Detection of Brake Disc Deformation With Adaptive Wavelet Neural Networks

Tamás Dózsa, Péter Őri, Gergő Ungvári, Zoltán Szabó, Alexandros Soumelidis, István Lakatos, and Péter Kovács

***Abstract*—In this work, we consider the problem of detecting the phenomenon of brake disc runout, which is usually caused by deformations of the brake disc in passenger vehicles. We introduce a novel measurement system that collects and processes information on the vibration profile of the vehicle, as well as the pressure of the brake fluid. Our novel processing method for the collected signals relies on model-driven adaptive wavelet neural networks. We demonstrate that the proposed signal processing methodology outperforms previous approaches to detect aberrations of the brake disc. In addition, the proposed neural network models are defined by interpretable parameters and have reduced complexity when compared to traditional and similar approaches. These characteristics make the proposed construction suitable for automotive applications.**

***Index Terms*—Brake disc, fault detection, wavelets.**

I. Introduction

RELIABLE detection of problems in the braking system of passenger vehicles is a crucial problem. In van Schoor et al. [1], note that faults in the braking system are the most significant contributor to accidents caused by mechanical faults. In commercial passenger vehicles, a disc-based braking system is usually employed. In such a system, the brake pads apply force to the rotating brake disc, increasing friction and therefore lowering the vehicle speed. While brake pads are routinely examined, deformations of the brake disc itself can remain undetected during regular maintenance checks. Severe deformations can cause a "bumpy" braking effect as well as an increase in braking distance. The early detection of brake disc deformations is therefore an important problem. In particular, such faults should be detected in a reliable and cost-efficient manner. Finally, we note that according to [2], faults in the braking system are more likely to appear as the vehicle ages.

The detection of faults in the braking system is a well-studied problem (see e.g., [3], [4], [5], [6], [7]). In particular, brake system health monitoring based on vibration signals is proposed in [4] and [5]. One interesting recent study in this field is [8], where the authors propose a measurement system based on visual input obtained from a high-speed camera. They propose a novel segmentation method of the images and use wavelet features to detect vibration displacement. We note that the method presented in this article is very different since the measurements are not acquired from cameras. In addition, in [8], a static Haar-wavelet-based discrete wavelet transformation is suggested to extract features from the measured signals. In contrast, our proposed method in this article relies on adaptively learned continuous wavelet coefficients, which allows for a reduced number of necessary features. While most previous results have been achieved under laboratory conditions, in our earlier work [9], brake disc deformation detection has been carried out using measurements obtained from a real vehicle operating on a test track. In [9], we propose the use of machine learning relying on interpretable feature transformations applied to the measured vibration signals. It should also be noted that nonpassenger vehicle-related brake system fault detection methods have also been recently developed in [10], [11], [12], and [13].

In this study, our main concern is the detection of so-called brake disc run-outs. The effectiveness of a braking system directly influences the ability of a vehicle to decelerate reliably under varying road and operational conditions. Within this system, the brake disc (or rotor) plays a pivotal role in converting kinetic energy into thermal energy via friction. Any deviation from its optimal geometric configuration, such as brake disc run-out, can compromise braking efficiency, generate noise and vibrations, and lead to premature wear of various braking components. Therefore, diagnosing brake

Received 28 October 2025; revised 9 February 2026; accepted 2 March 2026. Date of publication 29 April 2026; date of current version 13 May 2026. This work was supported in part by the Swiss Government Excellence Scholarship under Grant 2025.0057; in part by European Union Within the Framework of the National Laboratory for Autonomous Systems under Grant RRF-2.3.1-21-2022-00002; and in part by the University Excellence Fund of Eötvös Loránd University (ELTE), Budapest, Hungary, through supported by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the K_23 "Országos Tudományos Kutatási Alap (OTKA)" Funding Schemes, under Project K146721. The work of Tamás Dózsa was supported by the Swiss Government Excellence Scholarship under Grant 2025.0057. The work of Péter Kovács was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The Associate Editor coordinating the review process was Dr. Chuang Sun. *(Corresponding author: Tamás Dózsa.)*

Tamás Dózsa is with the Faculty of Mathematics and Computer Science, UniDistance Suisse/FernUni Schweiz, 3900 Brig, Switzerland (e-mail: tamas.dozsa@unidistance.ch).

Péter Őri is with the Department of Vehicle Maintenance and Diagnostics, Széchenyi István University, 9026 Győr, Hungary.

Gergő Ungvári is with the HUN-REN Institute For Computer Science and Control, Systems and Control Laboratory, 1111 Budapest, Hungary, and also with the Department of Numerical Analysis, Eötvös Loránd University, 1053 Budapest, Hungary.

Zoltán Szabó is with the with the HUN-REN Institute For Computer Science and Control, Systems and Control Laboratory, 1111 Budapest, Hungary, and also with the Research Center of Vehicle Industry (SZE-JKK), Széchenyi István University, 9026 Győr, Hungary.

Alexandros Soumelidis is with the HUN-REN Institute For Computer Science and Control, Systems and Control Laboratory, 1111 Budapest, Hungary.

István Lakatos is with the Department of Vehicle Maintenance and Diagnostics, Széchenyi István University, 9026 Győr, Hungary.

Péter Kovács is with the Department of Numerical Analysis, Eötvös Loránd University, 1053 Budapest, Hungary.

Digital Object Identifier 10.1109/TIM.2026.3674239

disc run-out is essential for ensuring the operational integrity, safety, and longevity of modern braking systems.

Brake disc run-out refers to the unintended lateral movement or oscillation of the disc surface as it rotates. This can manifest in two main forms.

1) *Lateral Run-Out (LRO):* a side-to-side wobble of the disc due to its surface not being perfectly perpendicular to the hub axis.
2) *Disc Thickness Variation (DTV):* uneven disc surface thickness that results in cyclical variations in braking force and torque during wheel rotation.

These phenomena typically arise due to mechanical tolerances, improper installation, thermal stress, corrosion, or wear patterns over time. Even a seemingly negligible deviation, such as a run-out exceeding 0.05 mm, can cause noticeable differences in driving experience.

The consequences of undiagnosed or ignored disc run-out are multifaceted and impact both vehicle safety and operational costs.

1) *Brake Judder and Pedal Pulsation:* One of the most common symptoms is brake judder—vibrations felt through the steering wheel or brake pedal during braking. These vibrations are often cyclical and correspond to disc rotation, caused by uneven contact with the brake pads due to lateral movement or thickness variation.
2) *Decreased Braking Efficiency:* inconsistent pad contact and braking torque reduce the system's ability to deliver uniform deceleration. This not only compromises stopping distance but also introduces potential safety hazards, especially in emergency situations.
3) *Increased Wear and Maintenance:* uneven disc surfaces promote accelerated and uneven pad wear, necessitating more frequent replacement of pads and discs. In some cases, run-out can damage the caliper or lead to sticking pistons due to the irregular movement.
4) *Thermal Stress and Cracking:* run-out leads to localized heating, known as hot spotting, which further exacerbates DTV and may eventually cause disc cracking under high-stress conditions.
5) *NVH (Noise, Vibration, and Harshness) Issues:* vibrations from the brake system propagate through the chassis and suspension, contributing to overall noise and reducing passenger comfort. Persistent NVH issues can degrade driver confidence and contribute to user dissatisfaction.

Given the critical nature of brake disc alignment, several diagnostic techniques have been developed to assess disc condition both in laboratory settings and advanced research environments.

1) *Dial Indicator Method:* A traditional and widely used method where a dial gauge measures run-out directly on the disc surface while it is rotated. This allows for identification of high and low spots and provides quantitative feedback on alignment.
2) *On-Car Versus Off-Car Measurements:* Measurements can be conducted with the disc mounted on the vehicle hub or on a precision fixture. On-car measurement accounts for hub conditions and mounting, making it more representative of real-world behavior.
3) *Laser and Optical Sensors [14]:* Modern diagnostic systems incorporate laser measurement tools and optical sensors that enable high-resolution, noncontact analysis of disc surface variations.
4) *3-D Scanning and Finite Element Modeling (FEM) [15]:* In research and development contexts, 3-D scanning of the brake assembly combined with FEM simulations can predict thermal and mechanical behavior under various conditions, helping engineers understand the implications of run-out in new designs.
5) *Vibration Analysis [9]:* Brake-induced vibration signatures, analyzed using accelerometers and frequency analysis, can indirectly indicate the presence of DTV or LRO.

In [9], we compare the performance of various well-known machine-learning (ML) algorithms using interpretable input features, which were a priori generated from the measured signals. The use of ML for this problem was justified because the connection between the measured vehicle vibration profile and the presence of brake disc deformations is highly nonlinear. In [9], we measure the vibration profile of the vehicle chassis by placing four accelerometers near the braking apparatus on a single wheel. Measurements were obtained by having the vehicle accelerate in a straight line on a test track from standstill to $\approx$ 90 km/h, then applying emergency braking. In order to gather appropriate training data, good quality and artificially deformed brake discs are employed in several experiments (see [9] and also Section II). The vibration measurements are then subjected to an intuitive feature transformation and processed using various ML algorithms. Among these, ML approaches, such as fully connected neural networks and convolutional networks, are also considered. In [9], we conclude that both convolutional neural networks and random forest classifiers using transformed measurement signals as input are capable of perfect brake disc fault recognition on the measured dataset.

