

Rational Gaussian Wavelets and Corresponding Model Driven Neural Networks

Attila Miklós Ámon , Kristian Fenech , Péter Kovács , and Tamás Dózsa 

Abstract—In this paper we introduce a highly adaptive continuous wavelet transform using Gaussian wavelets multiplied by an appropriate rational term. The zeros and poles of this rational modifier act as free parameters and their choice highly influences the shape of the mother wavelet. This allows the proposed construction to approximate signals with complex morphology using only a few wavelet coefficients. We show that the proposed rational Gaussian wavelets are admissible and provide numerical approximations of the wavelet coefficients using variable projection operators. In addition, we show how the proposed variable projection based rational Gaussian wavelet transform can be used in neural networks to obtain a highly interpretable feature learning layer. We demonstrate the effectiveness of the proposed scheme through a number of numerical experiments including biomedical applications, and the detection of abnormal road surface based on tire sensor signals.

Index Terms—Gaussian wavelets, continuous wavelet transform, rational functions, variable projection, variable projection networks, ECG.

Received 23 November 2024; revised 14 June 2025; accepted 11 July 2025. Date of publication 24 July 2025; date of current version 19 August 2025. This work was supported by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the NVKDP-2021 funding scheme under Grant C1748701; financed under the TKP2021-NVA and the K_23 “OTKA” funding schemes under Grant TKP2021-NVA-29 and Grant K146721, in part by the 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 under Grant EKÖP-KDP-24, in part by the European Union within the framework of the National Laboratory for Autonomous Systems under Grant RRF-2.3.1-21-2022-00002, and in part by the University Excellence Fund of Eötvös Loránd University, Budapest, Hungary (ELTE). 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 of this article and approving it for publication was Che Lin. (*Corresponding author: Péter Kovács.*)

Attila Miklós Ámon is with the Faculty of Informatics, Department of Numerical Analysis, Eötvös Loránd University, 1117 Budapest, Hungary and also with Siemens Mobility Kft, R&D Office, 1117 Budapest, Hungary (e-mail: ze3vjn@inf.elte.hu).

Kristian Fenech is with the Faculty of Informatics, Department of Artificial Intelligence, Eötvös Loránd University, 1117 Budapest, Hungary (e-mail: fenech@inf.elte.hu).

Péter Kovács is with Faculty of Informatics, Department of Numerical Analysis, Eötvös Loránd University, 1117 Budapest, Hungary (e-mail: kovika@inf.elte.hu).

Tamás Dózsa is with the Faculty of Informatics, Department of Numerical Analysis, Eötvös Loránd University, 1117 Budapest, Hungary, and also with the HUN-REN Institute for Computer Science and Control, Systems and Control Laboratory, H-1111 Budapest, Hungary, and also with the Vehicle Industry Research Center, University of Győr, 9026 Győr, Hungary (e-mail: dozsa.tamas@hun-ren.sztaki.hu).

Digital Object Identifier 10.1109/TSP.2025.3592099

I. INTRODUCTION

REPRESENTING signals using wavelet coefficients has proved beneficial in a variety of fields. Even though the modern notion of the wavelet transform is usually dated from [1] written by Alex Grossmann and Jean Morlet, historical accuracy requires that previous works now included in broader wavelet theory are also mentioned. These include results achieved by Haar, Franklin, Littlewood and Paley and others (we recommend [2] for a thorough overview).

The widespread use of wavelets in various signal and image processing applications is no coincidence. In addition to important mathematical properties, the wavelet transform provides the so-called time-scale representation of signals, which is closely associated with time-frequency representations. In other words, the wavelet transform allows us to study the frequency profile of signals in certain time windows. In this way, representing signals using wavelet transforms inherently provides interpretable information about their behavior. This property has been widely exploited in fields related to signal processing such as fault detection, where sudden changes in the frequency profile of signals can indicate the appearance of faults [3]. Contrary to classical time-frequency approaches such as the Gábor transform (see, e.g. [4]), wavelet representations allow for different time localization properties at different scales (frequencies), which is also well exploited in applications. Finally, the introduction of multiresolution analysis (see [4] and e.g. [5]) allowed for a practically realizable way to construct orthogonal wavelets. In turn, quick algorithms realizing the discrete wavelet transform were developed [6] allowing for many applications in signal and image processing [3], [7], [8], [9]. For a more comprehensive review of wavelets and their history, the authors recommend [10].

In most applications, a fixed so-called mother (or analyzing) wavelet is used to perform wavelet analysis of the signals. These may have certain general beneficial properties such as a compact support, symmetry or smoothness. In addition, a mother wavelet may be selected due to its morphological similarity to the signals to be analyzed, for example so-called Ricker wavelets are frequently employed to represent ECG signals [11], [12]. In practical applications however, signals are always represented by a finite set of wavelet coefficients and fixing a mother wavelet may not be able to capture the behavior of the signal to a satisfactory degree. In order to overcome this limitation, a variety of adaptive wavelet transforms have been introduced (see e.g. [13], [14], [15]).

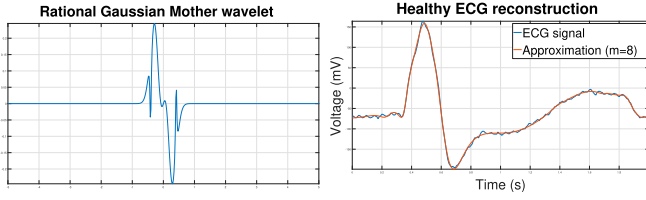


Fig. 1. LEFT: A rational gaussian motherwavelet. Here $n = 3$ poles and $p = 10$ zeros were used. RIGHT: Reconstruction of ECG signal using $m = 8$ wavelet coefficients.

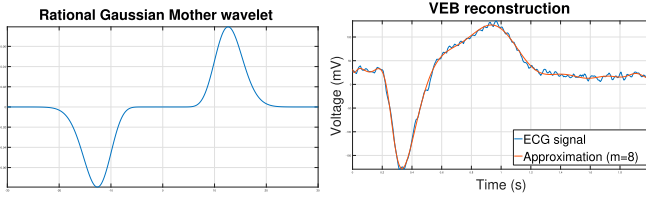


Fig. 2. Approximation of QV signals with RGW coefficient based reconstructions. LEFT: optimized RGW mother wavelet. RIGHT: approximation of a ventricular ectopic heartbeat ECG (VEB) with $m = 8$ wavelet coefficients. The number of poles and zeros of the RGW mother wavelet matched that of the healthy case (see Fig. 1).

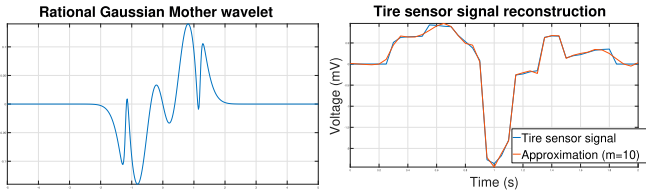


Fig. 3. Approximation of a signal obtained from force sensor implanted into the tire of a passenger vehicle. For a single tire revolution, these are QV signals with a main lobe centered around the point, where sensor is closest to the ground (and thus tire deformation acting on the sensor is maximal). to obtain this approximation, $m = 10$ wavelet coefficients were used, where RGW wavelets defined by $n = 1$ poles and $p = 3$ zeros were considered.

In this work, we present a new class of mother wavelets referred to as rational Gaussian wavelets (RGW). Instead of previous approaches, where parameters of existing wavelets were optimized (e.g. [15]), the proposed wavelet class has arbitrary degrees of freedom. This allows for the construction of highly unusual mother wavelet morphologies (see e.g. Figs. 1–3), while guaranteeing smoothness and symmetry. Admissibility is proved for the RGW class. In addition, every parameter of the proposed class of wavelets can be optimized using gradient based methods which allows the proposed transform to be easily included in modern machine learning (ML) architectures.

ML methods have become increasingly important in signal processing [16], [17]. This is in large part due to their ability to approximate highly nonlinear operators based solely on data. Deep learning methods have achieved remarkable results in image processing [18], natural language processing [19] and biomedical signal processing [20], [21], [22] among other disciplines. One of the most cited limitations of ML methods, however, is that well-performing models usually consist of a large number of parameters, which are difficult to interpret by

humans. For example, the popular GPT-3 model contains ≈ 175 billion parameters [23]. These factors make the deployment of such models expensive in the case of real-time, and especially safety critical applications (such as autonomous driving and biomedical signal processing).

To overcome this issue, several recent research directions have been considered. The taxonomy of explainable artificial intelligence methods proposed in [24] identifies several important, distinct ways to approach the question of explainability. One of these, the construction of so-called transparent ML methods, attempts to construct partially interpretable ML models. Recent examples of transparent ML schemes include physics informed neural networks [25], [26] which incorporate classical physical models with meaningful parameters into ML models. It is important to mention that the use of the wavelet transform has also been considered previously as a means of achieving transparent ML models. In particular, the recently introduced WaveletKernelNet [11] (WKN) exploits the relationship between convolution and wavelet coefficients to create a partially interpretable neural network relying on the wavelet transform. Nevertheless in Section III-B we provide a qualitative comparison between the chosen model architecture and WKN.

In this work, of particular interest to us are so-called variable projection (VP) based ML methods [20], [27], [28]. VP operators implement adaptive orthogonal projections in Hilbert spaces and can be used to obtain interpretable and sparse signal representations. They have been successfully integrated into various ML models including neural networks [20] and support vector machines [21]. In our recent work [29], we proposed the approximation of continuous wavelet coefficients using VP operators. In fact, a VP-neural network architecture was considered, where the first layer was responsible for representing the input using only a few wavelet coefficients. The learnable parameters of the layer consisted of the scale and translation parameters associated with the wavelet coefficients. The output of the layer was the collection of wavelet coefficients defined by these learned parameters. In this way, the proposed model learned which time and scale parameters provided the most crucial information for the classification of the input signals. In the current study, this idea is expanded upon by introducing a variable projection based continuous wavelet transform which uses the rational Gaussian wavelet class that we introduce in this paper. In this approach, not only translation and scale parameters, but also parameters governing the morphology of the mother wavelet can be learned. As shown in our experiments in Section IV, this allows for sparser signal representations and clearer explainability than similar alternative ML methods.

In summary, the main novel contributions of our paper are the following:

- 1) The introduction of rational Gaussian wavelets (RGW). Through an arbitrary number of parameters, these wavelets allow for the manipulation of the mother wavelet's morphology.
- 2) We show that the RGW class is admissible, which allows for signal reconstruction based on the continuous wavelet coefficients (in $L_2(\mathbb{R})$ norm).

- 3) We show that parameters of an RGW mother wavelet can be optimized using gradient based methods. This allows for simple inclusion into the variable projection framework and subsequently ML models utilizing backpropagation.
- 4) We provide an upper bound on the error of wavelet coefficient estimates acquired using variable projection operators for signals with a compact support.
- 5) We provide four different experiments showcasing the utility of the proposed RGW continuous wavelet transform and the subsequent variable projection based RGW ML schemes. In our first experiment, we show how the adaptive nature of RGW can be leveraged to construct meaningful time-scale representations of non-stationary oscillating signals. Our second experiment verifies that variable projection based RGW coefficients can be used to effectively compress ECG signals. In this experiment we also compare the performance of our method to various other adaptive wavelet transforms. Our third experiment showcases a neural network with interpretable RGW layers, which can be used to detect ventricular ectopic beats (a common arrhythmia) in ECG. Finally, in our last experiment we recognize abnormalities on the road surface such as bumps and potholes using a similar, RGW based neural network. The implementation of the proposed methods and experiments is available in [30], to ensure reproducibility.

II. MOTIVATION FOR ADAPTIVE CONTINUOUS WAVELET TRANSFORMS

A. Wavelet Transforms

Consider the family of functions

$$\psi_{\lambda,\tau}(t) = \frac{1}{\sqrt{|\lambda|}} \psi(1/\lambda \cdot (t - \tau)), \quad (t, \lambda, \tau \in \mathbb{R}, \lambda \neq 0),$$

where the function ψ is referred to as the “mother” or “analyzing” wavelet. The wavelet transform of a function f with respect to ψ is defined as a function of two variables

$$W_\psi f(\lambda, \tau) := \int_{-\infty}^{\infty} f(t) \overline{\psi}_{\lambda,\tau}(t) dt, \quad (1)$$

whenever the integral exists. In Eq. (1), $\overline{\psi}$ denotes the complex conjugate of ψ . We note that the existence criteria is fulfilled for example if $f, \psi \in L_2(\mathbb{R})$ by Hölder’s inequality. For a fixed scale-translation pair $(\lambda, \tau) \in \mathbb{R} \setminus \{0\} \times \mathbb{R}$, the number $W_\psi f(\lambda, \tau)$ will also be referred to as a *wavelet coefficient* corresponding to (λ, τ) .

The admissibility theorem (see e.g. [4]) describes a set of conditions on ψ which allow for signal reconstruction in L_2 norm.

