



Detection and Early Detection of Pathological Motor Evoked Potentials: A Novel Approach for Intraoperative Neuromonitoring Assistance

Alberto Azzari¹ · Alessandro Boaro² · Federica Basaldella³ · Sonia Nunes² · Francesco Sala² · Carlo Combi¹ · Pietro Sala¹ · Manuele Bicego¹

Received: 6 June 2025 / Revised: 18 June 2026 / Accepted: 18 June 2026 /

Published online: 10 July 2026

© The Author(s) 2026

Abstract

Intraoperative monitoring of motor evoked potentials (MEPs) plays a crucial role in detecting potential neurological complications during surgical procedures. Traditional methods rely on real-time human interpretation, which can introduce delays or, worse, errors in response. This study proposes an early detection framework for pathological MEPs, leveraging an anomaly detection approach based on an extension of isolation forests—one of the most effective anomaly detection models. By modeling the temporal evolution of MEPs, our approach not only identifies pathological deviations but also anticipates their occurrence, enabling proactive intervention. We validated our methodology on a dataset of intraoperative MEP recordings, demonstrating improved detection performance and enhanced early warning capabilities compared to conventional approaches. The findings suggest that this framework may support real-time intraoperative monitoring and serve as a foundation for future decision-support tools, although prospective clinical validation is still required before any deployment can be considered.

Keywords Isolation Forest · Ensemble Learning · Intraoperative monitoring · Surgery · Neurosurgery

1 Introduction

Intraoperative monitoring (IOM) of motor evoked potentials (MEPs) enables real-time assessment of the functional integrity of motor pathways. It has become a valuable tool in neurosurgery, orthopedics, and vascular surgery procedures [1, 2]. Minimizing the risk of disabling motor deficits is of primary importance during these surgeries. However, achieving optimal outcomes without compromising the patient's

Extended author information available on the last page of the article

functional status presents a major challenge. MEP alterations should serve as critical warning signs, prompting the surgical team to intervene or pause to prevent irreversible injury to the nervous system. Established warning criteria include MEP disappearance, amplitude reduction, threshold elevation, and morphology simplification [3]. Unlike D waves, which have stable properties and limited clinical indications, muscle MEPs are widely used but inherently polysynaptic, nonlinear, and highly variable. This variability complicates the establishment of robust warning criteria and makes MEPs unusually sensitive to intraoperative physiological changes [4]. While there is consensus on the clinical interpretation of MEP signal alterations, cut-off values, such as a reduction in amplitude greater than 50% during brain surgery or 80% during spinal cord surgery, are often derived empirically, leading to potential inconsistencies in clinical decision-making [5].

Recent advancements in biomedical data analysis, particularly in machine learning and deep learning, have demonstrated their potential to extract and summarize clinically meaningful information across medical imaging, histology data, and biological signals [6–8]. Specifically, several machine learning techniques have been successfully applied to classify cardiac arrhythmias from electrocardiograms (EKG), detect seizures from electroencephalograms (EEG), analyze muscular activity from surface electromyography (sEMG) for robotic prosthesis control, and recently, for the accurate characterization of MEPs based on the muscles that generate them [9–13]. These approaches, when adapted to intraoperative MEP analysis, could refine current warning criteria by unveiling new, actionable signal features, ultimately strengthening their predictive and diagnostic value [14].

One of the key unexplored challenges in intraoperative neurophysiology is the real-time detection of anomalous motor evoked potentials, which plays a crucial role in surgical guidance. Identifying pathological MEP patterns can provide immediate warnings to the surgeon, allowing for timely intervention when abnormal events occur. Unlike traditional time series anomaly detection models, which primarily focus on point, contextual, or global anomalies [15], this problem involves detecting abnormal time series patterns, as defined in [16]. However, there is a significant gap in existing models capable of handling this specific type of anomaly.

In clinical practice, MEP monitoring naturally generates a sequence of discrete evoked responses: following each stimulation, a complete MEP waveform is recorded over a short time window and becomes immediately available to the neurophysiologist. Consequently, the real-time monitoring task consists of assessing each newly acquired MEP waveform individually, as soon as it is recorded, to determine whether it deviates from expected physiological behavior. This approach is limited by the potential error in manual amplitude measurements, the time required to perform them, and fatigue during long procedures [17, 18].

Furthermore, explainability is a critical requirement in this domain, as highlighted in [14]. Deep learning models, while powerful, were unsuitable for our approach due to both the need for interpretability and the challenges associated with training deep networks on small datasets. Instead, we leveraged insights from [12], where

ensemble-based methods, particularly Random Forests [19], demonstrated strong performance in classifying MEPs. Building on this evidence, we propose a novel forest-based anomaly detection model tailored to the unique constraints and requirements of intraoperative MEP monitoring.

In this paper, we propose the Canonical Isolation Forest (CIsoF), an extension of Isolation Forest [20] specifically designed for anomaly detection on time series. This model isolates MEPs effectively by leveraging both the temporal structure and statistical properties of these biomedical signals. Our method computes CATCH22 time series features [21], a comprehensive set of statistical and structural descriptors on random intervals of each time series. This effectively captures the characteristics of MEP signals. Additionally, our method is patient-specific, meaning that the model is trained and tested on the same patient, allowing it to account for individual variability in MEP patterns. This patient-specific design ensures robust detection even with the small datasets typical of intraoperative settings. Furthermore, this is the first anomaly detection forest specifically designed for identifying abnormal time series patterns, addressing a gap in existing methods [15]. From an algorithmic perspective, CIsoF continuously assigns a continuous *anomaly score* to each incoming MEP waveform, quantifying its deviation from a patient-specific baseline distribution established at the beginning of the surgical procedure. This formulation naturally leads to an unsupervised anomaly detection setting, which, together with the limited availability of patient data and the clinical requirement for explainable decision mechanisms, motivates the use of forest-based approaches over recent deep learning models in this context. Finally, in addition to standard detection, where the goal is to flag the current MEP as anomalous, we also consider an *early detection* setting. In this scenario, the objective is to provide an anticipatory warning before a clinically confirmed pathological MEP occurs. This is evaluated using the same anomaly scoring mechanism by analyzing whether MEPs immediately preceding a pathological event already exhibit elevated anomaly scores. The rest of the paper is structured as follows. Section 2 discusses related work, including the application of machine learning in intraoperative monitoring and time series anomaly detection, as well as existing forest-based methods for time series. Section 3 presents our proposed method, including the medical problem and its machine-learning formalization, the detection tasks, and details on the Canonical Isolation Tree and Forest. Section 4 provides an extensive experimental evaluation, analyzing different variants of our model, its detection capabilities, and comparisons with alternative methods. Finally, Section 6 concludes the paper, summarizing key findings and potential future directions.

2 Related Work

2.1 Machine Learning in Intraoperative Monitoring

The integration of machine learning (ML) into intraoperative monitoring (IOM) has significantly advanced the ability to analyze and interpret neurophysiological

data. Traditional methods for monitoring motor evoked potentials (MEPs) rely on empirical thresholds and expert interpretation, which can lead to inconsistencies in decision-making. ML-based approaches offer the potential to enhance the accuracy and robustness of MEP analysis, reducing subjectivity and improving patient safety. A broader review of ML applications in IOM [14] emphasized the role of explainable AI techniques in neurophysiological monitoring. The study underscored the importance of models that not only achieve high accuracy but also provide interpretable decision-making processes, which is essential for clinical adoption. Beyond classification, ML has also been applied to model the effects of anesthetic agents on somatosensory evoked potentials (SSEPs) [22]. By leveraging signal processing and ML techniques, researchers have characterized and predicted SSEP waveform changes under different anesthetic conditions. This suggests that ML models could similarly be employed to predict intraoperative MEP variations and improve anomaly detection. Recent studies have demonstrated the feasibility of ML models for intraoperative MEP classification [12, 23]. For instance, Boaro et al. in [12] developed a machine learning framework that successfully classified MEPs based on the muscles that generated them, surpassing expert-level performance. This work highlights the potential of ML models to assist the surgical team in interpreting MEP alterations more objectively and efficiently while maintaining explainability and reducing complexity. The work of Boaro et al. [12] identified RF as the best-performing model for MEP classification, demonstrating its ability to capture complex patterns within neurophysiological signals.

2.2 Time Series Anomaly Detection and Forest-Based Models

2.2.1 Time Series Anomaly Detection

Anomalies in time series data can be broadly categorized into three types: point anomalies, where individual observations deviate significantly from expected values; contextual anomalies, where an observation is anomalous in a specific context but not in others; and collective anomalies, where a group of observations collectively deviates from normal patterns [15]. These anomalies can arise due to equipment failures, environmental changes, or underlying physiological conditions in medical applications. Over the last two decades, the field of time series anomaly detection has been widely explored [24]. This problem can be addressed from five different perspectives: clustering-based, density-estimation-based, distance-based, reconstruction-based, and forecasting-based techniques [15]. Earlier approaches have primarily relied on traditional machine learning [20, 25] and statistical tools [26]. At the same time, more recent methods have explored deep learning (DL) techniques [27, 28], which have been particularly effective in handling high-dimensional time series [16]. However, DL-based methods often require large amounts of training data, making them unsuitable for our problem, where the number of available samples is minimal. In addition to data require-

ments, deep learning approaches typically offer limited interpretability, which represents a critical limitation in intraoperative neurophysiological monitoring [29–31].

