A General Framework for the Assessment of Detectors of Anomalies in Time Series

Andriy Enttsel, *Graduate Student Member, IEEE*, Silvia Onofri, *Graduate Student Member, IEEE*, Alex Marchioni, *Member, IEEE*, Mauro Mangia, *Member, IEEE*, Gianluca Setti, *Fellow, IEEE*, and Riccardo Rovatti, *Fellow, IEEE*

***Abstract*—Anomalies are rare events, and this affects the design flow of detectors that monitor systems that behave normally most of the time but whose failure may have serious consequences. This limitation is particularly evident in the detector performance evaluation: it requires an abundance of normal and anomalous data but realistically faces a scarcity of the latter. To address this, in this article, we develop a framework comprising a set of abstract anomalies modeling the effects real-world failures and disturbances have on sensor readings. In addition, we devise synthetic generation procedures for these anomalies. Given a dataset of normal tracks from the actual application, one may apply such procedures to produce anomalous-like time series for a comprehensive detector assessment. We show that this framework can anticipate the detector performing best with real-world anomalies in the context of human and structural health monitoring, also highlighting that, in these cases, the best detector is not the most complex.**

***Index Terms*—Model selection, outlier detection, second-order statistics, sensor faults, synthetic anomalies, time series.**

Manuscript received 5 January 2024; revised 24 April 2024; accepted 16 May 2024. Date of publication 3 July 2024; date of current version 7 October 2024. This work was supported in part by the FAIR—Future Artificial Intelligence Research and in part by the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1555 11/10/2022, PE00000013). Paper no. TII-24-0058. *(Corresponding author: Andriy Enttsel.)*

Andriy Enttsel, Silvia Onofri, Alex Marchioni, and Mauro Mangia are with the Department of Electrical, Electronic, and Information Engineering, University of Bologna, 40136 Bologna, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy (e-mail: andriy.enttsel@unibo.it; s.onofri@unibo.it; alex.marchioni@unibo.it; mauro.mangia@unibo.it).

Gianluca Setti is with the CEMSE, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia, on leave from the Department of Electronics and Telecommunications (DET), Politecnico di Torino, 10129 Torino, Italy (e-mail: gianluca.setti@kaust.edu.sa).

Riccardo Rovatti is with the Department of Electrical, Electronic, and Information Engineering, University of Bologna, 40136 Bologna, Italy, also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy, and also with the Alma Mater Research Institute for Human-Centered AI University of Bologna, 40015 Bologna, Italy (e-mail: riccardo.rovatti@unibo.it).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TII.2024.3413359.

Digital Object Identifier 10.1109/TII.2024.3413359

I. Introduction

NOWADAYS, sectors such as healthcare, industry, and structural engineering commonly engage in extensive data collection. Sensor-equipped systems produce time-series data that are crucial for monitoring, controlling, and optimizing the physical environment. In this context, anomaly detection (AD) or identification of novelties/outliers plays a crucial role in detecting system malfunctions, failures in data chains, or unauthorized intrusions. The literature offers various AD algorithms [1], and often the challenge lies in selecting the most suitable algorithm for specific applications. An effective assessment of a detector requires knowing the potential anomalies and having a substantial number of their occurrences. However, in most applications, this requirement is hardly ever met due to the scarcity of anomalous data, especially at design time.

This problem falls into the broader task of *model selection* that consists of choosing a statistical model according to the given data [2], [3]. In the specific case of AD in time series, this problem has been systematically addressed only recently in [3], [4], [5], and [6], showing that a promising solution consists of generating synthetic anomalies.

Nevertheless, to adopt this strategy, a designer needs to identify the most common anomalies to the application; characterize and model the identified anomalies; and devise a synthetic generation procedure. The literature still lacks a unique framework that includes and automates all these steps.

A. Related Work

This direction has already been explored for image data, albeit in a different context than AD. Hendrycks and Dietterich [7] developed a framework to generate common image corruptions with varying intensities for the assessment of a classifier.

In the case of time series, most studies have focused only on specific sources of abnormality [8], [9], [10], [11], and only in some cases, outline a systematic procedure for generation [3], [12], [13], [14], [15], [16].

Some investigations have considered anomalies affecting monitoring devices. For instance, the authors in [8], [9], and [10] analyze the most common sensor faults, such as spike, stack-at, low battery, and classify them into a few categories such as *short*, *noise*, and *constant*. On the same line, Jan et al. [11] model a limited number of faults to investigate the detection capabilities of one particular class of detectors [17]. Instead,

TABLE I
EXAMPLES OF DATA CORRUPTIONS AND SYSTEM ALTERATIONS

Sensor faults

Type	Description
Noise faults	Occur due to hardware failure, environmental conditions, or battery supply issues, resulting in unexpected high variance or noise in sensor values [8].
Short fault/spike	Manifest as an instantaneous increase in the rate of change of sensor values and is caused by hardware or connection failures [11], [18].
Clipping or saturation	Result in a clipping of the extreme values in sensor readings, either due to physical limitations of certain sensor types or calibration issues [19], [20].
Stuck-at faults	Refer to situations when the output shows low variance with samples concentrated around a constant value [21].
Constant	Due to improper calibration or drift of calibration parameters over time, sensor readings may manifest a consistent deviation by a constant value [22].
Narrow-band interference	Considers spurious narrow-band signals contaminating the band of the signal of interest [23].
Dead-zone	Is a nonlinearity that manifests when a sensor or actuator fails to provide a non-null output value [24].

System alterations

Type	Description
Aging	Occurs due to the natural and unavoidable variation of the physical properties of the system over time [25], [26].
Wearing	Alterations caused by degradation of the system's components due to repeated environmental or mechanical stimuli, e.g., erosion, friction [27].
Abrupt changes	Sudden variations in the system's behaviour. Examples of such alterations include damages in a civil structure (tendon/strand breakages [28] and earthquakes [29]); irregular heart rate, irregular rhythm, and ectopic rhythm during heart activity [30].

Enttsel et al. [16] focus on the malfunctioning of the monitored system and proposes two models of anomaly, named *spectrum* and *basis*, describing the statistical deviation from the normal behavior.

Alternatively, the authors in [12], [13], and [14] detach from the source of abnormality to classify anomalies according to their effect on the signal. Lai et al. [12] model the normal signal as the combination of two components, namely *seasonality* and *trend*, and generate anomalies as alterations of these two components. The authors in [13] and [14] consider a more general model for the normal signal but limit their analysis to a few anomalies such as *global*, *local*, and *dependence* anomalies.

Finally, the authors in [3] and [15] offer a wide range of anomalies covering some of the effects that real anomalous signals might have on normal data. Although nonnegligible, the set of anomalies is still nonexhaustive and the framework lacks a shared parameter connected to anomaly severity, which is crucial to allow a comprehensive quantitative evaluation of detectors across different anomaly types.

B. Our Contribution

We propose a framework based on synthetic anomalies for a systematic assessment of a detector. We first summarize in Table I the anomalies that are prevalent in time series, ranging from sensor faults to aging and wearing of the monitored system.

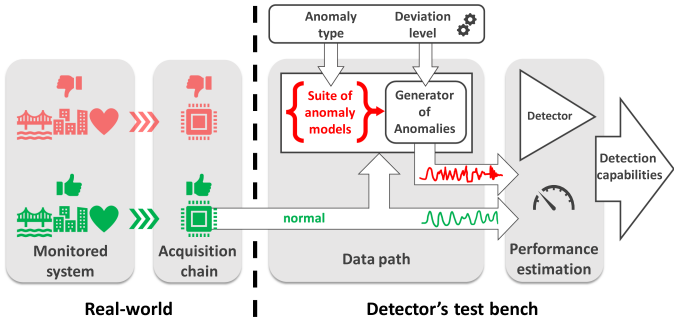


Fig. 1. Framework that allows testing a detector aimed at identifying unknown anomalies coming from different sources.

Then, we address the issue of anomaly characterization by identifying the effects of the anomalies on the signal, enabling the development of models that are independent of the source. This approach allows us to expand on the characterizations available in the literature, encompassing all possible anomalies outlined in Table I and potentially others.

Anomalies are grouped into three families according to their effect on the signal's power: increase, invariance, and decrease. The former family includes superimposed disturbances, power invariant anomalies account for deviations affecting the main source of information and the latter case covers nonlinear distortions causing a power decrease. Since the anomalies are modeled with a parameter that decides the degree of deviation from the signal, we devise a versatile procedure for synthetic generation that allows for a consistent assessment of the detector.

As a result, we set up a flexible testing suite,[1] depicted in Fig. 1, for assessing anomaly detectors within any specific context. With historical data and domain-specific knowledge, designers can tailor this test suite to their applications. On this line, we adapted our framework to human health and structural health monitoring in which detectors analyze electrocardiogram (ECG) and acceleration signals. These signals were synthetically corrupted to create anomalous datasets that have been used to assess the performance of various anomaly detectors. Such assessment is then matched with the response of the detectors to real anomalies confirming the validity of the proposed framework.

To summarize, we propose a framework aimed at assessing detectors using synthetic anomalies that is

1) *versatile*: can be applied in any monitoring scenario;
2) *enabling*: requires access to normal data only;
3) *comprehensive*: covers most of the effects real anomalies have on a signal;
4) *consistent*: features anomalies that stem from a common model and that can be controlled by a single common parameter to set their intensity.