Theorem 1 (Admissibility property): Suppose the wavelet $\psi \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ satisfies

- 1) $\int_{-\infty}^{\infty} \frac{1}{|\xi|} \cdot \left| \widehat{\psi}(\xi) \right|^2 d\xi = M < \infty$,
- 2) $\|\widehat{\psi}\|_2 = 1$,

where $\widehat{\psi}$ denotes the Fourier transform of ψ . Then, for any $f \in L_2(\mathbb{R})$ the following results hold:

- 1) Energy conservation:

$$\frac{1}{M} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |W_\psi f(\lambda, \tau)|^2 \frac{d\lambda d\tau}{\lambda^2} = \int_{-\infty}^{\infty} |f(t)|^2 dt.$$

- 2) Define the function f_ε as

$$f_\varepsilon(t) := \frac{1}{M} \int_{|\lambda| \geq \varepsilon} \int_{-\infty}^{\infty} W_\psi f(\lambda, \tau) \psi_{\lambda,\tau}(t) \frac{d\lambda d\tau}{\lambda^2}.$$

Then, f can be reconstructed from the wavelet coefficients $W_\psi f$ in the sense that $f_\varepsilon \rightarrow f$ in $L_2(\mathbb{R})$ as $\varepsilon \rightarrow 0^+$.

In Section II we state a theorem that guarantees admissibility for the proposed family of wavelets. For further discussion of reconstruction formulas and a proof of theorem 1, we refer to [4], [31].

In applications, wavelet coefficients are computed in a numerical fashion. This means that the parameters (λ, τ) have to be restricted to a (finite) subset of $\mathbb{R} \setminus \{0\} \times \mathbb{R}$. As mentioned in the introduction, different strategies have been developed to calculate the required wavelet coefficients $W_\psi f(\lambda_j, \tau_k)$.

A popular approach is to construct the family ψ_{λ_n, τ_m} so that they form an orthogonal basis in $L_2(\mathbb{R})$ using multiresolution analysis. The so-called discrete wavelet transform (DWT) algorithm uses such orthogonal wavelets and has the benefit, that discrete wavelet coefficients can be used to achieve perfect signal reconstruction.

Even though fast DWT methods [6] have been developed relying on digital filtering, the use of orthogonal wavelets has its disadvantages. One problem is that, for some signal analysis tasks, a finer resolution of the coefficients is desirable [32]. In addition, DWT algorithms are not translation invariant by default which makes their use difficult for some signal processing tasks. It should be noted that certain translation invariant (and from the reconstruction point of view redundant) variations, such as the stationary wavelet transform, [33] exist. Finally, when constructing orthogonal wavelets there is a trade off between “desirable” properties. For example, Daubechies’ wavelets are r times differentiable, however their support increases with r meaning that smoother mother wavelets become less well localised in time. Another example is that the often desired symmetry property can only be approximated for orthogonal wavelets.

For the above reasons, continuous wavelet transform (CWT) algorithms are also frequently used for signal analysis [32]. These approximate the wavelet coefficients (1) using numerical quadrature formulas. CWT algorithms cannot achieve perfect signal reconstruction. On the other hand, CWT algorithms retain the important translation invariance property, and can be used with mother wavelets which were not constructed using multiresolution analysis. This allows for the use of wavelets with desirable properties such as Gaussian wavelets [34], which can be acquired by differentiating the Gaussian function. For example, Ricker’s wavelet has enjoyed popularity for biological signal processing and fault detection applications [11], [12]. The adaptive wavelet family proposed in Section III is also closely connected to Gaussian wavelets.

B. Adaptive Wavelet Transforms

Choosing the correct analyzing wavelet for the CWT algorithm can be a challenging problem. Fig. 5 shows the scalograms (absolute values of wavelet coefficients) computed by MATLAB's "cwt" method for a non-stationary signal consisting of different sinusoids in different time intervals. Clearly, the choice of mother wavelet has a great influence on the quality of the time-scale representation.

Next, we consider some concrete adaptive wavelet transforms in more detail. In [35], the so-called Empirical wavelet transform (EWT) was introduced. EWT can be interpreted as an adaptive discrete wavelet transform. The objective of EWT is to develop a wavelet-like filter bank, where the supports of individual filters are well separated in the frequency domain. It should be noted, that the resulting filters usually cannot be used to construct an orthogonal wavelet basis in $L_2(\mathbb{R})$ in the classical sense. The EWT method first attempts to identify well separated empirical frequency bands in the frequency domain which characterize the signal. This is usually carried out by applying various smoothing operations followed by a search for local maxima in the signal spectrum. A set of bandpass filters is then constructed that are sensitive in the identified empirical frequency bands. In this way, they can be used to decompose the input signal in a multiresolution analysis-like manner. EWT was later modified using Bessel-Fourier expansions in [36]. In particular, the frequency bands for which the adaptive filters are to be constructed have been identified using the Bessel-Fourier spectrum instead of the classical trigonometric approach in [36]. Using this modification, better separation of the frequency components can be achieved, which is especially useful when dominant frequency components of the signal appear close to each other.

Another adaptive wavelet transform relevant for comparison with the results in this study is the tunable Q-factor wavelet transform (TQWT) [37]. Similarly to EWT, the TQWT is a DWT-like approach which constructs a redundant frame whose so-called Q-factor defined by $Q := \frac{\psi_0}{\Delta\psi}$ is a free parameter. This parameter measures the ratio between the center frequency ψ_0 of the analyzing filter and the bandwidth $\Delta\psi$. Selecting a desired Q-factor influences the time-scale (or time-frequency) tiling of the transformation. Once the Q factor is set, a DWT-like algorithm is used to decompose the signal, that is, appropriate low- and highpass filters are constructed to obtain approximation and detail coefficients. The filter passbands are heavily influenced by the selected Q-factor. For a more thorough review of TQWT, we refer to [37].

There are several key differences between the EWT, TQWT and similar methods and the adaptive Rational Gaussian wavelets presented in this study. From a theoretical point of view, our proposed transformation is a wavelet transformation in the classical sense, that is, we present a large class of functions for which the admissibility property (see theorem 1) is proved. In addition, RGW wavelets are smooth and symmetric. Importantly, while adaptive DWT-like methods present a way to decompose the signal into well-behaved components in the frequency domain (e.g. separated supports for EWT, adaptive time-scale tiling for TQWT, etc.), the construction proposed

in this article provides a way to obtain high quality signal reconstruction *using very few wavelet coefficients*. Admittedly, while DWT and DWT-like methods yield perfect signal reconstruction, they achieve this by computing at least N wavelet coefficients, where N denotes the sampling points of the signal. In contrast, our proposed method aims to recover a high quality reconstruction using only $m \ll N$ wavelet coefficients (i.e. it can be used to extract significant features and compress the signals). In addition, DWT based methods do not exhibit time shift invariance, which makes analyzing groups of signals (e.g. quasi periods of a phenomenon) difficult. In contrast, finding an appropriate set of parameters for an RGW mother wavelet can effectively describe groups of signals which makes RGW features appropriate for various detection tasks as demonstrated in Section IV.

III. RATIONAL GAUSSIAN WAVELETS

Gaussian wavelets are defined by

$$\psi_n(t) := C_n \cdot \frac{d^n}{dt^n} e^{-t^2/2}, \quad (2)$$

where $n \in \mathbb{N}$ and the constant satisfies $C_n = 1 / \left\| \frac{d^n}{dt^n} e^{-t^2/2} \right\|_2$. These types of wavelets are commonly used in signal and image processing, especially since they exhibit good localisation properties on both time and frequency space. In particular, Ricker wavelets (the $n = 2$ case in Eq. (2)) have been successfully applied for fault detection [11], and biological signal processing [12].

In this paper, we consider the case $n = 0$ and multiply the Gaussian $e^{-t^2/2}$ with an appropriate rational function, whose poles and zeros will act as free parameters in our construction. If the rational term satisfies the following criteria, the resulting family of wavelets will satisfy the conditions of theorem 1. Consider first, the polynomials

$$r_z(t) := (t - z)(t + z)(t - \tilde{z})(t + \tilde{z}) \quad (z \in \mathbb{C}, \Im(z) \neq 0, t \in \mathbb{R}), \quad (3)$$

where $\Im(z)$ denotes the imaginary part of the complex parameter z and $\tilde{z} = -\Re(z) + i\Im(z)$. We note that polynomials of the form $r_z(x) = (x - z)(x - \bar{z})$ may also be used if $\Re z = 0$ for the proposed construction. We denote the class of polynomials satisfying Eq. (3) by $\mathcal{P}_e$. Clearly, the elements of $\mathcal{P}_e$ are real valued even functions, who do not vanish on the real line. Using the class $\mathcal{P}_e$ we can introduce the class of rational functions

$$\mathcal{V} := \left\{ v(t) = \frac{1}{\prod_{k=0}^{n-1} q_k(t)} : q_k \in \mathcal{P}_e, n \in \mathbb{N} \right\}. \quad (4)$$

Then, the rational functions $v \in \mathcal{V}$ are also real and do not vanish anywhere on $\mathbb{R}$. Consider now the polynomial

$$P(t) := t \cdot \prod_{k=1}^p (t - t_k)(t + t_k) \quad (t_k \in \mathbb{C} \setminus \{0\}, p \in \mathbb{N}). \quad (5)$$

With the conditions stated in Eq. (5), P is an odd polynomial satisfying $P(t) = -P(-t)$ ($t \in \mathbb{R}$). Now, it is possible to

state the following definition. We note that for the experiments and examples detailed in this work, we were particularly interested in real-valued wavelets (see Section IV). For this reason, and for simplicity we assumed $\Im(t_k) = 0$ in our experiments. It should also be noted, that complex variations of rational Gaussian wavelets are possible, as long as P remains an odd function and $\mathcal{V}$ contains even functions, who do not vanish on the real line (see the proof of admissibility in the appendix).

Definition 1 (Rational Gaussian wavelet): The function $\Psi \in L_2(\mathbb{R})$ is called a rational Gaussian wavelet (RGW), if

$$\psi^\eta(t) := C(\eta) \cdot P^\eta(t) \cdot v^\eta(t) \cdot e^{-t^2/2}, \quad (t \in \mathbb{R}, \quad \eta \in \mathbb{C}^{p+n}) \quad (6)$$

where P^η is a polynomial satisfying Eq. (5), $v^\eta \in \mathcal{V}$ and η contains the poles z of v^η and the non-zero roots of P^η . Moreover, the constant $C(\eta) \in \mathbb{R}$ is chosen such that $\|\psi^\eta\|_2 = 1$ holds.

Mother wavelets defined according to Eq. (6) clearly belong to $L_2(\mathbb{R})$, in fact the following theorem holds.

Theorem 2 (Admissibility of rational Gaussian wavelets): Suppose $\psi \in L_2(\mathbb{R})$ is defined according to Eq. (6). Then, there exists a number $M \in \mathbb{R}_+$, such that

$$\int_{-\infty}^{\infty} \frac{1}{|\xi|} \left| \widehat{\psi}(\xi) \right|^2 d\xi = M$$

holds.

The proof of theorem 2 can be found in the appendix. The admissibility property guarantees that the proposed RGW wavelet transform can be inverted with respect to the L_2 norm. It should be noted, that theorem 2 also holds, if we replace the set of possible rational terms $\mathcal{V}$ with the linear space $\mathcal{W} := \left\{ \sum_{j=0}^k c_k \cdot v_j(x) : v_j \in \mathcal{V}, k \in \mathbb{N} \right\}$, however, for simplicity we state the results using the definition above. As seen in the appendix, the proof relies on two properties of the RGWs: symmetry (they are odd functions), and the fact that the denominator of v in Eq. (6) does not disappear anywhere. These properties however are also satisfied if v is chosen from $\mathcal{W}$ instead of $\mathcal{V}$.

RGW mother wavelets depend on a number of free parameters. The modification of these parameters greatly influences the morphology of the analyzing wavelet. For example, moving the poles of v^η close to the real line leads to the appearance of sharp peaks in the mother wavelet. This spike-like morphology resembles a number of real life signals which appear in fault detection and biomedical applications (see e.g. Figs. 1–3).

It is important to note, that despite the sharp oscillations shown in Figs. 1–3, RGWs retain smoothness and symmetry. This adaptive property of the proposed wavelets allows for the capture of important signal behavior using only a few wavelet coefficients. That is, a sparse wavelet coefficient representation of a large class of signals may be obtained by optimizing the poles of v^η and the zeros of P^η with respect to an objective function. This makes the continuous wavelet transform utilizing RGWs a powerful trainable feature extractor. Since RGWs are also differentiable with respect to the poles of v^η and zeros of P^η , gradient based methods (such as backpropagation algorithms) may be used to optimize these parameters. This allows for an easy integration of the scheme into existing machine learning models such as neural networks.

