Spike detection with continuous wavelet transform based model driven machine learning methods

Attila Miklós Ámon[1,2]
Eötvös Loránd University
Department of Numerical Analysis[1]
and Siemens Mobility Kft.[2]
1117, Budapest Pázmány Péter sétány 1/C
Email: ze3vjn@inf.elte.hu

Bram Cornelis[3]
Siemens Industry Software NV[3]
Interleuvenlaan 68, Leuven, Belgium

Péter Kovács[1]
1117, Budapest Pázmány Péter sétány 1/C
Email: kovika@inf.elte.hu

Tamás Dózsa[1,4]
Institute for Computer Science and Control
Systems and Control Laboratory[4]
1111 Budapest, Kende u. 13-17.
Email: dozsa.tamas@hun-ren.sztaki.hu

***Abstract*—In this paper, we detect spike-like measurement errors in accelerometer sensor signals using continuous wavelet transform based machine learning methods. The wavelet coefficients are computed in automatic feature extraction layers (CWT layers), producing sparse representations of the input signals. The proposed methods ensure low model complexity, which allows real-time application. We complement a previously proposed variable projection based method to estimate wavelet coefficients with a numerical quadrature based approach. We present a qualitative and quantitative comparison of the CWT layers. To demonstrate the generality of the method, we introduce support vector machines supplemented with CWT layers in addition to previously used neural networks. Sensor fault detection experiments are conducted on real measurements using a low-cost accelerometer. The results of the experiments show that the proposed method achieves perfect accuracy on our dataset while outperforms previously used approaches in terms of interpretability and online usage. In addition to convincing classification accuracy, our results illustrate the interpretability of the proposed model driven machine learning framework.**

I. Introduction

During measurement campaigns, monitoring data quality is essential. An important preprocessing step on the measurement data is the detection and removal of various measurement errors [1]. These are often invisible or very difficult to detect for the operators performing the test. Due to the high sampling rate, tests can generate large amounts of data in a few seconds. Therefore, the use of automatic detection methods is necessary to solve the problem. The methods are expected to be able to run in real time on low-power target hardware, thus reducing the cost of performing the tests, while meeting the requirements of industrial applications. Online detection of faults reduces the time spent on the experiments by allowing for immediate intervention by test operators.

In this paper, we detect so-called 'spike' measurement errors in one-dimensional accelerometer data. These appear as abnormal peaks in the data that arise and disappear suddenly, as opposed to peaks caused by physical effects that result in oscillatory waveforms. Figure 1 shows side by side a normal peak caused by a physical effect and a spike caused by hardware malfunction. Such errors can arise in the signal for various reasons. One reason may be the use of a low-cost sensor, which is common in rough environments in automotive testing. Another cause may be cable failure in the test setup [2]. In some cases, the detection of these errors is difficult because spikes can appear on physical peaks or extremely close next to a physical peak. Examples of these types of errors can be seen in Fig. 2.

Several classical signal processing methods for solving this problem can be found in the literature. In [3] a Bayesian spike detection algorithm is proposed. In [4] Deuschle et al. proposed a method using the Dynamic Time Warping (DTW) algorithm. These methods, because they do not use supervised learning models, do not require the generation of large amounts of labeled data, which is often difficult to produce. On the other hand, machine learning algorithms like shallow neural networks or a support vector machine (SVM) could have running time performance benefits on low-cost target hardware.

Machine learning (ML) has become popular in recent years in the signal processing domain. ML algorithms enable the accurate classification of non-linearly separable datasets which is often required to solve signal processing tasks. However, deep learning (DL) approaches use expensive target hardware hence violate the need for cost-efficient real time usage. DL methods also usually use millions of parameters which are difficult for humans to interpret. The AI regulation of the European Union (EU AI Act [5]) makes the use of trustworthy and transparent methods mandatory in the fields of safety-critical engineering applications like biological signal processing or autonomous driving. Explainable AI (XAI) research aims to develop methods to meet the needs of safety-critical

signal processing applications. Post-hoc interpretable methods examine the behavior of previously trained DL models [6]. Model driven [7] or physics-informed [8] ML methods supplement classical AI algorithms with interpretable mathematical transformations. The parameters of these methods describe physical phenomena like frequency information of the signal hence the parameters of the transformations are interpretable and the developed models exhibit transparent architectures.

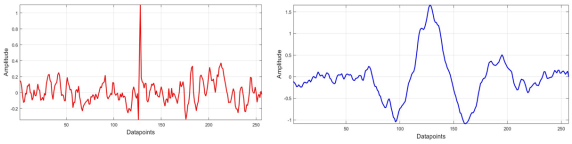

Fig. 1. An abnormal spike and a normal peak.