Among the major paradigms, distance-, reconstruction-, and forecasting-based methods explicitly account for the temporal structure of time series data. Distance-based approaches detect anomalies by measuring distances between time series. The most commonly used distance is Dynamic Time Warping (DTW) [32]. Reconstruction-based methods, often leveraging autoencoders or principal component analysis, identify anomalies as instances that cannot be effectively reconstructed from a learned representation. Forecasting-based techniques predict future values and flag anomalies when the actual observations significantly deviate from the expected trajectory [15]. Given the constraints of our datasets, we focus on traditional approaches rather than deep learning solutions due to limited data availability.

Recent research has explored streaming and online anomaly detection frameworks for non-stationary data streams, including adaptive tree-based approaches such as STREAMRHF [33], online detectors designed to handle missing values and concept drift [34], and comprehensive surveys of streaming detection paradigms [35]. These methods are designed for continuously evolving data streams in which the model must update incrementally as new observations arrive and no stable training baseline is available. In contrast, the intraoperative setting studied here is characterized by a well-defined, patient-specific baseline acquired before surgery begins, on which the model is trained once and then kept fixed throughout the procedure. This architectural difference makes streaming update mechanisms unnecessary and potentially harmful. Nevertheless, for settings in which a stable pre-surgical baseline is unavailable, for example, in emergent or unplanned procedures, streaming anomaly detection frameworks represent a meaningful direction for future research.

Building on this background, the following subsection narrows the focus from general time series anomaly detection to the specific methodological foundations most relevant to our approach. Section 2.2.2 reviews forest-based models developed for time series classification, such as Time Series Forest and Canonical Interval Forest, which exploit interval-based feature extraction to capture temporal structure. Although these approaches have shown strong performance in supervised settings, there is still a lack of forest-based methods specifically designed for unsupervised anomaly detection in time series. This limitation motivates the development of CISO_F, which combines the interval-based feature extraction philosophy of canonical forests with an isolation-based anomaly detection strategy, adapted to the patient-specific setting required for intraoperative MEP monitoring.

2.2.2 Random Forests for Time Series

In the context of time series classification, forest-based methods have demonstrated strong performance. Time Series Forest (TSF) [36] utilizes intervals of time series

to extract features such as mean, standard deviation, and slope, which are then used in decision trees for classification. Canonical Interval Forest (CIF) [37] extends this idea by employing a broader set of summary statistics, improving robustness across different types of time series data. While these models excel in classification tasks, they are not inherently designed for anomaly detection, as they need labeled datasets to be trained.

Despite the effectiveness of existing forest-based methods for classification, a dedicated forest-based anomaly detection model for time series remains absent. In particular, for intraoperative motor evoked potential (MEP) monitoring, anomalies are not merely statistical outliers but deviations from expected physiological signal evolution. Our work fills this gap by proposing a novel forest-based anomaly detection model specifically tailored for time series data, integrating domain-specific adaptations to improve accuracy in this critical application.

3 Proposed Methods

In this section, we provide a detailed explanation of our approach to intraoperative MEP anomaly detection. First, we formally define the detection and early detection tasks, establishing the problem formulation both from the medical (see Section 3.1) and computational (see Section 3.2) points of view. Then, in Section 3.3, we introduce our Canonical Isolation Forest (CIsoF), which serves as the foundation for our anomaly detection framework. To avoid ambiguity, we emphasize that the proposed approach is an *unsupervised* task and that the problem belongs to anomaly detection. CIsoF assigns a continuous *anomaly score* to each newly acquired MEP. Throughout the paper, the term “decision” refers to this scoring step.

3.1 Medical Problem

During a major surgical procedure, the patient is under general anesthesia, which eliminates voluntary muscle movement. In those types of surgeries in which monitoring the integrity of the neuromuscular system is needed, such as neurosurgery, vascular surgery or thyroid surgery, the neurophysiologist is responsible for continuously assessing the patient’s motor function by stimulating specific regions of the brain or spinal cord and recording the corresponding muscle responses. These responses, known as MEPs, are collected using electrodes placed on the patient’s skin or in the muscles. At the beginning of the procedure, before any surgical intervention takes place, the neurophysiologist establishes a reference measurement by collecting baseline MEP signals. This baseline represents the expected response of the patient’s motor pathways under normal conditions and serves as a point of comparison for subsequent MEP recordings. It is crucial to ensure that the baseline is stable and reflects the patient’s preoperative neurological status. As the surgery progresses, the neurophysiologist periodically collects new MEPs

and compares them to the baseline. The goal of this comparison is to detect any deviations that may indicate potential neurological damage. If the neurophysiologist detects significant changes in the MEP signals, they inform the surgical team. The definition of “significant deviation” mainly refers to signal amplitude reduction. There is no widespread consensus on a specific value, as it varies across studies and depending on the type of surgery, from 20% to complete disappearance. This warning allows the surgical team to take corrective actions, such as temporarily pausing the surgery, modifying patient positioning, adjusting anesthesia parameters, or investigating potential causes of the signal changes. Rapid intervention can help prevent permanent neurological deficits, making intraoperative MEP monitoring a crucial component of patient safety. In the next section (Section 3.2), we will formalize the computational aspects of this detection problem, defining how MEP signals are processed and analyzed algorithmically to achieve robust anomaly detection and early detection. Importantly, intraoperative monitoring produces *discrete* MEP waveforms: after each stimulation, a full MEP trace is acquired within a short time window and becomes immediately available for analysis.

3.2 Machine Learning Formalization

We define a dataset of intraoperative motor evoked potentials (MEPs) as:

$$\mathcal{D} = \{x_1, x_2, \dots, x_N\} \quad (1)$$

where N is the number of MEPs in $\mathcal{D}$. Each x_i represents a MEP signal:

$$x_i = \{v_1^i, v_2^i, \dots, v_{l^i}^i\} \quad (2)$$

where v_j^i denotes the value of the i -th signal at time step j , and l^i denotes the length of the i -th time series.

In our setting, each x_i corresponds to the full waveform of one evoked response (one stimulation), that is, a univariate time series of fixed duration recorded immediately after stimulation.

The dataset $\mathcal{D}$ represents the patient-specific *baseline* and consists of MEP waveforms recorded before a reference time point λ , that is, before the surgical procedure begins (baseline acquisition phase). CISO is trained once on $\mathcal{D}$ and kept fixed for the remainder of the procedure.

In the next part of this section, we will introduce both the Detection and Early Detection tasks.

After λ , a new time series is recorded:

$$x_{\text{TEST}} = \{v_1, v_2, \dots, v_{l_{\text{TEST}}}\} \quad (3)$$

where l^{TEST} is the length of the newly observed MEP time series.

The objective is to determine whether x_{TEST} exhibits anomalous behavior when compared to the learned distribution of normal MEPs in $\mathcal{D}$. Specifically, we seek to develop a model m that, given $\mathcal{D}$ and x_{TEST} , outputs a binary decision:

$$m(\mathcal{D}, x_{TEST}) = \begin{cases} 1, & \text{if } x_{TEST} \text{ is an anomalous MEP,} \\ 0, & \text{if } x_{TEST} \text{ is a normal MEP.} \end{cases} \quad (4)$$

An anomalous time series is defined as one that exhibits a significant deviation from the learned distribution of normal MEPs in $\mathcal{D}$, technically as a reduction of at least 20% in signal amplitude relative to the learned baseline. Such deviations are intended to capture potentially pathological intraoperative events that may indicate neurological injury. Our method aims to identify such anomalies early, with the potential to provide real-time alerts to neurophysiologists and surgeons for timely intervention.

This example in Fig. 1 illustrates a real but straightforward case where the anomaly is visually evident. However, in practical scenarios, distinguishing anomalies is more challenging due to the high variability of MEP signals. Hence, developing robust models capable of accurate detection is crucial for intraoperative neurophysiological monitoring.