A. Wavelet Transform Based Machine Learning Models

Feature extraction schemes relying on the wavelet transform have been studied before (see e.g. [3], [11], [12], [29], [38]). In many works, the wavelet transform (discrete or continuous) is used as a static feature extractor [3], [12], [38]. That is, a wavelet coefficient representation of the input signals is obtained, which is then passed to a machine learning algorithm to solve a regression or classification problem. The feature extraction step in this case is conducted separately from the training of the ML model. Since wavelet coefficients represent information about the signals' time-frequency profile, the magnitude of coefficients corresponding to fixed scale and translation parameters conveys interpretable information about signal behavior at given times and frequencies.

Notably, so-called Deep-Wavelet Neural Networks (DWNNs) were successfully used for various medical imaging problems, including the recognition of breast cancer from thermographic images [39], [40], [41], [42]. The DWNN method leverages Mallat's discrete wavelet transformation algorithm [43] of decomposing input signals and images by iteratively applying low and high pass filters and down sampling the filtered signals. Finally, in a DWNN, pooling-like synthesis operators are employed to obtain a sparse feature representation of the input images.

In most CWT based machine learning methods however, a large number of wavelet coefficients is computed, then passed to the underlying machine learning algorithms. A more adaptive approach was proposed in [11], where the authors computed wavelet coefficients using convolution with the kernel ψ_{λ_k, τ_k} . The learnable parameters of these wavelet kernel convolution layers were the scales λ_k and initial translation τ_k ($k = 1, \dots, M$). Then, wavelet coefficients were approximated by computing the (discrete) convolutions $\mathbf{f} * \psi_{\lambda_k, \tau_k}$ ($k = 1, \dots, M$), where $\mathbf{f} \in \mathbb{R}^N$ and $\psi_{\lambda_k, \tau_k} \in \mathbb{R}^N$ denote N -point equidistant samplings of f and ψ_{λ_k, τ_k} respectively. This approach means, that the input signals are represented with a large number of wavelet coefficients, therefore, in wavelet kernel net pooling layers are utilized to obtain a sparse signal representation. We note that the proposed RGW wavelets could easily be adapted for use with wavelet kernel net proposed in [11] and may perform well for certain applications.

The main benefit of optimizing the zeros and poles of the rational term $P^\eta(x) \cdot v^\eta(x)$ in Eq. (6), is that mother wavelet morphologies similar to the input signals may be obtained. This is beneficial, because in this way, precise reconstructions of the input signals may be obtained using only a few wavelet coefficients (see Section III-B and Figs. 1–3). This property is better demonstrated when computing RGW coefficients with the proposed variable projection operators instead of convolutions.

B. Variable Projection and Wavelet Coefficients

In practical applications, signals are usually only available in a discretely sampled form. That is, instead of $f \in L_2(\mathbb{R})$, we may only consider its (equidistant) sampling $\mathbf{f} \in \mathbb{R}^N$ ($N \in \mathbb{N}$). Consider an $m \ll N$ dimensional subspace in $\mathbb{R}^N$ spanned

by the linearly independent basis vectors $\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathbb{R}^N$. Denote this subspace by $U := \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_m\} \subset \mathbb{R}^N$ and by $\hat{\mathbf{f}} \in U$ the element which minimizes $\|\mathbf{f} - \mathbf{u}\|_2^2$ ($\mathbf{u} \in U$). It is well known that $\hat{\mathbf{f}}$ exists uniquely and may be expressed by the orthogonal projection $\hat{\mathbf{f}} = \mathcal{P}_U(\mathbf{f}) := \Phi\Phi^+ \mathbf{f}$ where the columns of $\Phi \in \mathbb{R}^{N \times m}$ coincide with the basis vectors $\mathbf{u}_k$ ($k = 1, \dots, m$) and Φ^+ denotes the Moore-Penrose pseudo inverse [27]. Suppose the basis vectors $\mathbf{u}_k$ depend on a real parameter vector $\boldsymbol{\eta} \in \mathbb{R}^q$ ($q \in \mathbb{N}$) and thus $U(\boldsymbol{\eta}) := \text{span}\{\mathbf{u}_1(\boldsymbol{\eta}), \dots, \mathbf{u}_m(\boldsymbol{\eta})\}$. Then, the projection operator $\mathcal{P}_{U(\boldsymbol{\eta})}$ also depends on $\boldsymbol{\eta}$ and is referred to as a *variable projection operator*. Given a family of subspaces $G(\boldsymbol{\eta})$ spanned by $\mathbf{u}_k(\boldsymbol{\eta})$ ($k = 1, \dots, m$), the best possible approximation (if it exists) of $\mathbf{f} \in \mathbb{R}^N$ can be found by minimizing

$$E_2(\boldsymbol{\eta}) := \|\mathbf{f} - \hat{\mathbf{f}}(\boldsymbol{\eta})\|_2^2 = \|\mathbf{f} - \Phi(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})^+ \mathbf{f}\|_2^2 \quad (7)$$

over $\boldsymbol{\eta} \in \mathbb{R}^q$. This nonlinear optimization problem is usually referred to as a separable nonlinear least squares problem (SNLLS), since for any fixed value of $\boldsymbol{\eta}$, minimizing Eq. (7) degrades to a least squares fitting task. SNLLS problems can be solved using gradient based optimization schemes due to the work of Golub and Pereyra, who in [27], [28] provided an explicit formula for the derivatives $\frac{\partial \Phi(\boldsymbol{\eta})^+}{\partial \boldsymbol{\eta}}$. Using this formula, the gradients of $E_2(\boldsymbol{\eta})$ may be calculated analytically. It is important to note however, that the differentiability of E_2 assumes that the partial derivatives $\frac{\partial \Phi(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}$ exist (see e.g. [27]). In other words, the basis functions $g_1(\boldsymbol{\eta}), \dots, g_m(\boldsymbol{\eta})$ should be differentiable with respect to the parameters $\boldsymbol{\eta}$.

In [29] we proposed to approximate continuous wavelet coefficients using variable projections. More precisely, let $m \in \mathbb{N}$ denote the number of wavelet coefficients used to represent the signal $\mathbf{f}$. Let $\boldsymbol{\eta} := [\lambda_1, \tau_1, \dots, \lambda_m, \tau_m] \in \mathbb{R}^{2 \cdot m}$ and let $\psi \in \mathbb{R}^N$ be an equidistantly sampled analyzing wavelet from $L_2(\mathbb{R})$. Suppose the sampling was done over the effective support (a subset of $\mathbb{R}$ over which $\psi(x)$ significantly differs from 0) of ψ . Define $\Psi(\boldsymbol{\eta}) \in \mathbb{R}^{N \times m}$ as the matrix, whose k -th column consists of the sampling of ψ_{λ_k, τ_k} ($k = 1, \dots, m$). Then, one has $W_\psi f(\lambda_k, \tau_k) \approx (\Psi(\boldsymbol{\eta})^+ \mathbf{f})_k$ ($k = 1, \dots, m$). We can provide the following error estimate for the obtained approximation if the input signal f is smooth and compactly supported. These properties can be assumed for a large class of signals encountered in applications (see e.g. Section IV).

Theorem 3 (Error of variable projection based continuous wavelet coefficients): Let $\psi \in L_2(\mathbb{R})$ be a fixed mother wavelet and suppose $f \in L_2(\mathbb{R})$ is continuously differentiable and compactly supported on $[a, b] \subset \mathbb{R}$. Let $a =: t_0 < t_1 < \dots < t_{N-1} := b$ be an equidistant sampling of $[a, b]$ with $h := t_1 - t_0$. In addition, consider $\boldsymbol{\eta} = (\lambda_1, \tau_1, \dots, \lambda_m, \tau_m) \in \mathbb{R}^{2m}$ with $\lambda_k > 0$ ($k = 1, \dots, m$). Finally, let $\mathbf{f}_k := f(t_k)$ ($k = 0, \dots, N-1$). Then we have

$$\begin{aligned} |W_\psi f(\lambda_k, \tau_k) - h \cdot (\Psi(\boldsymbol{\eta})^+ \mathbf{f})_k| &< h \cdot \frac{M_1(b-a)}{2} \\ &+ h \cdot \|\mathbf{f}\|_\infty \|\Psi^*(\boldsymbol{\eta})\|_\infty \cdot \left(\frac{\kappa(G(\boldsymbol{\eta}))}{\|G(\boldsymbol{\eta})\|_\infty} + 1 \right), \end{aligned} \quad (8)$$

where $M_1 := \max_{\xi \in [a, b], k=1, \dots, m} |f'(\xi) \cdot \overline{\psi_{\lambda_k, \tau_k}(\xi)}|$, the columns of $\Psi(\boldsymbol{\eta})$ contain the samplings of

ψ_{λ_k, τ_k} ($k = 1, \dots, m$) over t_j ($j = 0, \dots, N-1$) and $G(\boldsymbol{\eta})$ is the Gram-matrix constructed from the columns of $\Psi(\boldsymbol{\eta})$. In Eq. (8), $\kappa(G(\boldsymbol{\eta}))$ denotes the condition number of $G(\boldsymbol{\eta})$ using the matrix infinity norm and $\Psi^*(\boldsymbol{\eta})$ denotes the adjugate of $\Psi(\boldsymbol{\eta})$.

The proof of theorem 3 is provided in the appendix. A consequence of this theorem, is that the proposed approximations tend to the actual wavelet coefficients as the sampling frequency of f increases.

There is a number of arguments which support the approximation of wavelet coefficients using the above-described variable projection based methodology in ML models. Firstly, in contrast to the Wavelet Kernel Net approach, there is no need for pooling and dropout operations to obtain a sparse representation of the input. Furthermore, by minimizing (7), one obtains wavelet coefficients which are optimal with respect to the reconstruction error in the least squares sense. To demonstrate the effectiveness of calculating wavelet coefficients in this way, Figs. 1-3 illustrate the reconstruction of various QV signals using RGW wavelets with VP operators. Finally, since $E_2(\boldsymbol{\eta})$ may be optimized using gradient based methods, variable projection operators can be included into neural network architectures (and other machine learning models [21]) as separate layers and trained using backpropagation schemes. These types of neural networks, often referred to as VP-Net, were introduced in [20] and have been successfully used with a number of different variable projection operators to solve environmental recognition [44], fault detection [21] and biomedical signal processing problems [20], [21].

C. Considered Signal Class

The proposed RGW wavelets can be used to analyze any signal for which the integral in Eq. (1) exists. This is satisfied for example, when the signal f belongs to $L_2(\mathbb{R})$. However, when the objective is to obtain a good approximation of f using *only a few wavelet coefficients*, the method performs best, when f belongs to a class of so-called quickly vanishing (QV) signals. For a more formal definition, we refer to [21], but for the purposes of this study, the set of QV signals can be interpreted as a subset of $L_2(\mathbb{R})$ containing functions with quasi-compact support. In other words, if f is QV, then $|f(t)|$ tends to 0 quickly as $|t|$ increases. Many real-world signals fall into this category, for example segmented ECG signals and visually evoked potentials (VEPs) [45]. The relevance of QV signals is well reflected by the fact that in addition to biomedical applications, they appear in a wide range of problems such as autonomous vehicle control [44] and fault detection [3], [21].

Some examples of QV signals approximated by various RGW expansions are illustrated in Figs. 1-3. In these figures, the parameters (zeros, poles, scales and translations) of the RGW system were optimized by minimizing the functional (Eq. (7)) with gradient descent. The required derivatives were computed using the formulas given in Section III-D, while the number of wavelets used m , the number of zeros p and poles n are hyperparameters of the transform and should be set apriori. We can observe, that a similar number of wavelet coefficients

can be used to approximate QV signals with highly different morphologies. For example in Figs. 1 and 2, a healthy and a deformed ECG signal was approximated using $m = 8$ RGW wavelet functions. The same number of poles and zeros were used ($n = 3$ and $p = 10$ respectively) in both cases. The position of these poles and zeros as well as the dilation-translation parameter pairs determining the considered 8 wavelet coefficients were optimized separately for both signals. We can observe, that the optimized mother wavelet mimics the behavior of the approximated signal in both cases. Although signal morphologies differ significantly, the free positioning of the zeros and poles in Eq. (6) allowed the method to yield high quality reconstructions.

D. The RGW-VP Layer

Suppose we would like to solve a classification (or regression) problem using supervised learning. Consider the mappings $\mathcal{P}_{\Psi(\eta)} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ and $\mathcal{C}_{\Psi(\eta)} : \mathbb{R}^N \rightarrow \mathbb{R}^m$

$$\begin{aligned}\mathcal{P}_{\Psi(\eta)}(\mathbf{f}) &:= \Psi(\eta) (\Psi(\eta)^+ \mathbf{f}), \\ \mathcal{C}_{\Psi(\eta)}(\mathbf{f}) &:= \Psi(\eta)^+ \mathbf{f},\end{aligned}\quad (9)$$

where

$$\begin{aligned}\eta &:= [\lambda_1, \tau_1, \dots, \lambda_m, \tau_m, t_1, \dots, t_p, \dots, a_1, b_1, \dots, a_n, b_n]^T \\ &\in \mathbb{R}^{2m+p+2 \cdot n}, \quad (m, p, n \in \mathbb{N})\end{aligned}\quad (10)$$

and the columns of $\Psi(\eta)$ are samplings of rational Gaussian wavelets defined according to Eq. (6). In Eq. (10), λ_k, τ_k denote the learnable scales and translations, while t_j stand for the zeros of P^η and a_i, b_i determine the real and imaginary parts of the poles of v^η in Eq. (6). The mapping $\mathcal{C}_{\Psi(\eta)}$ in Eq. (9) will be referred to as an ‘‘RGW-VP’’ layer henceforth.