Unfortunately, the results in [9] suffer from several limitations. First, in our initial experiments, a total of 4 vehicle test runs were considered. Using a time window of 0.1 ms, this resulted in a total of 3750 data samples, which could be used to train ML models. In addition, the proposed feature transformations (albeit interpretable) are static and cannot adapt to different vehicle types and vibration profiles. Even though the convolutional neural network model considered in [9] does not suffer from this issue, its lack of interpretable parameters and model complexity leaves room for improvement. Finally, the methods introduced in [9] rely solely on measurements obtained from accelerometers attached to the vehicle chassis. While detecting changes in the vehicle's vibration profile is certainly a useful tool for the recognition of brake disc deformations, considering information obtained from different types of sensors improves the robustness of the measurement system.

In this work, we significantly improve our brake disc deformation approaches proposed in [9] and remedy the abovementioned limitations. First, a larger dataset, including a more diverse set of braking scenarios, is collected in order to measure the generalization capabilities of the proposed signal processing methods. Importantly, in the measurements conducted for this study, the accelerometer signals considered in our previous work are complemented by brake fluid pressure information. The inclusion of this information leads to considerable performance improvement as demonstrated by our experiments in Section V. Finally, for the first time, model-driven ML methods were employed to detect brake disc deformations. That is, instead of the static feature transformations proposed in [9], so-called model-driven neural networks are considered. In particular, low-complexity neural networks whose first layer uses wavelet kernels to extract interpretable information from the input signals are used. This construction is a modification of the methodology proposed in [16] and [17] extended with intuitive operations adapted to the brake disc run-out detection problem. It is important to emphasize that even though wavelet transform-based neural networks have been successfully used for other types of fault detection problems (e.g., bearing fault recognition [16], [17]), in this work, we significantly modify and extend these approaches to make them suitable for brake disc deformation detection. Model-driven neural network constructions in general (see e.g., [16], [17], [18], [19]) have been successfully applied recently to various engineering and medical signal processing problems. The key advantage of replacing traditional neural network layers (such as discrete convolution operators) with mathematical transformations that consider a priori information about the ML task is the emergence of physically interpretable parameters. In fact, in the proposed construction, the learned wavelet filters correspond to frequency ranges in the measured vibration and brake fluid pressure data, which characterize brake disc deformations. In contrast, the convolutional model proposed in [9] cannot produce physically interpretable learned features from the data.

Below, we summarize the main novelties of our current study.

1) The acquisition of a larger dataset containing a diverse set of braking scenarios.
2) The inclusion of brake fluid pressure information in our measurements.
3) The adaptation and use of wavelet transformation-based neural networks for brake disc deformation detection. The use of these model-driven ML approaches allows for interpretable (but adaptive) behavior and reduces model complexity.
4) Experiments using real measurements to demonstrate the effectiveness of the proposed fault detection methods. In order to ensure the reproducibility of our results, all code and data proposed in this study have been made publicly available (see [20]).

The rest of this article is organized as follows. In Section II, we specify the problem of brake disc deformation detection and describe in detail our data acquisition process. The methods introduced here can be regarded as the physical part of the proposed measurement system. Section III provides an overview of the characteristics of the measured signals and describes our preprocessing pipeline. In Section IV, we describe the proposed adaptive wavelet transformation-based neural network models used to detect brake disc deformations. This serves as the signal processing part of the proposed measurement system. Section V details our experiments and analyzes the results. Finally, in Section VI, we summarize our results and describe future research directions.

Fig. 1. Experiments were carried out using the depicted 2020 VW e-Golf, on the ZalaZone test track [21].

II. Measurement Apparatus

We carried out our braking tests on the ZalaZONE brake measurement surfaces platform [21], a test track designed specifically for collecting braking measurements. The total length of the test surface is 1100 m, the acceleration section is 750 m, the length of the braking surfaces is 200 m, and there is a safety zone of 150 m. The width of each lane is 4.5 m. There are eight braking surfaces with different coarsenesses (friction coefficient “0.1” to “1.0”) in the test area. In this study, a test surface with an approximate friction coefficient of “1.0” to was selected collect measurements. The braking surface was dry for each experiment.

The brake disc runout test was performed on a Volkswagen 2020 e-Golf, shown in Fig. 1. The measurements were carried out with two different brake discs mounted on the right front wheel. In each case, the brake disc runout was measured with a dial gauge, which was fixed using a magnetic stand as shown in Fig. 2.

The two considered brake discs had runout values of 0.02 mm and 0.12 mm. Measurement scenarios carried out using the brake disc with 0.02 mm runout shall be referred to as good condition, while those carried out with the 0.12 mm runout as bad condition measurements. Measurements were acquired using both low and high brake forces. Thus, a total of 4 different measurement scenarios were considered.

1) *Measurement:* Good brake disc, Low brake force.
2) *Measurement:* Good brake disc, High brake force.
3) *Measurement:* Bad brake disc, Low brake force.
4) *Measurement:* Bad brake disc, High brake force.

Three software tools are considered for data collection and evaluation. First, PicoDiagnostics is employed to record acceleration and brake pressure fluid values. In addition, CAN King is used to record acceleration data in three axial directions, which are determined using GPS data. To record the vibrations, we rely on the Pico Scope 4823, an 8-channel automotive oscilloscope, shown in Fig. 3.

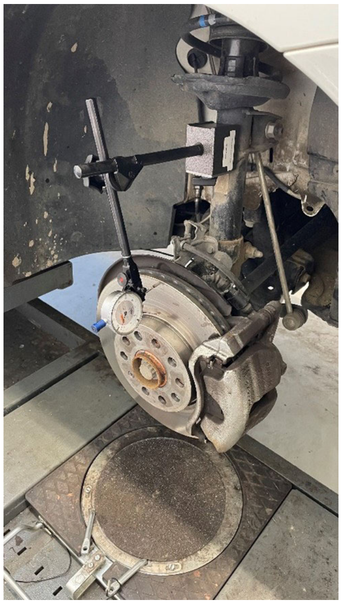

Fig. 2. Brake disc runout measurement with dial gauge.

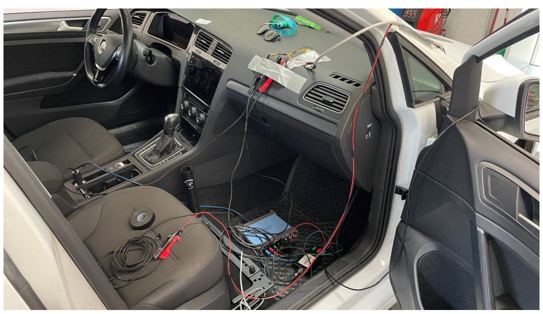

Fig. 3. Pico scope 4823 oscilloscope used to collect measurements.

In our previous work [9], we developed a vibration signal-based brake runout detection method using a similar measurement setup. In this previous study, a single three-axis accelerometer was attached to the front right arm (see Fig. 4). A novelty of this study is the introduction of a brake fluid pressure sensor attached to the front right brake fluid line (see Fig. 5).

In contrast to our previous work [9], we reduce the number of accelerometers collecting vibration signals to only one. This change is well supported by our findings in [9], where we verify that the addition of new accelerometers does not improve the performance of the measurement system significantly. The reduction of sensors also reduces the necessary computing capacity and the cost of the measurement system.

On the other hand, the current measurement setup introduces a pressure sensor whose performance can be compared with that of the acceleration sensor. Thus, we propose a multimodal measurement system in order to increase the robustness of the brake disc deformation detection schemes.

III. Signal Description and Preprocessing

As described in Section II, the proposed measurement system collects data along 4 channels. At a given time instance, 3 values corresponding to lateral, longitudinal, and vertical acceleration (ms/s^2) are collected as well as the current brake fluid pressure (Pa). The values collected for a single experiment are illustrated in Fig. 6.

Fig. 4. Three axis accelerometer used to collect vibration data.

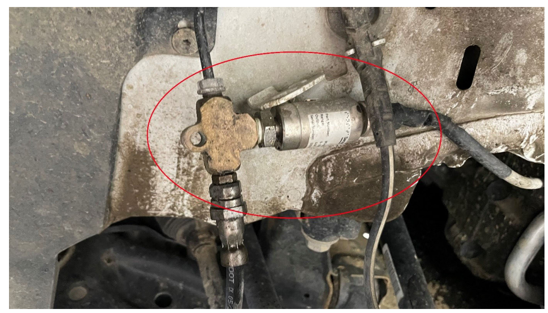

Fig. 5. Pressure sensor attached to the front right brake line.