In [9] Kovács et al. introduced a variable projection (VP) based model driven neural network (VP-Net) whose first layer acts as an adaptive automatic feature extractor. More precisely, the first layer of a VP-Net expresses the (1-D) input signals by projecting them onto a finite dimensional subspace spanned by a parameterized orthogonal system in an appropriate Hilbert-space. The parameters of the VP based automatic feature extraction layer trained together with the weights of the neural network. In [10] Dózsa et al. extended the method to SVM classifiers (VP-SVM). Implementations on low-cost microcontroller of VP-SVM [10] and VP-Net [11] were made in C/C++ languages.

These methods enable the solution of regression or classification problems where the input signal has quasi-periodic behavior. That is, a single input example is expected to correspond to a single "quasi" period of some underlying time process. In [10] we tested VP-SVM in ECG arrhythmia detection and detection of the aforementioned spike type measurement errors.

Hermite function system is commonly used in biological signal processing because of the shape similarities of various signals like ECG heartbeats [12], [13]. Since tire sensor signals are quasi periodic and have similar morphological properties to ECG signals, the so-called adaptive Hermite function system was also effectively used as feature extractors [14]. However, spike type measurement errors do not share morphological similarities with ECG signals and various generalized Hermite expansions were unable to capture significant properties of the data. Instead, a preprocessing step was made to solve this problem in [10]. Continuous wavelet transform (CWT) coefficients of the peaks at high frequencies were pre-computed using the Morse wavelet. Then summing the columns of the generated CWT coefficient matrix resulted signals with clearly distinguishable morphologies between physical peaks and spikes. The transformed peak signals can be approximated properly with adaptive Hermite functions. These transformed signals passed to the adaptive Hermite based VP-SVM which solved the task successfully.

In [15] we introduced CWT based automatic feature extraction layer with neural networks (CWT-VP) where the

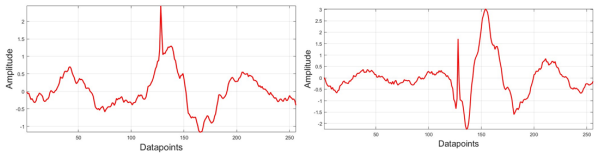

Fig. 2. A spike appearing on top of a normal peak and a spike close to a normal peak, respectively.

VP based approximation of CWT coefficients of the input signals were computed. In this paper, we detect spikes with CWT based automatic feature extraction layers based on our previous results in [10] and in [15]. The proposed method significantly simplifies the fault detection scheme proposed in [10], because the CWT based preprocessing task can be omitted. This reduces the computational cost and increases the interpretability of the method.

In Sec. II we introduce novel quadrature based approximation of CWT coefficients in the feature extraction layer (CWT-Q), in Sec. III compare the proposed method to CWT-VP. In addition we test CWT-VP and CWT-Q with model driven variations of neural networks and SVM classifiers.

II. Continuous wavelet transform based automatic feature extraction layers

Suppose $f, \psi \in L_2(\mathbb{R})$. The continuous wavelet transform of f respect to ψ is

$$W_\psi f(\tau, \lambda) := \int_{\mathbb{R}} f(t)\overline{\psi_{\tau,\lambda}(t)}dt, \tag{1}$$

where $\tau, \lambda \in \mathbb{R}, \lambda \neq 0$, $\overline{\psi(t)}$ denotes the complex conjugate of $\psi(t)$, f denotes the input signal and ψ is the so-called mother wavelet or analyzing wavelet. The wavelet transform is a similarity measure of the input signal with respect to a τ translated and λ dilated wavelet function, i. e.

$$\psi_{\tau,\lambda}(t) = |\lambda|^{-1/2}\psi(\lambda^{-1} \cdot (t - \tau)). \tag{2}$$

Low λ scale values correspond to high frequency components and high λ scale values correspond low frequencies in the input signal. That is, a CWT coefficient gives frequency information at scale λ and at translation τ in time. We note, although by definition we could use any $\psi \in L_2(\mathbb{R})$ as an analyzing wavelet, using functions that are well localized in the time and frequency domain have many advantages from an application point of view [16]. These so-called admissible wavelet functions conserve the energy of the input signal and guarantee the reconstruction of the input from an appropriate number of wavelet coefficients. For example, Ricker and Morlet wavelets satisfy the aforementioned properties and are widely used in signal processing applications [16].