We would like to optimize η together with the weights of an underlying neural network. This would ensure that features (wavelet coefficients) of the input signal are learned by the RGW-VP layer which are ideal from the point of view of the underlying classification or regression problem. In addition, this would enable the model to learn the appropriate mother wavelet to perform the transformation with by tuning the parameters t_j, a_i and b_i in η . In [20], so-called VP-NETs were proposed, which consist of a variable projection layer (performing adaptive feature extraction), followed by a number of fully connected linear layers (or any other layer type associated with common neural network architectures). This is possible since the partial derivatives satisfy $\frac{\partial \mathcal{C}_{\Psi(\eta)}}{\partial \eta} = \frac{\partial \Psi(\eta)^+}{\partial \eta} \mathbf{f}$ and $\frac{\partial \Psi(\eta)}{\partial \eta}$ can be expressed analytically [27]. The required formulas were first obtained by Pereyra and Golub in [27] (see also [20], [21] for their application in a machine learning context). Thus, it suffices to calculate the partial derivatives $\frac{\partial \Psi(\eta)}{\partial \eta}$ in order to supplement a neural network (or other ML model trained by backpropagation) with an RGW-VP layer.

First, we consider the derivatives $\partial \psi_{\lambda_k, \tau_k}^\eta / \partial \lambda_k$ and $\partial \psi_{\lambda_k, \tau_k}^\eta / \partial \tau_k$. Regardless of the mother wavelet ψ^η , these are given by

$$\frac{\partial \psi_{\lambda_k, \tau_k}^\eta}{\partial \lambda_k} = -\frac{1}{2} \cdot \lambda_k^{-3/2} \cdot \psi_{\lambda_k, \tau_k}^\eta(t) - \frac{t - \tau_k}{\lambda_k^2} \cdot \left(\psi_{\lambda_k, \tau_k}^\eta(t) \right)'$$

$$\frac{\partial \psi_{\lambda_k, \tau_k}^\eta}{\partial \tau_k} = - \left(\psi_{\lambda_k, \tau_k}^\eta(t) \right)' / \lambda_k, \quad (11)$$

where $\left(\psi_{\lambda_k, \tau_k}^\eta(t) \right)'$ denotes the derivative of $\psi_{\lambda_k, \tau_k}^\eta$ with respect to t .

In order to give the required derivatives with respect to the zeros of P^η and the poles of v^η , first consider that the elementary polynomials from Eq. (3) may be written as $r_z(t) = t^4 + t^2 \cdot 2(b^2 - a^2) + a^4 + 2a^2b^2 + b^4$, where $a := \Re(z)$ and $b := \Im(z)$. Recall, that in Section III, we required that $b > 0$. One way to ensure this, is to introduce $\hat{b} := b^2 + \varepsilon$, for some $\varepsilon > 0$ and modify the formulas presented here accordingly. Thus, using the notation $z_k := a_k + ib_k$, we obtain $\frac{\partial r_{z_k}(t)}{\partial a_k} = -4 \cdot (a_k t^2 - a_k^3 - a_k b_k^2)$ and $\frac{\partial r_{z_k}(t)}{\partial b_k} = 4 \cdot (b_k t^2 + b_k^3 + a_k^2 b_k)$. Notice that the v^η can be written as the finite product of $r_{z_k}(t)^{-1}$ terms. The partial derivatives of these terms are given by $\frac{\partial r_{z_k}(t)^{-1}}{\partial a_k} = -r_{z_k}(t)^{-2} \cdot \frac{\partial r_{z_k}(t)}{\partial a_k}$ and $\frac{\partial r_{z_k}(t)^{-1}}{\partial b_k} = -r_{z_k}(t)^{-2} \cdot \frac{\partial r_{z_k}(t)}{\partial b_k}$. Using this, we obtain the following partial derivatives for v^η with respect to a_k and b_k :

$$\begin{aligned}\frac{\partial v^\eta(t)}{\partial a_k} &= \frac{\partial r_{z_k}(t)^{-1}}{\partial a_k} \cdot \prod_{j=0, j \neq k}^{n-1} r_{z_j}(t) \\ \frac{\partial v^\eta(t)}{\partial b_k} &= \frac{\partial r_{z_k}(t)^{-1}}{\partial b_k} \cdot \prod_{j=0, j \neq k}^{n-1} r_{z_j}(t).\end{aligned}\quad (12)$$

Finally, using Eq. (12) the formula for the partial derivatives of the mother wavelet can be given as

$$\begin{aligned}\frac{\partial \psi_{\lambda_i, \tau_i}^\eta(t)}{\partial a_k} &= P^\eta(t) \frac{\partial v^\eta(t)}{\partial a_k} e^{-t^2/2} \\ \frac{\partial \psi_{\lambda_i, \tau_i}^\eta(t)}{\partial b_k} &= P^\eta(t) \frac{\partial v^\eta(t)}{\partial b_k} e^{-t^2/2}.\end{aligned}\quad (13)$$

Note that in Eq. (13), for the sake of simplicity, we abstained from adding the effect of the normalizing factor $C(\eta)$. The partial derivatives also considering this factor can be easily calculated using the theorem of parametric integration and the above equations, however we found (see Section IV), that in applications good results could be achieved without it. The partial derivatives of $\psi_{\lambda_i, \tau_i}^\eta$ with respect to the zeros of P^η can also be easily calculated:

$$\frac{\partial \psi_{\lambda_i, \tau_i}^\eta(t)}{\partial t_k} = - \prod_{j=1, j \neq k}^p (t - t_j)(t + t_j) \cdot 2t_k \cdot v^\eta(t) \cdot e^{-t^2/2}. \quad (14)$$

Equations (13) and (14) can be used to implement the backpropagation step for the RGW-VP layer. Fig. 4 shows a schematic of the proposed neural network architecture using the RGW-VP layer for automatic feature extraction. The main benefit of approximating wavelet coefficients using variable projection is that for any fixed η parameter vector, the computed coefficients minimize the error of the reconstructed signal in the least-squares sense. In order to ensure high quality approximations of the input signals and to avoid the problem of vanishing gradients, in VP-based ML methods the following regularization

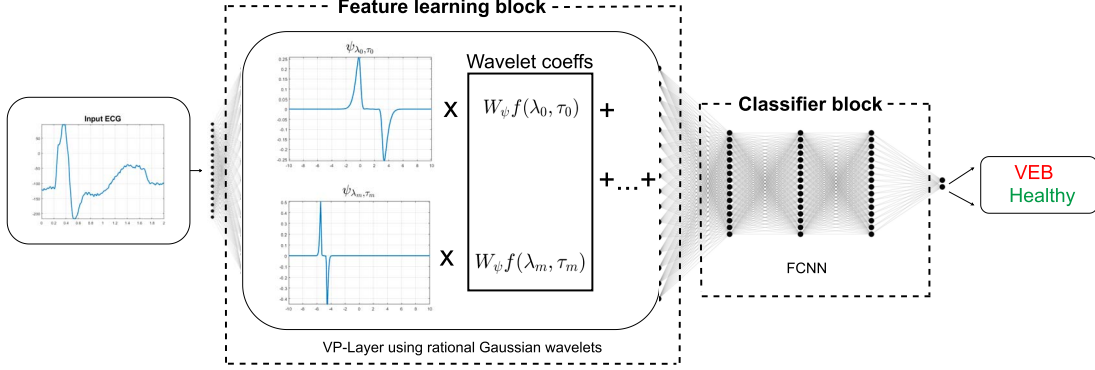


Fig. 4. Schematic of the proposed, rational gaussian wavelet based neural network architecture. The first layer acts as an automatic feature learning block passing the learned wavelet coefficients to an underlying fully connected network.

term is added to the loss function during training (see [20], [29]):

$$J_{VP}(\boldsymbol{\eta}) := \frac{\alpha}{s} \sum_{i=1}^s \frac{\|\mathbf{f}_i - \mathcal{P}_{\Psi(\boldsymbol{\eta})} \mathbf{f}_i\|_2^2}{\|\mathbf{f}_i\|_2^2}. \quad (15)$$

Here $\mathcal{P}_{\Psi(\boldsymbol{\eta})} \mathbf{f}$ is defined according to Eq. (9), $\alpha > 0$ is a hyper-parameter and s denotes the total number of input signals in the training set.

Now we can show how ML models supplemented with RGW-VP layer can be optimized with backpropagation. According to the original VP-Net approach in [20], let us model a neural network with RGW-VP layer as the composition of functions $F_{\Theta}(x) = g_{\theta_L}^L(\dots g_{\theta_1}^1(g_{\theta_0}^0(x)) \dots)$, where $g_{\theta_0}^0(x)$ denotes the RGW-VP layer with its parameter vector $\theta_0 = \boldsymbol{\eta}$, the remaining functions denotes common neural network layers such as linear layers or activation functions with its parameters.

Let us denote the cost function with $J = J_C + J_{VP}$, where J_C is a common loss function, in our experiments the Binary Cross Entropy loss, and J_{VP} is the regularization term of the VP-layer Eq. (15). The backpropagation algorithm optimize its layer parameters based on the chain rule as

$$\theta_l = \theta_l - \alpha \frac{\partial J}{\partial \theta_l}, \quad \frac{\partial J}{\partial \theta_l} = \frac{\partial J}{\partial g_l} \cdot \frac{\partial g_l}{\partial \theta_l} \quad (16)$$

where $\alpha > 0$ is the learning rate. We would like to highlight that RGW-VP layers seamlessly can be substituted into Eq. (16) without the modification of the optimization process. For example if we want to compute $\partial C_{\Psi(\boldsymbol{\eta})} / \partial \boldsymbol{\eta}$ in $g_{\theta_0}^0$ then $\frac{\partial g_0}{\partial \theta_0} = \frac{\partial \Psi(\boldsymbol{\eta})^+}{\partial \boldsymbol{\eta}} \cdot \mathbf{f}$. Therefore during training a low number of parameters needs to be optimized in the feature extraction layer compared to common convolution or feedforward layers. The computational complexity of an inference step of RGW-VP layers matches that of a linear layer, that is, it is for a single $\mathbf{f} \in \mathbb{R}^N$ input signal we have $O(m \cdot N)$, where m denotes the number of computed wavelet coefficients. Indeed, once the training is finished (i.e. $\boldsymbol{\eta}$ is fixed), $T := \Psi(\boldsymbol{\eta})^+$ can be precomputed. Hence the formalization of the trained layer will be $Y_0 = T \cdot F$, where $F \in \mathbb{R}^{B \times C \times N}$, B is the batch size, C is the number of channels and N is the length of the input signal. $Y_0 \in \mathbb{R}^{B \times C \times m}$ where m is the sparse number of computed CWT coefficients.

IV. EXPERIMENTS

We now detail several experiments which demonstrate the utility of the proposed RGW wavelet family. We note that all results discussed here are fully reproducible and publicly available at [30].