To achieve this, we have performed the following.

1) *Characterization*: We analyzed the most prevalent anomalies in time series to identify the effects they have on the normal signal.

[1]The source code can be found in the GitHub repository: https://github.com/SSIGPRO/wombats.

2) *Modeling*: We designed a set of abstract anomalies that simulate the effects of real-world anomalies, categorizing them based on their effect on the signal's power.
3) *Generation*: We devised a procedure to generate the anomalies that pose an equivalent degree of challenge to a detector.
4) *Validation*: We deployed the framework in two realistic scenarios and verified its effectiveness in selecting the most suitable detector for a given anomaly.

The rest of this article is organized as follows. In Section II, normal and anomalous signals are modeled. In Section III, the entire framework for detector selection is described. Section IV focuses on the setup of the numerical experiments, whose results are shown and commented in Section V. Finally, Section VI concludes this article.

II. Models of Normal and Anomalous Signals

In this section, we outline the model and assumptions we adopt to characterize normal observations, and we describe the general model that defines anomalous behaviors.

A. From Time Series to Time Instances

Data collected and recorded continuously over an interval of time are known as time series. Formally, it can be defined as a set S of pairs each comprising a vector of samples of m simultaneously monitored quantities $\boldsymbol{s}_i = (s^{(0)}, \dots, s^{(m-1)})$ acquired at a time instant, or timestamp t_i: $S = \{(\boldsymbol{s}_i, t_i) | i \in \mathbb{N}, t_k < t_l \text{ if } k < l\}$. When only one quantity is tracked, i.e., $m = 1$, the time series S is called univariate, while if $m > 1$, S is a multivariate time series. Since the time interval between two consecutive measurements usually remains constant, the information on time can be dropped. Moreover, in this work, we will focus on univariate series only, which ultimately can be seen as a vector $\boldsymbol{s}$ of ordered measurements $\boldsymbol{s} = (s_0, \dots, s_k, \dots)$. From now on, we will treat $\boldsymbol{s}$ as a signal sampled with a time period T.

Considering a realistic scenario where the signal $\boldsymbol{s}$ has to be processed in a real-time fashion, one cannot consider $\boldsymbol{s}$ as a whole. We model $\boldsymbol{s}$ as a concatenation of not overlapping signal instances each made of n consecutive samples: $\boldsymbol{s} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_k \dots)$, with $\boldsymbol{x}_j = (s_{jn}, \dots, s_{j(n+1)-1})$. We also suppose that the length of the instance n is large enough so that $\boldsymbol{x}^j$ captures the main statistical features of the entire signal $\boldsymbol{s}$, and that the dependence between $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ with $k \neq j$ is negligible. This allows us to focus on one $\boldsymbol{x} = (x_0, \dots x_{n-1})$ instance at a time.

B. Model of Normal Signals

With these considerations, the samples of the acquired signal are collected in a vector $\boldsymbol{x} \in \mathbb{R}^n$ that can be either an instance of a source of normal readings $\boldsymbol{x}^{\text{ok}} \in \mathbb{R}^n$ or of a source of anomalous readings $\boldsymbol{x}^{\text{ko}} \in \mathbb{R}^n$. The two sources are assumed to have different statistics which should help telling anomalous instances from normal ones. With no loss of generality, we assume $\boldsymbol{x}^{\text{ok}}$ to have zero mean $\mathbf{E}[\boldsymbol{x}^{\text{ok}}] = 0$ and covariance matrix $\boldsymbol{\Sigma}^{\text{ok}} = \mathbf{E}[\boldsymbol{x}^{\text{ok}} \boldsymbol{x}^{\text{ok}\top}]$. We also assume $\boldsymbol{x}^{\text{ok}}$ normalized to unit power so that $\frac{1}{n}\mathbf{E}[\|\boldsymbol{x}^{\text{ok}}\|^2] = 1$, where $\|\cdot\|$ indicates the l_2 norm.

Real-world signals tend to distribute power unevenly in the signal space. This can be observed in the spectral decomposition $\boldsymbol{\Sigma}^{\text{ok}} = \boldsymbol{U}^{\text{ok}} \boldsymbol{\Lambda}^{\text{ok}} \boldsymbol{U}^{\text{ok}\top}$ with $\boldsymbol{U}^{\text{ok}}$ being the $n \times n$ orthonormal matrix containing the eigenvectors of $\boldsymbol{\Sigma}^{\text{ok}}$ and $\boldsymbol{\Lambda}^{\text{ok}}$ being the $n \times n$ diagonal matrix listing the corresponding eigenvalues $\lambda_0^{\text{ok}} \geq \lambda_1^{\text{ok}} \geq \dots \geq \lambda_{n-1}^{\text{ok}} \geq 0$. Such eigenvalues are not equal, and given their decreasing order, one may find an $\bar{n}$ such that the fraction of power contained in the first $\bar{n}$ components is

$$\gamma = \frac{1}{n} \sum_{j=0}^{\bar{n}-1} \lambda_j^{\text{ok}} \gg \frac{\bar{n}}{n} \tag{1}$$

and thus, identifies the so-called $\bar{n}$ principal components of the signal. Signal components beyond the principal ones are usually dominated by noise and carry little information on the features of the normal signal.

C. Model of Anomalous Signals

Anomalies are often described according either to their cause or to their effect on the system in each specific application. Some examples of faults afflicting sensors, as well as of other anomalies afflicting the system under monitoring in different scenarios, are shown in Table I.

Although such domain knowledge is always valuable, its generalization into an assessment framework requires abstracting from causes and from general effects to focus on the changes that the anomalies cause on the acquired signals.

To this end, we propose a dictionary of anomalies derived from a common mathematical model that is specialized to focus on different signal features that are commonly altered by real-world anomalies. We concentrate on anomalies that are *variations* of the normal signal in the sense that anomalous waveforms can be written starting from normal waveforms as

$$\boldsymbol{x}^{\text{ko}} = c\left(\boldsymbol{x}^{\text{ok}}\right) + \boldsymbol{d} \tag{2}$$

where $c : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a potentially nonlinear function modeling how the anomaly changes the normal signal for which we assume $\mathbf{E}[c(\boldsymbol{x}^{\text{ok}})] = 0$, and $\boldsymbol{d} \in \mathbb{R}^n$ contains the samples of a disturbance waveform independent of $\boldsymbol{x}^{\text{ok}}$ for which we assume power $\frac{1}{n}\mathbf{E}[\|\boldsymbol{d}\|^2] = a^2$ for a given parameter a. Independence between $\boldsymbol{d}$ and $\boldsymbol{x}^{\text{ok}}$ implies $\mathbf{E}[\boldsymbol{x}^{\text{ok}\top} \boldsymbol{d}] = \mathbf{E}[\boldsymbol{d}^\top c(\boldsymbol{x}^{\text{ok}})] = 0$ so that the two components in (2) give separate power contributions. Depending on the anomaly, c may be a simple scaling, a full linear mapping identified by an $n \times n$ matrix, or a nonlinearity.

To effectively span the anomaly space, we apply variations to the normal signal separately, i.e., if $a^2 > 0$, then c amounts only to signal scaling, while if $a^2 = 0$, we either redistribute signal power by means of a linear transformation, or corrupt signal samples by means of a nonlinear c.

To quantify how much the anomalous waveforms are expected to be different from normal ones, we make use of *deviation* [16]

$$\delta = \frac{1}{n}\mathbf{E}\left[\left\|\boldsymbol{x}^{\text{ok}} - \boldsymbol{x}^{\text{ko}}\right\|^2\right] = 1 + \frac{1}{n}\text{tr}\left(\boldsymbol{\Sigma}^{\text{ko}}\right) - \frac{2}{n}\mathbf{E}\left[\boldsymbol{x}^{\text{ok}\top} c\left(\boldsymbol{x}^{\text{ok}}\right)\right] \tag{3}$$

where $\boldsymbol{\Sigma}^{\mathrm{ko}} = \mathbf{E}[\boldsymbol{x}^{\mathrm{ko}}\boldsymbol{x}^{\mathrm{ko}\top}]$ is the correlation matrix of the anomalous signal and we have exploited that $\frac{1}{n}\mathrm{tr}(\boldsymbol{\Sigma}^{\mathrm{ok}}) = 1$, as well as the uncorrelatedness of $\boldsymbol{x}^{\mathrm{ok}}$ and $\boldsymbol{d}$.

Deviation will be used to parameterize the difficulty of the AD task so that detectors can be tested against different anomalies that, in principle, offer the same level of challenge.