In many use cases, the equidistant discrete sampling of the input signal can be accessed. Although CWT is a continuous model, numerical approximations of the method exist in the literature, for example, the Matlab implementation of the CWT algorithm [17]. We note that other methods like the discrete

wavelet transform (DWT) exist in the literature to compute the time-scale information of the input vector [16], but CWT has finer resolution and DWT methods are not translation invariant [16]. The latter property will be essential in our use case of spike detection.

Classical applications of the CWT routine compute a large number of coefficients at time points $t_i \quad (i = 0, \dots N - 1, N \in \mathbb{N})$ and at $M \in \mathbb{N}$ number of scales, i. e. a CWT coefficient matrix with size $M \times N$ is calculated which usually contains many more entries than what is necessary to solve the underlying signal processing problem. Thus, this type of approach requires significant computational resources and is usually accompanied by the use of deep learning models for image processing tasks [18].

In [19], the WaweletKernelNet (WKN) architecture was proposed where the CWT coefficient matrix was computed with wavelet kernel based convolutions acting as an automatic feature extraction layer of a deep CNN model. In WKN, the scale and the translation parameter of convolutional kernels were optimized. With convolution we can compute CWT coefficient for every t_i at a given scale. This means, that the number of computed features is usually large for a WKN and subsequent pooling and dropout operations are used to achieve a sparse signal representation.

A. CWT-VP layer

In [15] we introduced CWT based variable projection (CWT-VP) neural networks (NN), where only a predefined, few number of coefficients were computed. The parameters of the coefficients, translation and dilation pairs, were optimized together with the weights of the NN during training. Hence the output of the layer resulted a sparse signal representation which do not require additional pooling or dropout operations. The CWT-VP layer computes features that have physical meaning and can be interpreted by humans. Namely, the CWT-VP Layer learns the n most important wavelet coefficients (corresponding to time and frequency information of the input) from the point of view of the underlying classification problem. CWT-VP layers can be used with any mother wavelet whose partial derivatives respect to the optimized parameters exist. We used the real Morlet wavelet in the CWT-VP layer and computed the partial derivatives analytically in [15].

The formulation of the CWT-VP layer is

$$g^{\boldsymbol{\eta}} : \mathbb{R}^N \to \mathbb{R}^n, \quad g^{\boldsymbol{\eta}}(\boldsymbol{f}) := \Psi(\boldsymbol{\eta})^+ \boldsymbol{f}. \tag{3}$$

where $n << N$ and $\Psi(\boldsymbol{\eta})^+$ denotes the Moore-Penrose pseudo inverse. We optimize the parameter vector $\boldsymbol{\eta}$ which contains (dilation,translation) pairs corresponding to CWT coefficients. $\Psi \in \mathbb{R}^{N \times n}$ contains the discrete sampling of the translated and dilated wavelets. Thus $\boldsymbol{\eta}$ defines the parameters of the subspace spanned by the column vectors of the transformation matrix Ψ. In [20] Golub and Pereyra derived the analytical form of the partial derivatives of the pseudo inverse matrix which allows for optimizing $\boldsymbol{\eta}$ using gradient based methods and a seamless integration of $g^{\boldsymbol{\eta}}$ into backpropagation schemes.

To avoid the problem of vanishing gradients, the training of VP-CWT layers usually requires adding the following regularizing term to the objective:

$$J_{VP}(\boldsymbol{\eta}) := \frac{\alpha}{s} \sum_{i=1}^{s} \frac{\|\boldsymbol{f}_i - \Psi\Psi^+ \boldsymbol{f}_i\|_2^2}{\|\boldsymbol{f}_i\|_2^2}, \tag{4}$$

where α is the penalty parameter and s is the batch size during training.

B. Proposed CWT-Q layer

In this section we introduce a new approach to calculate CWT coefficients. Henceforth assume, that $f \in L_2(\mathbb{R})$ is compactly supported in $[a, b] \subset \mathbb{R}$. In the proposed method, we approximate wavelet coefficients using Riemann sums, that is a CWT coefficient is equal to

$$c_i := \int_a^b f(t)\overline{\psi_{\tau_i,\lambda_i}(t)}dt \approx \sum_{j=0}^{N-1} f(t_j)\overline{\psi_{\tau,\lambda}(t_j)}, \tag{5}$$

where N is the number of sampling points and $a =: t_0 < t_1 < \ldots < t_{N-1} := b$ is an equidistant partitioning of the support of f. Then, we can define the proposed CWT-Q layer as the mapping

