IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 March 2024, accepted 24 March 2024, date of publication 28 March 2024, date of current version 19 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3382931

RESEARCH ARTICLE

Road Type Classification Using Time-Frequency Representations of Tire Sensor Signals

TAMÁS DÓZSA[1,2], VEDRAN JURDANA[3], SANDI BARESSI ŠEGOTA[3], JÁNOS VOLK[4], JÁNOS RADÓ[5], ALEXANDROS SOUMELIDIS[2], (Member, IEEE), AND PÉTER KOVÁCS[1]

[1]Department of Numerical Analysis, Faculty of Informatics, Eötvös Loránd University, 1117 Budapest, Hungary
[2]Systems and Control Laboratory, HUN-REN Institute for Computer Science and Control, 1111 Budapest, Hungary
[3]Department of Automation and Electronics, Faculty of Engineering, University of Rijeka, 51000 Rijeka, Croatia
[4]Nanosensors Laboratory, HUN-REN Centre for Energy Research, 1121 Budapest, Hungary
[5]Physics Department, Lakehead University, Thunder Bay, ON P7B 5E1, Canada

Corresponding author: Tamás Dózsa (dozsa.tamas@sztaki.hu)

This work was supported in part by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, through the NVKDP-2021, K_23 "OTKA," and TKP2021-NVA-03 Funding Schemes, under Project C1748701, Project K146721, and Project TKP2021-NVA-29; in part by the University of Rijeka under Project uniri-mladi-tehnic-23-2; and in part by the European Union within the framework of the National Laboratory for Autonomous Systems. under Grant RRF-2.3.1-21-2022-00002. The work of Péter Kovács was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

ABSTRACT Smart tire technologies offer a novel sensing methodology for vehicle environment perception by providing direct measurements of tire dynamics parameters. This information can be utilized in advanced driver assistance systems as well as autonomous vehicle control to enhance vehicle performance and safety. Considering these criteria, we develop algorithms for categorizing road types based on tire sensor signals. Road differentiation is a complex task due to the non-linear and non-stationary nature of the measured tire signals. To address this challenge, we fuse time-frequency distributions with machine learning approaches. Maintaining the robustness of the predictions, we integrate our own measurement system into a Nissan Leaf test vehicle and collect data involving diverse environmental factors and operational conditions, mimicking real-world scenarios. We showed that by experiments our predictions strongly correlate with road quality, which can be utilized in automatic vehicle control, such as intelligent speed adaptation.

INDEX TERMS Vehicle environment perception, smart tires, road type estimation, time-frequency representation, variable projection, deep learning.

I. INTRODUCTION

Principles of sustainability and the advancement of autonomous vehicle technology are revolutionizing the automotive industry, paving the way for eco-friendly and intelligent transportation. To achieve these goals, next-gen driving solutions are integrating state-of-the-art technologies, such as artificial intelligence, sensor fusion, and connectivity. In this progression, the digital evolution of tires has contributed significantly to enhancing vehicle performance and safety. In fact, smart tire technologies, such as tire pressure monitoring systems (TPMS) are already in commercial use, which help to avoid accidents and tire defects by warning the driver about losing tire pressure. Besides TPMS, there is a great potential in intelligent tire technologies that can be utilized in various vehicle applications including tire force estimation, tire condition monitoring, vehicle state prediction (e.g., skidding, hydroplaning), road condition characterization (e.g., road quality classification, snow detection).

Numerous researchers are actively experimenting with the development and implementation of novel sensing systems in vehicle tires [1]. Optical measurement systems consist of a light emitter focusing on a detector, which are placed on the inner liner and on the rim of the tire, respectively. These two components must be aligned precisely to provide reliable measurements of tire deformation. However, the tire can shift on the rim due to sudden speed changes caused by hard braking. Therefore, recalibration is required from time to time, which is difficult to implement in

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

practice [2]. The lateral deflection of the tire's contact patch can be determined using accelerometers, which are both energy-efficient and inexpensive sensors. However, apart from the lateral acceleration, the measured signal includes undesired components resulting from rotational, vibrational, and gravitational accelerations [1]. Strain-based measurements are directly related to tire deflection, which can be obtained by various methodologies including piezo electric materials [3], electromagnetism [4], acoustic wave reflection [5], and foil strain gauges [6]. These strain measurements are less sensitive to the previously mentioned noise components [7]; however, extensive tire deformation can result in their detachment from the inner liner. To address these limitations, we recently developed a piezoresistive Micro Electro-Mechanical Systems (MEMS) sensor [8], and showed that there is a strong correlation between the measured electrical resistance signal and the direction as well as the magnitude of the mechanical forces acting on the tire.

Besides the tire sensing technologies employed, it is equally important to process the measured data. However, this requires the design of novel signal processing methods capable of capturing the nonlinearities and the complexities inherent in tire structure. In order to minimize the production cost of tire sensing systems, computationally simple algorithms are preferred that can be implemented on embedded hardware. Although this criterion limits the use of resource-intensive deep learning (DL) methods in tire sensing systems, there are already pioneering works aiming to utilize recent advances in machine learning for smart tire applications. In fact, Maurya et al. [6] used a small scale two-layer feed-forward neural network (NN) with ten neurons in its hidden layer to estimate the tire pressure based on 3D printed strain sensors. Then, Strano et al. [9] utilized a physical model [10] for strain-based intelligent tires to generate virtual data for training supervised machine learning methods. This alleviates the data demand of DL methods, although their approach lacked numerical or experimental evaluation. Recently, Xu et al. [11] introduced an intelligent tire system equipped with a tri-axial acceleration sensor. To estimate tire-road forces they trained a simple NN with a single hidden layer by using measurements observed on a test platform. Note that previous works have involved simulated data and laboratory measurements, typically conducted under limited operational and environmental conditions. In contrast, our sensing system [8] is fully integrated into a vehicle, enabling real-world measurements to be conducted on public roads.

In this research, we progress toward practical implementations of our intelligent tire sensing systems [8] by categorizing road types according to the frequency of detected surface anomalies. We will show that there is a strong correlation between the number of positive predictions and the road quality. This information can be employed in advanced driver assistance systems, advising a reduction in vehicle speed when encountering low-quality road conditions. To this end, we combine time-frequency distributions (TFDs) with

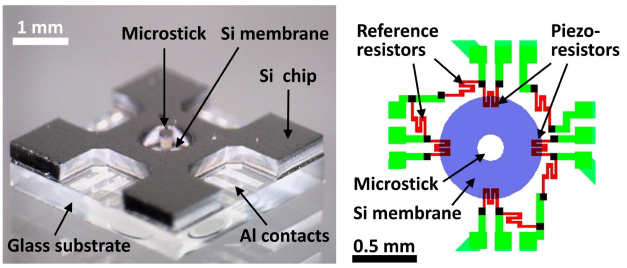


FIGURE 1. Digital microscope image (a) and corresponding circuit layout (b) of the piezoresistive force sensor used in the experiments.

lightweight machine learning methods that are suitable for analyzing non-stationary signals and can be implemented on embedded hardware [12]. In contrast to our previous work [8], we designed new measurements to investigate the generalization ability of the proposed NN models. This way, environmental factors (e.g., temperature, humidity) and operational conditions (e.g., tire pressure, speed) may vary between the training and test sets, mimicking a real-world use case.

To summarize, the main contributions outlined in this paper are as follows:

- Designed new measurements using diverse road data.
- Combined time-frequency distributions with thresholding and machine learning for anomaly detection.
- Demonstrated strong correlation between (true/false) positive predictions and road quality.

Finally, we investigate the potential to enhance prediction accuracy by leveraging more sophisticated end-to-end deep learning models, capable of learning representations from raw data without manual feature engineering. To do this, we employ the VGG16, a deep convolutional neural network architecture. Our experiments indicate that the proposed time-frequency (TF) features and the corresponding classifier are close to optimal in terms of accuracy and computational complexity.

