Linear Unit (ReLU) was employed as an activation function.
- The pooling layer moves the filter through the output feature map of the previous convolutional layer. The filter size was 2×2 and max-pooling approach was used.
- The number of epochs is the number of times for the training data that is displayed to the neural network.
- The batch size is the number of samples submitted to the neural network.
- The learning rate describes the step-size for a neural network model to achieve a function with minimum loss.
- Dropout size describes a technique that prevents the neural network over-fitting.
- Kernel size parameter describes the size of the filter, determining the size of the 2D convolution window.
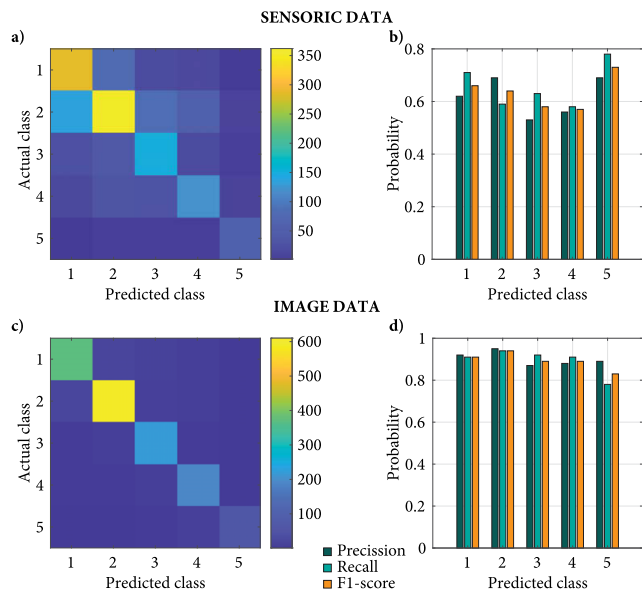

FIGURE 10. Graphical results representation obtained from individual CNN branches. (a) and (b) Fiber sensor datasets. (c) and (d) Image datasets.

TABLE 8. Results obtained from two independent CNN branches. Results for a) sensor and b) image datasets.

	SENSORIC DATA						IMAGE DATA				
	Confusion matrix										
a)	1	2	3	4	5	**b)**	1	2	3	4	5
1	**287**	83	18	15	3	1	**371**	17	11	6	1
2	136	**362**	84	56	10	2	20	**610**	10	6	2
3	26	40	**151**	18	5	3	5	7	**220**	5	3
4	15	34	29	**122**	10	4	4	5	8	**191**	2
5	2	5	5	6	**62**	5	2	2	5	9	**62**
	Classification results										
	Precision	Recall	F1-score				Precision	Recall	F1-score		
1	0.62	0.71	0.66			1	0.92	0.91	0.91		
2	0.69	0.59	0.64			2	0.95	0.94	0.94		
3	0.53	0.63	0.58			3	0.87	0.92	0.89		
4	0.56	0.58	0.57			4	0.88	0.91	0.89		
5	0.69	0.78	0.73			5	0.89	0.78	0.83		

TABLE 9. Obtained results from hybrid CNN.

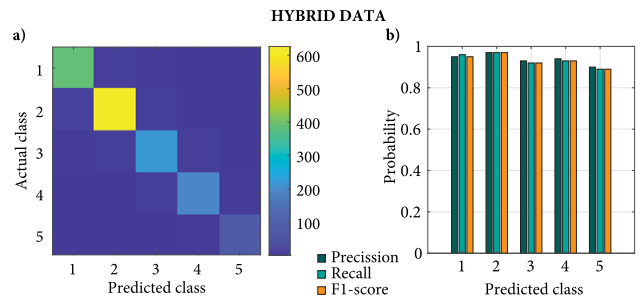| HYBRID DATA					
Confusion matrix					
	1	2	3	4	5
1	**389**	8	5	3	1
2	12	**626**	7	2	1
3	4	7	**221**	6	2
4	2	3	6	**195**	4
5	1	2	2	4	**71**
Classification results					
	Precision	Recall	F1-score		
1	0.95	0.96	0.95		
2	0.97	0.97	0.97		
3	0.93	0.92	0.92		
4	0.94	0.93	0.93		
5	0.90	0.89	0.89		

FIGURE 11. Graphical representation of obtained results using hybrid CNN.