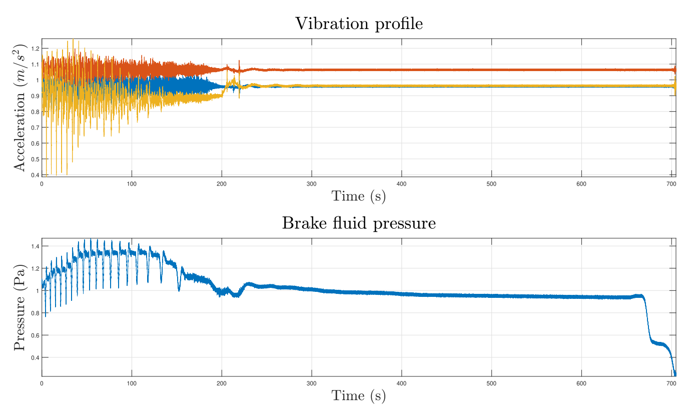

Fig. 6. Acceleration and pressure values collected from a single experiment.

These 1-D signals were then segmented into 0.1 ms long windows. Due to the sampling frequency of our measurement equipment, this meant that each data sample consists of an $\boldsymbol{f} \in 4 \times 3750$ matrix, where each row corresponds to data acquired from a different sensor. A limitation of this study is that our measurements do not include any precise information on vehicle speed or wheel angle. Because of this, it is not possible to segment the measurements into (varying-length) full

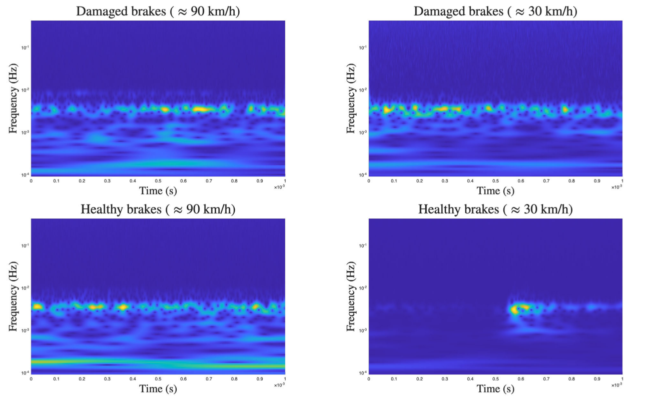

Fig. 7. Time–frequency (time-scale) representation of vibration signals obtained using continuous wavelet transform with Morlet mother wavelet. Top row: damaged brake disc at different approximate vehicle speeds. Bottom row: undamaged brake disc with different approximate vehicle speeds.

tire revolutions. In a real-life application, such a segmentation would be preferable, since point-like deformations of the brake disc may not influence the measured values at every wheel angle. Regardless, as we verify in Section V, this limitation can be overcome by the introduction of appropriate adaptive signal processing methods.

In [9], we observe that noticeable differences exist in the frequency profile of acceleration signals obtained from healthy and deformed brake discs. This observation remains true in our current study, as illustrated in Fig. 7. The tire fluid pressure signal also exhibits similar behavior (see Fig. 7). It should also be noted, however, that frequency components indicating brake disc deformations change with the speed of the vehicle (and possibly other factors, such as tire pressure and suspension characteristics). For this reason, considering the time–frequency representation of the data samples is appropriate when trying to detect brake disc deformations. In particular, brake disc deformations seem to manifest themselves as nonstationary but generally high-frequency components in the obtained signals. This makes the wavelet transform particularly well-suited for their detection, since it can provide improved time localization of such components when compared to the short-time Fourier transform. This observation forms the intuition behind the classification model introduced in Section IV. It should be noted, however, that the function that maps the measured data samples to faulty or healthy classes remains highly nonlinear, which warrants the use of the proposed ML methods. The nonlinear nature of the mapping to be learned is also well reflected by our experimental results detailed in Section V. In particular, we will show that linear classifiers are unable to achieve meaningful results on our dataset.

IV. Adaptive Wavelet Neural Networks

The ML-based classifier models considered in this study can be thought of as parameterized mappings of the form

$$F_\theta : \mathcal{X} \to \mathcal{Y} \tag{1}$$

where $\mathcal{X} \subset \mathbb{R}^{4\times N}$ $(N \in \mathbb{N})$, $\mathcal{Y} := \{0, 1\}$ and $\theta \in \mathbb{R}^p$ $(p \in \mathbb{N})$. That is, we consider mappings dependent on a vector of learnable parameters θ that map input data samples $x \in \mathcal{X}$ to a faulty or healthy label in $\mathcal{Y}$. In a typical ML problem, we assume that there exists an ideal map $F : \mathcal{X} \to \mathcal{Y}$, whose behavior is represented by the collected training data. The objective of the ML method is the numerical optimization of the parameters θ, such that $F_\theta \approx F$ in a chosen metric.

We now detail the proposed classification model. First, we specify a number of conditions that our model must satisfy.

1) The classifier must have at least some physically meaningful components (parameters) in θ.
2) The classifier must have as few parameters as possible.
3) The classifier must achieve a high fault detection accuracy (preferably on average $\approx 99\%$) on every conducted experiment. This performance should be achieved even when the input data samples were not seen by the model during training. We note that in this context, we refer to accuracy in the machine learning sense, that is, we would like to ensure the above-mentioned accuracy for the entire dataset rather than reflecting a probability for individual data samples. The difference is important to emphasize in order to avoid confusion with terminology more usual for instrumentation and measurement systems [22].

In order to achieve the above-stated objectives, we consider so-called model-driven neural network architectures. Neural networks in general can be thought of as a composition chain of linear and nonlinear transformations, where the matrices representing the linear mappings are defined by θ. That is,

$$F_\theta(x) := g_M\left(F_M^{\theta_M}\left(g_{M-1}\left(F_{M-1}^{\theta_{M-1}}\left(\dots g_1\left(F_1^{\theta_1}(x)\right)\dots\right)\right)\right)\right) \tag{2}$$

where $g_k \circ F_k^{\theta_k}$ is referred to as the kth layer of the network defined by the parameters $\theta_k \in \mathbb{R}^{p_k}$, $(p_k \in \mathbb{N}, k = 1, \dots, M \in \mathbb{N})$. We shall assume that $F_k^{\theta_k}$ is a linear map, while g_k denotes a nonlinear so-called activation function for $k = 1, \dots, M$. It has been observed that linear layers $F_k^{\theta_k}$ with special structures can yield very beneficial signal representations. In particular, if $F_k^{\theta_k}$ is represented by an appropriate Toeplitz matrix, the mapping is referred to as a (discrete) convolution layer. Tuning the parameters θ_k in this case can yield very useful signal representations. For example, certain features of the data sample $x \in \mathcal{X}$ can be filtered, which makes classifying it easier [23]. The power of convolution layers has been well exploited in various image and signal processing problems in recent years [24].

One drawback of using convolution layers is that the tuned parameters θ_k usually do not carry any physically meaningful information and are not interpretable by humans. This makes the deployment of deep convolutional networks difficult in safety-critical applications.

In order to overcome this limitation, in this work, we propose the use of adaptive wavelet neural networks to detect brake disc deformations. The key idea behind this construction is that convolution layers are further restricted to use scaled mother wavelets as convolution kernels. In this way, the output of the convolution will correspond to a set of wavelet coefficients that describe the time-scale (or time–frequency) profile of the signal. The free parameters of such a wavelet convolution layer will consist of the scales defining each wavelet convolution kernel. Thus, after training, the learned parameters of the layer will be human-interpretable: they will correspond to pseudo frequencies (scales), which characterize faults. In [16] and [17], wavelet transformation-based ML methods have been successfully applied for the detection bearing faults as well as medical signal processing problems (see also [18], [25] for nonwavelet transformation-based model-driven solutions). In this work, for the first time, we apply such a signal processing pipeline for the detection of faults in the braking system. In addition, (as explained below) we extend the idea of the wavelet kernel convolution layer with an intuitive pooling mechanism, which helps to identify key features of the input measurements.

Formally, the wavelet transform of a function f defined over the real line with respect to a so-called mother or analyzing wavelet ψ is defined as

$$W_\psi f(\lambda,\tau) := \int_{\mathbb{R}} f(t)\,\overline{\psi_{\lambda,\tau}}(t)\,dt \tag{3}$$

where $\overline{\psi}$ denotes the complex conjugate of ψ and

$$\psi_{\lambda,\tau}(t) := \frac{1}{|\lambda|}\psi\left(\frac{t-\tau}{\lambda}\right) \quad (t,\tau\in\mathbb{R},\ \lambda\in\mathbb{R}\setminus\{0\})$$

whenever the integral in (3) exists. This is satisfied, for example, when $f,\psi\in L_2(\mathbb{R})$, which is the class we will be focusing on in the current study. The parameters λ and τ in (3) are referred to as scales and translations, respectively. For a fixed parameter pair, the number $W_\psi f(\lambda,\tau)$ will be referred to as a wavelet coefficient, and it describes the similarity between f and $\psi_{\lambda,\tau}$. Wavelet coefficients provide a so-called time-scale representation of the signal f. This is closely related to time–frequency representations. In particular, for a fixed mother wavelet ψ, the scales always correspond to specific so-called pseudo frequencies defined by the center frequency of ψ. Furthermore, if ψ satisfies the admissibility property (see e.g., [26]), then f can be reconstructed (at least in L_2-norm) from its wavelet coefficients.