II. MEASUREMENT SETUP

In this work, data collection was done by the measurement system proposed in [8]. That is, a piezoresistive force sensor was implanted into the inner wall of the front tires of a Nissan Leaf test vehicle using a patented implantation method [13]. The sensor and the tire implantation method were developed at the Nano Sensors laboratory of the Hungarian Research Network's Institute for Physics and Material Science (HUN-REN MFA). The sensor used in our experiments is illustrated in Fig. 1. The sensor produces signals along four bridges, which represent change in electrical resistance in response to mechanical forces applied to the "microstick" mounted in the middle of the sensor. In ideal, laboratory environments, a linear combination of the sensor outputs can produce the 3-D force applied to the microstick (see e.g. [8]).

Unfortunately, the sensor needs to be covered in multiple protective layers when implanted into the tire [8]. In addition

FIGURE 2. The Nissan Leaf test vehicle that was used to collect the measurements for this study.

to this, when used in our measurement setup, the sensor was excited by tire wall deformations, which are very difficult to model [8], [12]. Thus, the relationship between the sensor's output and the mechanical forces acting between the tires and the road surface are highly nonlinear. Because of this nonlinear behavior, solving autonomous driving related problems such as recognizing road surface abnormalities, estimating road roughness, or estimating the tire contact forces based on the sensor's output requires the use of nontrivial signal processing tools.

In our experiments, the sensor was implanted into the two front tires of a Nissan Leaf test vehicle (see Fig. 2) provided by the Hungarian Research Network's Institute for Computer Science and Control (HUN-REN SZTAKI).

Wireless read out electronics were developed to access the sensor's measurements in real time using a nominal sampling rate of 1000 Hz. For a more thorough specification of the measurement hardware, we refer to [8]. Data was collected along the four bridges from a single sensor, therefore from the two front tires altogether 8 1-D signals were recorded. To simplify our models, in this work, we only relied on the signals recorded along the second bridge of the sensor implanted into the front left tire. The experimental results in Section V correspond the use of these signals. We repeated a number of our experiments using signals recorded along different bridges and acquired very similar results.

Since only a single sensor was present in the tire, the recorded signals presented "quasi compact" and "quasi periodic" behavior during each tire revolution. This means, that the signals contained the most information, when the sensor was closest to the ground and tire deformation around the sensor was maximal. In turn, the data points recorded during the part of each cycle when tire deformation was minimal, did not contain useful information for estimating road roughness. For this reason, the recorded measurements were segmented into single tire revolutions. This approach ensured, that each data segment (corresponding to a full revolution) contained a part, when the tire was significantly deformed and useful information could be deduced from it. Since the test vehicle was not equipped with sensors capable of measuring wheel angle, we relied on a software based segmentation method. In particular, we adapted the ECG segmentation method in [14] for use with the tire sensor data according to [8]. Fig. 3 illustrates "raw" recorded signals by

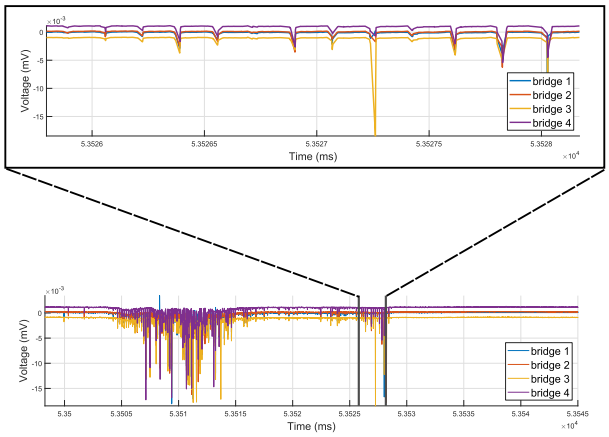


FIGURE 3. BOTTOM: signals recorded by the tire sensor. TOP: A portion of the entire recording. The signals exhibit "quasi periodic" behavior, with each period having "quasi compact" support.

TABLE 1. Number of normal and abnormal (recorded on rough surface).

	Normal	Abnormal
Number	282	235
Percentage	54.55%	45.45%

the sensor and a segmented "quasi periods" representing a tire revolutions.

Our data was recorded on the public roads of Budapest. In total, 12 measurements were recorded on roads with varying amount of abnormalities. Based on the number of abnormalities encountered on a given road, each measurement was assigned to a road type class. To ensure a fair comparison, the test vehicle stayed within the speed range of 0 to $\sim$ 50 km/h on all considered road classes. The read-out electronics used to collect the data employed a 1000 Hz sampling frequency. Considering a speed of 50 km/h and the fact that the test vehicle was equipped with tires having a circumference of 199 cm, this resulted in a single tire revolution being represented by 143 data points. Thus, even at the highest encountered velocity, enough data was acquired during each tire revolution to apply the proposed signal processing methods (see e.g. [8]). Two of these measurements were studied in more detail, namely tire revolutions were labeled as "normal" or "abnormal" depending on whether the vehicle encountered any abnormality during the revolution. The labeling was done according to [8] and the results were visually confirmed. These measurements were used to optimize parameterized models which were able to recognize road abnormalities [8] such as bumps and potholes. In this work, we focus on estimating road roughness in general instead of individual abnormalities. We achieve this by training predictive models on the aforementioned two measurements to detect abnormal tire revolutions. Then, we apply the trained models to the remaining 10 measurements and compare the frequency of detected abnormalities with the roughness of the surface (see Section V-B). The two measurements used to construct the training data contained 517 tire revolutions in total using roughly the same amount

of normal and abnormal signals [8]. Specifically, the ratio of ''normal'' and ''abnormal'' revolutions is shown in Table 1.

The remaining 10 measurements were used as the test data for the investigated methods. Similarly to the training data, these measurements were segmented into single tire revolutions, however the segmented data was left unlabeled. This approach was chosen because of the following reasons. Firstly, since all measurements were recorded on the public streets of Budapest very precise labeling of each individual tire revolution signal as normal and abnormal is difficult. For the training data 1, we used the method [15] to label the revolutions followed by a meticulous visual examination of the signals. For the remaining 10 measurements however, such a thorough examination was not possible due to the diverse nature of the surfaces and the length of the measurements. Secondly, the objective in this paper was not to recognize individual road abnormalities of the road surface, rather to propose robust methods capable of describing the road quality in general. To this end, we described each of the 10 measurements in the test set with a single number that describes the expected ratio of abnormal tire revolutions in it (see Section V for specifics). We then compared these ratios with the frequency of the

III. SIGNAL DESCRIPTION

Our current measurements were acquired with a single sensor implanted into each tire. Since the recorded signals correspond to tire deformations as the wheels of the vehicle roll, they exhibit very characteristic morphology. That is, the signals corresponding to single tire revolutions are quasi periodic (with each tire revolution representing a period) and quasi compactly supported. The signals contain a single peak corresponding to the part of the revolution when the sensor was closest to the ground and the tire deformation is maximal. Clearly, the recorded signals contain the most information about the road surface around this peak. In order to ensure that each input signal segment to our models has meaningful information about the environment, we segmented the measurements into single tire revolutions using the methodology discussed in [8]. We plan to equip our test vehicle with wheel angle measuring devices, therefore software based segmentation methods will not be needed in our future experiments.

IV. LEARNING FRAMEWORK

A. FEATURE EXTRACTION

1) TIME-FREQUENCY DISTRIBUTIONS

In the study of real-life non-stationary signals, the presence of multiple components and noise contamination necessitates the utilization of sophisticated analytical techniques that extend beyond conventional time-domain or frequency-domain analysis. Spectral analysis based on the Fourier transform (FT) offers insights into the frequency components present within a signal. However, the FT does not retain information regarding the temporal localization of these components. To address this limitation and provide a comprehensive understanding of both the frequency content and its temporal characteristics, advanced tools have been developed to enable concurrent signal analysis in the joint time-frequency domain, which has led to the development of two-dimensional TFDs [16], [17].

Diverse variants of TFD methods have been employed in the processing of non-stationary signals. The choice of these methods is contingent upon the particular application, with the objective of maximizing resolution and concentration of the useful components, while effectively suppressing interference [16], [17].

The spectrogram, $SP(t,f)$, has traditionally been the preferred choice in practical applications due to its simplicity. It visually represents the square modulus of the short-time Fourier transform (STFT), where the FT is applied to a signal $s(\tau)$ that has been localized using a sliding window $h(\tau)$ [18]:

$$STFT(t,f) = \int_{-\infty}^{\infty} s(\tau)h(\tau - t)e^{-j2\pi f\tau}d\tau, \tag{1}$$

$$SP(t,f) = \left|STFT(t,f)\right|^2. \tag{2}$$

The trade-off between time and frequency resolution is controlled with the window length – shorter windows provide better localization in time, while longer windows increase frequency resolution. Typically, Hamming window is used in practice [16].