CIsoF operates under two distinct tasks: Detection and Early Detection. While the Detection task follows the standard formulation described in Eq. 4, the Early Detection task extends this concept by aiming to predict future anomalies before they manifest. Operationally, the proposed framework assigns a continuous anomaly

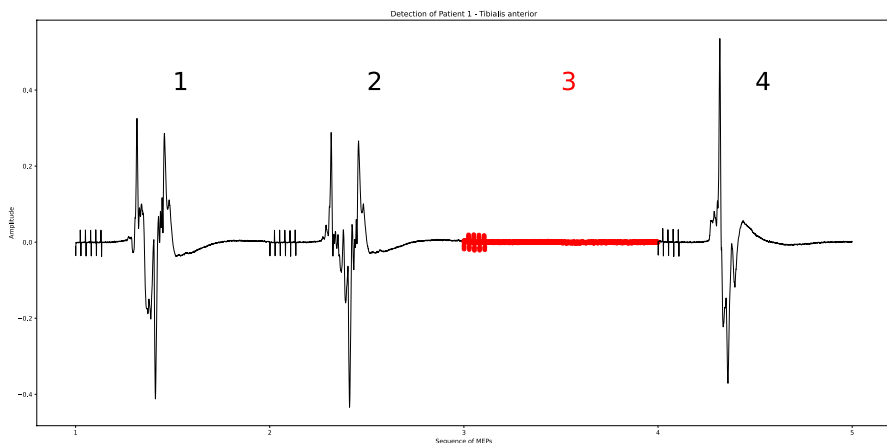


Fig. 1 Example of MEP sequences in the testing phase. Each MEP is numbered according to its temporal order. The MEP labeled as #3 (highlighted in red and with a larger line width) is an anomaly. In the detection task, the goal is to determine whether MEP #3 is anomalous

score to each newly acquired MEP waveform. In the Early Detection task, rather than determining whether the current MEP time series is anomalous, the goal is to predict whether the **next MEP time series** will be anomalous. Formally, given an observed MEP time series x_t , the model aims to predict the anomaly status of the subsequent time series x_{t+1} , thereby enabling proactive intervention. In this manuscript, we treat early detection as an *anticipatory warning* task evaluated by shifting the ground-truth one step earlier in the MEP sequence (see Section 4.1). Operationally, early detection can be instantiated by thresholding the anomaly score computed on the current MEP x_t and interpreting a high score as a warning that x_{t+1} may become anomalous.

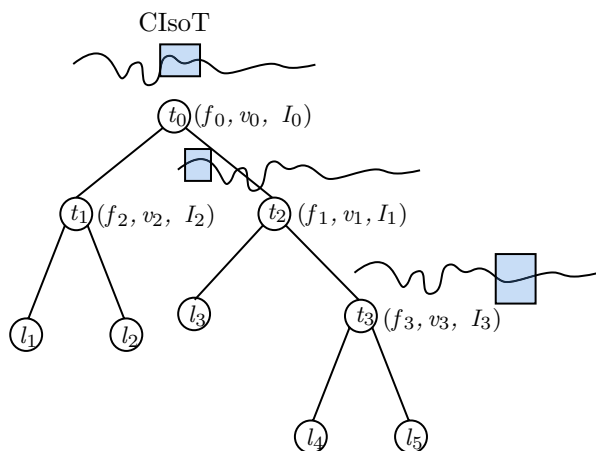
$$m'(\mathcal{D}, x_t) = \begin{cases} 1, & \text{if } x_{t+1} \text{ is expected to be anomalous,} \\ 0, & \text{if } x_{t+1} \text{ is expected to be normal.} \end{cases} \quad (5)$$

The early detection task is particularly crucial in intraoperative monitoring, as it could allow for timely intervention before significant neurological deterioration occurs. By leveraging the temporal structure of MEP sequences, the model can anticipate pathological deviations, reducing the risk of irreversible damage.

3.3 Canonical Isolation Tree and Forest

This section presents the forest-based approach: we first introduce how a tree works and then how to build a forest of this specific tree. The Canonical Isolation Forest (CIsoF) represents an ensemble of Canonical Isolation Trees (CIsoT). A CIsoT is a binary tree structure in which each node has either no children – making it a leaf – or exactly two children – making it an internal node (an illustration is provided in Fig. 2). Internal nodes are associated with tests that determine the routing of a time series: depending on the outcome, the time series is directed to either the left or the

Fig. 2 The Canonical Isolation Tree. In each node n the decision is taken i) by extracting the interval I_n from the time series (highlighted in light blue); ii) by extracting from it the feature f_n ; iii) by checking if $x_{I_n}^{f_n} < v_n$, making the time series follow the left branch if true and the right one if false



right child node. More formally, a test t on a time series $x = \{v_1, v_2, \dots, v_l\}$ is characterized by a tuple $t = (f, v, I)$, which operates as follows:

- It extracts from the time series x the subseries corresponding to the interval defined by I . This is the same idea used in Time Series Forest [36] and Canonical Interval Forest [37], where their investigations on classification tasks have shown that using intervals improves predictive performance. Assuming that all time series share the same length, the interval I can be specified using absolute positions: $I = [I_s, I_e]$, where I_s and I_e denote the start and end points of the interval, respectively. In cases where the time series may have varying lengths, the interval can instead be defined relative to the series length – for example, by specifying a range from 30% to 35% of the time series. Let us denote as x_I the extracted subseries corresponding to the given interval.
- The feature f is extracted from the interval x_I , where f corresponds to one of the predefined measures (outlined in Table 11). We denote the resulting value as x_I^f .
- If $x_I^f < v$, the time series x proceeds to the left child; otherwise, it is directed to the right child.

3.3.1 Features

The feature f is chosen from a set of 25 measures that can be computed from a time series. Specifically, the first 22 features are selected from CATCH22, a well-established collection of 22 descriptive features tailored for time series analysis [21]. For completeness, a summary of the CATCH22 features is provided in the Appendix (see Table 11). This choice is consistent with the approach used in the *Canonical Interval Forest* [37], a recent Random Forest-based method for time series classification. The remaining three features – the mean, the standard deviation, and the slope of the least squares regression line – are adapted from the *Time Series Forest* classifier proposed by Deng et al. [36].

3.3.2 Training of a CIsoT

The Canonical Isolation Tree (CIsoT) is constructed using a standard top-down, recursive approach. Beginning with the full training set $\mathcal{D} = \{x_1, \dots, x_N\}$, consisting of N time series, a test $t = (f, v, I)$ is selected for the root node. This test partitions $\mathcal{D}$ into two subsets, $\mathcal{D}_L$ and $\mathcal{D}_R$, based on the outcomes of the test for each time series. These subsets are then used to recursively define the tests for the left and right child nodes. The recursion continues until a predefined maximum tree height is reached or until a node contains fewer than a minimum number of samples. Since searching for the "best" question cannot be performed using traditional criteria (such as the Gini criterion in classification trees [38]) due to the absence of labels, we implement two distinct split methods. The first follows the standard random selection approach

used in the Isolation Forest [20], where a feature and a split threshold are chosen at random. The second method aims to minimize variance, as proposed in [39]. Specifically, we select a feature and evaluate a limited number of candidate thresholds (as commonly done in the RF paradigm [19]), choosing the one that minimizes the sum of variances in the resulting partitions $\mathcal{D}_L$ and $\mathcal{D}_R$. This variance-minimizing approach divides the data into more homogeneous subsets. More precisely, in our CIsOT, given a set $\mathcal{D}_n$ of time series reaching the node n , the test $t = (f, v, I)$ of that node is determined in the following way:

- The interval I is selected from a group of randomly generated intervals, with a different group created for each tree. Intervals are generated by randomly choosing starting points and lengths¹.
 - The feature f is randomly chosen from a randomly selected subset of the 25 pre-defined features.
 - A split threshold v in the range (f_{min}, f_{max}) is selected differently based on the split method chosen. Once defined $f_{min} = \min_{j=1 \dots N} (x_j)_I^f$ and $f_{max} = \max_{j=1, \dots, N} (x_j)_I^f$, we select v in two ways²:
 - randomly between (f_{min}, f_{max}) ,
 - minimizing the variance of the children by $\arg \min_{v \in (f_{min}, f_{max})} (\sigma_{\mathcal{D}_L}^2 + \sigma_{\mathcal{D}_R}^2)$.
- . The pseudocode is provided in Algorithm 1.

Algorithm 1 *MinVarianceSplit*: Obtaining best split threshold

Require: Training Set $\mathcal{D}_I^f$
Ensure: Best split threshold

```

1:  $minVariance \leftarrow \infty$ 
2: for all  $v \in \mathcal{D}_I^f$  do
3:    $\mathcal{D}_L \leftarrow \{x \in \mathcal{D} \mid x_I^f < v\}$ 
4:    $\mathcal{D}_R \leftarrow \{x \in \mathcal{D} \mid x_I^f \geq v\}$ 
5:    $variance \leftarrow |\mathcal{D}_L| \cdot \text{Var}(\mathcal{D}_L) + |\mathcal{D}_R| \cdot \text{Var}(\mathcal{D}_R)$ 
6:   if  $variance < minVariance$  then
7:      $v_{best} \leftarrow v$ 
8:      $minVariance \leftarrow variance$ 
9:   end if
10: end for
11: return  $v_{best}$ 

```

¹ Note that certain CATCH22 features require the interval to have a minimum length of 20.

² $(x_j)_I^f$ represents the value of feature f computed on the subseries corresponding to interval I , extracted from the time series x_j .

The complete training procedure for the CIsoT is provided in Algorithm 2.