Table 8 and Figure 10 show the results of the individual CNNs, separately for the sensor branch (sensor data only) and separately for the image branch (image data only), while Tab. 9 and Fig. 11 summarize the overall performance of the hybrid CNN. The proposed two-input-model of CNN shows a significant improvement in the performance (precision accuracy) compared to conventional CNNs (*AlexNet*, *GoogleNet*, *ResNet-50*, and *ResNet-101* - see results in Tab. 4 and our single-input CNN based on sensor data). First, from the retrieved results, it becomes apparent that the hybrid CNN of combined datasets yields much better results than the CNN with sensor or image data only.

As a result, precision for the vehicle classification reached extraordinary high values, ranging between 90% and 97%, with respect to the particular vehicle class. Moreover, comparing precision results obtained by this novel hybrid (combined image and sensor) CNNs with those previously obtained via conventional CNNs (*GoogleNet*, *ResNet-50*, *ResNet-101*, and *AlexNet* - shown in Tab. 4) and sensor data only, we can observe a considerable enhancement in the final classification results. Nominally, these improvements are between 35% - 58% with respect to *AlexNet*, 20% - 39% with respect to *GoogleNet*, and 19% - 33% with respect to *ResNet-50*. Obtained results (confusion matrix and evaluation criteria) are graphically represented in Fig. 11. Numbers in

each cell of the graphical representation of the confusion matrix refer to the percentage value of correctly and incorrectly assigned objects. To show the percentage values of individual objects of the actual class, we used color for each cell. Here, the dark-blue color indicates 0% incorrectly assigned vehicles, and then the color continually changes to the light-yellow, indicating 100% correctly assigned vehicles from all tested data.

The main advantage of the proposed hybrid CNN over conventional approaches is the combination of image and sensoric datasets. This solution has potential to afford comprehensive representation of the vehicle under test, with an in-built back-up verification. As a result, predicted accuracy is higher, as we demonstrated in this work. By prioritizing the individual branches, we can manage classification accuracy and improve the robustness of the system. The overall classification can be increased as both data sources are originally independent, and through the hybrid solution, they can trade-off mutual shortcomings of individual techniques. On the other hand, there are few challenges to be addressed with such hybrid CNN solution. The effective integration from both sources can be problematic in some situations. This is due to the fact that, in general case, the image and sensoric datasets can have contradictory characteristics, scales, and requirements for complex processing in one CNN system. Moreover, the categorization results can be affected by image quality and by the number of images used in each class and each branch. By comparing datasets examples from Fig. 4 (sensoric dataset) and 6 (image dataset), respectively, the image quality is different, which may impact the overall results as well. Nevertheless, improved datasets for both branches can facilitate further enhancement in the final classification product.

From an application point of view, the proposed solution for vehicle categorization can be potentially employed for road transport monitoring within Smart cities, performing an automatic counting or separation. This also includes dynamic management to optimize different vehicle classes on the road or road resources. Moreover, another potential application field can be found in vehicle speed monitoring or weight-in-motion function, where heavy vehicles can be diverted from the main road towards a destined parking lots or to evaluate the vehicle overload in- or on-side.

IV. CONCLUSION

In summary, we demonstrated a novel neural network architecture to improve vehicle classification. More specifically, we used clustered datasets from a hybrid technological platform based on a conventional CCTV system and FBG sensor array. A novel clustered CNN-based classification system (with both sensor and image-based datasets) improved vehicle detection accuracy, obtaining precision levels between 90% - 97%, and including a back-up verification with respect to each technology we used. Collecting data from two technologically independent platforms can also trade-off the intrinsic shortcoming of both conventionally decoupled

solutions. The hybrid CNN concept substantially enhances the detection accuracy for correct vehicle classification and opens up a way for effective vehicle classification by leveraging available and inexpensive technologies.

DATA AVAILABILITY
The data that support the findings of this work are available from the authors upon reasonable request. Data contact person: Dr. Patrik Kamencay (patrik.kamencay@uniza.sk).

REFERENCES

[1] R. Buyya and A. V. Dastjerdi, *Internet of Things: Principles and Paradigms*, 1st ed. San Francisco, CA, USA: Morgan Kaufmann, 2016.

[2] M. D. F. F. Domingues and A. Radwan, "Optical fiber sensors in IoT," in *Optical Fiber Sensors for IoT and Smart Devices*. New York, NY, USA: Springer, 2017, pp. 73–86, doi: 10.1007/978-3-319-47349-9_5.

[3] S. Aleksic, "A survey on optical technologies for IoT, smart industry, and smart infrastructures," *J. Sensor Actuator Netw.*, vol. 8, no. 3, p. 47, Sep. 2019, doi: 10.3390/jsan8030047.