In light of the constraints posed by the SP, the quadratic time-frequency distributions (QTFDs) have been utilized in many applications. The QTFDs under consideration in this study are categorized within Cohen's class, and they leverage the concept of the analytic signal. The analytic signal, denoted as $z(t)$, of the initial real-valued signal $s(t)$, is established using the Hilbert transform, $\mathcal{H}$, as [16]:

$$z(t) = s(t) + j\mathcal{H}\{s(t)\}. \tag{3}$$

The fundamental QTFD of Cohen's class is the Wigner-Ville distribution (WV) defined as a FT of the instantaneous auto-correlation function of the analytic signal $z(t)$ [19]:

$$WV(t,f) = \int_{-\infty}^{\infty} z\left(t + \frac{\tau}{2}\right) z^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau}d\tau, \tag{4}$$

where $*$ denotes the complex conjugate. The WV distribution offers high-resolution time-frequency representation of signal components (auto-terms) but, due to quadratic nature, introduces undesirable interference terms (cross-terms) that hinder visual interpretation [16]. To mitigate these interference terms, effective smoothing of the WV using specific kernels is essential. These kernels are often conveniently designed within the Doppler-lag (ν, τ) domain. The use of the Doppler-lag domain enables the application of kernels to be conceptualized as a filtering operation. TFDs resulting from the application of kernel smoothing to the WV are commonly referred as reduced-interference distributions (RIDs) [16].

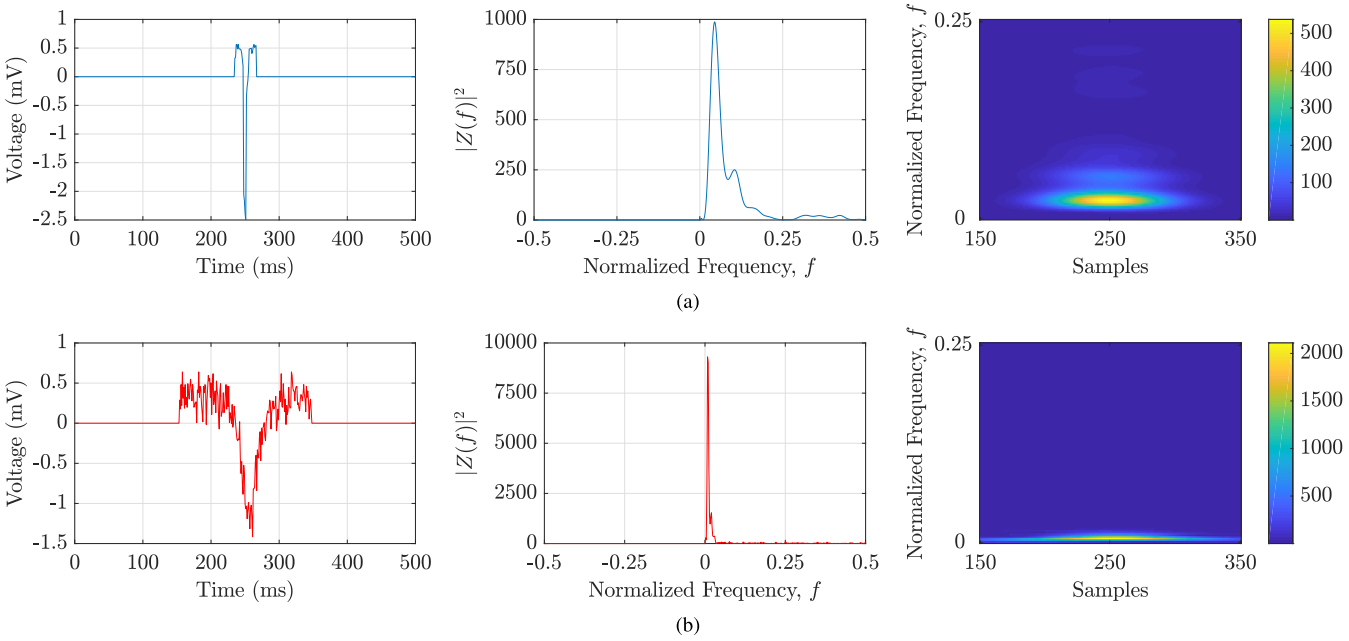


FIGURE 4. Time representation, energy spectrum and time-frequency representation using MWSP with $K = 1$ of a preprocessed signal segment labeled as: (a) "normal"; (b) "abnormal."

The cross-terms can be attenuated by using the smoothed-pseudo Wigner-Ville distribution (SPWV), which allows independent smoothing of the WV in both time and frequency domains by choosing the lengths of the windows $h(t)$ and $g(t)$ as follows [20], [21]:

$$SPWV(t,f) = \int_{-\infty}^{\infty} h(\tau) \int_{-\infty}^{\infty} g(u-t) z\left(u+\frac{\tau}{2}\right) \cdot z^*\left(u-\frac{\tau}{2}\right) du\, e^{-j2\pi f\tau} d\tau. \quad (5)$$

The RIDs constitute a group of TFDs defined with specialized kernel functions to effectively mitigate interference artifacts. Within this group, the Choi-Williams distribution (CW) [22] is notable for its utilization of an exponential kernel, as follows [23]:

$$CW(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\sqrt{\sigma}}{2\sqrt{\pi}|\tau|} e^{\frac{-u^2\sigma}{16\tau^2}} \cdot z\left(t+u+\frac{\tau}{2}\right) z^*\left(t+u-\frac{\tau}{2}\right) du\, e^{-j2\pi f\tau} d\tau. \quad (6)$$

The impact of CW on the trade-off between auto-term resolution and suppression of interference terms is governed by the scaling parameter $\sigma \in [0.1, 10]$ [22]. Specifically, higher σ values enhance auto-term resolution, bringing CW closer to the classical WVD, while lower values aid in reducing cross-terms.

The Butterworth distribution (BUD) represents an enhancement to the CW, realized through the implementation of a two-dimensional low-pass filter in the ambiguity domain, characterized by adjustable pass-band and transition region parameters. The BUD is defined as [24] and [25]:

$$BUD(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\sqrt{\sigma}}{2|\tau|} e^{\frac{-|u|\sqrt{\sigma}}{|\tau|}} z\left(t+u+\frac{\tau}{2}\right) \cdot z^*\left(t+u-\frac{\tau}{2}\right) du\, e^{-j2\pi f\tau} d\tau. \quad (7)$$

The Zhao-Atlas-Marks distribution (ZAM) employs a cone-shaped kernel to achieve a high time and frequency resolution while concurrently mitigating interference artifacts. The ZAM is defined as [26] and [27]:

$$ZAM(t,f) = \int_{-\infty}^{\infty} h(\tau) \int_{t-\frac{|\tau|}{2}}^{t+\frac{|\tau|}{2}} z\left(u+\frac{\tau}{2}\right) \cdot z^*\left(u-\frac{\tau}{2}\right) du\, e^{-j2\pi f\tau} d\tau. \quad (8)$$

However, the smoothing operation of the RIDs usually introduce bias and variance of the signal's auto-terms with respect its ideal position in the TF domain [16]. The reduction of both bias and variance while preserving the favorable attributes for non-stationary signals can be achieved through the adoption of TFD techniques incorporating the multi window (MW) approach with orthogonal Hermite window functions [28].

Hermite functions have gained acclaim due to their highly advantageous properties, including orthogonality and optimal concentration within the time-frequency domain for stable bias and variance estimations, and ease of implementation. The Hermite TFD emerges as a weighted sum of K spectrograms [28]:

$$MWSP(t,f) = \sum_{k=0}^{K-1} d_H(t) SP(t,f), \quad (9)$$

where $d_H(t)$ are weight coefficients. The spectrogram $SP(t,f)$ is obtained according to the Hermite window function $h_k(t)$ of kth order, given as [29] and [30]:

$$h_k(t) = \frac{1}{\|H_k\|_2} H_k(t)\sqrt{g(t)}, \qquad k \in \mathbb{N}, \tag{10}$$

where k denotes the order of the Hermite function ($k = 0, 1, \ldots, K$), $g(t) = e^{-t^2}$ is the Hermite weight function and $\|H_k\|_2$ is the usual L_2 norm of H_k.