Notice that for a fixed scale λ, one can express wavelet coefficients using the convolution operation

$$W_\psi f(\lambda,\tau) = (f*\psi_\lambda)(\tau) := \int_{\mathbb{R}} f(t)\,\overline{\psi_\lambda}(t-\tau)\,dt \tag{4}$$

where

$$\psi_\lambda(t) := \frac{1}{|\lambda|}\psi\left(\frac{t}{\lambda}\right).$$

For practical applications, f and ψ are usually available as discretely sampled vectors. That is, assuming that both f and ψ are quasi-compactly supported with their effective support falling in some interval $[a,b]\subset\mathbb{R}$, we consider $\boldsymbol{f}\in\mathbb{R}^N$ and $\boldsymbol{\psi}\in\mathbb{R}^N$ $(N\in\mathbb{N})$ such that

$$\begin{aligned}\boldsymbol{f}_k &= f(t_k)\\ \boldsymbol{\psi}_k &= \psi(t_k),\ (t_k\in[a,b],\ k=1,\dots,N).\end{aligned}$$

For simplicity, we henceforth assume that

$$t_{k+1}-t_k = h \tag{5}$$

that is, the signals have been sampled in an equidistant manner. Using such discrete signals, it is possible to approximate wavelet coefficients by discretizing (4)

$$W_\psi f(\lambda,k\cdot h) := (f*\psi_\lambda)(k\cdot h) \approx h\cdot\left(\boldsymbol{f}*\boldsymbol{\psi}_\lambda\right)[k] \tag{6}$$

where

$$(\boldsymbol{f}*\boldsymbol{\psi}_\lambda)[k] := \sum_{j=1}^{N}\boldsymbol{f}_j\left(\boldsymbol{\psi}_\lambda\right)_{k-j}.$$

In this work, we assume that $\boldsymbol{f}$ and $\boldsymbol{\psi}_\lambda$ have been zero-padded to make the above discrete convolution feasible.

Consider a set of wavelet filters $\boldsymbol{\psi}_{\lambda_1},\boldsymbol{\psi}_{\lambda_2},\dots,\boldsymbol{\psi}_{\lambda_m}$, where $\lambda_k>0$ $(k=1,\dots,m\in\mathbb{N})$. The key idea behind the proposed methodology is to treat the scales λ_k as free parameters and (through numerical optimization) obtain an ML model which extracts time-scale components from the input signals that can be used to recognize break disc deformations.

Similar constructions have been used in [10] and [17] for the detection of bearing faults. We now briefly describe these methods and highlight the differences between them and our proposed construction. The methodology proposed in this article differs significantly from [17]. In [17], a neural network architecture was proposed (called VP-WaveletNet), where the first layer learned exactly m approximate wavelet coefficients. The free parameters of this layer were defined as the $(\lambda_k,\tau_k)\in\mathbb{R}_+\times\mathbb{R}$ $(k=1,\dots,m)$ dilation–translation pairs. The benefit of such an approach is that it provides a very sparse representation of the input signals computing only m wavelet coefficients, instead of the convolution-based approach, where m wavelet filters produce an output matrix of size $m\times N-m$. The drawback of this approach, however is that it assumes a "close to periodic" behavior of the input signals when it is used for fault detection. Indeed, since the translation parameters τ_k are also learned, the fault has to appear at relatively the "same" time in each input sample. Since in this study, we cannot assume such quasi-periodic behavior of our measurements (because fixed time windows were used for segmentation), we have to resort to a convolution-based approach.

In [16], so-called wavelet kernel networks (WKNs) were introduced. The authors propose the use of parameterized discrete convolutions similar to (6) as layers in neural networks. The proposed network architecture is similar to the one used in this article, here we point out some key differences. First, in Li et al. [16], propose the use of discrete wavelet convolution layers of the form

$$F_\theta(\boldsymbol{f}) := \boldsymbol{f}*\boldsymbol{\psi}_{\lambda,\tau} \quad (\theta := [\lambda,\tau]) \tag{7}$$

where both $\lambda>0$ and $\tau\in\mathbb{R}$ are treated as learnable parameters. That is, each wavelet convolution map (in addition to the scale parameter learned by our proposed network) also learns an initial translation. In this work, we propose wavelet convolution layers, where only the scales are kept as learnable parameters. In Appendix, we show that omitting initial translation values from the parameter set greatly reduces model complexity without significantly impacting the quality of the obtained wavelet coefficient features. In particular, for the purpose of brake disc deformation, we propose the use of wavelet convolution layers defined by

$$F_\theta(\boldsymbol{f}) := \boldsymbol{f} * \boldsymbol{\psi}_{\lambda,0} = \boldsymbol{f} * \boldsymbol{\psi}_\lambda \quad (\theta := \lambda)\,. \tag{8}$$

Comparing (7) and (8), we notice that in this work, the translation parameter has been removed as a learnable entity from the wavelet transformation layer. We already established that choosing $\tau = 0$ provides a good approximation to any possible optimal τ whenever the sampling frequency of the input is high. Only optimizing the wavelet convolution kernels for scales, however, has additional benefits as well. An obvious benefit is that the number of parameters to be optimized is cut in half. On the other hand, the output of the proposed wavelet convolution kernel in (8) is more interpretable than the original WKN method [see (7)]. Indeed, usually, the θ parameter of (8) contains more than a single scale, that is $(\theta := [\lambda_1, \ldots, \lambda_m] \in [0, +\infty)^m)$. In this case, having a separate τ_k initial translation for each wavelet kernel can obfuscate the time localization of the individual convolution outputs. In contrast, the proposed transformation from (8) ensures that if θ contains m scales, that is,

$$F_\theta(\boldsymbol{f}) \in \mathbb{R}^{(N-m)\times m}, \quad (\theta := [\lambda_1, \ldots, \lambda_m] \in [0, +\infty)^m) \tag{9}$$

then the rows of $F_\theta(\boldsymbol{f})$ correspond to identical translations of the computed coefficients for each scale λ_k.

Finally, it should be mentioned that in order to simplify the proposed model and reduce the number of parameters, dimension reduction-based feature extraction methods are frequently performed in ML. Traditional convolution layers are frequently followed by so-called pooling operations to achieve this. Maximum pooling [23], for example, performs a dimension reduction to $\mathbb{R}^Q$ $(Q << N)$ by selecting the maximum values of the input in a sliding window. Similar average pooling operations are also common. It has been shown that applying such operations on the outputs of convolution layers reduces model complexity and improves generalization capabilities [23].

Applying traditional pooling operations to the proposed wavelet convolution output defined in (9) would be counterintuitive. Indeed, for example, taking the Q largest (in absolute value) wavelet coefficients and disregarding which scales they represented would lead to degradation of interpretability. In order to avoid this, for the brake disc deformation detection problem, we propose the use of so-called scale-sensitive pooling

$$\mathrm{pool}\,(F_\theta(\boldsymbol{f}))_j = \mathrm{top}_k\left(\left|F_{\lambda_j}(\boldsymbol{f})\right|\right) \quad (j = 1, \ldots, m) \tag{10}$$

where $\mathrm{top}_k(|F_{\lambda_j}(\boldsymbol{f})|)$ collects the k largest absolute value elements from $F_{\lambda_j}(\boldsymbol{f}) := (\boldsymbol{f} * \boldsymbol{\psi}_{\lambda_j})$. Thus, $Q = mk$ wavelet coefficients are extracted in total.

Intuitively, (10) means that after the wavelet kernel convolution is performed, the proposed transformation extracts the k most relevant wavelet coefficients corresponding to each scale. Because these coefficients remain determined by their corresponding scale, the interpretability of the transformation is preserved.

We are now ready to specify the proposed partially interpretable ML model to detect deformations of the brake disc. The exact architecture (e.g., number of wavelet coefficients and k parameter of the pooling operation) has been identified using a large grid search where we tried to maximize the performance of the model for the experiment detailed in Section V.

Individual input data samples are matrices belonging to $\mathbb{R}^{4\times 3750}$. Each row of this input matrix contained 1-D measurements obtained from a single sensor. The first three rows correspond to vibration measurements obtained using accelerometers, while the last row corresponds to the brake fluid pressure signal. Wavelet kernel transformations of the form defined in (9) are applied to each row separately. In particular, we propose the use of wavelet convolution kernels with the Morlet wavelet defined as

$$\psi(t) := \cos(\omega t) \cdot e^{-t^2/2}, \quad (t \in \mathbb{R}, \omega \in [-\pi, \pi)) \tag{11}$$

where ω was the center frequency defining the mother wavelet. Our choice of the Morlet mother wavelet is well supported by its previous uses in the analysis of vibration profiles [27] and was also backed up by our experimental results (see Section V). It should be mentioned that the center frequency parameter ω was not fixed in our model; it was also left as a free, learnable quantity. This is another noteworthy difference between the wavelet convolution layers presented here and in [16].