[4] D. Gil, M. Johnsson, H. Mora, and J. Szymański, "Review of the complexity of managing big data of the Internet of Things," *Complexity*, vol. 2019, Feb. 2019, Art. no. 4592902, doi: 10.1155/2019/4592902.

[5] R. Muñoz, R. Vilalta, N. Yoshikane, R. Casellas, R. Martínez, T. Tsuritani, and I. Morita, "Integration of IoT, transport SDN, and edge/cloud computing for dynamic distribution of IoT analytics and efficient use of network resources," *J. Lightw. Technol.*, vol. 36, no. 7, pp. 1420–1428, Apr. 1, 2018, doi: 10.1109/JLT.2018.2800660.

[6] H. Baali, H. Djelouat, A. Amira, and A. F. Bensaali, "Empowering technology enabled care using *IoT* and smart devices: A review," *IEEE Sensors J.*, vol. 18, no. 5, pp. 1790–1809, Mar. 2018, doi: 10.1109/JSEN.2017.2786301.

[7] M. Piccarozzi, B. Aquilani, and C. Gatti, "Industry 4.0 in management studies: A systematic literature review," *Sustainability*, vol. 10, no. 10, p. 3821, Oct. 2018, doi: 10.3390/su10103821.

[8] D. Zhang and J. He, "Super-resolution reconstruction of low-resolution vehicle plates: A comparative study and a new algorithm," in *Proc. 7th Int. Congr. Image Signal Process.*, Oct. 2014, pp. 359–364, doi: 10.1109/CISP.2014.7003806.

[9] C.-Y. Hsieh and Y.-S. Wang, "Traffic situation visualization based on video composition," *Comput. Graph.*, vol. 54, pp. 1–7, Feb. 2016, doi: 10.1016/j.cag.2015.07.007.

[10] L. Wenjun, Z. Yiming, and A. L. Wenbin, "A wireless and real-time monitoring system design for car networking applications," *Int. J. Antennas Propag.*, vol. 2013, Nov. 2013, Art. no. 298543, doi: 10.1155/2013/298543.

[11] Y. Huang, L. Wang, Y. Hou, W. Zhang, and Y. Zhang, "A prototype IoT based wireless sensor network for traffic information monitoring," *Int. J. Pavement Res. Technol.*, vol. 11, no. 2, pp. 146–152, Mar. 2018, doi: 10.1016/j.ijprt.2017.07.005.

[12] L. Xiao and Z. Wang, "Internet of Things: A new application for intelligent traffic monitoring system," *J. Netw.*, vol. 6, no. 6, p. 887, Jun. 2011, doi: 10.4304/jnw.6.6.887-894.

[13] P. A. Shinde and Y. B. Mane, "Advanced vehicle monitoring and tracking system based on raspberry pi," in *Proc. IEEE 9th Int. Conf. Intell. Syst. Control (ISCO)*, Jan. 2015, pp. 1–6, doi: 10.1109/ISCO.2015.7282250.

[14] S. Dhingra, R. B. Madda, R. Patan, P. Jiao, K. Barri, and A. H. Alavi, "Internet of Things-based fog and cloud computing technology for smart traffic monitoring," *Internet Things*, vol. 14, Jun. 2021, Art. no. 100175, doi: 10.1016/j.iot.2020.100175.

[15] Q. Li, H. Cheng, Y. Zhou, and G. Huo, "Road vehicle monitoring system based on intelligent visual Internet of Things," *J. Sensors*, vol. 2015, pp. 1–16, Jul. 2015, doi: 10.1155/2015/720308.

[16] X.-Y. Lu, P. Varaiya, R. Horowitz, Z. Guo, and J. Palen, "Estimating traffic speed with single inductive loop event data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2308, no. 1, pp. 157–166, Jan. 2012, doi: 10.3141/2308-17.

[17] S.-T. Jeng and L. Chu, "Tracking heavy vehicles based on weigh-in-motion and inductive loop signature technologies," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 632–641, Apr. 2015, doi: 10.1109/TITS.2014.2333003.

[18] C. Xu, Y. Wang, X. Bao, and F. Li, "Vehicle classification using an imbalanced dataset based on a single magnetic sensor," *Sensors*, vol. 18, no. 6, p. 1690, 2018, doi: 10.3390/s18061690.

[19] V. Markevicius, D. Navikas, M. Zilys, D. Andriukaitis, A. Valinevicius, and M. Cepenas, "Dynamic vehicle detection via the use of magnetic field sensors," *Sensors*, vol. 16, no. 1, p. 78, Jan. 2016, doi: 10.3390/s16010078.