Algorithm 2 *BuildCIsoT*: Fit a CIsoT

Require: Training Set $\mathcal{D}$, Current Height h , Height Limit hl , CATCH22 Feature Subset F , Interval Subset IS , Split Method m

Ensure: Fitted CIsoT.

```

1: if  $h \geq hl$  or  $|\mathcal{D}| < 2$  then
2:   return leaf{size :  $|\mathcal{D}|$ }
3: else
4:    $\mathcal{D}_L \leftarrow \emptyset$ 
5:    $\mathcal{D}_R \leftarrow \emptyset$ 
6:   Select a random interval  $I \in IS$ 
7:   Select a random feature  $f \in F$ 
8:   Let  $\mathcal{D}_I^f$  be the set composed by computing  $x_I^f$  for every  $x \in \mathcal{D}$ 
9:   if  $m = \text{"minvariance"}$  then
10:     $v = \text{MinVarianceSplit}(\mathcal{D}_I^f)$ 
11:   else
12:     $\min_v = \min(\mathcal{D}_I^f)$ 
13:     $\max_v = \max(\mathcal{D}_I^f)$ 
14:     $v = \text{RandomSample}(\min_v, \max_v)$ 
15:   end if
16:   Perform the split using  $v$  as threshold
17:    $\mathcal{D}_L \leftarrow \{x \in \mathcal{D} \mid x_I^f < v\}$ 
18:    $\mathcal{D}_R \leftarrow \{x \in \mathcal{D} \mid x_I^f \geq v\}$ 
19:   return node  $\left\{ \begin{array}{l} \text{left : } \text{BuildCIsoT}(\mathcal{D}_L, h + 1, hl, F, IS, m), \\ \text{right : } \text{BuildCIsoT}(\mathcal{D}_R, h + 1, hl, F, IS, m), \\ \text{split\_feature : } f, \\ \text{split\_value : } v, \\ \text{interval : } I \end{array} \right\}$ 
20: end if

```

3.3.3 Training of a CIsoF

The CIsoF (see Algorithm 3) represents an ensemble of CIsoTs. Following the classical procedure of Isolation Forest, each CIsoT is created using the procedure described previously, starting from a random subsampling of the training set. We set n_f , the number of CATCH22 features randomly sampled at each split. This approach aligns with the common practice in the random forest paradigm, where each tree is trained on a random subsample of features to enhance diversity and robustness. Finally, n_I denotes the number of intervals generated for each tree.

Algorithm 3 *CIsoFBuild*: Fit a CIsoF

Require: Training Set $\mathcal{D}$, Number of trees n_t , Number of intervals n_I , Number of features n_f , Split Method m
Ensure: Fitted CIsoF.

```

1:  $hl \leftarrow \log_2(|\mathcal{D}|)$ 
2:  $Forest \leftarrow []$ 
3: for  $j \leftarrow 0$  to  $n_t$  do
4:    $\mathcal{D}_j \leftarrow \text{Subsample}(\mathcal{D})$ 
5:   {generate a subsample of  $\mathcal{D}$ }
6:    $F_j \leftarrow \text{SampleFeatures}(n_f)$ 
7:    $I_j \leftarrow \text{SampleInterval}(n_I)$ 
8:    $Forest[j] \leftarrow \text{BuildCIsoT}(\mathcal{D}_j, 0, hl, F_j, I_j, m)$ 
9:   {build a tree and add it to the forest}
10: end for
11: return  $Forest$  {return the fitted forest}

```

3.3.4 Detection of Anomalies

Once the model is trained, we can detect anomalies by thresholding time series according to their anomaly scores, and anomalies are time series that produce a higher score. We define the anomaly score (AS; see Eq. 6) as in the work of Liu et al. [20]. The isolation mechanism exploits the fact that anomalies are easier to isolate than normal instances. Isolation is achieved by recursively partitioning the data using the test splits. Since anomalies are few and different, they tend to be isolated in fewer splits, leading to shorter path lengths in the trees. In contrast, normal instances require more splits due to their higher density and similarity to other points. This fundamental property allows the model to distinguish anomalies based on their expected path length in the constructed trees.

$$AS(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (6)$$

where x is a time series, n represents the size of the subsample set used in the training phase, $E(h(x))$ is the average depth of x from a CIsoF, and $c(n)$ is the average path length of an unsuccessful search in a binary search tree [40].

3.3.5 Early Detection of Anomalies

To better understand how the mechanism works for early detection, as detailed in Eq. 5, we retrieve the anomaly score of the MEP that precedes the actual anomaly. By doing so, we aim to distinguish between the anomaly scores of a normal MEP and those of MEPs that occur just before a true anomaly. In this way, we make an early prediction, attempting to anticipate the anomaly before it occurs, as detailed in Section 4.1. Early detection performance depends on the degradation modality of the

signal. With gradual signal degradation, the MEP recorded just before the confirmed anomaly is already different from baseline. This means that an anomalous signal can be detected earlier, before it is fully confirmed. In contrast, when the pathological event occurs abruptly, no detectable precursor is present, making anticipatory detection inherently more difficult. The early detection score should therefore be interpreted as a soft warning signal requiring clinical confirmation, not an autonomous alarm.

The proposed approach also differs from forecasting-based methods, which estimate future anomaly scores by fitting a model to the history of past scores. In the intraoperative setting, the pre-anomaly baseline is too short and non-stationary for reliable forecasting. Instead, an elevated anomaly score on x_t directly reflects how much the current physiological state deviates from the patient baseline, providing a natural and interpretable precursor signal for x_{t+1} without requiring a separate predictive model.

4 Experimental Details

4.1 Dataset Creation

The dataset used in this study consists of data collected from four patients undergoing neurosurgical procedures for intramedullary spinal cord tumor removal. The protocol used to elicit MEPs was transcranial electrical stimulation (TES). MEPs were evoked by applying transcranial anodal electrical stimulation with corkscrew electrodes placed on the scalp at C1 and C2 according to the 10–20 International System. The high-frequency short-train technique was used, applying a train of five pulses with an interstimulus interval of 4 ms. Intensity ranged between 50 and 150 mA. MEPs were recorded from the same target muscles in all patients, with pairs of subdermal needle electrodes placed in the muscle belly of the abductor pollicis brevis (APB) for upper limbs and the tibialis anterior (TA) and abductor hallucis (AH) for lower limbs. All procedures were performed in compliance with institutional guidelines and were approved by the appropriate institutional committee. The study obtained IRB approval with protocol ID: CRMS 23038 on June 05, 2024. These muscles were chosen because they were the only ones consistently monitored across all patients, ensuring uniformity in the dataset.

For each patient–muscle pair, we consider the chronological sequence of MEPs acquired during the procedure. Each MEP waveform corresponds to one stimulation and is treated as one univariate time series instance (see Section 3.2). The goal is to train a patient-specific baseline model using only the initial, stable portion of the sequence and then evaluate detection performance on the subsequent MEPs acquired during surgery.

The methodology for splitting the signals was designed to closely simulate real-world conditions in the operating room. The first anomaly within the sequence of MEP signals recorded during the procedure was identified. Using this anomaly as a reference point, the dataset was divided into two parts: the MEP signals preceding the first anomaly constituted the training dataset, while those following it, including the anomaly itself, formed the testing dataset. This approach preserves the natural

chronological order of events, ensuring that the training and testing datasets reflect real-time clinical scenarios.

4.1.1 Early detection dataset

To evaluate whether CISO_F provides a warning signal in advance of a clinically confirmed anomalous MEP, we construct an “early detection” version of each dataset by shifting the anomaly label one step earlier in the sequence. Concretely, for each anomaly occurring at time index t , we label the preceding MEP at index $t-1$ as positive (early warning), and exclude the MEP at t from the evaluation set for that specific event. In practice, this is implemented by removing the first MEP in the detection test set and replacing it with the last baseline MEP, and then repeating the procedure sequentially for each subsequent anomaly.

In Tables 1 and 2, a summary of the datasets is reported for both the Detection and Early Detection tasks, respectively. For each cell, the first number is the number of samples in the training set, the second number is the number of samples in the test set, and the third number is the proportion of anomalies in the test set.

4.2 CISO_F in Detail

In this section, we outline the experimental setup used to evaluate our proposed method, detailing the different configurations, split methods, evaluation metric, and repetition strategy employed. We designed three configurations for CISO_F to evaluate the trade-off between computational efficiency and predictive performance. We used the **Heavy**, **Light**, and **UltraLight** configurations:

- In the **Heavy** configuration, we set the number of trees (n_t) to 500, the number of features (n_f) to 4, and the number of intervals (n_I) to $\sqrt{N}$, where N is the sequence length.
- In the **Light** configuration, $n_t = 200$, $n_f = 4$, and $n_I = \log_2 N$.
- In the **UltraLight** configuration, $n_t = 100$, $n_f = 4$, and $n_I = \log_{10} N$.

Table 1 Detection Cardinality (Train-Test samples and percentage of anomalies in Test set)

	AH	APB	TA
1	130-366(0.16)	25-94(0.06)	130-111(0.12)
2	6-249(0.17)	6-332(0.17)	33-616(0.01)
3	7-25(0.36)	7-151(0.01)	19-77(0.12)
4	11-868(0.13)	41-1571(0.04)	7-153(0.29)

Table 2 Early Detection Cardinality (Train-Test samples and percentage of anomalies in Test set)

	AH	APB	TA
1	129-308(0.19)	24-89(0.07)	129-99(0.13)
2	5-207(0.21)	5-275(0.21)	32-609(0.01)
3	6-17(0.53)	6-150(0.01)	18-69(0.13)
4	10-757(0.15)	40-1505(0.04)	6-109(0.41)

We evaluated two different split methods (both described in Section 3.3):

1. **Random**: standard method in Isolation Forest [20].
2. **MinVariance**: unsupervised split technique that can be optimized.