In this work, feature extraction performance has been tested on the above defined TFDs. Fig. 4 depicts two example measurements from the dataset outlined in Table 1, showcasing their representations in the time domain, frequency domain, and combined time-frequency domains.

2) TIME, FREQUENCY AND TIME-FREQUENCY FEATURES

This research study delves into the comparative analysis of various sets of feature extraction techniques. To derive signal-related features, we expand upon both time-domain (t-domain) and frequency-domain (f-domain) characteristics to create composite features in the combined (t,f) domain. This approach harnesses the supplementary insights offered by TFDs. These features are derived from a discrete TFD denoted as $\rho[n,k]$, where n corresponds to discrete time samples, while k signifies discrete frequency bins. The dimensions of $\rho[n,k]$ are represented as $N \times M$, with N representing the length of the signal and M denoting the total count of frequency domain data points of $Z[k]$ ($Z[k]$ being the FT of the analytic associate $z[n]$ of a real signal $s[n]$. Drawing inspiration from encouraging outcomes in prior research [31], [32], [33], [34], [35], we now explore a range of signal-related features given in Table 2.

B. DEEP LEARNING CLASSIFICATION APPROACH

The signal processing approach outlined earlier yields three distinct datasets derived from a total of 517 signals given in Table 1. The first dataset comprises 8 combined features in both t- and f-domains, the second dataset consists of 8 features extracted from a TFD, and the third dataset comprises 517 TFDs. Since the extracted features for each signal are represented as vectors, while TFDs are represented as two-dimensional matrices, two distinct classification approaches are employed: a multilayer perceptron (MLP) and a deep convolutional neural network (CNN), specifically VGG16. Furthermore, to compare the performance of these models with our former work [8], we utilized a VP-NET variation designed for road anomaly detection (see e.g., Fig. 5).

1) CLASSIFICATION BASED ON T-DOMAIN, F-DOMAIN AND (T, F) FEATURES

An MLP artificial neural network (ANN) is used for creating t- and f-domain, and (t,f) classification models. MLP is a type of feed-forward ANN, consisting of an input layer, with the number of neurons equal to the dimension of the input vector (eight in the presented research); one or more hidden layers with a number of neurons; and a single output neuron which represents the output [36]. In the presented research all layers had an equal number of neurons. The output value of these neurons is determined by summing up the weighted outputs of neurons from the preceding layer, which is then processed by an activation function. The weights are adjusted during the training phase by minimizing a loss function. Hyperparameters of the neural network, such as the number of layers, neurons, and the choice of activation functions, are determined using the grid search procedure. In this method, potential values for the hyperparameters are set, and all possible combinations are tested [37]. The complete set of hyperparameters are given in the Table 3.

TABLE 2. Time, frequency and time-frequency features.

Feature	Formula
Standard mean (T_1)	$m_{(t)} = \frac{1}{N}\sum_n s[n]$
Standard variance (T_2)	$\sigma^2_{(t)} = \frac{1}{N}\sum_n \left(s[n] - m_{(t)}\right)^2$
Skewness (T_3)	$\gamma_{(t)} = \frac{1}{N\sigma^3_{(t)}}\sum_n \left(s[n] - m_{(t)}\right)^3$
Kurtosis (T_4)	$k_{(t)} = \frac{1}{N\sigma^4_{(t)}}\sum_n \left(s[n] - m_{(t)}\right)^4$
Coefficient of variation (T_5)	$c_{(t)} = \frac{\sigma_{(t)}}{m_{(t)}}$
Standard mean (TF_1)	$m_{(t,f)} = \frac{1}{NM}\sum_n\sum_k \rho[n,k]$
Standard variance (TF_2)	$\sigma^2_{(t,f)} = \frac{1}{NM}\sum_n\sum_k \left(\rho[n,k] - m_{(t,f)}\right)^2$
Skewness (TF_3)	$\gamma_{(t,f)} = \frac{1}{NM\sigma^3_{(t,f)}}\sum_n\sum_k \left(\rho[n,k] - m_{(t,f)}\right)^3$
Kurtosis (TF_4)	$k_{(t,f)} = \frac{1}{NM\sigma^4_{(t,f)}}\sum_n\sum_k \left(\rho[n,k] - m_{(t,f)}\right)^4$
Coefficient of variation (TF_5)	$c_{(t,f)} = \frac{\sigma_{(t,f)}}{m_{(t,f)}}$
Spectral flux (F_1)	$F_{(f)} = \sum_{k=1}^{M/2} \lvert Z_r[k] - Z_{r-1}[k]\rvert$
Spectral flatness (F_2)	$Fl_{(f)} = M\frac{\prod_{k=1}^{M}\lvert Z[k]\rvert^{M^{-1}}}{\sum_{k=1}^{M}\lvert Z[k]\rvert}$
Spectral entropy (F_3)	$Se_{(f)} = -\sum_{k=1}^{M}\frac{\lvert Z[k]\rvert^2}{\sum_{k=1}^{M}\lvert Z[k]\rvert^2}\log_2\frac{\lvert Z[k]\rvert^2}{\sum_{k=1}^{M}\lvert Z[k]\rvert^2}$
(t,f) flux (TF_6)	$F_{(t,f)} = \sum_{k=1}^{N-1}\sum_{k=1}^{M-1}\lvert\rho[n+1,k+1] - \rho[n,k]\rvert$
(t,f) flatness (TF_7)	$Fl_{(t,f)} = MN\frac{\prod_{n=1}^{N}\prod_{k=1}^{M}\lvert\rho[n,k]\rvert^{\frac{1}{NM}}}{\sum_{n=1}^{N}\sum_{k=1}^{M}\rho[n,k]}$
(t,f) Rényi entropy (TF_8), $\alpha = 3$	$R_{(t,f)} = \frac{1}{1-\alpha}\log_2\sum_{n=1}^{N}\sum_{k=1}^{M}\left(\frac{\rho[n,k]}{\sum_{n=1}^{N}\sum_{k=1}^{M}\rho[n,k]}\right)^{\alpha}$

TABLE 3. Tuned hyperparameters for MLP models with their possible values.

Hyperparameter	Possible Values	Count
Number of layers	1, 2, 3, 4, 5	5
Neurons per layer	8, 16, 32, 64, 128, 256	6
Activation Function	ReLU, Tanh, Sigmoid, Identity	4
Optimizer	adam, lbfgs	2
Learning Rate Type	constant, adaptive, inverse-scaling	3
Initial Learning Rate	0.5, 0.1, 0.01, 0.001, 0.00001	5

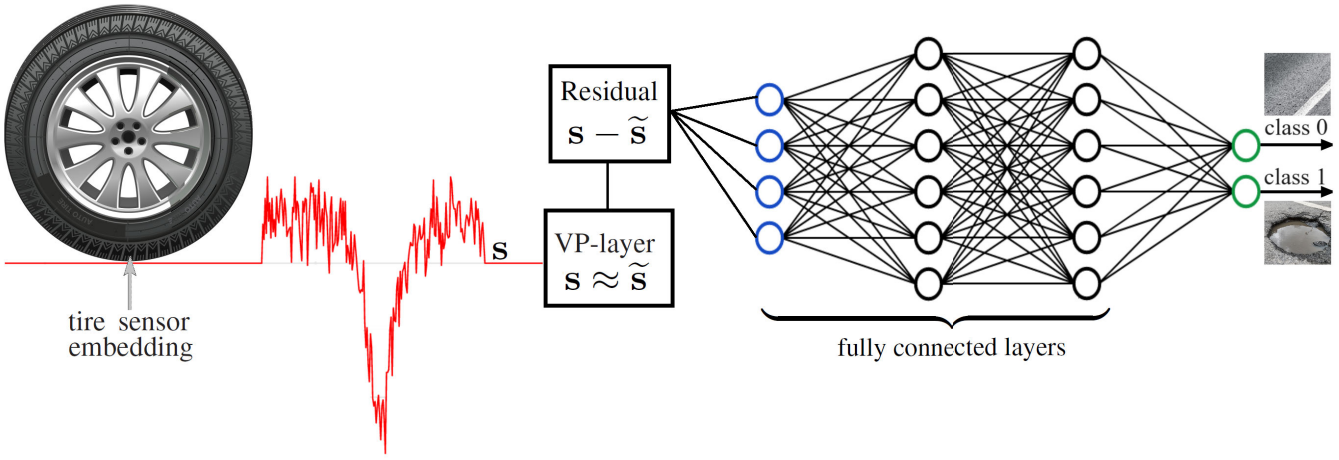


FIGURE 5. The VP-NET architecture that was trained for detecting road anomalies, such as bumps and potholes [8].