In case $a^2 = 0$ and $c(\boldsymbol{x}^{\mathrm{ok}}) = \boldsymbol{C}\boldsymbol{x}^{\mathrm{ok}}$ for some matrix $\boldsymbol{C}$, we may set

$$\boldsymbol{C} = \boldsymbol{U}^{\mathrm{ko}}\sqrt{\boldsymbol{\Lambda}^{\mathrm{ko}}}\sqrt{\boldsymbol{\Lambda}^{\mathrm{ok}}}^{-1}\boldsymbol{U}^{\mathrm{ok}\top} \tag{4}$$

that leverages the spectral decomposition also of the anomalous signals $\boldsymbol{\Sigma}^{\mathrm{ko}} = \boldsymbol{U}^{\mathrm{ko}}\boldsymbol{\Lambda}^{\mathrm{ko}}\boldsymbol{U}^{\mathrm{ko}\top}$ and on the straightforward definition of the square root of a diagonal matrix. The linearity of the mapping between normal and anomalous signals implies that the last expectation in (3) becomes

$$\mathbf{E}\left[\boldsymbol{x}^{\mathrm{ok}\top}c\left(\boldsymbol{x}^{\mathrm{ok}}\right)\right] = \mathrm{tr}\left(\boldsymbol{C}\boldsymbol{\Sigma}^{\mathrm{ok}}\right) = \mathrm{tr}\left(\boldsymbol{U}^{\mathrm{ko}}\sqrt{\boldsymbol{\Lambda}^{\mathrm{ko}}}\sqrt{\boldsymbol{\Lambda}^{\mathrm{ok}}}\boldsymbol{U}^{\mathrm{ok}\top}\right). \tag{5}$$

III. Framework for Detector Selection

In this section, we describe the proposed framework, depicted in Fig. 1, for detector selection. Once the target detectors have been trained on the normal training data, the designer can use a test set of normal data to generate its anomalous counterpart. The latter depends on the type of anomaly selected from a defined suite and on the intensity of the anomaly set by the deviation parameter. Subsequently, the detectors are tested with both normal and anomalous sets to assess their discriminating capabilities.

To describe the procedure in more detail, we present the anomalies suite categorized according to the effect anomalies have on the power of the signal; show how anomalies can be implemented; provide deviation expressions for each anomaly; and define a metric for detector's evaluation.[2]

A. Anomalies That Increase Signal Power

In this case, we assume that c is the identity and $a^2 > 0$, so the anomalous signal has power $\frac{1}{n}\mathbf{E}\|\boldsymbol{x}^{\mathrm{ko}}\|^2 = 1 + a^2$. We will consider a few possible disturbances.

1) Constant: $d_j = \pm a$ for $j = 0, \ldots, n-1$.

2) Step:

$$d_j = \pm a \begin{cases} r\sqrt{n/\bar{\jmath}} & j = 0, \ldots, \bar{\jmath}-1 \\ (1-r)\sqrt{n/(n-\bar{\jmath})} & j = \bar{\jmath}, \ldots, n-1 \end{cases} \tag{6}$$

where $r \in \{-1, 1\}$ decides whether we have a rising or a falling step and $\bar{\jmath} \in \{0, \ldots, n-1\}$ sets the step position within the time window.

3) Impulse:

$$d_j = \pm a \begin{cases} \sqrt{n} & \text{if } j = \bar{\jmath} \\ 0 & \text{otherwise} \end{cases}, \quad 0 \le j < n \tag{7}$$

where $\bar{\jmath} \in \{0, \ldots, n-1\}$ sets the pulse position.

[2]In Section V of the supplementary material, we discuss the computational considerations associated with the proposed framework.

4) Gaussian White Noise (GWN): $\boldsymbol{d}$ is a simple White Gaussian noise.

5) Gaussian Narrowband Noise (GNN): The disturbance is a Gaussian signal with band $[f_0 - B/2, f_0 + B/2]$ with $0 \le f_0 \le 1/2$ as center frequency and $0 \le B \le \min\{f_0, 1/2 - f_0\}$ as bandwidth.

B. Anomalies That Do Not Change Signal Power

In this case, anomalies alter the normal signal by redistributing its power in the signal space. Redistribution can be between the normal signal and an independent disturbance or between the components of the normal signal itself.

1) Mixing With a Disturbance: When a disturbance with power $a^2 \le 1$ is present, we model mixing by setting $c(\boldsymbol{x}^{\mathrm{ok}}) = \sqrt{1-a^2}\boldsymbol{x}^{\mathrm{ok}}$. As a disturbance, we consider only constant $\boldsymbol{d}$ and white noise $\boldsymbol{d}$.

2) Intrasignal Mixing: When power redistribution affects only the components of the normal signal, we assume this happens by linear mapping, i.e., by some matrix $\boldsymbol{C}$ yielding power invariance, i.e., causing $\frac{1}{n}\mathrm{tr}(\boldsymbol{\Sigma}^{\mathrm{ko}}) = 1$.

a) Time warping [31]: This anomaly alters the normal signal by local decelerations. It can be modeled assuming that the entries in $\boldsymbol{x}^{\mathrm{ok}}$ are samples at rate f_{s} taken from a continuous-time waveform $\boldsymbol{x}^{\mathrm{ok}}(t)$ subject to a time-to-time mapping $w(t)$ yielding an anomalous waveform $\boldsymbol{x}^{\mathrm{ko}}(t) = \boldsymbol{x}^{\mathrm{ok}}(w(t))$ that is sampled into the vector $\boldsymbol{x}^{\mathrm{ko}}$. With this, defining the sampling instants $t_j = \frac{j}{f_{\mathrm{s}}}$ for $j = 0, \ldots, n-1$, we have

$$\boldsymbol{x}_j^{\mathrm{ko}} = \boldsymbol{x}^{\mathrm{ko}}(t_j) = \boldsymbol{x}^{\mathrm{ok}}(w(t_j)) = \sum_{k=0}^{n-1} \boldsymbol{x}_k^{\mathrm{ok}}\eta(w(t_j)f_{\mathrm{s}} - k)\mathrm{d}w_k \tag{8}$$

where $\eta(\cdot)$ is the interpolating function allowing waveform reconstruction from its samples and $\mathrm{d}w_j$ accounts for local time warping to guarantee power invariance. Clearly, in this case, $\boldsymbol{C}_{j,k} = \eta(w(t_j)f_{\mathrm{s}} - k)\mathrm{d}w_k$.

Further to power redistribution along time, we have to consider power redistribution in the spectrum, i.e., from the point of view of the spectral decomposition of the correlation matrix $\boldsymbol{\Sigma}^{\mathrm{ok}} = \boldsymbol{U}^{\mathrm{ok}}\boldsymbol{\Lambda}^{\mathrm{ok}}\boldsymbol{U}^{\mathrm{ok}\top}$, in which the first $\bar{n}$ components are the principal ones and characterize normality. We may think of two alterations.

b) Spectral alteration: This anomaly affects how power is distributed among the principal components. It can be modeled by altering the eigenvalues on the diagonal of $\boldsymbol{\Lambda}^{\mathrm{ok}}$ obtaining $\boldsymbol{\Lambda}^{\mathrm{ko}}$ and thus $\boldsymbol{C}$ from (4).

c) Principal subspace alteration: This anomaly changes the principal components. It can be modeled by altering the columns of $\boldsymbol{U}^{\mathrm{ok}}$ to obtain a new orthonormal $\boldsymbol{U}^{\mathrm{ko}}$, and thus, $\boldsymbol{C}$ from (4).

C. Anomalies That Decrease Signal Power

1) Saturation: This implies that the largest samples in a window are clipped to a maximum value x_{SAT} as in

$$x_j^{\mathrm{ko}} = \begin{cases} x_{\mathrm{SAT}}\mathrm{sign}\left(x_j^{\mathrm{ok}}\right) & \text{if } \left|x_j^{\mathrm{ok}}\right| > x_{\mathrm{SAT}} \\ x_j^{\mathrm{ok}} & \text{otherwise.} \end{cases} \tag{9}$$

2) Dead-Zone: Dead-zone is the opposite of saturation and sets to 0 the samples smaller than x_{DZ} as in

$$x_j^{\mathrm{ko}} = \begin{cases} 0 & \text{if } \left|x_j^{\mathrm{ok}}\right| < x_{\mathrm{DZ}} \\ x_j^{\mathrm{ok}} & \text{otherwise.} \end{cases} \tag{10}$$

D. Anomaly Implementation

Considering the taxonomy of anomalies presented previously, we provide a definition of how the random components of the mathematical models are mapped into actual instances.