To evaluate performance, we used the Receiver Operating Characteristic area under the curve (ROC-AUC), a widely used metric for anomaly detection problems [20, 41]. For all statistical tests, we set $\alpha = 0.05$. Finally, to assess the impact of randomization on model performance, we conducted all experiments ten times for each dataset, configuration, and split method, and report the average over these repetitions. This repetition allows us to measure how inherent randomization affects the performance of our proposed method.

5 Results and Discussion

5.1 Analysis of ClsoF Variants

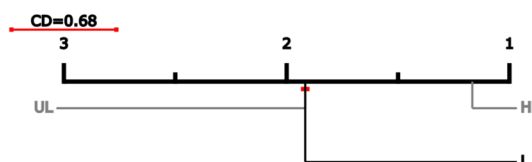
As an initial analysis, we compared the different configurations of ClsoF to evaluate their performance and computational efficiency. Table 3 presents a comparison of the three configurations in terms of their performance in detection and early detection tasks; results are averages of ROC-AUC across patients and muscles. The **Heavy** configuration achieves the highest performance across both tasks, while the **UltraLight** and **Light** configurations exhibit similar scores. The **Light** configuration provides a balance between computational efficiency and predictive performance, improving upon the **UltraLight** configuration on the detection task while maintaining a reduced computational burden compared to the **Heavy** configuration.

To assess the statistical significance of our comparisons, we conducted a Friedman test, followed by a post-hoc pairwise analysis using Nemenyi's test for multiple comparisons of mean rank sums [42]. The Friedman test determines whether there are significant overall differences in the rankings of the evaluated methods across different datasets. If a significant difference is found, Nemenyi's test is applied to examine pairwise differences in mean ranks. The results are then visualized using a critical difference diagram (see Fig. 3), where the mean ranks of the methods across

Table 3 Mean ROC-AUC comparison of configurations (UL, L, H) for detection and early detection

	UL	L	H
Detection	0.6681	0.6781	0.7824
Early Detection	0.5403	0.5257	0.6077

Fig. 3 Critical difference diagram comparing the ROC-AUC of the three configurations: UL, L, and H over all datasets



datasets are displayed. The red lines connect rankings that do not present statistically significant differences.

Figure 3 shows the critical difference diagram, where configurations are ranked based on their average performance across datasets. The numbers 1–3 indicate the average ranks, increasing from best to worst (left to right). Two configurations must differ by at least 0.68 (critical difference) to be considered statistically different. The **Light** and **UltraLight** configurations, which yield the lowest performance, do not differ in a statistically significant manner. In contrast, the **Heavy** configuration is statistically distinct from both. Overall, the **Heavy** configuration represents a suitable compromise when performance is the primary concern.

Given this statistically superior performance, it is important to assess whether the **Heavy** configuration remains compatible with real-time intraoperative deployment. Inference is highly efficient and performed exclusively at test time, while model training is carried out offline. On a consumer-grade laptop CPU (Intel® Core™ i9-10980HK), CIsoF requires approximately 2–3 ms to score a single MEP, enabling the processing of several hundred MEPs per second. In practical terms, batch inference over 10–50 MEPs achieves throughputs between approximately 300 and 430 MEPs/s, which is orders of magnitude faster than the rate at which MEPs are acquired during surgery. These results demonstrate that the **Heavy** configuration can be safely employed in real-time intraoperative monitoring scenarios.

In the next part, we present the performance of CIsoF models for both detection and early detection tasks, taking the best among the configurations. Two tables summarize the results: Table 4 for detection, and Table 5 for early detection; each table includes both split methods described above. We denote the Random and MinVariance variants as CIsoF_R and CIsoF_{MV}, respectively. Each table evaluates the models over four patients, denoted as rows 1–4, and the three muscle groups described above. Overall, the proposed model performs well, achieving an ROC-AUC greater than 0.80 in 12 out of 24 datasets in the Detection task, as shown in Table 4. Additionally, the MinVariance variant demonstrates strong performance, achieving ROC-AUC scores above 0.80 across 8 datasets. In the next section, we provide a detailed analysis of these results to further examine the model's behavior across different scenarios. Table 4 (left) reports the Random split variant, CIsoF_R.

Table 4 CIsoF ROC-AUC for detection under the two split methods

	AH	APB	TA
Random split			
1	0.9652	0.7078	0.9827
2	0.9619	0.4922	0.5389
3	0.7028	0.3466	0.9757
4	0.7539	0.5360	0.7736
MinVariance split			
1	0.9707	0.8985	0.9817
2	0.9382	0.5852	0.8909
3	0.7931	0.8634	0.9871
4	0.9596	0.6102	0.6411

Table 5 CIsoF ROC-AUC for early detection under the two split methods

	AH	APB	TA
Random split			
1	0.8262	0.6036	0.7188
2	0.6891	0.5616	0.5516
3	0.6292	0.6402	0.5426
4	0.5855	0.5431	0.5675
MinVariance split			
1	0.8136	0.6675	0.7262
2	0.6441	0.5282	0.6058
3	0.6611	0.9672	0.6919
4	0.6620	0.4956	0.5204

The highest scores, on average, are observed for AH, with values ranging from 0.7028 to 0.9652. TA shows high average performance but variable performance across patients. However, APB shows lower scores, particularly in patients 2 and 4, where it falls around 0.50. Table 4 (right) reports the MinVariance split variant, CIsoF_{MV}, aimed at improving the quality of child nodes. This modification leads to substantial performance gains across most metrics, particularly for APB. For instance, APB improves from 0.3466 for patient 3 to 0.8634. For TA, performance improves in two of four patients, with a mean ROC-AUC of approximately 0.88. These results suggest that minimizing the variance in splits better captures the structure of the underlying data, leading to improved detection performance. Table 5 evaluates the models' ability to perform early detection. The left subtable reports the Random split variant. While AH and TA maintain reasonable scores (e.g., 0.8262 and 0.7188 for patient 1), the performance for APB is relatively low, particularly in patients 2 and 4, with scores around 0.55.

The right subtable reports the MinVariance split criterion for early detection. Compared to the Random split, APB improves in two of four patients (e.g., 0.6402 to 0.9672 in patient 3), while TA improves in three of four patients.

Compared to the detection task, where our model achieved strong performance with high ROC-AUC scores, the early detection task proved to be more challenging, as expected. The Random split variant exceeds a ROC-AUC of 0.60 in 6 out of 12 datasets, while the MinVariance variant does so in 9 out of 12 cases. These results are less consistent than in the detection setting but still indicate promising predictive capability in this more difficult scenario, where identifying anomalies at an earlier stage is inherently challenging.

We compared the two different split methods of CIsoF, namely *Random* and *MinVariance*. Specifically, we averaged the results across the three previously described configurations and all datasets. These results are presented in Table 6. To determine whether the observed differences were statistically significant, we applied the Wilcoxon

Table 6 Mean ROC-AUC comparison of split methods (Random and MinVariance) for detection and early detection

	Random	MinVariance
Detection	0.6645	0.7546
Early Detection	0.5485	0.5673

signed-rank test [43], a non-parametric test that evaluates whether two paired samples drawn from related distributions differ significantly. The test results indicate a significant difference between the *MinVariance* and *Random* splitting strategies. Moreover, *MinVariance* consistently outperformed *Random*, confirming its statistical superiority.

5.2 Comparison with Literature Alternatives

In this section, we present the results of our experiments evaluating different approaches [41] for abnormal time series detection. Since there is no standard method for solving this problem, we compare our proposed approach against alternative strategies. We found the problem described in the literature in [16], but, as already explained in the introduction, deep learning techniques are unsuitable for our case. Thus, we explored two approaches to tackling this problem. The first involves transforming time series into feature vectors using statistical and temporal descriptors, in our case CATCH22 [21], and then applying standard anomaly detection models. The second approach relies on computing distances between time series — in our case, using constrained Dynamic Time Warping (cDTW) with a constraint value of 5% of the time series length — and subsequently employing a distance-based anomaly detection model. The selected models cover a diverse range of anomaly detection methodologies. One-Class Support Vector Machine (OC-SVM) [44] follows a boundary-based approach, learning a hyperplane that encapsulates normal data distribution in a high-dimensional space. Subspace Outlier Detection (SOD) [45] is designed for high-dimensional datasets, detecting anomalies within relevant subspaces. LOF-cDTW [46, 47] extends the classic Local Outlier Factor (LOF) by incorporating cDTW as a dissimilarity measure for time series data. Histogram-Based Outlier Score (HBOS) [48] is a non-parametric method that models feature distributions using histograms and assigns anomaly scores based on deviations from expected densities. Empirical Cumulative Distribution Outlier Detection (ECOD) [49] is a fully unsupervised approach that estimates the empirical cumulative distribution function and detects anomalies based on extreme deviations. Finally, Minimum Covariance Determinant (MCD) [50] is a robust statistical technique that identifies anomalies by computing the most representative subset of the data and detecting points that significantly deviate from it. We utilized the implementations available in the PyOD Python library (<https://pyod.readthedocs.io/en/latest/index.html>), which provides a comprehensive suite of anomaly detection tools.

Table 7 presents the ROC-AUC comparison between our proposed model and the other anomaly detection methods on the Detection Task. Our approach outperforms all competitors in 10 out of 12 cases, demonstrating its superior performance across most scenarios. Among the alternative models, ECOD achieves the second-best results overall, performing best in 2 cases.