A. Time-Scale Representations

We now examine how the proposed RGW wavelets can be used to analyze non-QV (see Section III-C) signals through a detailed example. In particular, let $I := [0, 2]$, $J := (0.9, 1.1)$ and consider the signal

$$f(t) := \begin{cases} \cos(2\pi \cdot 3t) & t \in I \setminus J \\ \cos(2\pi \cdot 3t) + 2 \cos(2\pi \cdot 80t) & t \in J. \end{cases} \quad (17)$$

This signal (depicted in Fig. 5) consists of a base oscillation and a rapid burst in the middle of the interval. In our experiments, the signal was sampled at 500 Hz for 2 seconds. We are going to compare the scalograms produced by MATLAB's "cwt" routine when using different types of mother wavelets. In particular, we are interested in demonstrating that the free parameters (Eq. (6)) of RGW can be leveraged to capture signal behavior in the time-scale domain. This is especially important for problems, when wavelet coefficients with only a limited number of scales may be computed (for example due to the need for on-line evaluation). An ideal wavelet transformation would produce a scalogram of the signal (17) which separates the high and low frequency components well and provides acceptable time localization. Scalograms can be given as the matrices $W \in \mathbb{R}^{M \times N}$, where $W_{j,k} := |W_{\psi}f(\lambda_j, \tau_k)|$ ($j = 1, \dots, M, k = 1, \dots, N$). In this experiment, we fixed a total of 80 scales chosen logarithmically between 10 and $10^{2.5}$. Although this choice is arbitrary, it serves to demonstrate the power of constructing time-frequency representations of signals with the proposed RGW class. Indeed, while usual continuous wavelet transform implementations consider signal characteristics and characteristics of the mother wavelet to determine which scales to consider when computing wavelet coefficients [32], the adaptive nature of RGW allows us to pose this problem in a different manner. Namely, when using RGW, we can settle

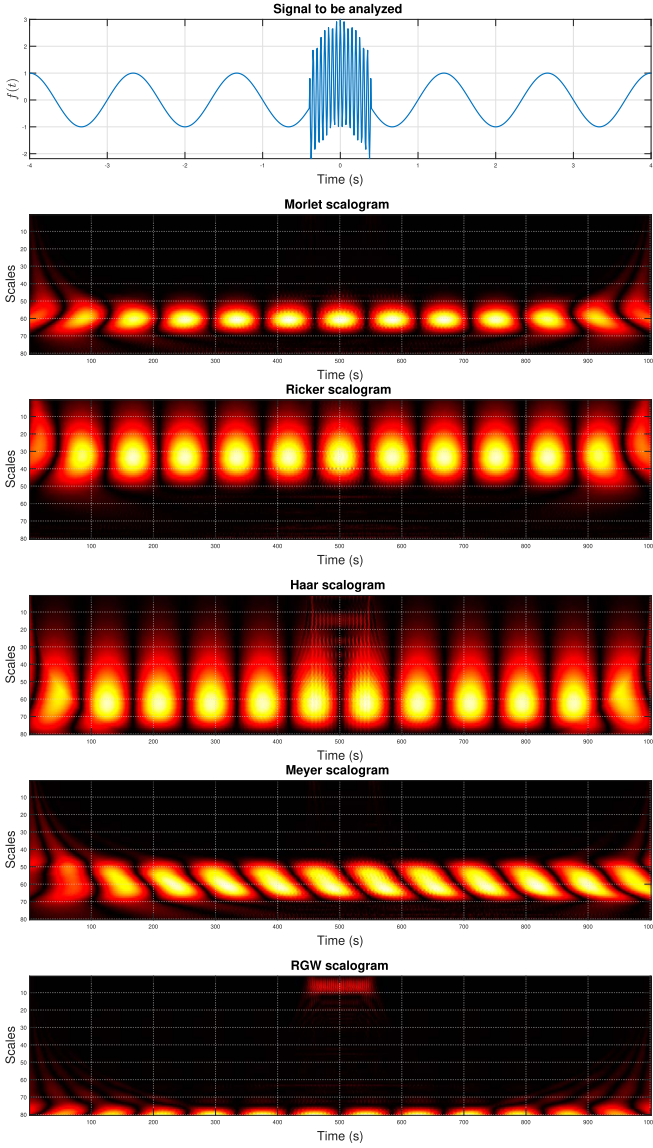


Fig. 5. An oscillating signal to be analyzed and corresponding scalograms generated using different mother wavelets. Images of the mother wavelets generating the above scalograms are available in our repository [30].

the question of which scales to use when computing the scalogram in a reverse manner: RGW determines what the mother wavelet should look like, if we want to obtain a meaningful scalogram given a fixed set of scales.

In this experiment, we did not compute RGW wavelet coefficients using the variable projection framework discussed in Section III-B, rather we relied on MATLAB’s “cwt” function. Accordingly, in contrast with the variable projection approach, for analyzing the signal in (17) we tuned the RGW parameters by minimizing an intuitive objective. Namely, for a signal $f \in \mathbb{R}^N$, denote by $\mathbf{W}(\boldsymbol{\eta}) \in \mathbb{R}^{M \times N}$ the matrix which satisfies $\mathbf{W}(\boldsymbol{\eta})_{jk} := W_{\psi\boldsymbol{\eta}} f(\lambda_j, \tau_k)$ ($j = 1, \dots, M$, $k = 1, \dots, N$), where $W_{\psi\boldsymbol{\eta}} f$ denotes the RGW wavelet coefficients (see Eq. (1)) and for this experiment $\boldsymbol{\eta} \in \mathbb{R}^{p+2 \cdot n}$ only defines the zeros and poles of the RGW mother wavelet. Let $\lambda_{\min}$ and $\lambda_{\max}$ be integers satisfying $1 \leq \lambda_{\min} < \lambda_{\max} \leq M$ and denote by $\widetilde{\mathbf{W}}(\boldsymbol{\eta})_{jk} := W_{\psi\boldsymbol{\eta}} f(\lambda_j, \tau_k)$ ($j = \lambda_{\min}, \dots, \lambda_{\max}$, $k =$

$1, \dots, N$). That is, $\widetilde{\mathbf{W}}(\boldsymbol{\eta}) \in \mathbb{R}^{(M - (\lambda_{\min} + \lambda_{\max})) \times N}$ is the matrix of “cropped” wavelet coefficients, where any scales below $\lambda_{\min}$ or above $\lambda_{\max}$ have been disregarded. We can now consider the objective

$$E_{\text{osc}}(\boldsymbol{\eta}) := \frac{\|\widetilde{\mathbf{W}}(\boldsymbol{\eta})\|_F}{\|\mathbf{W}(\boldsymbol{\eta})\|_F}, \quad (18)$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm. The objective functional in Eq. (18) describes what percentage of the total energy of wavelet coefficients computed in $\mathbf{W}(\boldsymbol{\eta})$ falls into $\widetilde{\mathbf{W}}(\boldsymbol{\eta})$. Minimizing Eq. (18) means finding a mother wavelet, which collects meaningful (high in absolute value) wavelet coefficients into scales below $\lambda_{\min}$ or above $\lambda_{\max}$. In this way, we wish to find an RGW mother wavelet which separates the frequency content of the analyzed signal well given the fixed scales used to compute the coefficients. In our experiment, using a trial and error approach, the parameters $\lambda_{\min} = 15$ and $\lambda_{\max} = 5$ were selected to generate the results discussed here.

To compare results, we also analyzed the signal in Eq. (17) using several traditional wavelets. More precisely, the Morlet wavelet, the Ricker wavelet, the Haar wavelet and the Meyer wavelet were considered. In each case, the “cwt” routine of MATLAB was used with the pre-set logarithmically placed scales discussed above. We emphasize that the purpose of this experiment was to demonstrate how the proposed RGW class of wavelets might be used for signal processing tasks other than the sparse representation of QV signals. The computed scalograms (see Fig. 5) confirm that optimizing the parameters of RGW yielded a better time-scale representation of the signal than traditional wavelets. In fact, using the given scales, the Morlet, Ricker and Meyer wavelet based scalograms could not identify the important high-frequency components of the signal at all. In the scalogram generated by the Haar wavelet, the high frequency component is visible, however frequency localization is very poor. On the other hand, by choosing the shape of the analyzing RGW wavelet appropriately, all important frequency components could be identified. It should also be noted, that the objective (18) was heuristic and a deeper investigation into defining criteria for “good” scalograms will be undertaken in a future work. The fact, that the full frequency profile of the signal could be captured using a very limited number of scales points towards possible real-life benefits of using RGW based computations.

B. QRS Compression

Electrocardiogram (ECG) is an effective and low-cost method for monitoring heartbeats. Detecting heart arrhythmia can sometimes be challenging due to the infrequent appearance of certain types of arrhythmias, which require long-term diagnostic periods. To aid cardiologists and people living in remote areas, the automatic processing and classification of heartbeat signals has been researched for decades [46]. A healthy ECG heartbeat consists of three segments: the P wave, the QRS complex, and the T wave, respectively [47]. In this section an example application of the proposed RGW class is given for the approximation of QRS segments. QRS compression is an

important and well researched problem [48], [49]. The QRS segment of a heartbeat carries important information which can be used to recognize a number of illnesses. Recently, the proliferation of wearable health monitoring devices [50] produced an increased need for efficient compression techniques.

The presented experiment had two objectives. First, we wanted to verify that the proposed RGW continuous wavelet transform, where wavelet coefficients are approximated using the variable projection operators discussed in Section III-B, provide quality approximations to QRS signals using very few wavelet coefficients. Traditionally, Hermite expansions have been employed to compress and approximate QRS segments [48], [51]. In addition to providing a proof-of-concept application of the proposed wavelets, the QRS approximation application provides a good opportunity to compare the performance of RGW transformations to other adaptive wavelet methods. In particular, we approximate QRS segments of ECG signals using three different adaptive wavelet transforms: the proposed RGW class, the empirical wavelet transform (EWT) [35] and the tunable Q-factor wavelet transform (TQWT) (see [37] and also Section II-B). The presented experiments have been implemented using MATLAB, which includes implementations of the EWT and TQWT algorithms.

For this experiment, the QRS dataset proposed in [20] has been considered. This is a subset of the MIT-BiH Arrhythmia dataset [52] and contains QRS segments associated with healthy heartbeats as well as so-called ventricular ectopic beats (VEBs). VEBs constitute a common arrhythmia and are associated with deformed QRS morphologies (see Figs. 1 and 2 for healthy and VEB ECG beats respectively). A total of 14960 QRS complexes have been considered, each of them recorded over 100 sample points. The dataset contained 7480 healthy and 7480 VEB signals. Each signal was approximated using the RGW, TQWT and EWT methods separately, using 2 to 10 wavelet coefficients. We then compared the average reconstruction error of the methods for the entire dataset. In the case of RGW, the approximation was obtained for each signal by minimizing Eq. (7), where the free parameters were given by (10). Notice, that the (λ_k, τ_k) scale-translation pairs defining each computed wavelet coefficient appear as free parameters of the optimization. In this way, the optimal wavelet coefficients to represent the signal are selected automatically through minimizing the functional (7). In addition, the pole and zero parameters are also optimized to find an appropriate RGW mother wavelet for the compression problem. In the case of EWT and TQWT, a different approach was necessary. Since these are DWT-like methods which successively apply (adaptively determined) low and highpass filters to the signals, they inherently compute a redundant signal representation. In this case, compression was performed by keeping only the m most significant coefficients (highest in absolute value) before reconstruction. Following the literature [48], [49], the error of the approximation was described by the normed percentage root mean square difference defined by

$$PRDN := \frac{\|\mathbf{f} - \hat{\mathbf{f}}\|_2}{\|\mathbf{f} - \bar{\mathbf{f}}\|_2} \cdot 100, \quad (19)$$

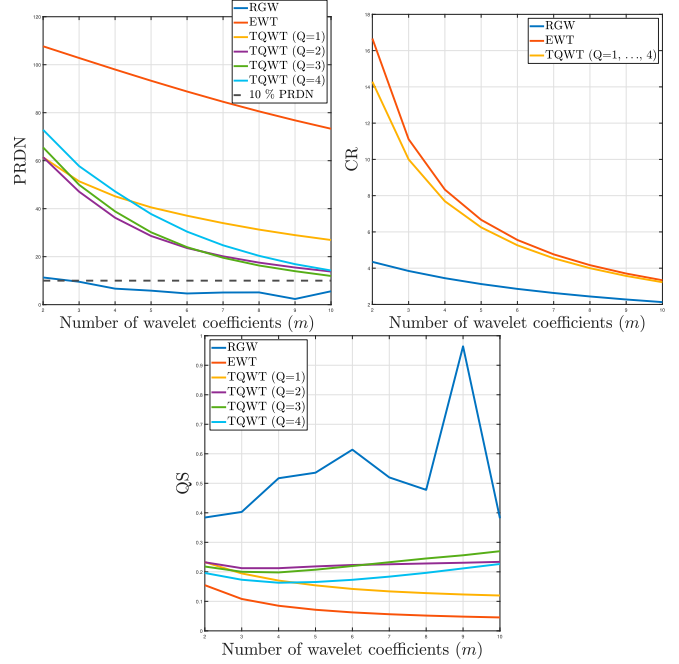


Fig. 6. QRS compression experiment results. TOP LEFT: Average error (PRDN, see Eq. (19)) of the reconstruction using increasing numbers of wavelet coefficients for EWT, TQWT and the proposed RGW wavelet transforms. TOP RIGHT: Compression ratio (Eq. (20)) of the different methods corresponding to different numbers of wavelet coefficients. BOTTOM: Quality score (QS, see Eq. (21)) for the different approaches.

where $\mathbf{f}$ and $\hat{\mathbf{f}}$ denote the signal and the approximation respectively and $\bar{\mathbf{f}}$ denotes the mean of $\mathbf{f}$. The subtraction in the denominator of Eq. (19) is understood component wise. Even though our main objective in this experiment is to show that using the same number of wavelet coefficients, RGW can provide significantly higher quality reconstructions of QRS than other adaptive wavelet based methods, in order to measure the true compression power of the algorithm we also considered the compression ratio (CR). This is defined by

$$CR := \frac{N}{|\eta|}, \quad (20)$$

where $N = 100$ denotes the number of sampling points in $\mathbf{f}$ and $|\eta|$ is the true number of parameters required for reconstruction. Indeed each wavelet coefficient is associated with two quantities that need to be stored (scale and translation) for each method. In addition, in the case of RGW the zeros and poles of the mother wavelet have to be known for reconstruction, while for TQWT the Q-factor also has to be considered. In order to measure the overall performance of each method, the so-called quality score (QS) is also considered. QS is the ratio of CR and PRDN:

$$QS := \frac{CR}{PRDN} \quad (21)$$

and thus measures the trade off between the reconstruction error and the compression ratio. Higher QS scores indicate a better overall compression algorithm. Fig. 6 illustrates the average PRDN, CR and QS scores respectively for each method when $m = 2, \dots, 10$ wavelet coefficients are considered.