For our current experiments, a total of $m = 10$ wavelet kernels are used on each row of the input matrix. This means that after the wavelet kernel convolutions have been performed, we obtain a total of 40 coefficient vectors corresponding to 10 different scales for each 1-D input signal. This output is subjected to batch normalization, then the pooling operation defined in (10) is applied to each coefficient vector with $k = 10$. This way, a total of 400 wavelet coefficients are kept. These coefficients are then passed through three fully connected layers with 256, 128, and 1 neurons, respectively. ReLu activation is employed between the first two layers, and sigmoid activation is applied to the output of the last layer. This ensures that integer rounding of the model output can be interpreted as the predicted label of the input measurements. A visual illustration of the proposed network is provided in Fig. 8.

In total, the proposed model consists of 142 180 trainable parameters. Considering floating point precision, we find that such a model requires less than 200 000 parameters $\times$ 4 bytes = 800 kB of memory to store the parameters, which is still feasible for microcontrollers. For example, the chips STM32F407ZGT6 and STM32H7A3RIT6 developed by STM microelectronics both offer > 1 mB flash memory and can be obtained for around 9–10 EUR at the time of writing this study. Convolutional neural networks of similar size have been shown to be appropriate for inference deployment on such microcontrollers (see e.g., [28]).

Finally, we note that neural network-based ML methods are most often trained using the backpropagation algorithm. This algorithm computes the gradient of a predefined loss function with respect to the trainable parameters, exploiting the chain rule of differentiation and the structure of neural networks [see (2)]. Then, it attempts to minimize said loss function by using a gradient-based update of the learnable parameters.

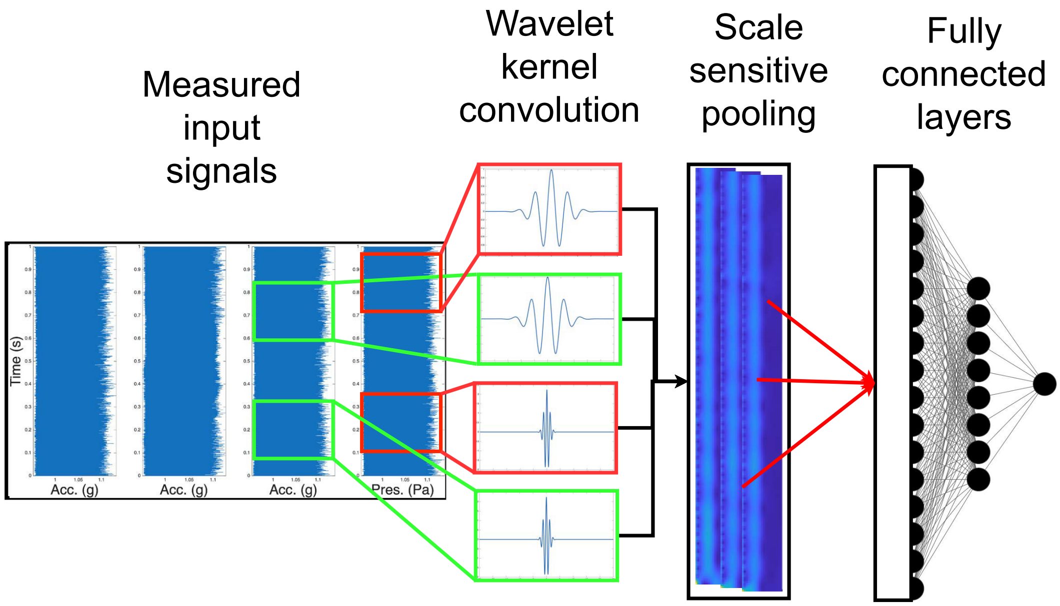

Fig. 8. Schematic of the proposed neural network architecture. The input signals are subjected to discrete convolution operators by wavelet kernels according to (8). Then, scale-sensitive pooling is applied to the wavelet coefficients [see (10)]. Finally, fully connected layers classify the extracted, interpretable wavelet features. The figure illustrates the network architecture, but the number of wavelet kernels and fully connected nodes has been reduced for illustration purposes. We emphasize that the scale parameters of the wavelet convolution layer and the weights of the underlying neural network are both learnable and trained together using backpropagation.

TABLE I

CONSIDERED MEASUREMENT SCENARIOS AND NUMBER OF SAMPLES

Scenario	**Number of data samples**	
	Healthy brake disc	**Damaged brake disc**
Mild braking	1597	1879
Strong braking	1696	1318
Total	3293 (≈ 51%)	3197 (≈ 49%)

Since the detection of brake disc deformations can be posed as a binary classification task, in this work, the so-called binary cross-entropy (BCE) loss is considered. For a given input sample $\boldsymbol{f} \in \mathcal{X}$, the BCE loss is given by

$$\begin{aligned}\ell\left(F_\theta\left(\boldsymbol{f}\right), F\left(\boldsymbol{f}\right)\right) = &\, F\left(\boldsymbol{f}\right) \log\left(F_\theta\left(\boldsymbol{f}\right)\right) \\ &+ \left(1 - F\left(\boldsymbol{f}\right)\right) \log\left(1 - F_\theta\left(\boldsymbol{f}\right)\right). \end{aligned} \tag{12}$$

The proposed wavelet convolution layer (9) and scale preserving pooling operation (10) are (sub) differentiable with respect to their input argument and learnable parameters. In other words, the subgradients of (12) exist with respect to the learnable scales. Thus, the proposed transformations can be easily integrated into neural networks and layers, and the optimization of their parameters can be performed together with the weights of the network similar to convolution layers. That is, in the proposed model the scale parameters of the wavelet convolution layer and the weights of the underlying neural network are both learnable and trained together using backpropagation. For further details, we refer to our implementation [20].

V. EXPERIMENTS

In order to verify the effectiveness of the wavelet kernel-based neural network proposed in Section IV, we analyze its performance on real input signals collected with the measurement system introduced in Section III. In addition, we compare its performance with several other classification schemes which have also shown promise for brake disc fault recognition previously [9]. Our dataset includes data samples from four different measurement scenarios. As mentioned in Section II, the initial measurements are subjected to a windowing operation with a static (1 ms) window size. Table I denotes each braking scenario and the number of input data samples acquired from them.

As can be observed from Table I, the acquired dataset was well balanced, containing a roughly equal number of data samples representing healthy and damaged brake discs in both experimental scenarios. For this reason, we will describe model performance in terms of classification accuracy.

A. Fivefold Cross Validation

In order to measure model performance, we employ a 5-fold cross-validation scheme. That is, each considered model is trained and tested 5 times on the dataset, each time using pairwise-distinct validation sets. The results of these experiments are shown in Table II. In addition to the wavelet kernel convolution-based model (WK-NN) introduced in Section IV, we perform the evaluation of a number of other ML methods. In particular, we compare WK-NN's performance with the static feature extraction-based methods proposed in [9].

In [9], we empirically show that changes in the vibration profile of the braking system can be captured by measuring

TABLE II
FIVEFOLD CROSS-VALIDATION RESULTS WITH THE CONSIDERED MODELS

Model	Mean Acc.	Min Acc.	Max Acc.
WK-NN (proposed)	99.2%	98.8%	**99.8%**
WK-NN-V	99.0%	98.2%	99.3%
WK-NN-P	98.1%	97.8%	98.6%
RF	90.2%	89.3%	91.6%
RF-V	92.8%	91.2%	93.9%
RF-P	92.9%	92.1%	93.9%
FCNN	89.9%	89.2%	91.1%
CNN	91.5%	90.9%	92.4%
TNN	89.8%	88.8%	91.2%
RESNET	**99.3%**	**98.9%**	99.5%

the standard deviation of the collected vibration signals. These deviation values are then passed to several classical ML methods to detect deformations of the brake discs. In this article, we use the same methodology to define baseline approaches. It should be noted, however, that there are several key differences between the dataset used in this article and the signals described in [9]. First, the static, standard deviation-based feature extraction step was developed for vibration data only; however, in the current experiment, it is also applied to the brake fluid pressure signal. In order to measure the effect of this, the models are also evaluated using only vibration or brake fluid pressure data. The extracted deviation features acted as input to a random forest (RF) model. The hyperparameters of each model (e.g., the size of the RF model) are identified using a trial-and-error-based approach. In addition, we expected a decrease in the performance of these models when compared to the reported performance in [9], because in [9], information about the velocity of the vehicle is also available. This allowed for the segmentation of the data in such a way that at least one full tire revolution is included in each data sample. No such guarantees could be established for the current dataset, therefore, the accuracy of ML models relying on nonadaptive feature transformations is expected to fall.

In addition to the static feature extraction-based methods, we were also interested in comparing the performance of the proposed wavelet convolution-based neural network with traditional neural network architectures. In particular, we wanted to verify that:

1) The presence of the proposed wavelet kernel convolution layers and subsequent pooling operation improved model performance.
2) In addition to the inherent interpretable nature of the proposed wavelet transformation layers, they could provide similar (or better) performance than traditional convolution layers.