1) $\pm$ is implemented either with a $+$ or $-$ with $\frac{1}{2}$ probability.
2) The rising parameter r is drawn uniformly from $\{0, 1\}$, and the position $\bar{\jmath}$ is set to $\lceil n/2 \rceil$.
3) When $\boldsymbol{d}$ is a Gaussian noise in the band $[f_0 - {}^{B}/_{2}, f_0 + {}^{B}/_{2}]$, we draw it as a Gaussian random vector with zero means and covariance $\boldsymbol{\Sigma}$ with [32]

$$\boldsymbol{\Sigma}_{j,k} = a^2 \cos\left(2\pi(j-k)f_0\right) \operatorname{sinc}\left((j-k)B\right) \tag{11}$$

for $j, k = 0, \ldots, n-1$. When white noise is needed, we simply set $\boldsymbol{\Sigma} = a^2 \boldsymbol{I}$.
4) Interpolation η uses third order cardinal splines [33].
5) The warping function is $w(t) = (1-\alpha)t$ with $\alpha \in [0, 1]$ controlling local deceleration.
6) We concentrate spectral alterations on the principal components and assume that the noise subspace is untouched, i.e., that $\lambda_j^{\mathrm{ko}} \neq \lambda_j^{\mathrm{ok}}$ for $j = 0, \ldots, \bar{n}-1$ but $\lambda_j^{\mathrm{ko}} = \lambda_j^{\mathrm{ok}}$ for $j = \bar{n}, \ldots, n-1$. The number of principal components $\bar{n}$ is computed according to the criterion defined in [34].
The first $\bar{n}$ eigenvalues of $\boldsymbol{\Lambda}^{\mathrm{ok}}$ are altered by defining $\boldsymbol{\ell}^{\mathrm{ko}} = \frac{1}{\sqrt{n\gamma}}(\sqrt{\lambda_0^{\mathrm{ko}}}, \ldots, \sqrt{\lambda_{\bar{n}-1}^{\mathrm{ko}}})^\top$ and $\boldsymbol{\ell}^{\mathrm{ok}} = \frac{1}{\sqrt{n\gamma}}(\sqrt{\lambda_0^{\mathrm{ok}}}, \ldots, \sqrt{\lambda_{\bar{n}-1}^{\mathrm{ok}}})^\top$. Since power invariance requires $\|\boldsymbol{\ell}^{\mathrm{ok}}\| = \|\boldsymbol{\ell}^{\mathrm{ko}}\| = 1$, we set $\boldsymbol{\ell}^{\mathrm{ko}} = \boldsymbol{R}_{\bar{n},\theta}\boldsymbol{\ell}^{\mathrm{ok}}$ for some random $\bar{n} \times \bar{n}$ matrix $\boldsymbol{R}_{\bar{n},\theta}$ encoding a rotation of an angle θ. The altered eigenvalues are then retrieved from $\boldsymbol{\ell}^{\mathrm{ko}}$ by squaring and scaling.
7) Principal subspace alteration sets $\boldsymbol{U}^{\mathrm{ko}} = \boldsymbol{R}_{n,\theta}\boldsymbol{U}^{\mathrm{ok}}$, where $\boldsymbol{R}_{n,\theta}$ is an $n \times n$ random matrix encoding a rotation of an angle θ applied to the columns of $\boldsymbol{U}^{\mathrm{ok}}$.
8) Random rotations are built assuming that n and $\bar{n}$ are even, with

$$\boldsymbol{R}_\theta = \boldsymbol{Q}\begin{pmatrix} \boldsymbol{r}_\theta & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{r}_\theta \end{pmatrix}\boldsymbol{Q}^\top, \quad \boldsymbol{r}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{12}$$

where $\boldsymbol{r}_\theta$ is the 2-D rotation matrix and $\boldsymbol{Q}$ is generated by orthonormalizing an instance of the Ginibre ensemble [35].
9) In clipping and dead-zone alterations, the thresholds are adapted in each window to match the desired δ in (18) and (19).

TABLE II
EXPRESSIONS OF DEVIATION δ FOR EACH ANOMALY

Anomalies	δ	
Increasing	a^2	(13)
Mixing	$2\left[1 - \sqrt{1-a^2}\right]$	(14)
Time warping	$2\left[1 - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{C}\boldsymbol{\Sigma}^{\mathrm{ok}}\right)\right]$	(15)
Spectral alt.	$2\gamma\left[1 - \cos(\theta)\right]$	(16)
Principal subspace alt.	$2\left[1 - \cos(\theta)\right]$	(17)
Saturation	$\sum_{\left\|x_j^{\mathrm{ok}}\right\| > x_{\mathrm{SAT}}} \left[x_j^{\mathrm{ok}} - x_{\mathrm{SAT}}\operatorname{sign}\left(x_j^{\mathrm{ok}}\right)\right]^2$	(18)
Dead-zone	$\sum_{\left\|x_j^{\mathrm{ok}}\right\| \leq x_{\mathrm{DZ}}} \left(x_j^{\mathrm{ok}}\right)^2$	(19)

10) In principle, different effects can be combined to create an anomaly that emulates real-world anomalies manifesting with multiple superimposed effects. More details can be found in Section IV of the supplementary material.

E. Deviations

For each type of anomaly, Table II reports the relationship between the parameters in the anomaly definition and the deviation δ. Those formulas allow to apply alterations posing a prescribed and uniform level of challenge to the detector to test.

In (15), the deviation is controlled by the parameter α inherent to $\boldsymbol{C}$, while in (16), deviation is caused by alterations $\boldsymbol{R}_{\bar{n},\theta}$. Considering that, principal subspace alterations caused by $\boldsymbol{R}_{n,\theta}$ are equivalent to rotating each $\boldsymbol{x}^{\mathrm{ok}}$ into the corresponding $\boldsymbol{x}^{\mathrm{ko}}$ by the same angle, (17) can be obtained by exploiting (4) and (16). Finally, (18) and (19) can be used to set, respectively, x_{SAT} and x_{DZ} so that the average deviation over the available dataset is the given δ.

F. Metrics

When the detector is fed with an instance $\boldsymbol{x}$ that is either ok or ko, the output of a detector can be considered as a function $z(\boldsymbol{x})$ that assigns a score to the instance, such that larger scores correspond to more anomalous behaviors [36].

To generate the final binary decision, the score must be compared with a threshold that is dependent on the specific needs of the application. To remain independent from the threshold choice, we rely on the area under the ROC Curve (AUC) measure [37]. When one anomalous and one normal vectors are randomly picked, AUC $\in [0, 1]$ is the probability that the score will be higher for the former than for the latter. What we use as a metric is a variation of AUC corresponding to the probability of correct detection

$$P_D = \begin{cases} \mathrm{AUC} & \text{if AUC} \geq 0.5 \\ 1 - \mathrm{AUC} & \text{if AUC} < 0.5 \end{cases}. \tag{20}$$

P_D accounts for the fact that, when a detector consistently scores normal instances higher than anomalous ones, i.e., AUC < 0.5, it is still useful for detection, if its output is interpreted in a reversed manner.

IV. Experimental Setup

To prove the effectiveness of the proposed anomaly models, we evaluate numerically the performance of a set of detectors on two different datasets: ECG signals coming from a Health Monitoring application and accelerometers' (ACC) waveforms used for Structural Health Monitoring.

A. Detection Methods

Our methodology is tested by deploying a variety of detectors based on different techniques that have been proposed in literature [1].

1) Principal Component Analysis (PCA) Based: PCA [39] allows to represent the signal in a lower $k < n$ dimensional principal subspace where most of the signal's average energy concentrates.[3] If such subspace is fit on the normal data, an anomaly can be identified by building a score that evaluates how well this representation describes each point. In particular, it is possible to define the following two different families of detectors based on the quantities [40].

1) Squared prediction error (SPE_k) that looks at the effects that an anomaly has on the residual space, defined by the eigenvectors not included in the principal components set. Formally, $z(\boldsymbol{x}) = \left\|\boldsymbol{x} - \boldsymbol{U}_k^{\text{ok}} \boldsymbol{U}_k^{\text{ok}\top} \boldsymbol{x}\right\|^2$, where $\boldsymbol{U}_k^{\text{ok}}$ is the matrix containing the first k columns of $\boldsymbol{U}^{\text{ok}}$.
2) Hotelling's T-squared test (T_k^2) that intends to identify the anomalies better highlighted in the principal space and it is defined as $z(\boldsymbol{x}) = \left\|(\boldsymbol{\Lambda}_k^{\text{ok}})^{-1/2} \boldsymbol{U}_k^{\text{ok}\top} \boldsymbol{x}\right\|^2$, with $\boldsymbol{\Lambda}_k^{\text{ok}}$ a $k \times k$ upper left submatrix of $\boldsymbol{\Lambda}^{\text{ok}}$.

2) Gaussian Distribution (GD) Based: Consider that the normal data comes from a multivariate GD. We analyze the two different approaches.

1) Mahalanobis distance (MD) [1] builds the score as a distance of the vector $\boldsymbol{x}$ from the center of the estimated distribution, weighted by the variance of each component
2) Autoregressive model (AR_p) [36] technique first fits a linear regression model that predicts the next samples, and then, during testing, uses the residual of the prediction as the score. More precisely, if the order of the model is $p \in [1, n)$, the score can be defined as $z(\boldsymbol{x}) = \frac{1}{n-p} \sum_{i=0}^{n-p-1} (x_{p+i} - y_i)^2$ with $\boldsymbol{y}$ being a vector containing the predictions of the last $n - p$ elements of $\vec{\boldsymbol{x}}$ provided by the AR model starting from the first p elements of $\vec{\boldsymbol{x}}$.

3) Machine Learning (ML) Based:

1) Local outlier factor (LOF_h) [41] is a technique based on nearest neighbors that defines the score of a vector as a ratio of the average density around each of the h neighbors and the density of the vector itself.
2) Isolation forest (IF_q) [42] is based on the concept that anomalies generally can be more easily separated (isolated) from the other vectors. Accordingly to this hypothesis, it proceeds to build q tree-like structures by randomly separating the vectors of a given dataset in a recursive fashion. The separability score is then related to the average depth of the vector among different trees.
3) One-class support vector nachine ($\text{OC}_{\text{kernel},\nu}$) [17] learns a suitable nonlinear mapping able to transform input data into a higher dimensional embedding, where a hyperplane can separate anomalous points from the normal ones. The score attributed to each point is then computed as the distance of such point from the origin of the embedding space. The main hyperparameters of this method are the *kernel* of the nonlinear mapping and ν, which is the lower bound of the fraction of vectors belonging to the training set used as support vectors.