In Fig. 6, we can observe that representing QRS segments with RGW wavelet coefficients computed using the $C_{\Psi(\eta)}$ variable projection operator (Eq. (9)) leads to an efficient compression scheme. In fact, the following relationship between PRDN and clinical relevance can be made. Compressed signals with $\text{PRDN} \leq 9\%$ can be considered a “very good” approximation and may be appropriate for some clinical use [48], [53].

Fig. 6 shows that out of the considered adaptive wavelet methods, only RGW was able to consistently achieve a sub 10% PRDN on the dataset. In fact, for $m \geq 3$ wavelet coefficients, the reconstruction error stayed strictly below 10%, while for $m \geq 6$ it performed near, or under 5%. The RGW scheme was also the only one of the considered algorithms to achieve an average PRDN score of 2.53% for $m = 9$, which can be considered an excellent quality average reconstruction error.

From the top right figure of 6, we can observe the true compression ratio for each considered method. As expected, since one needs to store not only the computed wavelet coefficients, but also the scale-translation pairs describing them as well as the parameters (zeros and poles) of the mother wavelet, a lower CR score can be achieved using RGW for a fixed coefficient number. In order to compare the compression quality of the considered methods, the quality score (see Eq. (21)) has also been computed for each considered wavelet coefficient number. Here again, RGW significantly outperforms other adaptive wavelet based approaches for all considered wavelet coefficient numbers.

In order to further emphasize this point, it is worthwhile to compare the results of the next-best performing TQWT method to RGW in more detail. In the presented experiment, the TQWT algorithm achieved its lowest PRDN score when signal reconstruction was carried out using the $m = 10$ most significant (highest in absolute value) TQWT coefficients. For the TQWT method this constituted a compression ratio of $\text{CR} = 3.23$, since in addition to the 10 wavelet coefficients the Q-factor (in this case $Q = 3$) also needs to be stored. With these parameters, the TQWT was able to achieve an average PRDN score of $\approx 11.95\%$ on the dataset. When considering the RGW scheme with only $m = 4$ wavelet coefficients, a slightly higher compression ratio of $\text{CR} = 3.45$ is achieved and the proposed approach was able to achieve an average PRDN score of 6.67%. Thus, we can observe that in this case, RGW provides better compression rates while maintaining a better reconstruction error. We emphasize once more that the main objective of the current experiment was to demonstrate the potential of the proposed RGW wavelet class for compressing QRS segments and to compare its performance with other adaptive wavelet based methods. The results presented here could (and based on the preliminary results should) be further improved. One way to do this for example, is to exploit the fact the morphology of ECG signals is highly patient dependent [38], [54]. In this way an RGW based compression scheme could be conceived, where the parameters of the RGW mother wavelet (zeros and poles) are optimized using only the first few beats from a single patient. Then, for the rest of the recorded heartbeats from the same patient, one could fix these parameters to achieve much higher compression ratios. For the remaining heartbeats only

optimizing the scale-translation pairs responsible for encoding individual QRS segments could be considered.

C. Arrhythmia Detection

In this section, we provide an example application of the proposed RGW wavelet based model driven neural networks to detect the presence of so-called ventricular ectopic beats (VEBs) in ECG recordings. For simplicity of notation, any numerical approximation of the integral (1) will be referred to as “CWT coefficients” henceforth, including the variable projection based approximation discussed in Section III-B.

There are other methods in the literature that propose the use of CWT coefficients to classify ECG heartbeats. In most approaches scalograms are constructed from the signals prior to training and are used as the input to machine learning models. For example, in [55], a static feature extraction scheme was applied on the input heartbeats. Namely, the scalograms with respect to three different mother wavelets were computed and represented as a tensor of size $3 \times M \times N$. Clearly, such a tensor can be thought of as an RGB image. This idea was exploited to pass the preprocessed signals (tensors) to a deep convolutional (CNN) network architecture, which expects RGB images as inputs. In [56], the Morse analytic mother wavelet was used to compute scalograms of groups of successive ECG heartbeats, then they were fed into a deep CNN model. Another approach used scalograms of heartbeats and interval features as input to a CNN in order to classify arrhythmias [57]. In this article, it was found that the Ricker wavelet was the most suitable for this task. The authors of [57] claim that this is due to the shape similarities between QRS complexes and the Ricker wavelets.

One common property of these methods is that they compute thousands of CWT coefficients providing a redundant signal representation, and in turn require deep CNNs to solve the classification task. CNNs extract features automatically and show state-of-the-art performance for various image or signal processing tasks. As mentioned earlier, one disadvantage of such deep learning models is that they use a large number of parameters and thus are difficult to interpret. CWT-VP layers achieve sparse signal representation by approximating only the CWT coefficients necessary for the classification task (that is only $m < M \cdot N$ coefficients are computed). Sparse signal representation also allows for lower model complexity. Additionally, CWT-VP layers do not require pooling or dropout layers to reduce input dimensions, because when using variable projection operators to approximate wavelet coefficients, by definition only m coefficients are computed (see Eq. (9)).

In [20] VP-Net was used with Adaptive Hermite functions to classify ECG heartbeats. Adaptive Hermite functions were utilized for ECG processing in [49], although various other parameterizations of Hermite functions have long been used for approximating these kinds of signals. Hermite-functions form a complete and orthonormal system in $L_2(\mathbb{R})$ and show similar morphology to healthy QRS complexes. For this reason, the first few Hermite-Fourier coefficients may be used to provide a good approximation of QRS complexes. Adaptive Hermite-functions consider the affine argument transforms of the classical Hermite

functions. In this case, the VP-Layer learns a single dilation and translation parameter pair, which describes the entire function system.

In our experiments, we classified VEB signals, which are the most common heartbeat irregularities [46]. We used heartbeats from the MIT-BIH arrhythmia database [52], provided by PhysioNet [58]. In this database, a total of 48 recordings from different patients can be found. We used the proposed datasets from [46] by De Chazal et al. for training and testing respectively. These datasets are unbalanced, since in real life scenarios, normal heartbeats are more common, than abnormal ones. The datasets contains approximately 50,000 heartbeats. We represent every heartbeat with 300 data samples, which includes the P segment, QRS complex and T segment. It is important to note, that many current approaches only consider the QRS complex (see [22]), however as we point out in Section IV-D, biological markers such as the shape of the P-wave are also used by cardiologists to recognize VEBs.

In our experiment we considered the proposed RGW-VP-Net architecture using $m = 10$ wavelet coefficients, $p = 3$ zeros and $n = 4$ poles. We compared the performance of the proposed model with a VP-Net using the adaptive Hermite system as introduced in [20]. We note that the Hermite expansion based VP-Net also received the entire heartbeat as input and used $m = 10$ Hermite-functions. In addition we considered a variable projection based wavelet network, where the VP-wavelet transformation from Eq. (9) was performed using the Ricker wavelet. This model architecture was included to measure the effect of the proposed RGW wavelets, where the shape of the mother wavelet can also be learned. In the Ricker-wavelet case, the learnable parameters of the VP-layer consisted of the dilation-translation parameter pairs associated with each wavelet coefficient. Again, $m = 10$ coefficients were considered. Finally, we considered the performance of well-known methods from [22]. The classification results are shown in Table I. In addition to overall accuracy, common performance metrics, sensitivity/precision (Se) and Positive predictivity/recall (+P), also calculated:

$$Se = \frac{TP}{TP + FN} \quad \text{and} \quad +P = \frac{TP}{TP + FP} \quad (22)$$

where TP, FN, FP are true positive, false negative and false positive, respectively. These metrics were considered, since the dataset was highly biased towards healthy signals. Based on the results in Table I, the proposed RGW-VP model outperformed the other considered approaches across all metrics. It should be noted, that it also performed at the level of the best methods from the literature [22], some of which used much more complex, deep learning models for classification. From the medical point of view, the VEB Se and VEB +P metrics are particularly important, as they provide information about how reliably the methods recognize diseased waveforms. Unfortunately, in this case the Ricker-VP found just half of the VEB signals. The Hermite function based VP-Net achieved better results than Ricker-VP and comparable results to RGW-VP, but RGW-VP outperforms Hermite-VP in the VEB +P metric by approximately 10%. It should also be emphasized,

TABLE I
VEB CLASSIFICATION RESULTS ON THE MIT-BIH ARRHYTHMIA DATASET

Layer	Total acc. (%)	Normal		VEB	
		Se (%)	+P (%)	Se (%)	+P (%)
RGW-VP	98.51	99.60	98.79	85.48	94.71%
Ricker-VP	95.69	99.37	96.1	51.68	87.27
Hermite-VP	97.42	98.65	98.56	82.75	83.63
Literature [22]	-	80-99	85-99	77-96	63-99

(see also Section IV-D), that the learned parameters of RGW-VP are highly interpretable when compared to non-wavelet based approaches.

In order to measure the effect of different feature learning layers, all models used the same underlying fully-connected architecture. The hyperparameters of the model were tuned using a large, manual grid search. Sparse signal representations achieved by the VP automatic feature extraction layers reach peak performance at low model complexity [20]. In the first layer we computed $m = 10$ coefficients in all cases. We tested Ricker and Hermite-VP with more coefficients, but the performance did not outperform RGW-VP. Each model used a single hidden linear layer with 15 neurons and with a ReLu activation function after the first layer. The one-dimensional output layer used the Sigmoid function for binary classification.

D. Interpretability of the Arrhythmia Detection RGW-VP-Net

In this section we examine the interpretability of the trained parameters of the first layer. According to the literature, cardiologist recognise VEBs from their wide and bizarre QRS complex morphologies and the absence of a P wave in the heartbeat [47]. This observation supports the use of the entire heartbeat as input to the models, unlike previous methods (see e.g. [20], [46]), which only considered QRS complexes. The right column of Fig. 7 illustrates a VEB heartbeat. In the left column of the same figure, we show the scalogram of the heartbeat computed using the RGW mother wavelet learned in our experiment. The scalogram was acquired by computing all wavelet coefficients with the RGW-VP layer at every time point and 25 scales between $\lambda_1 = 0.1$ and $\lambda_{25} = 1.5$. We can see that there are dominant coefficients at lower frequencies (higher scales) because of the wide QRS complex of the heartbeat. If we examine the positions of the trained coefficients - marked with green dots - we conclude that they correspond to lower frequency information of VEB signals. At the QRS complex time segment there are also trained coefficients at high frequency values (i.e., at low scales) to aid the classification of healthy ECG heartbeat signals. Fig. 8 illustrates the learned dilated and translated RGW wavelets used to extract the features from the ECG data. The figure shows that the wavelets correspond to segments which cardiologist examine to classify VEB signals. The dark blue RGW wavelet corresponds to the P wave, light green wavelet corresponds to a wide QRS complex of a VEB signal, orange and light blue wavelets correspond to a QRS complex of a healthy ECG heartbeat. In Fig. 9 we can see a healthy and an unhealthy ECG heartbeat. In the bottom row, the value of the

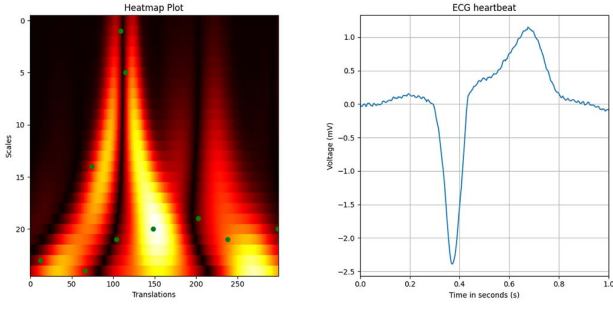


Fig. 7. CWT coefficients of a VEB signal represented by their scalogram and a VEB signal plot, respectively. Trained coefficients marked with green dots on the scalogram.

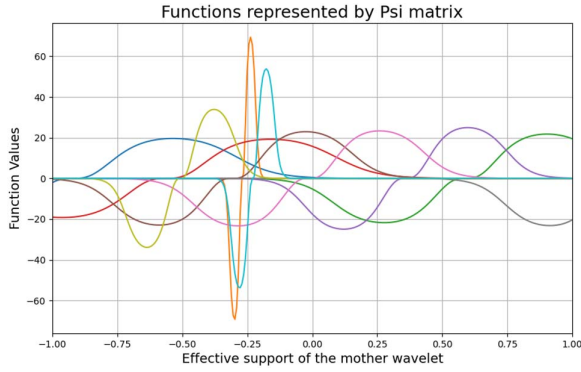


Fig. 8. Dilated and translated RGW wavelets used to extract features from ECG signals.