TABLE 4. Tuned hyperparameters for VGG16 model and their possible values.

Hyperparameter	Possible Values	Count
Optimizer	adam, nadam, adagrad, rmsprop	4
Epochs	10, 20, 30, 40, 50	5
Batch size	16, 32, 64	3

2) CLASSIFICATION BASED ON TFDS AS IMAGES

TFDs are represented by 2D matrices, facilitating the use of a convolutional neural network (CNN) for classification. A convolution layer makes use of several filters (i.e., kernels) to extract features from the input image(s). The filters in question store the weights of the CNN, and they are adjusted during the training process [38]. These types of networks are mainly developed to work with image data, commonly exhibiting higher performance compared to methods that have not been specifically designed in this manner [39]. CNNs are usually applied with the transfer learning methodology in which the model used for classification is not trained from randomized connection weights, but instead reuses weights that are the result of training the network on a separate dataset. We chose the VGG16 architecture [40], because this model provides a balance between size and expected performance. VGG16 has proven to be efficient in similar tasks which involve the processing of spectrograms [41], [42], [43]. Another advantage of the VGG16 architecture is its smaller size compared to large CNN-based models with similar performance, such as VGG19 and Xception [44]. Just as with MLPs, we utilized a grid search to determine the hyperparameters of the model, with the possible values detailed in Table 4.

C. MODEL EVALUATION

To ensure a fair comparison with the findings of [8], we adopted a 5-fold cross-validation approach [45]. Each fold comprised identical training and validation sets of signals as delineated in Table 1 of [8]. It's noteworthy that the training sets, consisting of 413 (80%) signals, and the test sets, comprising the remaining 104 (20%) signals, were mutually exclusive, with no data overlap between the two phases. Each combination of hyperparameters was evaluated in this manner. As the signal dataset from Table 1 is balanced, model performance was assessed using the accuracy (ACC) metric, defined as the ratio of correct predictions to all predictions made:

$$ACC = \frac{TN + TP}{TN + TP + FN + FP}, \tag{11}$$

where TP represents true positives (examples in class '1' correctly classified as '1'), TN denotes true negatives (examples in class '0' correctly classified as '0'), FP signifies false positives (examples in class '0' misclassified as '1'), and FN indicates false negatives (examples in class '1' misclassified as '0').

TABLE 5. Examined TFDs and their accuracy averaged over all threshold-based classification schemes using (t, f) features.

Distribution	Mean Accuracy	Highest	Lowest
WV	66.26%	70.61%	60.57%
ZAM	77.33%	81.99%	72.03%
BUD	80.64%	85.15%	75.39%
CW	83.01%	87.53%	78.13%
SPWV	87.92%	91.33%	83.78%
SP (Hamming window)	88.06%	91.39%	84.25%
MWSP ($K = 1$)	**88.32%**	**91.74%**	**84.51%**

V. EXPERIMENTS

Initially, we assessed the performance of TFDs as detailed in Section IV, employed for a threshold-based classification scheme applied to the extracted (t, f) features. Table 5 showcases the classification accuracy (11), specifically the mean accuracy achieved on the test set across all folds, alongside the lowest and highest accuracy scores observed within a single fold. These accuracy scores were derived by averaging threshold-based classification scores across all the 8 (t, f) features listed in Table 2. The outcomes underscore the influence of different TFDs on threshold-based classification

TABLE 6. Accuracy of threshold-based classification schemes.

Classifier	Mean Accuracy	Highest	Lowest
T_5	55.92%	61.50%	47.63%
TF_7	59.76%	67.31%	55.34%
T_1	64.01%	70.20%	60.22%
F_1	65.08%	72.11%	59.25%
T_3	67.72%	71.24%	61.20%
T_4	78.02%	80.61%	72.83%
TF_1	86.09%	90.29%	79.81%
Standard deviation (t-domain) [8]	86.92%	90.38%	79.81%
T_2	87.50%	93.21%	79.85%
TF_4	89.94%	92.31%	86.41%
TF_3	89.94%	93.27%	87.38%
Standard deviation using Hermite functions (t-domain) [8]	93.85%	96.15%	90.38%
TF_5	94.01%	95.19%	91.35%
TF_2	94.21%	98.06%	88.46%
TF_8	95.17%	97.09%	94.17%
F_3	95.62%	99.01%	90.44%
F_2	96.17%	98.11%	92.35%
TF_6	**97.30%**	**100%**	**95.19%**

TABLE 7. Accuracy of machine learning classification schemes.

Classifier	Mean Accuracy	Highest	Lowest
SVM [8]	93.65%	96.15%	92.31%
FCNN [8]	96.13%	99.02%	94.23%
CNN [8]	96.52%	97.12%	93.20%
MLP (t- and f-domain)	97.56%	**99.82%**	97.31%
VP-NET [8]	97.68%	99.04%	96.12%
MLP ((t,f) features)	98.12%	99.24%	97.65%
VGG16 (MWSP)	**98.27%**	99.23%	**97.99%**

accuracy, with the employment of the MWSP yielding the highest threshold-based classification accuracy across all metrics, i.e., mean, highest, and lowest. Consequently, the MWSP with $K = 1$ is selected for subsequent analyses and tests in the study.

Moreover, Tables 6 and 7 provide a detailed breakdown of the accuracy scores for each classification model employed in this study, allowing for a direct comparison with the results reported in [8]. Notably, when utilizing a computationally inexpensive threshold-based classifier, the advantages of employing spectral and the proposed time-frequency representations of our signals over those based solely on the time domain become evident. Actually, the incorporation of (t,f) flux features even outperforms computationally intensive approaches such as SVM, FCNN, and CNN across all accuracy metrics. Furthermore, the implementation of the proposed MLP and CNN (VGG16) deep networks yields further enhancements in accuracy scores, thereby emphasizing the efficacy of time-frequency signal representation in this particular application domain.

A. AI TRAINING RESULTS

The training was performed on two separate architectures. The MLPs were trained using a computation node consisting of AMD Epyc 7532 CPU (32C/64T@2.4 GHz) and 128 GB of RAM each; while the VGG16 models were trained using five NVIDIA Quadro RTX 6000 GPUs (576 Tensor cores, 24GB VRAM) on a node with Intel Xeon Gold 6240R (24C/48T@2.4 GHz) and 768 GB of RAM.

TABLE 8. Average times and standard deviations per sample for used classification models.

Method	Average time per sample [s]	Standard deviation [s]
MLP (t- and f-domain)	$5.06 \cdot 10^{-6}$	$2.77 \cdot 10^{-6}$
MLP ((t,f) features)	$2.78 \cdot 10^{-6}$	$1.69 \cdot 10^{-6}$
VGG16 (MWSP)	0.43	0.08

To assess the complexity of each classifiers, we consider the number of trainable parameters and the inference time. The layerwise structure for the classification model utilizing t- and f-domain features is specified as (8, 128, 128, 1), and for the model based on (t,f) features, the configuration involves (8, 32, 32, 32, 32, 1) neurons across its layers. Taking into account the weights and biases in each fully connected layer, the total number of trainable parameters amounts to 17, 920 for the first MLP model (utilizing t- and f-domain), and 3, 520 for the second MLP model (using (t,f) features). As for the VGG16 architecture, it encompasses a total of 138, 357, 544 trainable parameters.

The inference was performed using the previously described models with optimal hyperparameters on an Intel(R) Core(TM) i5-9400 CPU @ 2.90GHz, employing a single thread. A total of 6235 samples were used for inference. The average times and standard deviations per sample are presented in Table 8 for the utilized classification models.