$$g^{\boldsymbol{\eta}} : \mathbb{R}^N \to \mathbb{R}^n, \quad g^{\boldsymbol{\eta}}(\boldsymbol{f}) := \Psi(\boldsymbol{\eta})^T \boldsymbol{f}, \tag{6}$$

where the columns of Ψ contains the sampling of translated and dilated wavelets. Then for the backpropagation step we have to compute only the partial derivatives respect to η of Ψ:

$$\left(\frac{\partial g^{\boldsymbol{\eta}}}{\partial \eta} f\right)_j = \frac{\partial \Psi}{\partial \eta_j} f \tag{7}$$

where $\left(\frac{\partial g^{\boldsymbol{\eta}}}{\partial \eta} f\right)_j$ denotes the j-th column of the Jacobian of g^{η}. In the proposed CWT transformation we have $\boldsymbol{\eta}_{2j} = \lambda_j$ and $\boldsymbol{\eta}_{2j+1} = \tau_j \quad (j = 1, \dots, n)$, hence $\partial\Psi \in \mathbb{R}^{N \times 2n}$, $\Psi \in \mathbb{R}^{N \times n}$. We regularize the parameters of the layer with L_2 regularization:

$$J_Q(\eta) := \alpha \sum_{i=1}^{n} c_i^2, \tag{8}$$

where α is the penalty term and n is the number of computed features. One benefit of the proposed approach is that given a compactly supported, continuously differentiable signal f, we can estimate the error of the numerically computed wavelet coefficients by

$$\left|W_\psi f(\lambda, \tau) - h \cdot \left(\Psi(\boldsymbol{\eta})^T \boldsymbol{f}\right)\right| < h \cdot \frac{M_1(b-a)}{2}, \tag{9}$$

where $h := t_1 - t_0$ and

$$M_1 := \max_{\xi \in [a,b], k=1,\dots,m} \left|f'(\xi) \cdot \overline{\psi_{\lambda_k,\tau_k}}(\xi)\right|.$$

For the VP based approximation, error formulas also exist [21]. One advantage of the quadrature based method that we do not have to compute the pseudo inverse of Ψ. CWT-Q requires less computation both at the forward and backward passes, but CWT-VP ensures the L_2-optimal reconstruction

of the input signals. The trained features also depend on the different regularization strategies. Both the quadrature and VP-based wavelet coefficient estimates can be useful depending on the requirements of a given classification or regression problem. We compare the performance of the two layer in the next section.

III. Experiments and results

In our experiments, we used data from a low-cost accelerometer sensor instrumented on a wheel hub of a car [3]. We extracted the peaks from the accelerometer signal by extending the peak location with its neighboring samples. The created dataset was labeled by an expert. The number of extracted spike-type errors was very low compared to extracted physical peaks resulting in a highly unbalanced dataset. We managed this issue by creating a balanced training set with the same number of spikes and physical peaks.

The architecture of the CWT based machine learning models were the following. In the first layer n number of CWT coefficients were computed with CWT-Q (Eq. 6) or CWT-VP (Eq. 3). After the automatic feature extraction step, the calculated features were passed to the underlying classification algorithm. In the case of CWT-Q-NN and CWT-VP-NN the features were passed to a neural network with additional hidden layers. ReLu activation functions were used between these layers. Radial Basis Function (RBF) kernel SVM was used for CWT-Q-SVM and CWT-VP-SVM after the first layer. For a precise description of the employed model architectures, we refer to [9] and [10].

In Table I we can see the results of the experiments. Using the aforementioned training and test sets, we classified the accelerometer signals using four different adaptive CWT based model driven ML schemes. We compared our results with an Hermite-expansion based VP-SVM model (requiring preprocessing), and a number of classical ML algorithms.

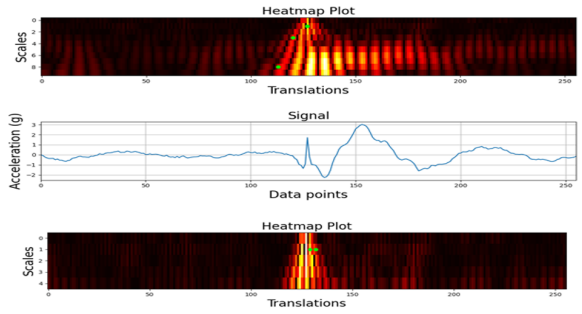

Fig. 3. Spike near a physical peak with its scalogram at high frequencies (low scales). Trained features marked with green dots. Top: trained CWT coefficients with CWT-Q layer. Bottom: two trained coefficients with CWT-VP at high frequencies, one coeff. corresponding to low frequrncy not displayed.