Table 8 presents the ROC-AUC comparison for the Early Detection Task, which is the most critical evaluation in our study. CISO_F achieves the highest performance in 8 out of 12 cases. In contrast, ECOD wins 3 times, while MCD achieves the highest score in a single case. This further reinforces the advantage of our approach over traditional anomaly detection methods. On average, CISO_{F_{MV}} achieves a mean ROC-AUC of 0.6653, which is approximately 6% higher than ECOD's mean score of 0.6081, further highlighting its superior effectiveness in early anomaly detection.

Table 7 ROC-AUC comparison for detection across models

		IForest	OCSVM	SOD	LOF-cDTW	HBOS	ECOD	MCD	CIsosFr	$\mathbf{GFR}_{IV}$
1	AH	0.8965	0.8908	0.2195	0.5062	0.8270	0.8551	0.7378	0.9652	0.9707
	APB	0.2767	0.4110	0.5663	0.3617	0.2841	0.6951	0.5718	0.7078	0.8985
	TA	0.9169	0.7245	0.6468	0.6601	0.8132	0.9474	0.9195	0.9827	0.9817
2	AH	0.7832	0.6885	0.4081	0.8395	0.4525	0.7230	0.6841	0.9619	0.9382
	APB	0.3435	0.5000	0.6330	0.5631	0.2263	0.8083	0.5654	0.4922	0.5852
	TA	0.6135	0.3639	0.3520	0.1053	0.8131	0.8106	0.4106	0.5389	0.8909
3	AH	0.4635	0.5000	0.4583	0.2292	0.7014	0.5139	0.5736	0.7028	0.7931
	APB	0.1374	0.5034	0.2148	0.0235	0.2114	0.8322	0.3440	0.3466	0.8634
	TA	0.6964	0.5441	0.6340	0.2304	0.4085	0.8431	0.4606	0.9757	0.9871
4	AH	0.4198	0.5354	0.5736	0.4115	0.3308	0.5514	0.6993	0.7539	0.9596
	APB	0.6076	0.6160	0.4414	0.5524	0.7327	0.7608	0.6969	0.5360	0.6102
	TA	0.6170	0.5048	0.4866	0.5321	0.6121	0.5918	0.5607	0.7736	0.6411
Mean		0.5643	0.5652	0.4695	0.4179	0.5344	0.7444	0.6020	0.7281	0.8433

Table 8 ROC-AUC comparison for early detection across models

		IForest	OCSVM	SOD	LOF-cDTW	HBOS	ECOD	MCD	CIsosFr	$\frac{CIsosFr}{CIsosFr_{IV}}$
1	AH	0.5900	0.6778	0.6072	0.6784	0.5664	0.6315	0.6571	0.8262	0.8136
	APB	0.5534	0.5843	0.6747	0.4297	0.5341	0.6847	0.6671	0.6036	0.6675
	TA	0.5794	0.4267	0.5653	0.5818	0.6011	0.6950	0.5699	0.7188	0.7262
2	AH	0.5982	0.5693	0.5067	0.4558	0.5125	0.6792	0.5732	0.6891	0.6441
	APB	0.3537	0.5000	0.6437	0.3983	0.4769	0.6920	0.6027	0.5616	0.5282
	TA	0.4342	0.3696	0.7169	0.4056	0.6624	0.6741	0.3044	0.5516	0.6058
3	AH	0.4542	0.5000	0.7778	0.6250	0.6597	0.5833	0.5319	0.6292	0.6611
	APB	0.0639	0.5034	0.4932	0.0456	0.0270	0.5541	0.2514	0.6402	0.9672
	TA	0.5561	0.5500	0.5352	0.4639	0.4444	0.5148	0.6626	0.5426	0.6919
4	AH	0.3959	0.4978	0.4482	0.2720	0.4208	0.4737	0.4980	0.5855	0.6620
	APB	0.5086	0.5281	0.4987	0.5363	0.5009	0.5764	0.5819	0.5431	0.4956
	TA	0.5137	0.5042	0.5396	0.4997	0.5528	0.5382	0.4963	0.5675	0.5204
Mean		0.4668	0.5176	0.5839	0.4493	0.4966	0.6081	0.5331	0.6216	0.6653

To statistically assess the performance of CIsoF, we selected its best variant (CIsoF_{MV}) and compared it against the best competitor ECOD. We then conducted a Wilcoxon signed-rank test to evaluate whether CIsoF_{MV} significantly outperforms the alternative. The test results confirmed that CIsoF_{MV} passed the significance threshold, demonstrating that it is statistically superior to the best competing model.

5.3 External Cohort Validation

To further validate the generalization capability of the proposed CIsoF method, we conducted a comprehensive evaluation on an independent external cohort of patients. The external validation cohort comprised MEP recordings from 13 patients who underwent neurosurgical procedures to the spinal cord. Importantly, these data were collected prospectively without prior knowledge of our study's specific methodological assumptions. The recordings were obtained from two consistently monitored muscles across all patients: the abductor pollicis brevis (APB) for upper limb monitoring and tibialis anterior (TA) for lower limb assessment. For completeness, the cardinality of the dataset used for both the detection and early detection (see Tables 12 and 13, respectively) tasks in this external cohort is reported in Appendix B.

5.3.1 Performance Analysis Across External Cohort

Table 9 presents the comparison of CIsoF configurations (UltraLight, Light, and Heavy) on the external cohort, stratified by task type. Results demonstrate that the Heavy configuration continues to achieve superior performance in the external setting, with a detection ROC-AUC of 0.7988 and an early detection score of 0.5172. This consistency with the primary analysis validates the appropriateness of the Heavy configuration across diverse clinical datasets.

The detection task achieved a mean ROC-AUC of 0.7988 for the Heavy configuration, representing strong anomaly detection capability in previously unseen patient populations. The early detection task, while more challenging as anticipated, achieved a mean ROC-AUC of 0.5172, demonstrating that the temporal predictive capability translates reasonably well to external data.

The comparison between Random and MinVariance split methods in the external cohort (Table 10) replicates the findings from the primary analysis. The MinVariance approach substantially outperformed the Random method for both detection (0.7923 vs. 0.7111) and early detection (0.5104 vs. 0.4868), yielding performance improvements. These results strongly validate that variance-minimizing split criteria provide superior partitioning in unsupervised anomaly detection, when applied to previously unobserved clinical data.

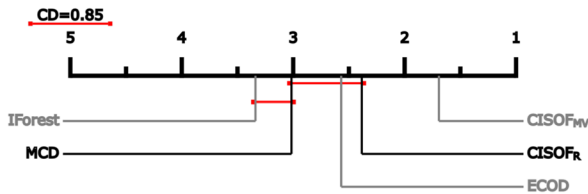
Figure 4 presents the critical difference diagram derived from Friedman test analysis and Nemenyi post-hoc pairwise comparisons, ranking the five best-performing models according to their mean ROC-AUC across all external cohort datasets.

Table 9 Mean ROC-AUC comparison of configurations (UL, L, H) for detection and early detection

	UL	L	H
Detection	0.7022	0.7541	0.7988
Early Detection	0.4971	0.4815	0.5172

Table 10 Mean ROC-AUC comparison of split methods (Random and MinVariance) for detection and early detection

	Random	MinVariance
Detection	0.7111	0.7923
Early Detection	0.4868	0.5104

**Fig. 4** Critical difference diagram comparing the top five anomaly detection models based on mean ROC-AUC across all external cohort datasets. Models are ranked from left (worst) to right (best) by average rank. CISO_{F_{MV}} achieves the best average rank and is statistically superior to all competing methods

First, CISO_F achieves the best average rank among all evaluated methods, demonstrating superior overall performance in the external validation setting. Specifically, the variance-minimizing variant (CISO_{F_{MV}}) consistently outperforms competing approaches, including OC-SVM, SOD, LOF-cDTW, HBOS, ECOD, and MCD. The statistical analysis confirms that CISO_F is statistically different from the competing methodologies at the $\alpha = 0.05$ level, indicating that the observed performance advantage is not attributable to random variation.

5.3.2 Detailed Performance Across Patient Populations

A granular analysis of the top five models' performance across individual patients and muscles is provided in Table 14 (Appendix Section C). CISO_{F_{MV}} achieves ROC-AUC exceeding 0.90 in 14 out of 26 detection datasets, demonstrating reliable high-performance anomaly detection in the majority of external cohort patients. Even in more challenging datasets, CISO_{F_{MV}} generally maintains competitive performance for detection tasks, falling below 0.60 in only one case.

Early detection performance, while more variable, remains above chance levels (0.50) in most datasets, with CISO_{F_{MV}} achieving a mean early detection ROC-AUC of 0.57 across the external cohort. The greater variability in early detection reflects the inherent difficulty of predicting future anomalies from current observations.

5.4 Limitations and Future Directions

From a clinical perspective, false positives and false negatives carry asymmetric costs during intraoperative monitoring. A missed detection (false negative) may delay intervention during neurological deterioration, potentially leading to irreversible motor deficits. Conversely, an excessive rate of false alarms (false positives) may reduce the practical usability of the system, increase the cognitive burden on the surgical team, and ultimately undermine trust in the warning signal.