In order to verify the above points, we consider two different neural network architectures. First, we consider only the fully connected part (FCNN) of the proposed WK-NN architecture described at the end of Section IV and remove the wavelet convolution layer. In this way, we can measure if the proposed wavelet convolution and pooling layers did in fact improve brake disc detection accuracy. In order to test the second point, we consider a traditional convolutional neural network (CNN). In a CNN, the values of the convolution kernels act as free parameters. Thus, the model learns an ideal signal representation, which is used by subsequent fully connected layers performing the classification. In the considered CNN, a total of 40 discrete convolutions were performed in the first layer of the network. In an identical fashion to the proposed WK-NN model, 10 convolutions were considered for each 1-D input signal (3 vibration signals, 1 fluid pressure signal). The length of the convolution kernel was chosen as 512. This is identical to the length of the proposed wavelet convolution kernels. Hence, the effect of replacing wavelet convolutions with traditional adaptive variants could be measured.

In addition to the CNN baseline, we compare the performance of WK-NN with two larger neural network models. In the first case, we consider a transformer neural network [29] (TNN). The TNN model interprets the input data samples $x \in \mathcal{X} \subset \mathbb{R}^{4\times 3750}$ (see Section III) as a set of $N = 3750$ tokens and tries to find associations between these using self-attention layers. The output of the self-attention layers is passed to a fully connected feedforward part, whose size is comparable to the feedforward component of the proposed WK-NN model. In total, the considered TNN model contained 509 489 parameters, ≈ 3.6 times more than the proposed WK-NN approach.

In addition, we also considered a RESNET-18 [30] deep neural network pretrained on the IMAGENET-1K dataset [31]. This is a large, deep neural network model used for image classification. In [32], a RESNET model is successfully applied to classify scalograms obtained from vibration signals to detect anomalies in bridges. We take a similar approach to apply RESNET for brake disc runout detection. In order to make RESNET compatible with the collected dataset, first, we generate the Morlet scalograms of the 1-D input signals obtained from the acceleration and pressure sensors. In particular, we approximate the Morlet wavelet coefficients of each 1-D signal using 64 logarithmically spaced scale parameters in $[0.5, 2.5] \in \mathbb{R}$. The logarithmic scale spacing and the bounds were determined through a trial-and-error approach to maximize model performance. Once the scalograms have been computed from the 1-D signals, they are normalized and resized to 224×224 matrices in order to match the input shape expected by RESNET. Denote by $V^{(n)} \in \mathbb{R}^{224\times 224}$ $(n = 1, \dots, 3)$, the scalogram matrix obtained from vibration measurements and by $P \in \mathbb{R}^{224\times 224}$, the scalogram obtained from the pressure signal corresponding to a single data sample. Then, the input to the RESNET model is given by the tensor $X \in \mathbb{R}^{224\times 224\times 3}$, where

$$X_{k,j,n} = V^{(n)}_{k,j} + \beta \cdot P_{k,j} \quad (k, j = 1, \dots, 224,\ n = 1, \dots, 3).$$

Here, we choose $\beta = 0.3$ through a trial and error approach. Hence, each input tensor X can be interpreted as an RGB image containing wavelet coefficient information about the vibration and pressure signals. For the RESNET approach, our implementation [20], uses the pywavelets [33] and OpenCV [34] libraries to generate scalograms and manipulate images. We note that using the RESNET model is computationally demanding, since, in addition to the large model size, the full scalograms have to be generated from the measurements in inference time. Indeed, in addition to the computational overhead due to scalogram generation, the considered RESNET-18 model is defined by 11 177 025 trainable parameters, which is approximately 79 times more than the size of the proposed WK-NN. In addition, we should already note that no interpretable learned parameters can be expected from a model of this size. Finally, it is important to mention that the RESNET-18 model was pretrained. We adapt it to our dataset by retraining the entire network for five epochs in each iteration of the fivefold cross-validation experiment. We refer to our implementation in [20] for further hyperparameters of the RESNET training process.

In Table II, the RF classifiers took as input the standard deviations of each measured 1-D signal (vibration and brake fluid pressure). The RF-V and RF-P notations indicate that the models were trained using only vibration (V) or pressure (P) data. In these cases, the standard deviation of the appropriate signal type was extracted and passed as input to the models. The results indicate that the RF classifier was unable to improve its performance when it had access to both pressure and vibration information. This is likely due to the lack of appropriate adaptive feature extraction schemes for the RF classifier. Indeed, we can compare the performances of RF-V and RF-P to the WK-NN-V and WK-NN-P models. These denote the proposed adaptive wavelet convolution-based methods restricted to using only vibration (WK-NN-V) or brake fluid pressure (WK-NN-P) data. We observe that both models solve the classification task with a high mean accuracy on the test set (over 98%) and the improvement over the respective RF methods is significant. We also observe that allowing the proposed WK-NN model access to both vibration and pressure information further increases model performance by 0.2%. The performance increase is even more significant when we consider the worst-case fold for the models. Indeed, the WK-NN model with access to both types of signals outperformed WK-NN-V by 0.6% and WK-NN-P by 1% in the worst performing fold. In this way, we conclude that, in addition to increasing robustness against sensor failure, the inclusion of brake fluid pressure information also improves the performance of the proposed fault detection algorithm.

The above results indicate that the proposed WK-NN model can significantly outperform similar ML methods when used for brake disc deformation detection. In fact, it achieves a mean accuracy of 99.2% on the test set, which is a more than 6% increase compared to the RF classifier. When considering model performance on individual folds, this difference increases even more. For example, the proposed fault detection scheme is able to detect the brake disc type correctly in 98.8% of the cases, even for the worst performing fold. It was also the only classifier that achieved the highest test accuracy on a single fold.

The decline in accuracy of the RF classifier compared to its reported performance in [9] was somewhat expected because of the above-mentioned differences in data segmentation. Regardless, all of the baseline classifiers achieved a near 90% mean accuracy. Importantly, applying the RF classification scheme to vibration or pressure data yields similar results, showing that the newly introduced brake fluid pressure signals contained meaningful information for the fault detection problem.

TABLE III

WK-NN's Ability to Generalize Between Mild and Strong Braking Conditions

Training set	**Test set**	Accuracy (on test set)
MH + MF	SH + SF	99.0%
MH + SF	SH + MF	99.3%
SH + MF	MH + SF	97.7%
SH + SF	MH + MF	99.0%

The investigated traditional neural networks (FCNN, CNN) significantly underperformed when compared to the proposed wavelet transformation-based approach. It should be once again emphasized that the proposed WK-NN method contains interpretable parameters defining wavelet coefficients (at least in the first layers), while the considered CNN model does not contain physically meaningful parameters. Despite this, and its worse performance, the CNN model contained significantly more free parameters.

Finally, comparing the performance of WK-NN to modern large deep neural networks, we can observe the following. The proposed WK-NN model significantly outperforms the considered transformer network (TNN). In fact, the TNN's accuracy on the test set falls below that of the more traditional CNN and FCNN architectures. In spite of the success of TNNs in other domains, this is not very surprising, because the effect of brake disc deformations manifests instantaneously in pressure and vibration measurements. In addition, (see also Section III), the signals are segmented into short, 0.1 ms long windows, which negates the self-attention layer's ability to search for associations between measurements corresponding to far away time-instances.

The large RESNET-18 deep learning model was the only one to slightly outperform the proposed construction in terms of accuracy. Indeed, it achieved a 0.1% increase in average accuracy on the test set during the fivefold cross-validation experiment. It should be emphasized that both models achieve near perfect ($\geq$ 99%) fault detection on average. In order to achieve this increase of 0.1% in accuracy, the RESNET model needed more than 10 000 000 additional parameters compared to the proposed WK-NN approach. Because of this, the RESNET model would require specialized hardware (not cost-effective microcontrollers) for any real-time application. In addition to the model size, the computational costs for inference are also increased, since the RESNET model requires the full scalograms to be generated (in real time) from the input measurements. Finally, we note that, unlike the WK-NN, none of the learned parameters or inferred values contain physically interpretable information.

B. Generalization Ability for Different Scenarios

Next, we consider an experiment to examine the generalization ability of the WK-NN model. In particular, we show that if the proposed model is trained using data obtained solely from mild or strong braking scenarios (see Table I), it is still able to detect faults appearing in both cases. To this end, let MH,

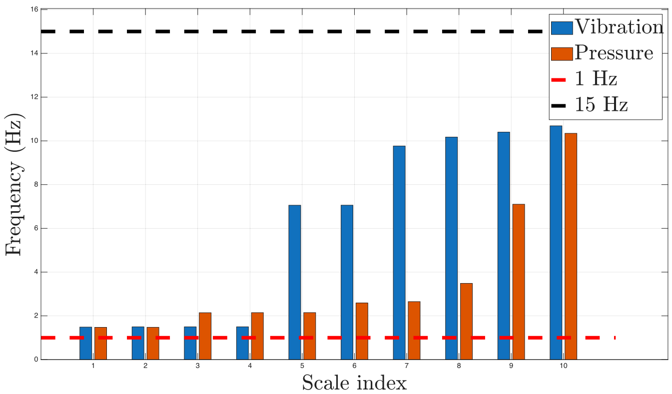

Fig. 9. Learned pseudo frequencies (scale parameters) used to compute wavelet coefficients from vibration and brake fluid pressure signals. The learned frequencies align well with the frequency range associated with brake disc runout from the literature.