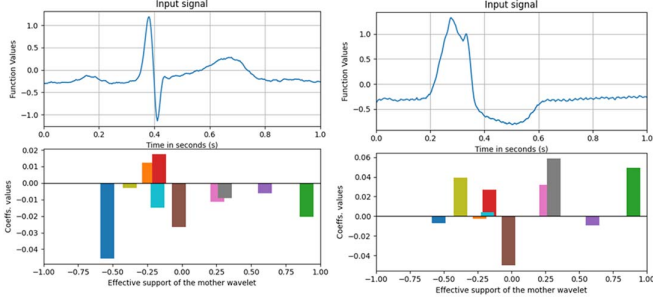


Fig. 9. A healthy and VEB ECG heartbeat signal. Below them are corresponding CWT coefficients plotted at maximum places of their corresponding wavelets. Colors corresponds to the wavelets in Fig. 8.

wavelet coefficients is also illustrated. In the case of a healthy ECG heartbeat, the coefficients which correspond to the P wave (in time) were relatively large. In turn, a VEB signal has a small dark blue coefficient value, because of the absence of the P wave. In the case of VEB signals, we can clearly see that the light green coefficient is bigger than coefficients corresponding to healthy QRS complexes.

E. Road Abnormality Recognition

In our final experiment, we demonstrate how RGW-VP nets can be used to solve a non-biomedical classification problem

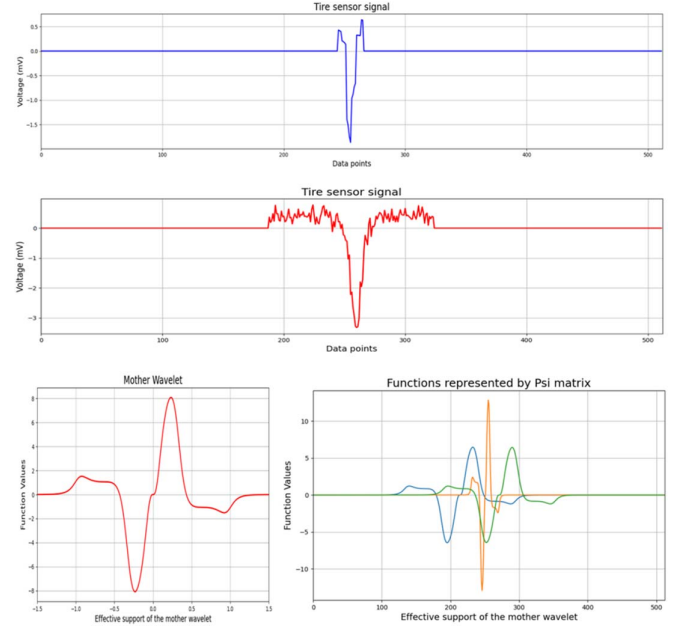


Fig. 10. Top: Preprocessed tire sensor signal segments labeled normal and abnormal, respectively. Bottom: Trained RGW mother wavelet morphology and trained wavelets of RGW-VP layer.

involving QV signals. We detect road abnormalities by classifying tire sensor signal segments. Road quality monitoring is an important problem, which, if automatized correctly can greatly reduce maintenance costs [44], [59]. In [44] a novel measurement apparatus was introduced, where a force sensor was implanted into the inner wall of the front tires of an electric test vehicle. The measured resistance of the sensors correspond to tire wall deformations. Wireless read-out electronics were employed to collect signals recorded by the sensor along four bridges. We used the data recording from the measurement of the second bridge of the sensor of the front left tire. The recording is a 1-D signal measured with a sampling rate of 1000 Hz. We created a quasi-periodic dataset containing single tire revolutions using the segmentation method suggested in [44]. A more detailed measurement setup and description of the signal preprocessing steps can be found in [44].

We note that signals corresponding to individual tire revolutions varied in length, since their duration depended on the speed of the vehicle. The recordings considered in this study were identical to the measurements used in [44] and were recorded on the public roads of Budapest, Hungary. We zero padded the tire revolution segments to the size of 512 data points to ensure they belong to the QV class. The created balanced dataset contained 282 normal and 235 abnormal data segments. The latter class included tire revolutions, where the vehicle encountered potholes and bumps on the road surface. Preprocessed tire revolution segments from a normal and abnormal road surface are depicted in Fig. 10.

We evaluated the performance of RGW-VP and Ricker-VP on the dataset. The baseline method was Hermite-VP which were proposed in the original experiments from [44]. The neural networks consisted of 3 fully connected layers with 64 neurons

TABLE II
TIRE SENSOR CLASSIFICATION RESULTS

Layer	Mean Accuracy (%)	Highest (%)	Lowest (%)
RGW-VP	99.23	100	97.09
Ricker-VP	98.84	100	98.06
Hermite-VP	97.68	99.04 %	96.12

and ReLu activations. An output layer with Sigmoid activation was used to classify the input measurements.

If we examine the morphologies of the signal segments in Fig. 10, we can see that because of the zero padding they have compact support. The support of the tire revolutions tends to zero quickly, hence they belong to the class of QV signals. The main difference between normal and abnormal signals is that tire revolutions on abnormal road surfaces have wider support and contain more noise. Wider supports are expected, as drivers tend to slow down when encountering rough surfaces. In order to compensate for this effect, several tire revolutions recorded on normal surfaces but at slow speeds were also included in the dataset. To extract these features from the input signals, we used the VP-layers as high pass filters, hence in this case we passed the residual vector of the input to the underlying classifier, defined as

$$\mathcal{R}_{\Psi(\eta)}(f) := f - \Psi(\eta) (\Psi(\eta)^+ f). \quad (23)$$

For Ricker and RGW VP-layers we computed the approximation of the input signal with $m = 3$ coefficients. The original model computed $m = 11$ coefficients. We also decreased the number of neurons to 20 in each layer. Table II contains the results of the experiments. We used 5-fold cross validation to test our models on the dataset. We can see that Ricker and RGW wavelets achieved comparable results to adaptive Hermite function system. In the case of RGW we used the same number of poles and zeros as in the ECG arrhythmia classification task. We would like to highlight that the trained RGW morphologies become similar to the shape of the tire revolution signals. As depicted in Fig. 10, two trained wavelets (blue and green wavelets) became similar to abnormal tire revolutions with a wider quasi compact support. The third orange wavelet because of the smaller scale parameter became highly similar to normal tire revolutions with narrower quasi compact support.

V. CONCLUSION

In this work, we proposed a new analytic family of wavelets which we referred to as rational Gaussian wavelets. The main idea behind the RGW wavelets is to multiply the Gaussian with an appropriate rational term, whose poles and zeros act as free parameters that govern the morphology of the mother wavelet. We demonstrated that using RGW, non-symmetric and sharp waveforms (such as ECG signals) can be reconstructed using only a few wavelet coefficients, provided that the mother wavelet morphology is chosen appropriately. This allows for a precise and sparse representation of a large class of signals. Furthermore, we proved the admissibility of the proposed

RGW wavelets and considered an appropriate numerical approximation of wavelet coefficients which minimize the error of the reconstructed signal. We showed that this method of approximating RGW coefficients using variable projection can be effectively included in neural network architectures to solve several real life signal processing problems. Our results show that the proposed methodology performs this task at the level of well-established alternatives from the literature, while providing interpretable and simple model architectures. Importantly, the proposed RGW-VP neural network was shown to learn wavelet coefficients, whose time parameter corresponds to ECG segments that are crucial in the detection of VEBs according to cardiology literature. In this way, the proposed transformation provides highly interpretable features while obtaining a high quality sparse representation of ECG data. Due to these advantages, RGW-VP was able to outperform other model-driven architectures, such as wavelet based neural networks using a fixed analyzing wavelet and variable projection networks using an adaptive Hermite expansion of the signals.

The results disclosed in this work demonstrate the potential in the proposed RGW-VP architecture. In our future work, we would like to find other potential applications of the RGW-VP scheme. Building on the results of our first experiment, we plan to further investigate the potential of RGW to generate traditional time-scale representations of signals. In addition, we would like to investigate the possibility of using RGW wavelets for other tasks as well, such as feature extractors for industrial fault detection problems. Finally, we would like to investigate whether this construction may be useful in other settings entirely, such as univariate function layers of Kolmogorov-Arnold Networks.

APPENDIX A

PROOFS OF STATEMENTS 2 AND 3

Proof of theorem 2: Recall that rational Gaussian wavelets were defined according to Eq. (6), that is $\Psi^\eta(t) := C(\eta) \cdot P^\eta(t) \cdot v^\eta(t) \cdot e^{-t^2/2}$, where P^η is an odd polynomial symmetric to 0 and v^η belongs to the class of rational functions given in Eq. (4). As explained in Section III, the parameter vector η fully defines the zeros and poles of $P^\eta \cdot v^\eta$. From Eq. (4), it is clear that $\lim_{t \rightarrow \pm\infty} v^\eta(t) = 0$, and since v^η has no real poles we have $|v^\eta(t)| < V < \infty$ ($t \in \mathbb{R}$). The Gaussian term $e^{-t^2/2}$ belongs to the class of Schwartz-functions denoted by $\mathcal{S}(\mathbb{R})$. It is well-known that $\mathcal{S}(\mathbb{R}) \subset L_1(\mathbb{R})$ and if $f \in \mathcal{S}(\mathbb{R})$, then $Q \cdot f \in \mathcal{S}(\mathbb{R})$, where Q is an arbitrary polynomial. From these observations we obtain

$$\begin{aligned} \int_{-\infty}^{+\infty} |t \cdot \psi^\eta(t)| dt &= C(\eta) \int_{-\infty}^{+\infty} |t \cdot P^\eta(t) v^\eta(t) e^{-t^2/2}| dt \\ &= C(\eta) \int_{-\infty}^{+\infty} |t \cdot P^\eta(t) \cdot e^{-t^2/2}| |v^\eta(t)| dt \\ &< C(\eta) \cdot V \cdot \int_{-\infty}^{+\infty} |Q^\eta(t) \cdot e^{-t^2/2}| dt < \infty, \end{aligned} \quad (24)$$

since $Q^\eta(t)$ is a polynomial, we have $t \cdot P^\eta(t) \cdot e^{-t^2/2} \in \mathcal{S}(\mathbb{R}) \subset L_1(\mathbb{R})$. In other words we find that $t \cdot \psi^\eta$ is in $L_1(\mathbb{R})$.

Notice, that ψ^η is an odd function, that is $\psi(t) = -\psi(-t)$ ($t \in \mathbb{R}$). This is so, because $e^{-t^2/2}$ and v^η are even and P^η is an odd polynomial satisfying Eq. (5). Using the oddity of ψ^η we obtain

$$\widehat{\psi^\eta}(0) = \int_{-\infty}^{+\infty} \psi^\eta(t) dt = \int_0^{+\infty} \psi^\eta(t) + \psi^\eta(-t) dt = 0. \quad (25)$$

Our claim is a simple consequence of Eqs. (24) and (25). Since ψ^η and $t \cdot \psi^\eta$ are both integrable, by the theorem regarding the differentiability of the Fourier transform (see e.g. Proposition 17.2.1 in [4]) we have $\widehat{\psi^\eta} \in C^1(\mathbb{R})$. However if this is the case, then the function $\widehat{\psi^\eta}(\xi)/\xi$ is continuous at 0. This is true, since in 0 by definition, the derivative of $\widehat{\psi^\eta}$ is expressed by $(\widehat{\psi^\eta})'(0) := \lim_{\xi \rightarrow 0} \frac{\widehat{\psi^\eta}(\xi) - \widehat{\psi^\eta}(0)}{\xi - 0} = \lim_{\xi \rightarrow 0} \frac{\widehat{\psi^\eta}(\xi)}{\xi}$, where we used Eq. (25) for the second equality. If $\xi \rightarrow \infty$, then the integral in theorem 2 certainly exists, since $\psi^\eta \in L_2(\mathbb{R})$. This concludes our proof of statement 2. $\square$

Proof of theorem 3: Since the support of f coincides with $[a, b]$, we have $W_\psi(\lambda, \tau)f = \int_a^b f(t) \overline{\psi_{\lambda, \tau}}(t) dt$ for all $(\lambda, \tau) \in \mathbb{R}_+ \times \mathbb{R}$. This integral can be approximated using the numerical quadrature $E(f, \lambda, \tau) := \sum_{j=0}^{N-1} f(t_j) \cdot \overline{\psi_{\lambda, \tau}}(t_j)$ with the well-known error estimate (following from the mean value theorem regarding integration)

$$|W_\psi(\lambda, \tau)f - h \cdot E(f, \lambda, \tau)| < h \cdot \frac{M_1(b-a)}{2}. \quad (26)$$