Many of the previously listed techniques have been generalized to neural networks [4], [43], [44], [45], [46] to capture nonlinear relationships in the underlying data. Other techniques have been developed to capture additional features, such as long-range temporal dependencies [47]. However, we do not include any of them here because we aim to show the effectiveness of our assessment procedure, not to test any particular detector.

4) Feature Based: Along with the most common detectors, we also track some of the signal's representative features that can be efficiently computed. Each of the features listed in Table III corresponds to a different score and its detection capabilities can highlight the nature of the modeled anomalies.

In Section V of the supplementary material, we provide details on the time complexity and memory occupation of the deployed detectors during the testing phase.

TABLE III
Signal Features and the Corresponding Score $z(x)$

Feature	$z(\boldsymbol{x})$	Description
pk-pk	$\max \boldsymbol{x} - \min \boldsymbol{x}$	Peak-to-peak value
energy	$\sum_{i=0}^{n-1} x_i^2$	Energy of the vector
TV	$\sum_{i=0}^{n-1} \lvert x_{i+1} - x_i \rvert$	Total variation [38]
ZC	$\frac{1}{2} \sum_{i=0}^{n-2} [1 - \text{sign}(x_i x_{i+1})]$	Number of zero crossings

B. Datasets

1) ECG Signals: The reference ECG signals have been generated using a realistic generator introduced in [48] with the setup described in [49]. In particular, the sampling rate is set to 256 sample/s and the heart-beat rate is uniformly drawn in the range 60–100 beat/min. We first generate 7.7×10^4 chunks of 2 s that are then randomly split into nonoverlapping windows of length $n = 256$ to create vectors $\boldsymbol{x}^{\text{ok}} \in \mathbb{R}^n$. GWN is superimposed with a power guaranteeing an SNR of 40 dB. The resulting number of principal components is $\bar{n} = 52$. Finally, the dataset is split so that, 10^4 vectors are allocated to performance

[3] In this work we select $k = \bar{n}$, and k corresponding to $\gamma = 0.999$.

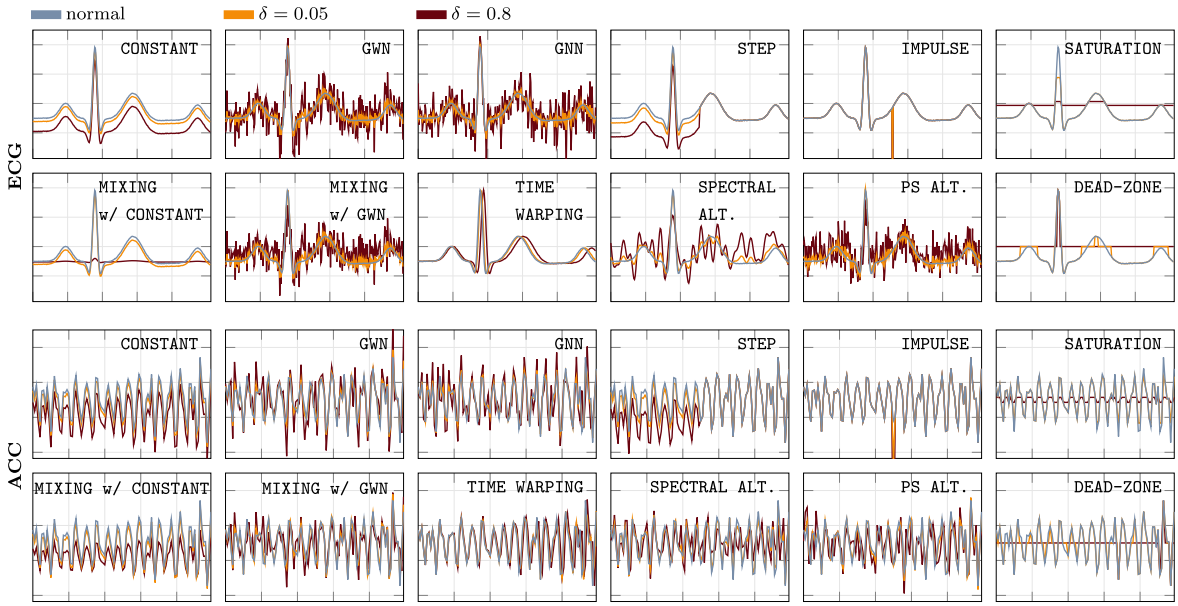


Fig. 2. Examples of ECG and ACC anomalies for two different values of deviation $\delta = \{0.05, 0.8\}$.

TABLE IV
PERFORMANCE OF DIFFERENT DETECTORS (COLUMNS) IN TERMS OF P_D, WORKING ON ECG ANOMALIES (ROWS) WITH A FIXED DEVIATION $\delta = 0.05$

		PCA-BASED				GD-BASED			ML-BASED			FEATURE-BASED			
	ANOMALY	SPE_{52}	SPE_{59}	T^2_{52}	T^2_{59}	AR_4	AR_{16}	MD	$OC_{poly,\,0.01}$	LOF_5	IF_{250}	energy	TV	ZC	pk-pk
INCREASING	GWN	1.00	1.00	0.55	0.63	1.00	1.00	1.00	0.50	0.98	0.59	0.56	1.00	1.00	0.70
	IMPULSE	1.00	1.00	0.54	0.62	1.00	1.00	1.00	0.50	0.98	0.54	0.56	0.76	0.66	0.76
	STEP	0.54	0.60	0.65	0.63	0.96	0.99	0.99	0.50	0.89	0.58	0.56	0.51	0.53	0.59
	CONSTANT	0.50	0.50	0.52	0.51	0.50	0.51	0.51	0.50	0.63	0.56	0.55	0.50	0.54	0.50
	GNN	0.93	0.92	0.55	0.59	0.93	0.93	0.93	0.50	0.97	0.59	0.56	0.95	0.94	0.67
INVARIANT	MIXING w/ GWN	1.00	1.00	0.50	0.59	1.00	1.00	1.00	0.59	0.98	0.55	0.50	1.00	1.00	0.62
	MIXING w/ CONSTANT	0.52	0.51	0.53	0.53	0.56	0.56	0.57	0.59	0.62	0.52	0.50	0.54	0.54	0.60
	SPECTRAL ALT.	0.50	0.50	0.62	0.60	0.59	0.55	0.55	0.59	0.97	0.54	0.50	0.73	0.80	0.51
	PRINCIPAL SUBSPACE ALT.	1.00	1.00	0.50	0.58	1.00	1.00	1.00	0.59	0.98	0.55	0.50	1.00	1.00	0.61
	TIME WARPING	0.52	0.52	0.51	0.51	0.70	0.73	0.72	0.50	0.52	0.50	0.50	0.51	0.51	0.50
DEC.	SATURATION	0.86	0.99	0.79	0.81	1.00	1.00	1.00	0.83	0.81	0.61	0.70	0.69	0.50	0.97
	DEAD-ZONE	0.92	0.98	0.50	0.55	1.00	1.00	1.00	0.73	0.93	0.53	0.56	0.50	0.86	0.50

assessment during the testing phase and the remaining vectors are employed during the training phase.[4]

2) ACC Signals: Acceleration signals utilized in this work have been collected as a result of the Structural Health Monitoring of a viaduct located in Italy [28], [29]. The entire dataset comprises readings of 90 three-axis accelerometers acquired at a sampling rate of 100 sample/s and which represent the elastic response of the structure to vehicles transit and environmental stimuli. Focusing only on a single axis of a particular sensor, we randomly split the readings into 3.77×10^5 windows of length $n = 100$ to create vectors $\boldsymbol{x}^{\text{ok}} \in \mathbb{R}^n$. The intrinsic noise of the dataset leads to $\bar{n} = 16$. Once again, 10^4 vectors have been allocated to the performance assessment, and the rest are dedicated to the training phase.

V. RESULTS

For every class of signals and type of anomaly, we obtain anomalous instances by corrupting the original input vectors as described in Section III. Considering the two use cases, Fig. 2 shows the effect of each type of anomaly on both a normal ECG instance (top) and a normal ACC instance (bottom). Each subfigure corresponds to a different anomaly type and displays both the normal ECG or ACC instance and the resultant anomalous signals for two levels of δ.[5] As expected, a higher deviation leads to a more pronounced anomaly.

Original and corrupted vectors have been used to test several detectors with two different levels of deviation. Results expressed in terms of P_D for each defined anomaly[6] are in Tables IV–VII where each column corresponds to a different detector. Achieved performance may identify the most effective

[4]In Section V of the supplementary material, we discuss how the complexity of the proposed methodology is affected by the number of training and test instances.

[5]The first section of the supplementary material shows examples of corrupted ECG/ACC with $\delta = \{0.05, 0.1, 0.3, 0.8\}$.