[20] B. Santoso, B. Yang, C. L. Ong, and Z. Yuan, "Traffic flow and vehicle speed measurements using anisotropic magnetoresistive (AMR) sensors," in *Proc. IEEE Int. Magn. Conf. (INTERMAG)*, Apr. 2018, pp. 1–4, doi: 10.1109/INTMAG.2018.8508869.

[21] D. Xu, S. Huang, Q. Lei, and X. Cheng, "Preparation, properties and application research of piezoelectric traffic sensors," *Proc. Eng.*, vol. 27, pp. 269–280, Jan. 2012, doi: 10.1016/j.proeng.2011.12.453.

[22] G. Guido, V. Gallelli, D. Rogano, and A. Vitale, "Evaluating the accuracy of vehicle tracking data obtained from unmanned aerial vehicles," *Int. J. Transp. Sci. Technol.*, vol. 5, no. 3, pp. 136–151, Oct. 2016, doi: 10.1016/j.ijtst.2016.12.001.

[23] I. Mardare, I. Tiţa, and R. I. Pelin, "Researches regarding a pressure pulse generator as a segment of model for a weighing in motion system," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 147, Aug. 2016, Art. no. 012060, doi: 10.1088/1757-899X/147/1/012060.

[24] W. Zhang, Q. Wang, and C. Suo, "A novel vehicle classification using embedded strain gauge sensors," *Sensors*, vol. 8, no. 11, pp. 6952–6971, Nov. 2008, doi: 10.3390/s8116952.

[25] R. Bajwa, E. Coleri, R. Rajagopal, P. Varaiya, and C. Flores, "Development of a cost-effective wireless vibration weigh-in-motion system to estimate axle weights of trucks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 6, pp. 443–457, Jun. 2017, doi: 10.1111/mice.12269.

[26] M. Al-Tarawneh, Y. Huang, P. Lu, and R. Bridgelall, "Weigh-in-motion system in flexible pavements using fiber Bragg grating sensors. Part A: Concept," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 5136–5147, Dec. 2020, doi: 10.1109/TITS.2019.2949242.

[27] M. Frniak, M. Markovic, P. Kamencay, J. Dubovan, M. Benco, and M. Dado, "Vehicle classification based on FBG sensor arrays using neural networks," *Sensors*, vol. 20, no. 16, p. 4472, Aug. 2020, doi: 10.3390/s20164472.

[28] H. Liu, J. Ma, T. Xu, W. Yan, L. Ma, and X. Zhang, "Vehicle detection and classification using distributed fiber optic acoustic sensing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1363–1374, Feb. 2020, doi: 10.1109/TVT.2019.2962334.

[29] F. Xiao, D. Meng, Y. Yu, Y. Ding, L. Zhang, G. S. Chen, W. Zatar, and J. L. Hulsey, "Estimation of vehicle-induced bridge dynamic responses using fiber Bragg grating strain gages," *Sci. Prog.*, vol. 103, no. 1, Jan. 2020, Art. no. 003685041987420, doi: 10.1177/0036850419874201.

[30] J. Nedoma, M. Fajkus, R. Martinek, J. Vanus, S. Kepak, R. Kahankova, R. Jaros, D. Cvejn, and M. Prauzek, "Analysis of the use of fiber-optic sensors in the road traffic," *IFAC-PapersOnLine*, vol. 51, no. 6, pp. 420–425, 2018, doi: 10.1016/j.ifacol.2018.07.117.

[31] M. Fridrich, M. Fajkus, P. Mec, J. Nedoma, M. Kostelansky, and E. Bednar, "Portable optical fiber Bragg grating sensor for monitoring traffic density," *Appl. Sci.*, vol. 9, no. 22, p. 4796, Nov. 2019, doi: 10.3390/app9224796.

[32] J. K. Sahota, N. Gupta, and D. Dhawan, "Fiber Bragg grating sensors for monitoring of physical parameters: A comprehensive review," *Opt. Eng.*, vol. 59, no. 6, p. 1, Jun. 2020, doi: 10.1117/1.OE.59.6.060901.

[33] M. Al-Tarawneh, Y. Huang, P. Lu, and D. Tolliver, "Vehicle classification system using in-pavement fiber Bragg grating sensors," *IEEE Sensors J.*, vol. 18, no. 7, pp. 2807–2815, Apr. 2018, doi: 10.1109/JSEN.2018.2803618.

[34] M. Fajkus, M. Fridrich, J. Nedoma, R. Kahankova, R. Martinek, E. Bednar, and J. Kolarik, "PDMS-FBG-based fiber optic system for traffic monitoring in urban areas," *IEEE Access*, vol. 8, pp. 127648–127658, 2020, doi: 10.1109/ACCESS.2020.3006985.