Let the parameter pairs (λ_k, τ_k) ($k = 1, \dots, m$) be fixed and notice, that for a fixed k , the value of $E(f)$ can be expressed by $E(f, \lambda_k, \tau_k) = (\Psi^* \mathbf{f})_k$, where $\mathbf{f}$ denotes the sampling of f over the points t_j ($j = 0, \dots, N-1$) and Ψ^* denotes the adjugate of $\Psi := \Psi(\boldsymbol{\eta})$. From this it is clear that the estimate in Eq. (26) only holds for the approximations obtained using variable projection, if Ψ is unitary, i.e. $\Psi^+ = \Psi^*$. If the columns of Ψ contain analytic wavelets (for example RGW), this cannot be assumed, therefore it is necessary to estimate $|(\Psi^* \mathbf{f})_k - (\Psi^+ \mathbf{f})_k|$ ($k = 1, \dots, m$). Notice that $\Psi^+ \mathbf{f}$ coincides with the solution to the Gaussian normal equation $\Psi^* \Psi \mathbf{c} = \Psi^* \mathbf{f}$. Here we exploited the fact that for pairwise different λ_k, τ_k parameters, Ψ will be full-rank. The solution is expressed by $\mathbf{c} = (\Psi^* \Psi)^{-1} \Psi^* \mathbf{f} = G^{-1} \Psi^* \mathbf{f}$, where $G(\boldsymbol{\eta}) = G = \Psi^* \Psi$. Using this, the triangle inequality and the consistency of the $\|\cdot\|_\infty$ matrix and vector norms, we obtain

$$\begin{aligned} \|\Psi^* \mathbf{f} - G^{-1} \Psi^* \mathbf{f}\|_\infty &< \|\Psi^* \mathbf{f}\|_\infty + \|G^{-1} \Psi^* \mathbf{f}\|_\infty \\ &< \|\Psi^*\|_\infty \|\mathbf{f}\|_\infty \cdot (\|G^{-1}\|_\infty + 1) \\ &= \|\Psi^*\|_\infty \|\mathbf{f}\|_\infty \left(\frac{\kappa(G)}{\|G\|_\infty} + 1 \right). \end{aligned} \quad (27)$$

We once more use the triangle rule together with Eqs. (26) and (27):

$$\begin{aligned} |W_\psi f(\lambda_k, \tau_k) - h(\Psi^+ \mathbf{f})_k| &= |W_\psi f(\lambda_k, \tau_k) - h(\Psi^* \mathbf{f} + \Psi^+ \mathbf{f} - \Psi^* \mathbf{f})_k| \\ &= |W_\psi f(\lambda_k, \tau_k) - h(\Psi^* \mathbf{f})_k - h(\Psi^+ \mathbf{f} - \Psi^* \mathbf{f})_k| \\ &< |W_\psi f(\lambda_k, \tau_k) - h(\Psi^* \mathbf{f})_k| + h \cdot |(\Psi^+ \mathbf{f} - \Psi^* \mathbf{f})_k| \end{aligned}$$

$$\begin{aligned} &< h \cdot \frac{M_1(b-a)}{2} + h \cdot \|\Psi^+ \mathbf{f} - \Psi^* \mathbf{f}\|_\infty \\ &< h \cdot \frac{M_1(b-a)}{2} + h \cdot \|\Psi^*\|_\infty \|\mathbf{f}\|_\infty \left(\frac{\kappa(G)}{\|G\|_\infty} + 1 \right). \end{aligned}$$

$\square$

REFERENCES

- [1] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. Anal.*, vol. 15, no. 4, pp. 723–736, 1984.
- [2] R. Polikar, "The story of wavelets," *Phys. Modern Topics Mech. Elect. Eng.*, pp. 192–197, 1999.
- [3] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. New York, NY, USA: Springer, 2006.
- [4] C. Gasquet and P. Witomski, *Fourier Analysis and Applications: filtering, Numerical Computation, Wavelets*, vol. 30. New York, NY, USA: Springer, 2013.
- [5] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [6] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM J. Math. Anal.*, vol. 29, no. 2, pp. 511–546, 1998.
- [7] L. Rajani Kumari, Y. Padma Sai, and N. Balaji, "R-peak identification in ECG signals using pattern-adapted wavelet technique," *IETE J. Res.*, vol. 69, no. 5, pp. 2468–2477, Jul. 2023.
- [8] J. Fan et al., "Le-LWTNET: A learnable lifting wavelet convolutional neural network for heart sound abnormality detection," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.
- [9] C. Tian, M. Zheng, W. Zuo, B. Zhang, Y. Zhang, and D. Zhang, "Multi-stage image denoising with the wavelet transform," *Pattern Recognit.*, vol. 134, 2023, Art. no. 109050.
- [10] I. Daubechies, "Ten lectures on wavelets," *Soc. Ind. Appl. Math.*, 1992.
- [11] 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.
- [12] M. Burke and M. Nasor, "ECG analysis using the Mexican-hat wavelet," in *Proc. 5th WSES Int. Conf. Circuits, Syst., Commun. Comput. (CSCC)*, Rethymno, Greece, CiteSeer, 2001, pp. 8–15.
- [13] R. L. Claypoole, R. G. Baraniuk, and R. D. Nowak, "Adaptive wavelet transforms via lifting," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP) (Cat. No. 98CH36181)*, vol. 3. Piscataway, NJ, USA: IEEE Press, 1998, pp. 1513–1516.
- [14] N. Erdol and F. Basbug, "Wavelet transform based adaptive filters: Analysis and new results," *IEEE Trans. Signal Process.*, vol. 44, no. 9, pp. 2163–2171, Sep. 1996.
- [15] H. Chen, P. S. Chua, and G. Lim, "Adaptive wavelet transform for vibration signal modelling and application in fault diagnosis of water hydraulic motor," *Mech. Syst. Sig. Process.*, vol. 20, no. 8, pp. 2022–2045, 2006.
- [16] S. Sahoo, M. Dash, S. Behera, and S. Sabut, "Machine learning approach to detect cardiac arrhythmias in ECG signals: A survey," *Irbm*, vol. 41, no. 4, pp. 185–194, 2020.
- [17] H. Dahrouj et al., "An overview of machine learning-based techniques for solving optimization problems in communications and signal processing," *IEEE Access*, vol. 9, pp. 74908–74938, 2021.
- [18] J. Gu et al., "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, Oct. 2018.
- [19] S. C. Fanni, M. Febi, G. Aghakhanyan, and E. Neri, "Natural language processing," in *Introduction to Artificial Intelligence*. New York, NY, USA: Springer, 2023, pp. 87–99.
- [20] P. Kovács, G. Bognár, C. Huber, and M. Huemer, "VPNet: Variable projection networks," *Int. J. Neural Syst.*, vol. 32, no. 1, 2022, Art. no. 2150054.
- [21] T. Dózsa, F. Deuschle, B. Cornelis, and P. Kovács, "Variable projection support vector machines and some applications using adaptive Hermite expansions," *Int. J. Neural Syst.*, vol. 34, no. 01, 2024, Art. no. 2450004, doi: 10.1142/S0129065724500047.
- [22] E. J. D. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, Dec. 2016.

- [23] J. K. Eshraghian et al., "Training spiking neural networks using lessons from deep learning," *Proc. IEEE*, vol. 111, no. 9, pp. 1016–1054, Sep. 2023.
- [24] 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.
- [25] K. Hammernik et al., "Physics-driven deep learning for computational magnetic resonance imaging: Combining physics and machine learning for improved medical imaging," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 98–114, Aug. 2023.
- [26] D. R. Pereira, M. A. Piteri, A. N. Souza, J. P. Papa, and H. Adeli, "FEMA: A finite element machine for fast learning," *Neural Comput. Appl.*, vol. 32, pp. 6393–6404, Mar. 2020.
- [27] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 413–432, 1973.
- [28] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Prob.*, vol. 19, no. 2, p. R1, 2003.
- [29] A. M. Ámon, K. Fenech, P. Kovács, and T. Dózsa, "Continuous wavelet transform based variable projection networks," in *Proc. Eur. Conf. Signal Process. (EUSIPCO)*, 2024, pp. 1906–1910.
- [30] A. Ámon, K. Fenech, P. Kovács, and T. Dózsa, "Codes for rational gaussian wavelets and corresponding model driven neural networks," 2025. [Online]. Available: <https://gitlab.com/AmonAttilaMiklos/ldrgwvp/>
- [31] C. Heil and D. F. Walnut, *Fundamental Papers in Wavelet Theory*. Princeton, NJ, USA: Princeton Univ. Press, 2006.
- [32] D. Jordan, R. Miksad, and E. Powers, "Implementation of the continuous wavelet transform for digital time series analysis," *Rev. Sci. Instrum.*, vol. 68, no. 3, pp. 1484–1494, 1997.
- [33] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, New York, NY, USA: Springer, 1995, pp. 281–299.
- [34] G. Ososkov and A. Shitov, "Gaussian wavelet features and their applications for analysis of discretized signals," *Comput. Phys. Commun.*, vol. 126, nos. 1–2, pp. 149–157, 2000.
- [35] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.
- [36] A. Bhattacharyya, L. Singh, and R. B. Pachori, "Fourier–Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals," *Digit. Signal Process.*, vol. 78, pp. 185–196, Feb. 2018.
- [37] I. W. Selesnick, "Wavelet transform with tunable Q-factor," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3560–3575, Aug. 2011.
- [38] T. Dózsa, G. Bognár, and P. Kovács, "Ensemble learning for heartbeat classification using adaptive orthogonal transformations," in *Proc. Int. Conf. Comput. Aided Syst. Theory*, New York, NY, USA: Springer, 2019, pp. 355–363.
- [39] V. A. De Freitas Barbosa, M. A. de Santana, M. K. S. Andrade, R. d. C. F. de Lima, and W. P. dos Santos, "Deep-wavelet neural networks for breast cancer early diagnosis using mammary thermographies," *Deep Learn. Data Analytics*, Cambridge, MA, USA: Elsevier, 2020, pp. 99–124.
- [40] M. A. de Santana, V. A. de Freitas Barbosa, R. de Cássia Fernandes de Lima, and W. P. dos Santos, "Combining deep-wavelet neural networks and support-vector machines to classify breast lesions in thermography images," *Health Technol.*, vol. 12, no. 6, pp. 1183–1195, 2022.
- [41] M. A. de Santana and W. P. dos Santos, "A deep-wavelet neural network to detect and classify lesions in mammographic images," *Res. Biomed. Eng.*, vol. 38, no. 4, pp. 1051–1066, 2022.
- [42] V. A. de Freitas Barbosa, A. Félix da Silva, M. A. de Santana, R. Rabelo de Azevedo, R. d. C. Fernandes de Lima, and W. P. dos Santos, "Deep-wavelets and convolutional neural networks to support breast cancer diagnosis on thermography images," *Comput. Methods Biomechanics Biomed. Eng.: Imag. Vis.*, vol. 11, no. 3, pp. 895–913, 2023.
- [43] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2091–2110, Dec. 1989.
- [44] 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.
- [45] 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.
- [46] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.
- [47] M. S. Thaler, *The Only EKG Book You'll Ever Need*. Chicago, IL, USA: Lippincott Williams & Wilkins, 2021.
- [48] A. Sandryhaila, S. Saba, M. Puschel, and J. Kovacevic, "Efficient compression of QRS complexes using Hermite expansion," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 947–955, Feb. 2012.
- [49] T. Dózsa and P. Kovács, "ECG signal compression using adaptive Hermite functions," in *Proc. ICT Innov. Emerg. Technol. Better Living 7*, New York, NY, USA: Springer, 2016, pp. 245–254.
- [50] P. Kamga, R. Mostafa, and S. Zafar, "The use of wearable ECG devices in the clinical setting: A review," *Current Emergency Hospital Med. Rep.*, vol. 10, no. 3, pp. 67–72, 2022.
- [51] L. Sörnmo, P. O. Borjesson, M.-E. Nygards, and O. Pahlm, "A method for evaluation of QRS shape features using a mathematical model for the ECG," *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 10, pp. 713–717, Oct. 1981.
- [52] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.
- [53] A. Němcová, R. Smíšek, L. Maršánová, L. Smital, and M. Vitek, "A comparative analysis of methods for evaluation of ECG signal quality after compression," *Biomed Res. Int.*, vol. 2018, no. 1, 2018, Art. no. 1868519.
- [54] P. Kovács, S. Fridli, and F. Schipp, "Generalized rational variable projection with application in ECG compression," *IEEE Trans. Signal Process.*, vol. 68, pp. 478–492, 2020.
- [55] M. M. Al Rahhal, Y. Bazi, M. Al Zuair, E. Othman, and B. BenJdira, "Convolutional neural networks for electrocardiogram classification," *J. Med. Biol. Eng.*, vol. 38, pp. 1014–1025, Mar. 2018.
- [56] H. Fujita and D. Cimr, "Decision support system for arrhythmia prediction using convolutional neural network structure without preprocessing," *Appl. Intell.*, vol. 49, pp. 3383–3391, Apr. 2019.
- [57] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, "Automatic ECG classification using continuous wavelet transform and convolutional neural network," *Entropy*, vol. 23, no. 1, p. 119, 2021.
- [58] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [59] R. Du, G. Qiu, K. Gao, L. Hu, and L. Liu, "Abnormal road surface recognition based on smartphone acceleration sensor," *Sensors*, vol. 20, no. 2, p. 451, 2020.