[6]For spectral alteration and principal subspace alteration, P_D is computed as the median value over 100 different realizations of $\boldsymbol{\ell}^{\text{ok}}$ and $\boldsymbol{U}^{\text{ko}}$, respectively.

TABLE V
PERFORMANCE OF DIFFERENT DETECTORS (COLUMNS) IN TERMS OF P_D, WORKING ON ECG ANOMALIES (ROWS) WITH A FIXED DEVIATION $\delta = 0.8$

		PCA-BASED				GD-BASED			ML-BASED			FEATURE-BASED			
	ANOMALY	SPE_{52}	SPE_{59}	T^2_{52}	T^2_{59}	AR_4	AR_{16}	MD	$OC_{poly,\,0.01}$	LOF_5	IF_{250}	energy	TV	ZC	pk-pk
INCREASING	GWN	1.00	1.00	0.94	1.00	1.00	1.00	1.00	0.51	1.00	0.99	0.99	1.00	1.00	0.99
	IMPULSE	1.00	1.00	0.93	1.00	1.00	1.00	1.00	0.53	1.00	0.55	0.99	1.00	0.66	1.00
	STEP	0.90	0.98	1.00	1.00	1.00	1.00	1.00	0.52	1.00	0.98	0.96	0.56	0.82	0.81
	CONSTANT	0.50	0.50	0.75	0.71	0.55	0.66	0.59	0.51	0.99	0.96	0.94	0.50	0.89	0.50
	GNN	0.94	0.93	0.70	0.74	0.96	0.96	0.97	0.50	1.00	0.99	0.97	0.99	0.97	0.96
INVARIANT	MIXING w/ GWN	1.00	1.00	0.52	0.96	1.00	1.00	1.00	1.00	1.00	0.85	0.51	1.00	1.00	0.52
	MIXING w/ CONSTANT	0.79	0.77	0.91	0.91	1.00	1.00	1.00	1.00	0.98	0.77	0.50	0.94	0.97	1.00
	SPECTRAL ALT.	0.52	0.51	0.96	0.94	0.91	0.81	0.85	0.99	1.00	0.73	0.53	1.00	1.00	0.72
	PRINCIPAL SUBSPACE ALT.	1.00	1.00	0.50	0.93	1.00	1.00	1.00	1.00	1.00	0.80	0.50	1.00	1.00	0.50
	TIME WARPING	0.59	0.58	0.54	0.55	0.77	0.79	0.78	0.50	0.57	0.50	0.50	0.54	0.56	0.50
DEC.	SATURATION	0.86	0.58	1.00	1.00	0.72	0.87	0.87	1.00	1.00	1.00	1.00	1.00	0.50	1.00
	DEAD-ZONE	1.00	1.00	0.99	0.61	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99

TABLE VI
PERFORMANCE OF DIFFERENT DETECTORS (COLUMNS) IN TERMS OF P_D, WORKING ON ACC ANOMALIES (ROWS) WITH A FIXED DEVIATION $\delta = 0.05$

		PCA-BASED				GD-BASED			ML-BASED			FEATURE-BASED			
	ANOMALY	SPE_{16}	SPE_{85}	T^2_{16}	T^2_{85}	AR_4	AR_{16}	MD	$OC_{RBF,\,0.01}$	LOF_5	IF_{500}	energy	TV	ZC	pk-pk
INCREASING	GWN	0.80	0.98	0.55	0.87	0.76	0.93	1.00	0.64	0.85	0.59	0.58	0.59	0.57	0.60
	IMPULSE	0.80	0.99	0.55	0.88	0.76	0.94	1.00	0.64	0.86	0.55	0.58	0.54	0.51	0.73
	STEP	0.82	0.97	0.50	0.91	0.85	0.90	0.93	0.64	0.86	0.59	0.58	0.50	0.70	0.54
	CONSTANT	0.81	0.51	0.50	0.84	0.85	0.90	0.83	0.64	0.84	0.59	0.58	0.50	0.72	0.50
	GNN	0.78	0.80	0.54	0.83	0.71	0.85	0.88	0.63	0.84	0.58	0.58	0.58	0.55	0.59
INVARIANT	MIXING w/ GWN	0.79	0.98	0.54	0.87	0.75	0.93	1.00	0.63	0.85	0.58	0.58	0.58	0.57	0.60
	MIXING w/ CONSTANT	0.81	0.51	0.51	0.84	0.85	0.90	0.83	0.64	0.84	0.58	0.58	0.51	0.72	0.51
	SPECTRAL ALT.	0.50	0.50	0.51	0.51	0.50	0.51	0.51	0.50	0.53	0.50	0.50	0.50	0.52	0.50
	PRINCIPAL SUBSPACE ALT.	0.58	0.83	0.50	0.63	0.55	0.72	0.94	0.51	0.60	0.50	0.50	0.50	0.51	0.51
	TIME WARPING	0.50	0.50	0.50	0.50	0.50	0.51	0.51	0.50	0.51	0.50	0.50	0.50	0.51	0.50
DEC.	SATURATION	0.54	0.69	0.56	0.53	0.55	0.59	0.90	0.56	0.54	0.55	0.55	0.55	0.50	0.68
	DEAD-ZONE	0.58	0.81	0.50	0.62	0.55	0.72	0.93	0.50	0.61	0.51	0.51	0.51	0.82	0.50

detector for each type of anomaly, thus validating the tool under scrutiny at design time.

In detail, Tables IV and V provide an overview of the detectability in case of ECG signals with a deviation level of $\delta = 0.05$ for Table IV and $\delta = 0.8$ for Table V. From these results, it is evident how each type of anomaly has a different effect on the normal signal, leading to different detection performances. In addition, some anomalies such as constant and time warping, are difficult to detect using any of the considered detectors.

Moreover, comparing the detection performance in Tables IV and V, as expected, increasing δ leads to better detection performance. This confirms the validity of δ as a parameter for generating anomalies with varying levels of severity.[7]

The detection results for ACC signals are summarized in Tables VI and VII for $\delta = 0.05$ and $\delta = 0.8$, respectively. As anticipated, the detection performances for the ACC signals are significantly different from those for ECG signals. For the same level of δ, P_D values are generally lower compared to the ECG case. Moreover, spectral alterations, time warping, and dead-zone are found to be the most challenging anomalies to detect in the ACC case. The results also indicate that feature-based detectors have inferior detection capabilities compared to more complex solutions.

To highlight the framework's approach in assessing a detector's performance within a specific signal class, we stress that the results in Tables IV–VII can be read column wise. Within each column, the values summarize the anticipated detection performance of the respective detector.

Fig. 3 further demonstrates the capability of our tool to predict the performance of detectors working on real-world anomalies.[8] On the left-hand side, we present four different detector score profiles (bottom), which were trained on a reference synthetic ECG signal. These scores vary based on the changes observed in a real-world ECG time series (top) that exhibits some anomalies [50]. While all four detectors can identify the anomaly

[7]This claim is confirmed in the second section of the supplementary material, where P_D for each detector/anomaly is plotted against δ.

[8]An extended version of Fig. 3 including the scores of all detectors can be found in the third section of the supplementary material.

TABLE VII
PERFORMANCE OF DIFFERENT DETECTORS (COLUMNS) IN TERMS OF P_D, WORKING ON ACC ANOMALIES (ROWS) WITH A FIXED DEVIATION $\delta = 0.8$

		PCA-BASED				GD-BASED			ML-BASED			FEATURE-BASED			
	ANOMALY	SPE_{16}	SPE_{85}	T^2_{16}	T^2_{85}	AR_4	AR_{16}	MD	$OC_{RBF,\,0.01}$	LOF_5	IF_{500}	energy	TV	ZC	pk-pk
INCREASING	GWN	0.96	1.00	0.76	0.99	0.97	1.00	1.00	0.89	0.99	0.81	0.80	0.82	0.67	0.83
	IMPULSE	0.96	1.00	0.77	1.00	0.98	1.00	1.00	0.89	0.99	0.58	0.80	0.64	0.51	0.95
	STEP	0.97	1.00	0.52	1.00	1.00	1.00	1.00	0.90	0.99	0.82	0.80	0.51	0.92	0.69
	CONSTANT	0.97	0.69	0.51	0.98	1.00	1.00	0.98	0.90	0.99	0.83	0.80	0.50	0.93	0.50
	GNN	0.96	0.89	0.67	0.97	0.93	0.97	0.98	0.88	0.98	0.80	0.79	0.77	0.57	0.81
INVARIANT	MIXING w/ GWN	0.96	1.00	0.70	0.99	0.96	1.00	1.00	0.86	0.99	0.77	0.76	0.77	0.69	0.79
	MIXING w/ CONSTANT	0.96	0.58	0.63	0.97	1.00	0.99	0.97	0.87	1.00	0.79	0.76	0.64	0.96	0.65
	SPECTRAL ALT.	0.56	0.51	0.63	0.57	0.55	0.58	0.57	0.56	0.74	0.52	0.51	0.53	0.66	0.53
	PRINCIPAL SUBSPACE ALT.	0.78	0.98	0.54	0.87	0.73	0.93	0.99	0.59	0.93	0.51	0.50	0.52	0.65	0.53
	TIME WARPING	0.61	0.50	0.52	0.59	0.51	0.50	0.57	0.52	0.67	0.50	0.50	0.51	0.54	0.50
DEC.	SATURATION	0.89	0.69	0.94	0.88	0.90	0.74	0.65	0.93	0.83	0.92	0.92	0.94	0.50	0.98
	DEAD-ZONE	0.57	0.91	0.72	0.66	0.50	0.81	0.98	0.63	0.77	0.78	0.71	0.87	1.00	0.51