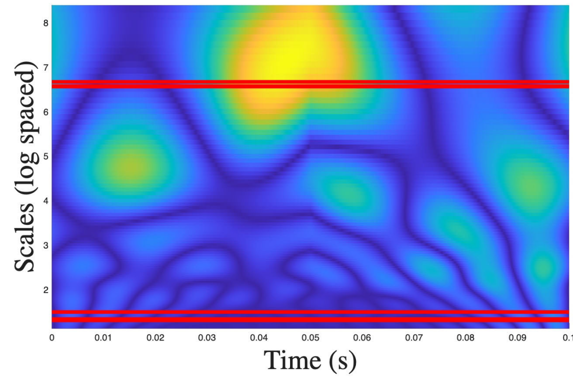

Fig. 10. Difference scalogram $|W_{\text{healthy}} - W_{\text{fault}}|$ with the learned scales indicated as horizontal lines.

SH, MF, and SF denote the set of all healthy signals measured under mild, healthy signals measured under strong, faulty signals measured under mild, and faulty signals measured under strong braking conditions, respectively. We trained the proposed WK-NN model using all possible combinations of healthy and faulty training sets and measured its performance on the remaining signals. Our results are given in Table III.

By Table III, we conclude that WK-NN can successfully recognize disc runout in signals, which correspond to braking intensities not present in the training set.

C. Interpretability of the Learned Parameters

We now examine the interpretability of the learned scale parameters of the proposed WK-NN model. We begin by noting that several studies and technical reports concluded (see e.g., [35], [36], [37]) that the presence of brake disc runout is felt in low frequencies of vibration signals. For example, in [36], the frequency range 10–20 Hz is shown to be relevant, but even lower frequency vibrations have been recorded due to brake disc runout phenomenon [35]. It is also important to mention that vehicle velocity can influence the relevant frequency band [37]. Much less research has been conducted on how brake disc runout influences changes in brake fluid pressure; however, since the effect of a point-like deformation can be measured once per wheel revolution, we expect similarly low-frequency components to indicate the fault. Hence, we expect that a trained WK-NN model will learn scales, whose associated pseudo-frequencies fall within the 1–20 Hz range for both signal types. Furthermore, we expect that the learned scales cover several frequency-bands, which correspond to different vehicle speeds observed during test scenarios (from near standstill to around 90 km/h).

In order to test the above hypothesis, we train a single WK-NN model with the same structure as the ones producing results for Table II. In this experiment, we only trained the model for 20 epochs, which already achieves an accuracy of over 99% on the training set. The connection between pseudo-frequencies ξ and scales λ can be written as

$$\xi = \frac{\omega}{2\pi\lambda} \tag{13}$$

assuming a real Morlet wavelet with center frequency ω [see (11)] and a scale parameter $\lambda > 0$. Using (13), we convert the learned scales used to compute Morlet wavelet coefficients from vibration and pressure signals. Fig. 9 shows the learned parameters.

Clearly, the learned scales (pseudo frequencies) for both input signal types fall within our expected frequency range. Furthermore, the learned scales appear in groups which correspond to very low $\approx$1–2 Hz, medium $\approx$4–8 Hz, and higher >10 Hz frequency bands. Thus, they match our a priori expectation that the learned scales would be spread out over the feasible frequency range to account for changes in vehicle speed.

Finally, we show that the learned scales correspond to regions of the scalogram, where significant differences between wavelet coefficients arise due to brake disc runout. To this end, first, we compute the Morlet wavelet coefficients of two vibration signals. Both of the signals were recorded during a strong braking scenario with the vehicle traveling at approximately 40 km/h. One of the signals was recorded with a healthy brake disc, while the other one corresponds to brake disc runout. The scalograms were computed over 100 logarithmically spaced scales. Denote the obtained coefficient matrices by $W_{\text{healthy}}, W_{\text{fault}} \in \mathbb{R}^{100\times 3750}$. In Fig. 10, we illustrate

$$W := \left|W_{\text{healthy}} - W_{\text{fault}}\right|$$

together with the learned scales shown as horizontal red lines. Even though many factors that affect the signals could not be controlled (such as exact vehicle speed and position on the track), we expect W to attain high values (representing significant differences between the healthy and damaged signals) in the lower frequency (higher scale) range. This is well reflected by Fig. 10, and we can conclude that the learned scales cover this region of interest well. We can also observe that some scales corresponding to higher frequency components are also taken into consideration. As mentioned previously, these can play a part when vehicle speed increases. In addition, they also account for detecting higher frequency vibrations caused by brake disc runout, which can happen according to [36]. In Fig. 11, we illustrate the Morlet wavelet filter bank corresponding to the learned scales.

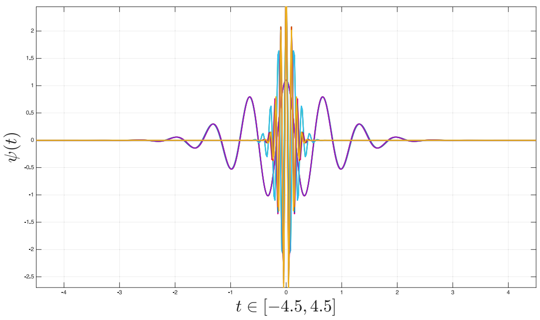

Fig. 11. Morlet wavelets corresponding to the learned scales.

VI. Conclusion

In this work, we propose a measurement system that can be used to detect brake disc deformations in passenger vehicles. In particular, we propose a measurement apparatus collecting information on the vibration profile of the braking system and the brake fluid pressure. Using the proposed measurement setup, we collected a dataset that can be used to conduct experiments for brake disc fault recognition. We propose the use of a novel wavelet kernel convolution-based neural network. Through experiments, we verify that:

1) The proposed measurement system is correct in the sense that the collected data can be used (with appropriate signal processing) to recognize brake disc deformations.
2) The proposed intuitive model-driven ML-based signal processing framework significantly outperforms previous and similar methods.
3) The parameters of the wavelet convolution layers in the proposed model contain physically interpretable information (wavelet coefficients), greatly improving the interpretability of the proposed model.

In the next steps of our research, we would like to exploit the fact that the proposed wavelet neural networks are compact in size, that is, they should be implemented on cost-efficient small computational devices (such as microcontrollers). Building on the experimental results discussed in this article, we would like to create a real-time application that can be used to continuously monitor the appearance of brake disc deformations.

Appendix A
Omitting Translation as a Learnable Parameter

In this section, we detail why omitting the translation parameter from the proposed wavelet convolution layer [see (8)] does not reduce the performance of our model. In this work, we consider each input $\boldsymbol{f}$ to be a sampling of some compactly supported function $f \in L_2(\mathbb{R})$. In addition, we assume the Lipshitz-continuity of the mother wavelet ψ with constant $K > 0$, that is, for any $t_1, t_2 \in \mathbb{R}$ we have

$$|\psi(t_1) - \psi(t_2)| < K\,|t_1 - t_2|\,.$$

Finally, we assume each $\boldsymbol{\psi}_{\lambda,\tau}$ is an equidistant sampling of $\psi_{\lambda,\tau}$ over the effective support of $\psi := \psi_{1,0} \in L_2(\mathbb{R})$. Choosing $\lambda > 1$ guarantees that $\boldsymbol{\psi}_{\lambda} := \boldsymbol{\psi}_{\lambda,0}$ was sampled over the effective support of $\psi_{\lambda,0}$. Notice, however, that leaving the translation $\tau \in \mathbb{R}$ as a free parameter can only lead to one of three cases.