The training times differ between different networks and hyperparameters. The total training time, across all hyperparameter combinations and folds was 47.32 hours for MLP networks. Considering the training was performed on five dataset folds and 3600 different hyperparameter combinations, this indicates that the average training time of a single MLP on the data in the presented research is approximately 48 seconds. Training of the considered VP-NET architecture was somewhat cheaper with an average of 36 seconds needed. Of course, this time will vary, with larger networks taking a longer time to train. Comparing this to the training of the VGG16 CNN, the total training time was 22 hours and 33 minutes across all folds, making the average time per epoch approximately 15.496 seconds.

B. ROAD TYPE ESTIMATION

We now examine the proposed methods' ability to estimate the type of the road surface based on the frequency of tire revolutions identified as "abnormal". We used the training set introduced in [8] containing a roughly equal number of normal and abnormal revolutions (see Table 1). We then applied the trained models to the unseen data measured on different road surfaces as discussed in Section II. For each of these measurements, we calculated the ratio of revolutions classified as abnormal and the total number of revolutions present in the measurement. That is, for each model and each measurement we calculated the following

TABLE 9. Measurement names and corresponding labels. These measurements were not used as training data.

Name	Label (2 classes)	Label (3 classes)
old cobblestone	1	1
normal asphalt	0	0.5
new (flat) cobblestone 1	0	0.5
new (flat) cobblestone 2	0	0.5
new (flat) cobblestone 3	0	0.5
smooth asphalt 1	0	0
rough surface	1	1
smooth asphalt 2	0	0
smooth asphalt 3	0	0
smooth asphalt 4	0	0

"hit percentage":

$$\text{hit}(\text{model, meas}) := \frac{P_{\text{model, meas}}}{N_{\text{meas}}}, \tag{12}$$

where $P_{\text{model, meas}}$ denotes the number of detected abnormal revolutions by the current model in the current measurement and N_{meas} is the total number of tire revolutions in the current measurement. Clearly, on a good quality road, a reliable model should provide a hit percentage close to 0. If the road contained a lot of abnormalities (bumps, potholes, etc.), then a good model will give a hit percentage near 1. For this reason, in our first experiment, we labeled the measurements discussed in Section II as 0 or 1, based on our empirical experience of the road type. In addition, to provide a more nuanced experiment, we also provided a three-class labeling of the road type measurements. In this case, measurements which were recorded on surfaces where occasional abnormalities were experienced (by the driver of the vehicle and through visual examination of the surfaces) were labeled with the value 0.5. The name of each measurement and the corresponding labels for each experiment are given in Table 9.

Ideally, the measurements would be conducted on roads, where the road quality is known with respect to the International Roughness Index (IRI) [46] or another well known standard measure for road profiles. The IRI roughness index measures the accumulated vertical movement of an ideal quarter-car model over a segment of the road. It is usually given in the units of slope (e.g., m/km). Unfortunately, due to the location of our test vehicle, we had to conduct our measurements on the public streets of Budapest. We were unable to find any freely available IRI (or any other recognized standard) labeling of these road profiles. We therefore had to rely on the more subjective labeling given in Table 9. It is important to mention that every measurement used in this study was recorded on various high-quality paved public roads in Budapest. Recognizing unpaved surfaces (such as dirt roads) could also be an important application of the presented methods. However, since the tire deformation profile varies significantly with pavement characteristics (see e.g. [47]), it is reasonable to assume that recognizing unpaved surfaces presents an easier challenge than the investigated cases.

TABLE 10. Examining the correlation between hit percentages and empirical road quality on our measurements using threshold-based classifiers that utilizes time-frequency features.

Feature (TF)	Accuracy (train)	Corr (2-class)	Corr (3-class)
Mean	86.27%	0.61	0.764
Coeff of var	95.55%	**1.0**	**0.802**
Skewness	92.65%	**1.0**	**0.802**
Kurtosis	92.84%	**1.0**	**0.802**
Rényi Entropy	95.16%	**1.0**	**0.802**
Entropy flatness	68.86%	−0.33	−0.12
Flux	**97.29%**	0.764	0.758

In order to evaluate the effectiveness of the proposed models, we measured the Pearson correlation coefficient (see e.g. [48]) between the the hit percentages given by each model and the vector of empirical measurement labels from Table 9. This coefficient provides information about the behavior of two random variables with respect to each other. If the two variables' behavior is close to identical, then the coefficient should obtain a positive value close to 1.

In our first experiment, we considered simple threshold based classification schemes, where each tire revolution was represented by a single time-frequency feature. After optimizing the threshold for accuracy on the entire training set (Table 1), we calculated the hit percentage (12) for each time-frequency threshold model and each measurement from Table 9. Then, we obtained the correlation coefficient between the hit percentage scores and the 2-class and 3-class label sequences corresponding to our measurements. Our results are given in Table 10. In Table 10, we also indicate the accuracy of the linear classifiers achieved on the training set. This score shows how well the classifier could distinguish between normal and abnormal tire revolutions in the training set.

From Table 10 we can observe that several time-frequency feature based threshold-based classifiers performed very well on the road type classification problem. Despite using only a single feature and a thresholding scheme to recognize abnormal tire revolutions, the frequency of the recognized abnormalities correlated very highly with our empirical labels for the different measurements. In particular the coefficient of variance, skewness, kurtosis and Rényi Entropy time-frequency features all showed perfect correlation with the binary labels and over 0.8 for the three class roughness labels. In addition, the threshold-based classifiers using these features also recognized abnormal tire revolutions with a high accuracy in the training set. That means, that extracting these features from the tire revolution time-frequency data separates abnormal and normal revolutions in a nearly linear manner. These findings are in line with the assumption in [8] that abnormal revolution data contains more noise, as the value of these features increases on noisy data. We note that even though the correlation decreases when we introduce another class for roads with medium roughness it still remains strong. Furthermore, our labeling scheme did not take into account the differences between different "medium quality" surfaces, as this would have introduced the even

TABLE 11. Machine learning models using multiple features and correlation coefficients with measurements in the test set.

Model (features + name)	Corr (2-class)	Corr (3-class)
t- and f-domain, VGG16	0.78	0.74
(t,f), VGG16	0.98	**0.83**
MWSP, VGG16	0.97	0.82
t-domain signals, VP-NET [8]	0.98	**0.83**
(t,f), SVM	**1.0**	0.80

more subjective point of view of the driver. The difference between two "medium quality" roads however could be significant, therefore we believe the achieved correlation scores of 0.802 closely support our claim that the proposed classifiers can distinguish different road types. We highlight that each of the classifiers in Table 10 uses a threshold approach, therefore real time implementation on low capacity hardware poses no difficulty.

Interestingly, the ability to distinguish between normal and abnormal revolutions in the training set does not guarantee a strong correlation between the hit percentage score and our empirical labels. This is well reflected in the last row of Table 10, corresponding to the flux feature in the time-frequency representation of the tire revolutions. It achieved the highest accuracy score of the investigated methods on the training set, yet showed significantly lower correlation with our empirical labels. TF flux measures the rate of change of the signal energy in the TF plane both along time and frequency axes. Due to the sum of TFD absolute value, it considers changes in the overall energy (magnitudes). Using this feature to transform the training data seems to have resulted in a threshold that missed abnormalities in other measurements, or in other words the time-frequency Flux feature based threshold-based classifier overfitted the training data.

In order to further investigate this phenomenon, we also applied the VP-NET [49] and VGG16 deep learning classifier to the road type detection problem. These models are much more capable than the single feature based simple threshold-based classifiers discussed thus far. In terms of abnormality recognition on the training set, they clearly outperform the simpler architectures, however we were interested in examining their generalization properties to identify different road types. Our results are summarized in Table 11

Even though VGG16 used every available time-frequency feature, the correlation with our empirical road roughness labels was slightly lower than some of the single threshold based features. It is worth noting, that when considering the 3 label experiment, the more sophisticated neural network based methods, especially VGG16 using time-frequency features and VP-NET as introduced in [8] outperformed the threshold based approaches. Thus, it is reasonable to conclude that the introduction of more road types results in a less linearly separable dataset, which makes the use of ML more efficient.