Models with CWT based automatic feature extraction layers achieved perfect accuracy while they do not need input transformation as in the case of Hermite based VP-SVM. Classical ML methods could not differentiate spikes from normal peaks perfectly. We used Morlet wavelet in CWT-VP and CWT-Q layers. In all cases we used three coefficients to match the model complexity of the Hermite expansion based VP-SVM. CWT based layers followed by a neural network with one hidden layer and sigmoid with one output layer for binary classification. For the CWT-VP-SVM we used the same optimized implementation as in the case of Hermite VP-SVM in [10]. After a hyperparameter search, we find that CWT-VP and CWT-Q layers achieved perfect accuracy with only 9 and 12 neurons, respectively. The public implementation of the methods can be found at [22].

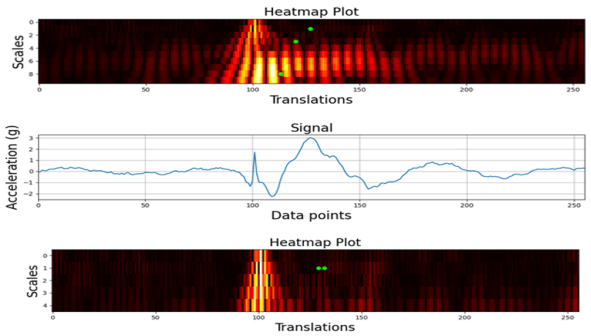

Fig. 4. Physical peak with its scalogram at high frequencies (low scales). Trained features marked with green dots. Top: trained CWT coefficients with CWT-Q layer. Bottom: two trained coefficients with CWT-VP at high frequencies, one coeff. corresponding to low frequrncy not displayed.

Model	Accuracy
CWT-VP-NN	100%
CWT-Q-NN	100%
CWT-VP-SVM	100%
CWT-Q-SVM	100%
Hermite VP-SVM (CWT-preproc.)	100%
FCNN	94.74%
SVM (RBF kernel)	97.37%
Random Forest	86.84%
Decision Tree	97.37%
Gaussian Naive Bayes	60.53%
Gradient Boost	86.84%

TABLE I
CLASSIFICATION RESULTS OF THE SPIKE DATASET

The interpretability of the trained CWT coefficients is underlined by comparing the scalograms of the input signals with the positions of the learned wavelet coefficients. On Figs. 3 and 4 different representations of a spike appearing close to a physical peak are illustrated. We generated the scalograms using MATLAB cwt routine at high frequencies, then plotted the wavelet coefficients learned by the proposed CWT-Q and CWT-VP layers as green points. In conclusion, both CWT layers learned meaningful information corresponding to significant scales and time instances which can be used to recognize the presence of spikes. Interestingly, although initial parameters were uniformly chosen for CWT-Q and CWT-VP based methods, the time-scale positions of the learned coefficients were somewhat different. This can be explained by the different regularization terms used by the two approaches: in order to produce good approximations of the input signals

(in L_2 norm), CWT-VP needed to learn a wavelet coefficient position corresponding to lower frequencies. Hence the third coefficient is not depicted on the truncated scalograms in Figs. 3 and 4. In contrast, the CWT-Q layer was not regularized to compute wavelet coefficients which provide a good approximation of the input signals. Hence, all 3 learned scale-time positions corresponded to high frequencies. Since the frequency profile of spikes and physical peaks differ most significantly at higher frequencies, the learned parameters have interpretable meaning.

IV. Conclusion

In this paper we used continuous wavelet transform based automatic feature extraction layers to detect spike measurement errors in accelerometer signals. We proposed a novel quadrature based approximation of the CWT coefficients (CWT-Q layer). We compared CWT-Q to a VP based method. In the conducted experiments both models achieved perfect accuracy while requiring less computation than earlier methods. We also tested CWT-VP and CWT-Q layers with SVMs. Interpretability of the learned features was also improved compared to earlier Hermite-expansion based approaches.

In future works we would like to extend the models in other application areas. The continuation of the research will include the examination of post-hoc explainability of the parameters of the underlying classifiers. We would like to define metrics that could assure the trustworthiness of ML algorithms used with CWT based feature extraction layers.