1) Let $\tau := s \cdot h$ ($s \in \mathbb{N}, s > 0$), where h is defined according to (5) and assume that s is small enough, that is, the effective support of $\psi_{\lambda,\tau}$ is contained in the effective support of $\psi_{1,0}$ (the sampling interval). Then, we have
$$\left(\boldsymbol{f} * \boldsymbol{\psi}_{\lambda,0}\right)_{n-s} = \left(\boldsymbol{f} * \boldsymbol{\psi}_{\lambda,\tau}\right)_n .$$
Thus, the output of the WKN layer proposed in [16] can be expressed as a shifted (by s) variant of the result obtained by performing the convolution with $\boldsymbol{\psi}_{\lambda,0}$.
2) Let $|\tau|$ be large, that is, suppose that the effective support of $\psi_{\lambda,\tau}$ is not nested inside the sampling interval (the effective support of $\psi_{1,0}$). In this case, $\boldsymbol{\psi}_{\lambda,\tau}$ was not sampled over its effective support, which leads to poor quality estimates of the wavelet coefficients [see (6)].
3) Assume that $0 < |\tau| < h$ is satisfied. In this case, the output of the WKN convolution layer from (7) does indeed differ from $\boldsymbol{f} * \boldsymbol{\psi}_{\lambda,0}$. Let t_k denote an arbitrary sampling point from our sampling interval. The assumption on the Lipschitz continuity of ψ means that
$$\begin{aligned}\left|\psi_{\lambda,\tau}(t_k) - \psi_{\lambda,0}(t_k)\right| &= \left|\left(\boldsymbol{\psi}_{\lambda,\tau}\right)_k - \left(\boldsymbol{\psi}_{\lambda,0}\right)_k\right| \\ &< K \cdot |\tau| < K \cdot h.\end{aligned}$$
From here, we get
$$\begin{aligned}&\left|\left(\boldsymbol{f} * \boldsymbol{\psi}_{\lambda,\tau}\right)_n - \left(\boldsymbol{f} * \boldsymbol{\psi}_{\lambda,0}\right)_n\right| \\ &= \left|\sum_{j=1}^{N} \boldsymbol{f}_j \left(\boldsymbol{\psi}_{\lambda,\tau}\right)_{n-j} - \sum_{j=1}^{N} \boldsymbol{f}_j \left(\boldsymbol{\psi}_{\lambda,0}\right)_{n-j}\right| \\ &= \left|\sum_{j=1}^{N} \boldsymbol{f}_j \left(\left(\boldsymbol{\psi}_{\lambda,\tau}\right)_{n-j} - \left(\boldsymbol{\psi}_{\lambda,0}\right)_{n-j}\right)\right| \\ &\le \sum_{j=1}^{N} \left|\boldsymbol{f}_j\right| \left|\left(\boldsymbol{\psi}_{\lambda,\tau}\right)_{n-j} - \left(\boldsymbol{\psi}_{\lambda,0}\right)_{n-j}\right| < Kh\,\|\boldsymbol{f}\|_1\end{aligned}$$
where $\|\cdot\|_1$ denotes the $p = 1$ vector norm on $\mathbb{R}^N$. It follows that if the sampling time h is small enough, the choice of τ does not make a significant difference in the output of the convolution layer.

Notice that we can safely assume the Lipshitz continuity of the mother wavelet as well as the compact support of f (since we assume f sampled over a compact interval). Hence, our proposed wavelet convolution layer in (8) can be interpreted as a modified version of the one proposed in [16], which halves the number of learnable parameters (and greatly reduces model complexity) without any significant impact on model performance.

Credit Author Statement

Tamás Dózsa: Methodology, software, writing–original draft. **Pé;ter Õri:** Data curation, writing–original draft. **Gergõ Ungvári:** Validation. **Zoltán Szabó:** Investigation, supervision. Alexandros Soumelidis: Project administration, formal analysis. **István Lakatos:** Conceptualization, resources. **Péter Kovács:** Formal analysis, funding acquisition, writing–review and editing.

References

[1] O. van Schoor, J. L. van Niekerk, and B. Grobbelaar, "Mechanical failures as a contributing cause to motor vehicle accidents—South Africa," *Accident Anal. Prevention*, vol. 33, no. 6, pp. 713–721, Nov. 2001.

[2] M. T. Haq, V.-M.-K. Ampadu, and K. Ksaibati, "An investigation of brake failure related crashes and injury severity on mountainous roadways in Wyoming," *J. Saf. Res.*, vol. 84, pp. 7–17, Feb. 2023.

[3] Z. Hou, C. K. M. Lee, Y. Lv, and K. L. Keung, "Fault detection and diagnosis of air brake system: A systematic review," *J. Manuf. Syst.*, vol. 71, pp. 34–58, Dec. 2023.

[4] R. Jegadeeshwaran and V. Sugumaran, "Method and apparatus for fault diagnosis of automobile brake system using vibration signals," *Recent Patents Signal Process.*, vol. 3, no. 1, pp. 2–11, Apr. 2013.

[5] R. Jegadeeshwaran and V. Sugumaran, "Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines," *Mech. Syst. Signal Process.*, vols. 52–53, pp. 436–446, Feb. 2015.

[6] Y. Yin, X. Xiao, J. Bao, Z. Li, and Y. Lu, "Feature extraction and pattern recognition on the frictional faults of disc brake," *J. Balkan Tribological Assoc.*, vol. 22, no. 2, pp. 1208–1227, 2016.

[7] X. Zhang, M. Zhang, Z. Xiang, and J. Mo, "Research on diagnosis algorithm of mechanical equipment brake friction fault based on MCNN-SVM," *Measurement*, vol. 186, Dec. 2021, Art. no. 110065.

[8] A. Shen, C. Liu, S. Wang, and L. Zhu, "A Laplacian image-guided semantic segmentation framework for disc brake vibration displacement measurement," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–15, 2025.

[9] T. Dózsa, P. Őri, M. Szabari, E. Simonyi, A. Soumelidis, and I. Lakatos, "Brake disc deformation detection using intuitive feature extraction and machine learning," *Machines*, vol. 12, no. 4, p. 214, Mar. 2024.

[10] J. Li, S. Jiang, M. Li, and J. Xie, "A fault diagnosis method of mine hoist disc brake system based on machine learning," *Appl. Sci.*, vol. 10, no. 5, p. 1768, Mar. 2020.

[11] G. Xu, D. Song, D. Zhang, X. Zhang, and S. Sha, "A novel mechanical design of disc brakes for fault diagnosis and monitoring positive braking pressure in mine hoist," *Adv. Mech. Eng.*, vol. 11, no. 4, p. 16, Apr. 2019, doi: 10.1177/1687814019842494.

[12] N. Giulietti, P. Chiariotti, F. Zanelli, N. Debattisti, and A. Cigada, "Combined use of infrared imaging and deep-learning techniques for real-time temperature measurement of train braking components," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–8, 2025.

[13] M. Zhang, R. Hu, and Q. Feng, "Cross-speed state monitoring of high-speed train brake pads using guided deep subdomain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–11, 2024.

[14] B. J. Murphy, M. S. Lebold, J. C. Banks, and K. Reichard, "Diagnostic end to end monitoring & fault detection for braking systems," in *Proc. IEEE Aerosp. Conf.*, Mar. 2006, pp. 1–8.

[15] A. Belhocine and O. I. Abdullah, "Finite element analysis (FEA) of frictional contact phenomenon on vehicle braking system," *Mech. Based Design Struct. Mach.*, vol. 50, no. 9, pp. 2961–2996, Sep. 2022.

[16] T. Li et al., "WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 4, pp. 2302–2312, Apr. 2022.

[17] A. M. Ámon, K. Fenech, P. Kovács, and T. Dózsa, "Continuous wavelet transform based variable projection networks," in *Proc. 32nd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2024, pp. 1906–1910.

[18] 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.

[19] 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.

[20] T. Dózsa, P. Őri, G. Ungvári, A. Soumelidis, and I. Lakatos. (2025). *Detection of Brake Disc Deformation With Adaptive Wavelet Neural Networks (Implementation)*. [Online]. Available: https://gitlab.com/tamasdzs/brakeWav

[21] H. Tschürtz, F. Wagner, W. Schröter, Z. Szalay, and Á. Török, "System of systems safety analysis and evaluation in ZalaZONE," *Periodica Polytechnica Transp. Eng.*, vol. 49, no. 4, pp. 317–323, Dec. 2021.

[22] S. Shirmohammadi and H. Al Osman, "Machine learning in measurement part 1: Error contribution and terminology confusion," *IEEE Instrum. Meas. Mag.*, vol. 24, no. 2, pp. 84–92, Apr. 2021.

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[25] T. Dózsa, C. Böck, J. Meier, and P. Kovács, "Weighted Hermite variable projection networks for classifying visually evoked potentials," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 7, pp. 12415–12428, Jul. 2025.

[26] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.

[27] D. Łuczak, "Mechanical vibrations analysis in direct drive using CWT with complex Morlet wavelet," *Power Electron. Drives*, vol. 8, no. 1, pp. 65–73, Jan. 2023.

[28] M. S. Diab and E. Rodriguez-Villegas, "Embedded machine learning using microcontrollers in wearable and ambulatory systems for health and care applications: A review," *IEEE Access*, vol. 10, pp. 98450–98474, 2022.

[29] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, Nov. 2022.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[32] H. Yousefpour, S. A. Amiri, and Z. Mohammadpoory, "Bridge anomaly detection via structure health monitoring with ResNet-152 and scalogram techniques on vibration data," *Signal, Image Video Process.*, vol. 19, no. 4, p. 310, Apr. 2025.

[33] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A Python package for wavelet analysis," *J. Open Source Softw.*, vol. 4, no. 36, p. 1237, Apr. 2019.

[34] G. Bradski, "The OpenCV library," *Dr. Dobb's J. Softw. Tools*, vol. 25, pp. 120–125, Nov. 2000.

[35] J.-s. Peng, X. Li, and X. Gao, "Analysis of vibration reduction for brake judder based on vibration transmission path," *J. Vibroengineering*, vol. 17, no. 6, pp. 3187–3195, 2015.

[36] A. de Vries and U. von Wagner, "*The brake judder phenomenon*," SAE Tech. Paper 920554, 1992.

[37] H. Jakobsson, "Aspects of disc brake judder," *Proc. Inst. Mech. Eng., Part D, J. Automobile Eng.*, vol. 217, no. 6, pp. 419–430, 2003, doi: 10.1243/095440703766518069.