[35] K. Yuksel, D. Kinet, K. Chah, and C. Caucheteur, "Implementation of a mobile platform based on fiber Bragg grating sensors for automotive traffic monitoring," *Sensors*, vol. 20, no. 6, p. 1567, Mar. 2020, doi: 10.3390/s20061567.

[36] U. Senkans, J. Braunfelds, I. Lyashuk, J. Porins, S. Spolitis, and V. Bobrovs, "Research on FBG-based sensor networks and their coexistence with fiber optical transmission systems," *J. Sensors*, vol. 2019, pp. 1–13, Nov. 2019, doi: 10.1155/2019/6459387.

[37] *Tire Size Explained: What the Numbers Mean | Les Schwab*. Accessed: Sep. 9, 2021. [Online]. Available: https://www.lesschwab.com/article/tire-size-explained-reading-the-sidewall.html

[38] M. Yamashita and S. Yamanaka, "Dust resulting from tire wear and the risk of health hazards," *J. Environ. Protection*, vol. 4, no. 6, pp. 509–515, 2013, doi: 10.4236/jep.2013.46059.

[39] L. Faes, X. Liu, S. K. Wagner, D. J. Fu, K. Balaskas, D. A. Sim, L. M. Bachmann, P. A. Keane, and A. K. Denniston, "A clinician's guide to artificial intelligence: How to critically appraise machine learning studies," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 7, Feb. 2020, doi: 10.1167/TVST.9.2.7.

[40] *Confusion Matrix—Clearly Explained | by Indhumathy Chelliah | Towards Data Science*. Accessed: Sep. 10, 2021. [Online]. Available: https://towardsdatascience.com/confusion-matrix-clearly-explained-fee63614dc7

PATRIK KAMENCAY (Member, IEEE) received the Ph.D. degree in telecommunications from the University of Žilina, Slovakia, in 2012. He is currently an Associate Professor with the Department of Multimedia and Information-Communication Technology, University of Žilina. His research interests include digital processing areas as diverse as image data collection, processing and machine learning methods, and neural networks, with a focus on image data segmentation, classification, 2-D/3-D recognition, and 3-D reconstruction.

MIROSLAV MARKOVIC received the Ph.D. degree in telecommunications from the University of Žilina, Slovakia, in 2009. He is currently an Assistant Professor with the Department of Multimedia and Information-Communication Technology, University of Žilina. His research interests include reservation protocols in high-speed transparent optical networks, optical communication networks and fiber-optic sensors, neural networks, smart cloud infrastructures, and web applications.

JOZEF DUBOVAN received the Ph.D. degree in telecommunications from the University of Žilina, Slovakia, in 2009. His research is oriented on the mitigation of polarization mode dispersion in high-speed optical networks. Since 2005, he has been an Assistant Professor with the Department of Multimedia and Information-Communication Technology, University of Žilina. His recent research interests include fiber-optics communication systems and fiber-based optical sensors.

MILAN DADO (Senior Member, IEEE) is currently a Full Professor with the Department of Multimedia and ICT, University of Žilina. He has been actively involved in European research and education programs (TEMPUS, COST, LEONARDO, Socrates, 5th, 6th, and 7th Framework Programs, H2020, European University association projects) and has managed national projects related to information and communication technologies, intelligent transportation systems, regional innovation strategies, and e-learning. The main milestones of his international activities were staying abroad, e.g., a two-month stay with York University, Canada, Northern Telecom and Bell, Canada, in 1993, a six-month stay with the Royal Institute of Technology, Stockholm, in 1990, and a six-month stay with the Vienna University of Technology, from 1981 to 1982. He has visited many other foreign institutions during the last 30 years.

DANIEL BENEDIKOVIC (Member, IEEE) received the Ph.D. degree in telecommunications from the University of Žilina, Slovakia, in 2015. During his study, he completed several educational and scientific internships with Aalto University, Finland, the University of Málaga, Spain, and the National Research Council Canada, Canada. From 2015 to 2020, he was a Postdoctoral Researcher with Centre de Nanoscience et de Nanotechnology, CNRS, Université Paris-Saclay, France. He is currently an Associate Professor with the Department of Multimedia and Information-Communication Technologies, University of Žilina. His current research interests include the development, design, and optical characterization of active and passive silicon photonic components, sub-wavelength grating nanostructures, fiber-optic communication systems, and optical sensors.

• • •