The proposed framework addresses this asymmetry through its continuous anomaly score output. Rather than imposing a fixed binary threshold, the system allows the operating threshold to be selected by the neurophysiologist at the beginning of the procedure, according to the risk profile of the specific surgical context. In high-risk procedures, a lower threshold (higher sensitivity) can be adopted to minimize missed detections at the cost of a higher false alarm rate; in lower-risk settings, a stricter threshold reduces false alarms. The ROC-AUC metric used throughout this work characterizes this trade-off across all possible operating points precisely, making it an appropriate evaluation criterion for a system intended to support, rather than replace, clinical judgment. In all cases, the anomaly score is presented to the neurophysiologist as an assistive signal, and all clinical decisions remain the sole responsibility of the surgical team.

Several limitations should nonetheless be acknowledged. First, the study is based on a limited number of patients, reflecting the intrinsic difficulty of collecting intraoperative MEP data. While our results are encouraging, larger and more diverse cohorts are required to fully assess robustness across surgical types, anatomical targets, and anesthetic protocols. Second, early detection is evaluated retrospectively through an anticipatory labeling protocol; prospective validation is therefore necessary to confirm the utility of early warning signals in real-time surgical decision-making. Third, before clinical implementation can be considered viable, the physiological meaning of the mathematical features involved should be further explored and defined.

Future work will focus on prospective clinical validation, integration with multi-modal intraoperative monitoring signals, adaptive threshold selection strategies, and robustness to non-pathological physiological variations. Addressing these directions is essential to bridge the gap between methodological development and the safe, effective deployment of the system in clinical practice.

6 Conclusions

This study presents a novel approach to the detection and early detection of pathological intraoperative motor evoked potentials by introducing the Canonical Isolation Forest (CIsoF), an extension of Isolation Forest specifically tailored to time series biomedical signals. Our results demonstrate that the proposed method effectively identifies deviations from patient-specific baseline MEP patterns and, in cases characterized by gradual neurological compromise, provides early warning signals prior to clinically confirmed pathological events.

A key strength of the proposed framework is its compatibility with real intraoperative constraints. CIsoF operates in a patient-specific, unsupervised manner, requires only baseline data acquired at the beginning of the procedure, and produces interpretable anomaly scores without relying on real-time labeling, online retraining, or automated decision-making. The proposed framework represents a promising step toward computational support tools for intraoperative neurophysiological monitoring, while additional prospective and clinical validation studies remain necessary before practical deployment can be considered.

Appendix A CATCH22 Features

Table 11 Simplified descriptions of the CATCH22 features grouped by category, based on [21] and following the table format proposed in [51]

Category	Description
Distribution	• Mode of z-scored distribution (5-bin histogram) • Mode of z-scored distribution (10-bin histogram)
Simple temporal statistics	• Longest period of consecutive values above the mean • Time intervals between successive extreme events above the mean • Time intervals between successive extreme events below the mean
Linear autocorrelation	• First $1/e$ crossing of autocorrelation function • First minimum of autocorrelation function • Total power in lowest fifth of frequencies in the Fourier power spectrum • Centroid of the Fourier power spectrum • Mean error from a rolling 3-sample mean forecasting
Nonlinear autocorrelation	• Time-reversibility statistic, $\langle (x_{t+1} - x_t)^3 \rangle_t$ • Automutual information, $m = 2, \tau = 5$ • First minimum of the automutual information function
Successive differences	• Proportion of successive differences exceeding 0.04σ [52] • Longest run of successive incremental decreases • Shannon entropy of two successive letters in equiprobable 3-letter symbolization • Change in correlation length after iterative differencing • Exponential fit to successive distances in 2-d embedding space
Fluctuation Analysis	• Proportion of slower timescale fluctuations that scales with DFA (Detrended Fluctuational Analysis) (50% sampling) • Proportion of slower timescale fluctuations that scales with linearly rescaled range fits
Others	• Trace of covariance of transition matrix between symbols in 3-letter alphabet • Periodicity measure, as defined in [53]

Appendix B External Cohort Cardinality

Table 12 Detection Cardinality (Train-Test samples and percentage of anomalies in Test set)

	APB	TA
1	39-976(0.03)	44-711(0.07)
2	21-457(0.14)	21-607(0.00)
3	6-274(0.32)	335-840(0.05)
4	25-1817(0.01)	7-586(0.08)
5	8-56(0.29)	7-41(0.37)
6	7-189(0.14)	7-526(0.02)
7	12-952(0.17)	457-1484(0.02)
8	122-58(0.03)	13-9(0.56)
9	97-188(0.11)	9-198(0.20)
10	34-39(0.15)	9-26(0.38)
11	7-173(0.12)	18-47(0.30)
12	14-75(0.09)	7-48(0.33)
13	15-162(0.18)	15-230(0.15)

Table 13 Early Detection Cardinality (Train-Test samples and percentage of anomalies in Test set)

	APB	TA
1	38-946(0.03)	43-660(0.08)
2	20-396(0.16)	20-606(0.00)
3	5-188(0.46)	334-796(0.06)
4	24-1807(0.01)	6-540(0.09)
5	7-41(0.39)	6-27(0.56)
6	6-163(0.17)	6-518(0.02)
7	11-789(0.21)	456-1455(0.02)
8	121-57(0.04)	12-5(1.00)
9	96-168(0.12)	8-159(0.25)
10	33-34(0.18)	8-17(0.59)
11	6-153(0.14)	17-34(0.41)
12	13-69(0.10)	6-33(0.48)
13	14-134(0.22)	14-197(0.17)

Appendix C External Cohort Complete Results

Table 14 reports the detailed performance results for each dataset, listing the five best-performing models. These results further confirm the robustness of CISO_F, which consistently achieves strong detection and early detection performance despite the increased variability introduced by the new patient cohort.

Table 14 Top 5 anomaly detection models

			MCD	SOD	ECOD	CIsoF _R	CIsoF _{MV}
1	APB	D	0.2767	0.5718	0.6951	0.7208	0.9011
		ED	0.5534	0.6671	0.6847	0.6757	0.6295
	TA	D	0.9121	0.9050	0.9274	0.9763	0.9814
		ED	0.6468	0.6208	0.7203	0.7521	0.7956
11	APB	D	0.4395	0.6621	0.5599	0.7473	0.7670
		ED	0.5161	0.5533	0.5369	0.6279	0.6860
	TA	D	0.9638	0.9685	0.9042	0.9732	0.9835
		ED	0.7964	0.8292	0.8033	0.8217	0.8137
13	APB	D	0.8019	0.7760	0.9307	0.8567	0.9380
		ED	0.5278	0.5080	0.5570	0.5285	0.4403
	TA	D	0.4090	0.3689	0.6683	0.5563	0.7819
		ED	0.4570	0.4912	0.5923	0.5783	0.6898
14	APB	D	0.6425	0.7023	0.5234	0.5359	0.6225
		ED	0.4490	0.4928	0.5150	0.4530	0.5893
	TA	D	0.6758	0.5615	0.5641	0.5872	0.6269
		ED	0.5706	0.6200	0.6722	0.4506	0.6300
17	APB	D	0.7623	0.6123	0.6808	0.7609	0.9010
		ED	0.4153	0.4350	0.4856	0.5306	0.5899
	TA	D	0.3753	0.3955	0.7182	0.5034	0.7299
		ED	0.2959	0.3239	0.4894	0.4978	0.4297
18	APB	D	0.7771	0.6811	0.5891	0.8163	0.9014
		ED	0.6436	0.6300	0.5195	0.6975	0.8162
	TA	D	0.7728	0.7224	0.9285	0.9874	0.9942
		ED	0.4683	0.6203	0.5804	0.6365	0.6375
19	APB	D	0.3364	0.4420	0.7232	0.4597	0.9405
		ED	0.3830	0.4479	0.6461	0.4017	0.6742
	TA	D	0.2631	0.7033	0.7194	0.6198	0.7498
		ED	0.3811	0.5998	0.5762	0.5446	0.4583
2	APB	D	0.0471	0.0098	0.8671	0.5860	0.8321
		ED	0.4456	0.5051	0.5410	0.4852	0.4076
	TA	D	0.6723	0.6758	0.5796	0.7918	0.7755
		ED	0.6212	0.5787	0.5082	0.6007	0.6257
21	APB	D	1.0000	1.0000	1.0000	0.9714	0.9973
		ED	0.5336	0.6118	0.3364	0.5818	0.5009
	TA	D	0.5650	0.4250	0.6000	0.8900	0.9400
		ED	0.0000	0.0000	0.0000	0.0000	0.0000
3	APB	D	0.9717	0.9217	1.0000	0.9798	0.9869
		ED	0.3702	0.4524	0.3810	0.4935	0.5327
	TA	D	0.5053	0.6325	0.8250	0.6656	0.7125
		ED	0.5343	0.2971	0.5000	0.5700	0.6100
5	APB	D	0.1678	0.8097	0.9123	0.8010	0.9787
		ED	0.3930	0.5710	0.7381	0.6969	0.8672
	TA	D	0.6011	0.6026	0.5801	0.6500	0.6530
		ED	0.5107	0.4725	0.4571	0.3729	0.3214
6	APB	D	0.6754	0.6103	0.7773	0.7920	0.8132
		ED	0.4675	0.5869	0.3779	0.5099	0.5486
	TA	D	0.6568	0.6859	0.5879	0.7004	0.7285
		ED	0.4559	0.5070	0.5257	0.5732	0.6379

Table 14 (continued)

			MCD	SOD	ECOD	CIs _{oF_R}	CIs _{oF_{MV}}
7	APB	D	0.6933	0.7351	0.8434	0.8857	0.9010
		ED	0.5842	0.5500	0.6007	0.4226	0.4306
	TA	D	0.8746	0.8216	0.8349	0.9338	0.9426
		ED	0.5404	0.5341	0.6285	0.5031	0.5034

Acknowledgements A.B. was supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNE-SYS (PE00000006) - "A Multiscale integrated approach to the study of the nervous system in health and disease" (DN. 1553 11.10.2022).