Acknowledgment

Supported by the EKÖP-KDP-24 university excellence scholarship program cooperative doctoral program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. Supported by the EKÖP-24 University excellence scholarship program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. The research was supported by the European Union within the framework of the National Laboratory for Autonomous Systems. (RRF-2.3.1-21-2022- 00002). Project no. K146721 have been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the K_23 "OTKA" funding scheme. Supported by the University Excellence Fund of Eötvös Loránd University, Budapest, Hungary (ELTE). The research was supported by the National Research, Development and Innovation Office (NKFIH) through the project 'Design of high performance safe autonomous vehicle systems via integrated robust control and learning-based methods' (2021-1.2.4-TÉT-2022-00065). This project was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

[1] L. Jia, B. Cornelis, and K. Gryllias, "Validation of time series anomaly detection methods for automated detection of measurement anomalies in noise and vibration testing."

[2] B. Cornelis, F. Deuschle, and K. Gryllias, "Performance study of dtw-based spike measurement anomaly detection in sensors on real world tests." in *Surveillance, Vibrations, Shock and Noise*, 2023.

[3] B. Cornelis and B. Peeters, "Online bayesian spike removal algorithms for structural health monitoring of vehicle components," in *EURODYN 2014: IX INTERNATIONAL CONFERENCE ON STRUCTURAL DYNAMICS*, vol. 2014. EUROPEAN ASSOC STRUCTURAL DYNAMICS, 2014, pp. 2295–2301.

[4] F. Deuschle, B. Cornelis, and K. Gryllias, "Robust sensor spike detection method based on dynamic time warping," in *Proc. 4th Int. Conf. Advances in Signal Processing and Artificial Intelligence (ASPAI)*, 2022, pp. 20–27.

[5] EU AI Act - http://data.europa.eu/eli/reg/2024/1689/oj.

[6] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery." *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[7] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *Proceedings of the IEEE*, vol. 111, no. 5, pp. 465–499, 2023.

[8] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[9] P. Kovács, G. Bognár, C. Huber, and M. Huemer, "Vpnet: Variable projection networks," *International Journal of Neural Systems*, vol. 32, no. 01, p. 2150054, 2022.

[10] T. Dózsa, F. Deuschle, B. Cornelis, and P. Kovács, "Variable projection support vector machines and some applications using adaptive hermite expansions," *International Journal of Neural Systems*, vol. 34, no. 1, 2024.

[11] T. Dózsa, A. Ámon, F. Braun, E. Simonyi, A. Soumelidis, J. Volk, and P. Kovács, "Towards intelligent tire development," 2023.

[12] P. Kovács, C. Böck, T. Dózsa, J. Meier, and M. Huemer, "Waveform modeling by adaptive weighted hermite functions," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1080–1084.

[13] T. Dózsa and P. Kovács, "Ecg signal compression using adaptive hermite functions," in *ICT Innovations 2015*, S. Loshkovska and S. Koceski, Eds. Cham: Springer International Publishing, 2016, pp. 245–254.

[14] T. Dózsa, J. Radó, J. Volk, Kisari, A. Soumelidis, and P. Kovács, "Road abnormality detection using piezoresistive force sensors and adaptive signal models," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[15] A. M. Ámon, K. Fenech, P. Kovács, and T. Dózsa, "Continuous wavelet transform based variable projection networks," in *European conference on signal processing (EUSIPCO)*, 2024.

[16] C. Gasquet and P. Witomski, *Fourier analysis and applications: filtering, numerical computation, wavelets*. Springer Science & Business Media, 2013, vol. 30.

[17] D. Jordan, R. Miksad, and E. Powers, "Implementation of the continuous wavelet transform for digital time series analysis," *Review of scientific instruments*, vol. 68, no. 3, pp. 1484–1494, 1997.

[18] M. M. Al Rahhal, Y. Bazi, M. Al Zuair, E. Othman, and B. BenJdira, "Convolutional neural networks for electrocardiogram classification," *Journal of Medical and Biological Engineering*, vol. 38, pp. 1014–1025, 2018.

[19] T. Li, Z. Zhao, C. Sun, L. Cheng, X. Chen, R. Yan, and R. X. Gao, "Waveletkernelnet: An interpretable deep neural network for industrial intelligent diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 4, pp. 2302–2312, 2021.

[20] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse problems*, vol. 19, no. 2, p. R1, 2003.

[21] A. M. Ámon, F. Kristian, K. Péter, and D. Tamás, "Rational Gaussian wavelets and corresponding model driven neural networks," *arXiv:2502.01282*, 2025.

[22] Implementation of the proposed methods: https://gitlab.com/AmonAttilaMiklos/CwtSpikeDet/.