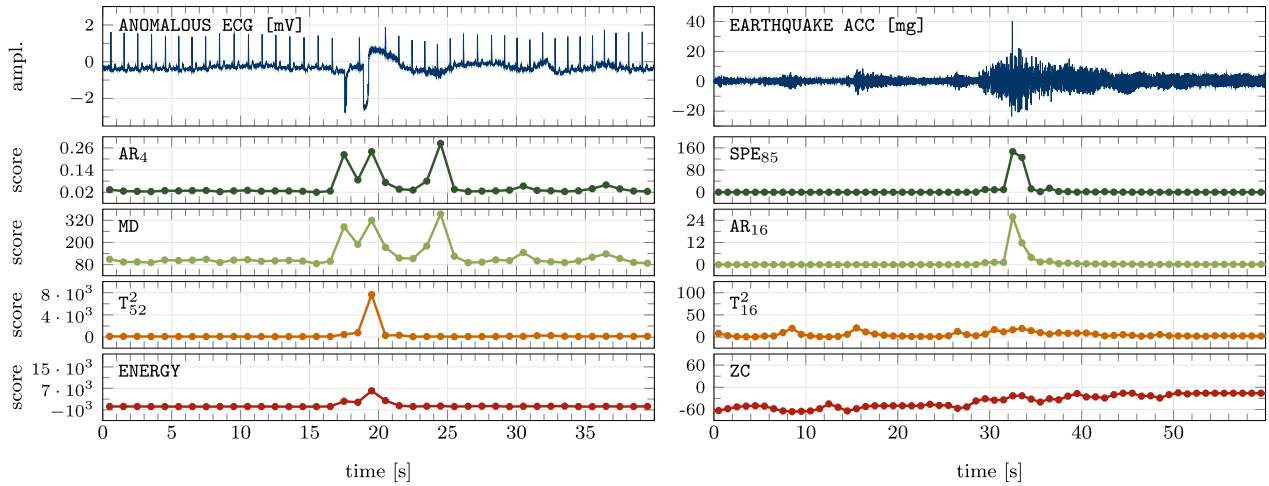


Fig. 3. Scores trends (bottom) for four detectors working on anomalous real ECG time-series (right) and for four detectors applied on an ACC time series (left) containing an earthquake event.

that occurs in the time range of 16–22 s, T^2_{52} and energy-based detector fail to detect a noise increase that does not alter signal's power between 23 and 26 s. Even when testing on a real ECG signal, this outcome could have been foreseen by referring to Table IV. Focusing on the identification of mixing with GWN, AR_4, and MD are ranked as the two most effective detectors, while T^2_{52} and the energy-based detector are ranked as the two least effective.

On the right-hand side of Fig. 3, we also demonstrate how Table VII can anticipate the performance of different detectors in identifying an earthquake occurring in the time interval of 28–38 s. As an earthquake tends to increase the energy of all signal components, it can be considered a GWN anomaly. Looking at Table VII, it is evident that SPE_{85} and AR_{16} detectors are better suited for detecting this anomaly, while T^2_{16} and ZC fail at the task. This result is consistent with the plot in the bottom right-hand side of Fig. 3. In fact, the score trends of SPE_{85} and AR_{16} show a sharper variation in correspondence with the earthquake, while the scores of T^2_{16} and ZC do not exhibit significant variations compared to their normal trends. Interestingly, in both scenarios, the best detector is not an ML-based one but rather a less computationally demanding solution.

VI. CONCLUSION

The assessment of the performance of anomaly detectors is often affected by the scarcity of anomalous data. In this article, we presented a framework allowing to systematically assess an anomaly detector working on time series. This tool used a dictionary of anomalies modeled as alterations to the normal signal that reflect nonnormal behaviors in the monitored system or acquisition chain, such as aging, wearing, abrupt changes in conditions, or hardware failures and disturbances. These models were characterized by a parameter that controls the deviation of the anomaly from normality, giving control of how difficult it is to detect the anomaly itself.

The effectiveness of the proposed framework was shown in two different real-world scenarios: human and structural health monitoring through the analysis of ECG and acceleration signals. The framework can assess different candidate detectors highlighting the kind of anomalies for which they perform best and worst. Empirical evidence showed that our framework was able to anticipate the performance on real-world anomalies, such as artifacts in ECG signals and an earthquake in the health monitoring of a bridge.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, Jul. 2009, pp. 1–58.

[2] J. Ding, V. Tarokh, and Y. Yang, "Model selection techniques: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 6, pp. 16–34, Nov. 2018.

[3] M. Goswami, C. I. Challu, L. Callot, L. Minorics, and A. Kan, "Unsupervised model selection for time series anomaly detection," in *Proc. 11th Int. Conf. Learn. Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=HJz6tiCqYm

[4] L. Ruff et al., "A unifying review of deep and shallow anomaly detection," *Proc. IEEE*, vol. 109, no. 5, pp. 756–795, May 2021.

[5] S. Schmidl, P. Wenig, and T. Papenbrock, "Anomaly detection in time series: A comprehensive evaluation," *Proc. VLDB Endowment*, vol. 15, no. 9, pp. 1779–1797, May 2022.

[6] R. Wu and E. J. Keogh, "Current time series anomaly detection benchmarks are flawed and are creating the illusion of progress," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 2421–2429, Mar. 2023.

[7] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," in *Proc. Int. Conf. Learn. Representations*, 2019.

[8] A. Sharma, L. Golubchik, and R. Govindan, "On the prevalence of sensor faults in real-world deployments," in *Proc. 4th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, 2007, pp. 213–222.

[9] K. Ni et al., "Sensor network data fault types," *ACM Trans. Sens. Netw.*, vol. 5, no. 3, pp. 1–29, 2009.

[10] A. B. Sharma, L. Golubchik, and R. Govindan, "Sensor faults: Detection methods and prevalence in real-world datasets," *ACM Trans. Sens. Netw.*, vol. 6, no. 3, pp. 1–39, 2010.

[11] S. U. Jan, Y.-D. Lee, J. Shin, and I. Koo, "Sensor fault classification based on support vector machine and statistical time-domain features," *IEEE Access*, vol. 5, pp. 8682–8690, 2017.

[12] K.-H. Lai, D. Zha, J. Xu, Y. Zhao, G. Wang, and X. Hu, "Revisiting time series outlier detection: Definitions and benchmarks," in *Proc. 35th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track (Round 1)*, 2021 [Online]. Available: https://openreview.net/forum?id=r8IvOsnHchr

[13] G. Steinbuss and K. Böhm, "Benchmarking unsupervised outlier detection with realistic synthetic data," *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 4, pp. 1–20, Apr. 2021.

[14] S. Han, X. Hu, H. Huang, M. Jiang, and Y. Zhao, "ADBench: Anomaly detection benchmark," in *Proc. 36th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track*, 2022. [Online]. Available: https://openreview.net/forum?id=foA_SFQ9zo0

[15] J. Paparrizos, Y. Kang, P. Boniol, R. S. Tsay, T. Palpanas, and M. J. Franklin, "TSB-UAD: An end-to-end benchmark suite for univariate time-series anomaly detection," *Proc. VLDB Endowment*, vol. 15, no. 8, pp. 1697–1711, Apr. 2022.

[16] A. Enttsel, F. Martinini, A. Marchioni, M. Mangia, R. Rovatti, and G. Setti, "Second-order statistic deviation to model anomalies in the design of unsupervised detectors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2023, pp. 1–5.

[17] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *Proc. 12th Int. Conf. Neural Inf. Process. Syst.*, Cambridge, MA, USA, 1999, pp. 582–588.

[18] H. Q. Zhang and Y. Yan, "A wavelet-based approach to abrupt fault detection and diagnosis of sensors," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 5, pp. 1389–1396, Oct. 2001.

[19] Q. K. Dang and Y. S. Suh, "Sensor saturation compensated smoothing algorithm for inertial sensor based motion tracking," *Sensors*, vol. 14, no. 5, pp. 8167–8188, 2014.

[20] C. Huang, B. Shen, L. Zou, and Y. Shen, "Event-triggering state and fault estimation for a class of nonlinear systems subject to sensor saturations," *Sensors*, vol. 21, no. 4, 2021, Art. no. 1242.

[21] N. Ramanathan et al., "The final frontier: Embedding networked sensors in the soil," *UCLA: Center for Embedded Netw. Sens.*, 2006.

[22] S. R. Moreno, L. d. S. Coelho, H. V. Ayala, and V. C. Mariani, "Wind turbines anomaly detection based on power curves and ensemble learning," *IET Renewable Power Gener.*, vol. 14, no. 19, pp. 4086–4093, 2020.

[23] S. Wenndt and A. Noga, "Narrow-band interference cancellation for enhanced speaker identification," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoustics.*, 1999, pp. 123–126.