VI. CONCLUSION

In this paper, we addressed the task of road type categorization using a piezoresistive force sensor integrated into the front tires of our test vehicle. The resulting sensing system provides measurements of tire deformation, enabling the detection of surface anomalies that can be utilized to estimate road quality. To test this assumption, we designed new measurements, incorporating data obtained from various types of roads. Subsequently, we developed anomaly detection methods by combining TFDs with threshold- and machine learning-based approaches. Our experiments indicate the significance of leveraging time-frequency representations of tire sensor signals, demonstrating notable improvements in detection accuracy compared to sole representation in the time domain. Moreover, we demonstrated that employing deep neural network models along with TFD features further enhances detection accuracy.

We applied the proposed classifiers to the new (unseen) measurement data and evaluated the number of (true/false) positive predictions for each road type. We found a strong correlation between the frequency of detected anomalies and the quality of the road, showcasing potential applications for estimating road conditions without requiring the retraining of classification models. These insights underscore the robustness and practicality of our approach in real-world scenarios, where diverse factors may influence sensor signals. This includes novel applications in autonomous driving, such as tire force estimation, which is part of our future work.

ACKNOWLEDGMENT

The technical support of Attila Nagy in the sensor packaging and implementation is greatly appreciated. *(Tamás Dózsa and Vedran Jurdana are co-first authors.)*

REFERENCES

[1] H. Lee and S. Taheri, "Intelligent tires—A review of tire characterization literature," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 2, pp. 114–135, Apr. 2017.
[2] A. J. Tuononen, "Optical position detection to measure tyre carcass deflections," *Vehicle Syst. Dyn.*, vol. 46, no. 6, pp. 471–481, Jun. 2008.
[3] J. Yi, "A piezo-sensor-based 'Smart Tire' system for mobile robots and vehicles," *IEEE/ASME Trans. Mechatronics*, vol. 13, no. 1, pp. 95–103, Feb. 2008.
[4] O. Yilmazoglu, M. Brandt, J. Sigmund, E. Genc, and H. L. Hartnagel, "Integrated InAs/GaSb 3D magnetic field sensors for 'The intelligent tir,'" *Sens. Actuators A, Phys.*, vol. 94, nos. 1–2, pp. 59–63, Oct. 2001.
[5] K. Eun, K. J. Lee, K. K. Lee, S. S. Yang, and S.-H. Choa, "Highly sensitive surface acoustic wave strain sensor for the measurement of tire deformation," *Int. J. Precis. Eng. Manuf.*, vol. 17, no. 6, pp. 699–707, Jun. 2016.
[6] D. Maurya, S. Khaleghian, R. Sriramdas, P. Kumar, R. A. Kishore, M. G. Kang, V. Kumar, H.-C. Song, S.-Y. Lee, Y. Yan, J.-M. Park, S. Taheri, and S. Priya, "3D printed graphene-based self-powered strain sensors for smart tires in autonomous vehicles," *Nature Commun.*, vol. 11, no. 1, pp. 1–10, Oct. 2020.
[7] *Final Report Including Technical Implementation Plan (Annex)*, Standard IST-2001-34372, APOLLO Consortium, 2001.
[8] T. Dózsa, J. Radó, J. Volk, Á. Kisari, A. Soumelidis, and P. Kovács, "Road abnormality detection using piezoresistive force sensors and adaptive signal models," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
[9] S. Strano, M. Terzo, and C. Tordela, "A supervised machine learning framework for smart tires," in *Proc. IEEE 6th Int. Forum Res. Technol. Soc. Ind. (RTSI)*, Sep. 2021, pp. 364–369.

[10] D. Garcia-Pozuelo, O. Olatunbosun, S. Strano, and M. Terzo, "A real-time physical model for strain-based intelligent tires," *Sens. Actuators A, Phys.*, vol. 288, pp. 1–9, Apr. 2019.

[11] N. Xu, H. Askari, Y. Huang, J. Zhou, and A. Khajepour, "Tire force estimation in intelligent tires using machine learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3565–3574, Apr. 2022.

[12] D. Tamás, A. Attila, B. Ferenc, S. Ernő, S. Alexandros, V. János, and K. Péter, "Towards intelligent tire development," in *Proc. 13th. Int. Conf. Transp. Sci.*, 2023, pp. 145–155.

[13] J. Radó, A. Nagy, and J. Volk, "Measuring device for kinetic dynamic state of a vehicle tyre gumiabroncs menetdinamikai állapotát mérő eszköz," S. Patent 1 900 189, Oct. 3, 2019.

[14] J.-S. Park, S.-W. Lee, and U. Park, "R peak detection method using wavelet transform and modified Shannon energy envelope," *J. Healthcare Eng.*, vol. 2017, pp. 1–14, Jul. 2017.

[15] R. Du, G. Qiu, K. Gao, L. Hu, and L. Liu, "Abnormal road surface recognition based on smartphone acceleration sensor," *Sensors*, vol. 20, no. 2, p. 451, Jan. 2020.

[16] B. Boashash, *Time-Frequency Signal Analysis and Processing, A Comprehensive Reference*, 2nd ed. Amsterdam, The Netherlands: Elsevier, Dec. 2016.

[17] L. Stankovic, I. Orovic, S. Stankovic, and M. Amin, "Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4562–4572, Sep. 2013, doi: 10.1109/TSP.2013.2271752.

[18] J. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-25, no. 3, pp. 235–238, Jun. 1977.

[19] J. Ville, "Théorie et applications de La notion de signal analytique," *Cables Transmiss.*, vol. 2, no. 1, pp. 61–74, 1948.

[20] P. Flandrin, "Some features of time-frequency representations of multicomponent signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 1984, pp. 266–269.

[21] A. K. Marnerides, P. Smith, A. Schaeffer-Filho, and A. Mauthe, "Power consumption profiling using energy time-frequency distributions in smart grids," *IEEE Commun. Lett.*, vol. 19, no. 1, pp. 46–49, Jan. 2015.

[22] H.-I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 6, pp. 862–871, Jun. 1989.

[23] F. Hlawatsch, T. G. Manickam, R. L. Urbanke, and W. Jones, "Smoothed pseudo-wigner distribution, choi-williams distribution, and cone-kernel representation: Ambiguity-domain analysis and experimental comparison," *Signal Process.*, vol. 43, no. 2, pp. 149–168, May 1995.

[24] A. Papandreou and G. F. Boudreaux-Bertels, "Generalization of the choi-williams distribution and the Butterworth distribution for time-frequency analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 1, p. 463, Jan. 1993.

[25] E. Sejdić, I. Djurović, and J. Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digit. Signal Process.*, vol. 19, no. 1, pp. 153–183, Jan. 2009.

[26] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1084–1091, Jul. 1990.

[27] L. Stankovic, M. Dakovic, and T. Thayaparan, *Time-Frequency Signal Analysis With Applications*. Norwood, MA, USA: Artech House, Feb. 2013.

[28] F. Cakrak and P. J. Loughlin, "Multiwindow time-varying spectrum with instantaneous bandwidth and frequency constraints," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1656–1666, Aug. 2001.

[29] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*. Oxford, England: OUP Oxford, 2004.

[30] G. Szegő, *Orthogonal Polynomials*. Providence, RI, usa: American Mathematical Society, 1975. [Online]. Available: https://books.google.hr/books?id=ZOhmnsXlcY0C

[31] O. Karabiber Cura, A. Akan, and H. Sabiha Ture, "Classification of epileptic and psychogenic nonepileptic seizures via time–frequency features of EEG data," *Int. J. Neural Syst.*, vol. 33, no. 9, Sep. 2023, Art. no. 2350045, doi: 10.1142/s0129065723500454.

[32] N. Ganapathy, Y. R. Veeranki, and R. Swaminathan, "Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features," *Expert Syst. Appl.*, vol. 159, Jun. 2020, Art. no. 113571.

[33] D. E. Hernández, L. Trujillo, E. Z-Flores, O. M. Villanueva, and O. Romo-Fewell, *Detecting Epilepsy in EEG Signals Using Time, Frequency and Time-Frequency Domain Features*. Cham: Springer, 2018, pp. 167–182.