[24] D. Liang, N. Sun, Y. Wu, G. Liu, and Y. Fang, "Fuzzy-sliding mode control for humanoid arm robots actuated by pneumatic artificial muscles with unidirectional inputs, saturations, and dead zones," *IEEE Trans. Ind. Inform.*, vol. 18, no. 5, pp. 3011–3021, May 2022.

[25] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, "A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life," *IEEE Trans. Ind. Inform.*, vol. 17, no. 5, pp. 3478–3487, May 2021.

[26] K. Liu, Y. Shang, Q. Ouyang, and W. D. Widanage, "A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3170–3180, Apr. 2021.

[27] E. Karakose, M. T. Gencoglu, M. Karakose, I. Aydin, and E. Akin, "A new experimental approach using image processing-based tracking for an efficient fault diagnosis in pantograph–catenary systems," *IEEE Trans. Ind. Inform.*, vol. 13, no. 2, pp. 635–643, Apr. 2017.

[28] A. Burrello, A. Marchioni, D. Brunelli, S. Benatti, M. Mangia, and L. Benini, "Embedded streaming principal components analysis for network load reduction in structural health monitoring," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4433–4447, Mar. 2021.

[29] A. Marchioni, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Subspace energy monitoring for anomaly detection, sensor or, edge," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7575–7589, Aug. 2020.

[30] H. Li and P. Boulanger, "A survey of heart anomaly detection using ambulatory electrocardiogram (ECG)," *Sensors*, vol. 20, no. 5, 2020, Art. no. 1461.

[31] J. Bonnel, A. Thode, D. Wright, and R. Chapman, "Nonlinear time-warping made simple: A step-by-step tutorial on underwater acoustic modal separation with a single hydrophone," *J. Acoustical Soc. Amer.*, vol. 147, no. 3, pp. 1897–1926, 2020.

[32] R. K. Sharma and J. W. Wallace, "Improved spectrum sensing by utilizing signal autocorrelation," in *VTC Spring IEEE 69th Veh. Technol. Conf.*, 2009, pp. 1–5.

[33] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[34] M. Gavish and D. L. Donoho, "The optimal hard threshold for singular values is $4/\sqrt{3}$," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 5040–5053, Aug. 2014.

[35] J. Ginibre, "Statistical ensembles of complex, quaternion, and real matrices," *J. Math. Phys.*, vol. 6, no. 3, pp. 440–449, 1965.

[36] C. C. Aggarwal, *Outlier Analysis*. Berlin, Germany, Springer 2017.

[37] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[38] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D, Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[39] I. T. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer, 2002.

[40] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

[41] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," *SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, 2000.

[42] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, 2008, pp. 413–422.

[43] G. Williams, R. Baxter, H. He, S. Hawkins, and L. Gu, "A comparative study of RNN for outlier detection in data mining," in *Proc. IEEE Int. Conf. Data Mining*, 2002, pp. 709–712.

[44] C. L. Giles, G. M. Kuhn, and R. J. Williams, "Dynamic recurrent neural networks: Theory and applications," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 153–156, Mar. 1994.

[45] L. Ruff et al., "Deep one-class classification," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.

[46] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep isolation forest for anomaly detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12591–12604, Dec. 2023.

[47] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," in *Int. Conf. Learn. Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=LzQQ89U1qm_

[48] P. McSharry, G. Clifford, L. Tarassenko, and L. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, Mar. 2003.

[49] M. Mangia, R. Rovatti, and G. Setti, "Rakeness in the design of analog-to-information conversion of sparse and localized signals," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 5, pp. 1001–1014, May 2012.

[50] 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. E215–20, 2000.

Andriy Enttsel (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees (with Honors) in electronic engineering, in 2018 and 2021, respectively, from the University of Bologna, Bologna, Italy, where he is currently working toward the Ph.D. degree in electronics, telecommunications, and information technologies engineering with the Statistical Signal Processing Group.

His main research interests include signal processing, anomaly detection, and information theory.

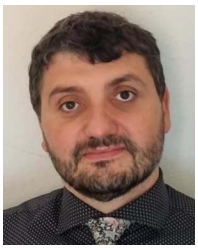
Silvia Onofri (Graduate Student Member, IEEE) received the M.Sc. degree in electronic engineering, in 2022, from the University of Bologna, Bologna, Italy, where she is currently working toward the Ph.D. degree in electronics, telecommunications, and information technologies engineering with the Statistical Signal Processing Group.

Her research interests include anomaly detection and predictive maintenance.

Dr. Onofri was a recipient of the Graduate Student Best Paper Award at the IEEE International Instrumentation and Measurement Technology Conference (I2MTC) 2024.

Alex Marchioni (Member, IEEE) received the B.Sc. and M.Sc. degrees (Hons.) in electronic engineering and the Ph.D. degree in electronic, telecommunication, and information technology from the University of Bologna, Bologna, Italy, in 2011, 2015, and 2022, respectively.

He is currently a Research Fellow with the Department of Electrical, Electronic, and Information Engineering (DEI), University of Bologna. He is also a Member of the Statistical Signal Processing Group, University of Bologna. His research interests include signal processing, machine learning, anomaly detection, compressed sensing, Internet of Things, and Big Data analytics.



Mauro Mangia (Member, IEEE) received the B.Sc. and M.Sc. degrees in electronic engineering and the Ph.D. degree in information technology from the University of Bologna, Bologna, Italy, in 2005, 2009, and 2013, respectively.

He was a Visiting Ph.D. student with the Ecole Polytechnique Federale de Lausanne in 2009 and 2012. He is currently an Assistant Professor with the Statistical Signal Processing Group, Department of Electrical, Electronic and Information Engineering, University of Bologna. He is also a Member of both the Advance Research Center for Electronic Systems (ARCES), University of Bologna and Alma Mater Research Institute for Human-Centered Artificial Intelligence, University of Bologna. His research interests include nonlinear systems, explainable artificial intelligence, machine learning and artificial intelligence, anomaly detection, Internet of Things, Big Data analytics, and optimization.

Dr. Mangia was the recipient of the 2013 IEEE CAS Society Guillemin-Cauer Award and of the 2019 IEEE BioCAS Transactions Best Paper Award. He was also the recipient of the Best Student Paper Award at International Symposium On Circuits and Systems (ISCAS) 2011. He was the Web and Social Media Chair for ISCAS2018.

Gianluca Setti (Fellow, IEEE) received Dr. Eng. degree (honors) and a Ph.D. degree in electronic engineering from the University of Bologna, in 1992 and in 1997, respectively.

He was with the University of Ferrara (1997–2017) and Politecnico di Torino (2017–2022) as a Professor of electronics, signal and data processing. He held also several positions as Visiting Professor/Scientist with École Polytechnique Fédérale de Lausanne, in 2002 and 2005, University of California San Diego in 2004, IBM in 2004 and 2007, and the University of Washington in 2008 and 2010. He is the Dean of the Computer, Electrical, Mathematical Sciences and Engineering Division and a Professor of electrical and computer engineering with the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. His research interests include recurrent neural networks, electromagnetic compatibility, compressive sensing and statistical signal processing, biomedical circuits and systems, power electronics, Internet of Things, circuits and systems for machine learning, and applications of artificial intelligence techniques for anomaly detection and predictive maintenance.

Dr. Setti served as the Editor-in-Chief for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—Part II (2006–2007) and of the IEEE Transactions on Circuits and Systems—Part I (2008–2009). Since 2019, he has been the first non US Editor-in-Chief of the Proceedings of the IEEE, the flagship journal of the IEEE. In 2010, he served as IEEE CAS Society President. In 2013–2014, he was the first non North-American Vice President of the IEEE for Publication Services and Products. He was the recipient of several awards, including the 2004 IEEE CAS Society Darlington Award, 2013 IEEE CAS Society Meritorious Service Award, 2013 IEEE CAS Society Guillemin-Cauer Award, and the 2019 IEEE Transactions on Circuits and Systems Best Paper Award.

Riccardo Rovatti (Fellow, IEEE) received the M.S. degree in electronic engineering and the Ph.D. degree in electronics, computer science, and telecommunications from the University of Bologna, Bologna, Italy, in 1992 and 1996, respectively.

He is currently a Full Professor of electronics with the University of Bologna. He has authored more than 300 technical contributions to international conferences and journals and two volumes. His research interests include mathematical and applicative aspects of statistical signal processing, on machine learning for signal processing, and on the application of statistics to nonlinear dynamical systems.

Dr. Rovatti was a Distinguished Lecturer of the IEEE CAS Society for the years 2017–2018. He was the recipient of the 2004 IEEE CAS Society Darlington Award, the 2013 IEEE CAS Society Guillemin-Cauer Award, and the 2019 IEEE BioCAS Transactions Best Paper Award. He was the recipient of the Best Paper Award at the European Conference on Circuit Theory and Design 2005 and the Best Student Paper Award at the EMC Zurich 2005 and International Symposium On Circuits and Systems (ISCAS) 2011. He is an IEEE fellow for his contribution to nonlinear and statistical signal processing applied to electronic systems.

Open Access funding provided by 'Alma Mater Studiorum - Università di Bologna' within the CRUI CARE Agreement