[34] B. Boashash, G. Azemi, and N. Ali Khan, "Principles of time–frequency feature extraction for change detection in non-stationary signals: Applications to newborn EEG abnormality detection," *Pattern Recognit.*, vol. 48, no. 3, pp. 616–627, Mar. 2015.

[35] B. Boashash and S. Ouelha, "Automatic signal abnormality detection using time-frequency features and machine learning: A newborn EEG seizure case study," *Knowl.-Based Syst.*, vol. 106, pp. 38–50, Aug. 2016.

[36] R. Lin, Z. Zhou, S. You, R. Rao, and C.-C. J. Kuo, "Geometrical interpretation and design of multilayer perceptrons," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1–15, Feb. 2022.

[37] P. Ganesh, P. Jagadeesh, and J. Samuel Raj. J, "Prediction of human activity recognition using convolution neural network algorithm in comparison with grid search algorithm," in *Proc. Int. Conf. Adv. Comput., Commun. Appl. Informat. (ACCAI)*, May 2023, pp. 1–5.

[38] S. Genc, "Parametric system identification using deep convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2112–2119.

[39] K. T. Islam, R. G. Raj, and A. Al-Murad, "Performance of SVM, CNN, and ANN with BoW, HOG, and image pixels in face recognition," in *Proc. 2nd Int. Conf. Electr. Electron. Eng. (ICEEE)*, Dec. 2017, pp. 1–4.

[40] D. Theckedath and R. R. Sedamkar, "Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks," *Social Netw. Comput. Sci.*, vol. 1, no. 2, pp. 1–7, Mar. 2020.

[41] K. K. Jena, S. K. Bhoi, S. Mohapatra, and S. Bakshi, "A hybrid deep learning approach for classification of music genres using wavelet and spectrogram analysis," *Neural Comput. Appl.*, vol. 35, pp. 1–26, Jan. 2023.

[42] E. A. Hadhrami, M. A. Mufti, B. Taha, and N. Werghi, "Ground moving radar targets classification based on spectrogram images using convolutional neural networks," in *Proc. 19th Int. Radar Symp. (IRS)*, Jun. 2018, pp. 1–9.

[43] T. M. Wani and I. Amerini, "Deepfakes audio detection leveraging audio spectrogram and convolutional neural networks," in *Proc. Int. Conf. Image Anal. Process.*, Cham, Switzerland: Springer, 2023, pp. 156–167.

[44] S. Kumaresan, K. S. J. Aultrin, S. S. Kumar, and M. D. Anand, "Deep learning-based weld defect classification using VGG16 transfer learning adaptive fine-tuning," *Int. J. Interact. Design Manuf. (IJIDeM)*, vol. 17, no. 6, pp. 2999–3010, Dec. 2023.

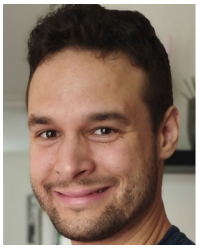
[45] C.-X. Jack Feng, Z.-G.-S. Yu, U. Kingi, and M. Pervaiz Baig, "Threefold vs fivefold cross validation in one-hidden-layer and two-hidden-layer predictive neural network modeling of machining surface roughness data," *J. Manuf. Syst.*, vol. 24, no. 2, pp. 93–107, Jan. 2005.

[46] P. Múč ka, "International roughness index specifications around the world," *Road Mater. Pavement Design*, vol. 18, no. 4, pp. 929–965, Jul. 2017.

[47] J. A. Hernandez and I. L. Al-Qadi, "Tire–pavement interaction modelling: Hyperelastic tire and elastic pavement," *Road Mater. Pavement Design*, vol. 18, no. 5, pp. 1067–1083, Sep. 2017.

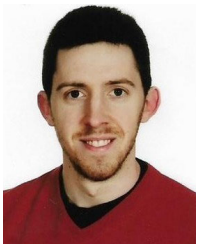
[48] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*, vol. 2, 1st ed., 2009, pp. 1–4.

[49] P. Kovács, G. Bognár, C. Huber, and M. Huemer, "VPNET: Variable projection networks," *Int. J. Neural Syst.*, vol. 32, no. 1, Jan. 2022, Art. no. 2150054.



TAMÁS DÓZSA received the M.Sc. degree in computer science from Eötvös Loránd University (ELTE), Budapest, Hungary, in 2020, where he is currently pursuing the Ph.D. degree in computer science.

Since 2020, he has been a Research Associate with the Systems and Control Laboratory, HUN-REN Institute for Computer Science and Control (SZTAKI), Budapest. During the B.Sc. and M.Sc. studies, he received numerous recognitions at the university and national level for participating in scientific competitions. He has authored 17 scientific publications thus far. His research interests include numerical analysis, signal processing, optimization, and system theory.



VEDRAN JURDANA received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Faculty of Engineering, University of Rijeka, Rijeka, Croatia, in 2015, 2017, and 2023, respectively.

From 2017 to 2023, he was a Research and Teaching Assistant with the Department of Automation and Electronics, Faculty of Engineering, University of Rijeka, where he has been a Postdoctoral Researcher, since 2023. His research interests include time-frequency signal analysis, statistical signal analysis and processing, compressive sensing, and information theory.

SANDI BARESSI ŠEGOTA received the B.Sc. and M.Sc. degrees in computer engineering from the Faculty of Engineering, University of Rijeka, Rijeka, Croatia, in 2017 and 2019, respectively.

Since 2019, he has been a Research and Teaching Assistant with the Department of Automation and Electronics, Faculty of Engineering, University of Rijeka. His research interests include industrial robotics, evolutionary computing, and data science.

JÁNOS VOLK received the Ph.D. degree in physics from Budapest University of Technology and Economics, Budapest, Hungary.

From 2005 to 2007, he was with the National Institute for Materials Science (NIMS), Tsukuba, Japan, as an ICYS Researcher. From 2013 to 2014, he was a HAESF Fellow with Virginia Commonwealth University (VCU), Richmond, VA, USA. He is currently the Head of the Nanosensors Laboratory (www.nems.hu), Centre for Energy Research, Budapest. He is the author of more than 60 scientific articles in peer-reviewed international journals. He has experience in thin film deposition, sensors, micro- and nanoelectromechanical systems (NEMS/MEMS), and semiconducting nanostructures.

JÁNOS RADÓ received the Ph.D. degree in material science from Óbuda University, Budapest, Hungary, in 2020.

From 2013 to 2022, he held a position with the Institute of Technical Physics and Materials Science, Budapest. Since 2022, he has been a Postdoctoral Fellow with Lakehead University, Thunder Bay, Canada. His current research interests include NEMS, MEMS devices, photoconductors, and X-ray detectors.

ALEXANDROS SOUMELIDIS (Member, IEEE) received the degree in electrical engineering from the Technical University of Budapest, Budapest, Hungary, in 1979, the Ph.D. degree from the Budapest University of Technology and Economics, in 2003, and the Habilitation degree from Széchenyi István University, Győ, Hungary, in 2013. In 1984, he joined as a Research Fellow with the HUN-REN Institute for Computer Science and Control, Budapest, where he is currently a Senior Researcher and participates in university education as a Habilitated Professor. His research interests include signal processing, sensing, measurements, and system identification. He is involved in several development areas, actually in the field of autonomous vehicle control.

PÉTER KOVÁCS received the Ph.D. degree in computer science from Eötvös Loránd University (ELTE), Budapest, Hungary, in 2016, and the Habilitation degree in computer science, in 2022. Since then, he has been an Associate Professor with the Department of Numerical Analysis, ELTE. In 2012, he was a Visiting Researcher with the Department of Signal Processing, Tampere University of Technology, Finland. From 2018 to 2020, he was a Postdoctoral Researcher with the Institute of Signal Processing, Johannes Kepler University Linz, Austria. With over 40 published scientific articles, his work on variable projection networks was recognized with the Hojjat Adeli Award by the *International Journal of Neural Systems*. His research interests include signal and image processing, numerical analysis, system identification, and explainable AI. He is a member of the public body of the Hungarian Academy of Sciences and the IEEE EMBS Young Professionals Committee. In 2016, he was honored with the Farkas Gyula Prize in applied mathematics from the János Bolyai Mathematical Society.