Author Contributions A.A. did the experiments and implementations. A.A., A.B., P.S., C.C. and M.B. wrote the manuscript. A.A., A.B., P.S., M.B. developed the main methodology. F.B., S.N collected and cleaned the data. All authors reviewed the manuscript.

Funding Open access funding provided by Università degli Studi di Verona within the CRUI-CARE Agreement. This research received no external funding.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest/Competing interests The authors declare no competing interests.

Ethics approval and consent to participate The study obtained IRB approval with protocol ID: CRMS 23038 on June 05, 2024.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Deletis V, Sala F (2008) Intraoperative neurophysiological monitoring of the spinal cord during spinal cord and spine surgery: A review focus on the corticospinal tracts. *Clin Neurophysiol* 119:248–264. <https://doi.org/10.1016/j.clinph.2007.09.135>
2. Sala F, Bricolo A, Faccioli F, Lanteri P, Gerosa M (2007) Surgery for intramedullary spinal cord tumors: the role of intraoperative (neurophysiological) monitoring. *Eur Spine J* 16(S2):130–139. <https://doi.org/10.1007/s00586-007-0423-x>
3. Sala F, Manganotti P, Tramontano V, Bricolo A, Gerosa M (2007) Monitoring of motor pathways during brain stem surgery: What we have achieved and what we still miss? *Neurophysiologie Clinique/ Clinical Neurophysiology* 37(6):399–406. <https://doi.org/10.1016/j.neucli.2007.09.013>

4. MacDonald DB, Skinner S, Shils J, Yingling C (2013) Intraoperative motor evoked potential monitoring - a position statement by the american society of neurophysiological monitoring. *Clin Neurophysiol* 124(12):2291–2316. <https://doi.org/10.1016/j.clinph.2013.07.025>
5. Neuloh G, Pechstein U, Cedzich C, Schramm J (2004) Motor evoked potential monitoring with supratentorial surgery. *Neurosurgery* 54(5):1061–1072. <https://doi.org/10.1227/01.neu.0000119326.15032.00>
6. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman MLD, Smith TR, Arnaout O (2018) Machine learning and neurosurgical outcome prediction: A systematic review. *World Neurosurgery* 109:476–4861. <https://doi.org/10.1016/j.wneu.2017.09.149>
7. Boaro A, Kaczmarzyk JR, Kavouridis VK, Harary M, Mammi M, Dawood H, Shea A, Cho EY, Juvekar P, Noh T, Rana A, Ghosh S, Arnaout O (2022) Deep neural networks allow expert-level brain meningioma segmentation and present potential for improvement of clinical practice. *Sci Rep* 12(1). <https://doi.org/10.1038/s41598-022-19356-5>
8. Goedmakers CMW, Lak AM, Duey AH, Senko AW, Arnaout O, Groff MW, Smith TR, Vleggeert-Lankamp CLA, Zaidi HA, Rana A, Boaro A (2021) Deep learning for adjacent segment disease at preoperative MRI for cervical radiculopathy. *Radiology* 301(3):664–671
9. Hannun AY, Rajpurkar P, Haghighpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY (2019) Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 25:65–69. <https://doi.org/10.1038/s41591-018-0268-3>
10. Di Nardo F, Nocera A, Cucchiarelli A, Fioretti S, Morbidoni C (2022) Machine learning for detection of muscular activity from surface EMG signals. *Sensors (Basel)* 22(9):3393. <https://doi.org/10.3390/s22093393>
11. Quitadamo LR, Cavrini F, Sbernini L, Riillo F, Bianchi L, Seri S, Saggio G (2017) Support vector machines to detect physiological patterns for EEG and EMG-based human-computer interaction: a review. *J Neural Eng* 14(1):011001. <https://doi.org/10.1088/1741-2552/14/1/011001>
12. Boaro A, Azzari A, Basaldella F, Nunes S, Feletti A, Bicego M, Sala F (2024) Machine learning allows expert level classification of intraoperative motor evoked potentials during neurosurgical procedures. *Comput Biol Med* 180:109032. <https://doi.org/10.1016/j.combiomed.2024.109032>
13. Lyu J, Yang Y, Zong Y, Leng Y, Zheng W, Ge S (2024) Novel sinusoidal signal assisted multivariate variational mode decomposition combined with Task-Related component analysis for enhancing SSVEP-Based BCI performance. *IEEE J Biomed Health Inform* 28(11):6474–6485. <https://doi.org/10.1109/JBHI.2024.3439391>
14. Park D, Kim I (2022) Application of machine learning in the field of intraoperative neurophysiological monitoring: A narrative review. *Appl Sci* 12(15):7943. <https://doi.org/10.3390/app12157943>
15. Mejri N, Lopez-Fuentes L, Roy K, Chernakov P, Ghorbel E, Aouada D (2024) Unsupervised anomaly detection in time-series: An extensive evaluation and analysis of state-of-the-art methods. *Expert Syst Appl* 256:124922. <https://doi.org/10.1016/J.ESWA.2024.124922>
16. Li G, Jung JJ (2023) Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion* 91:93–102. <https://doi.org/10.1016/j.inffus.2022.10.008>
17. Mandir A, Ebinger K, DeBruyn L, Kenney K (2024) A prospective look at the prevalence of setup Electrode-Swap errors across over 450 intraoperative neuromonitoring cases. *The Neurodiagnostic Journal* 64(3):96–111
18. Guzzi G, Ricciuti RA, Della Torre A, Lo Turco E, Lavano A, Longhini F, La Torre D (2024) Intraoperative neurophysiological monitoring in neurosurgery. *J Clin Med* 13(10):2966
19. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
20. Liu FT, Ting KM, Zhou Z-H (2008) Isolation forest. In: 2008 8th IEEE International Conference on Data Mining, pp 413–422. <https://doi.org/10.1109/icdm.2008.17>
21. Lubba CH, Sethi SS, Knaute P, Schultz SR, Fulcher BD, Jones NS (2019) catch22: CAnonical time-series CHaracteristics: Selected through highly comparative time-series analysis. *Data Min Knowl Disc* 33:1821–1852. <https://doi.org/10.1007/s10618-019-00647-x>
22. Wilson JPI, Kumbhare D, Ronkon C, Guthikonda B, Hoang S (2023) Application of machine learning strategies to model the effects of sevoflurane on Somatosensory-Evoked potentials during spine surgery. *Diagnostics (Basel)* 13(21):3389. <https://doi.org/10.3390/diagnostics13213389>
23. Wermelinger J, Parduzi Q, Sariyar M, Raabe A, Schneider UC, Seidel K (2023) Opportunities and challenges of supervised machine learning for the classification of motor evoked potentials according to muscles. *BMC Med Inform Decis Mak* 23(1):198
24. Han S, Hu X, Huang H, Jiang M, Zhao Y (2022) Adbench: Anomaly detection benchmark. *Advances in Neural Information Processing Systems (NeurIPS)* 35:32142–32159

25. Jin B, Chen Y, Li D, Poolla K, Sangiovanni-Vincentelli A (2019) A One-Class support vector machine calibration method for time series change point detection. In: 2019 IEEE International Conference on Prognostics and Health Management (ICPHM), pp. 1–5. <https://doi.org/10.1109/ICPHM.2019.8819385>
26. Yaacob AH, Tan IKT, Chien SF, Tan HK (2010) ARIMA Based Network Anomaly Detection, pp. 205–209. <https://doi.org/10.1109/ICCSN.2010.55>
27. Malhotra P, Vig L, Shroff G, Agarwal P et al (2015) Long short term memory networks for anomaly detection in time series. In: Esann, vol 2015, p 89
28. Audibert J, Michiardi P, Guyard F, Marti S, Zuluaga MA (2020) Usad: Unsupervised anomaly detection on multivariate time series. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp 3395–3404
29. Combi C, Amico B, Bellazzi R, Holzinger A, Moore JH, Zitnik M, Holmes JH (2022) A manifesto on explainability for artificial intelligence in medicine. *Artif Intell Medicine* 133:102423. <https://doi.org/10.1016/J.ARTMED.2022.102423>
30. Wu Z, Wu D, Luo B, Jiang S, Zou X, Chen Y (2026) Medical time series classification based on multi-feature fusion self-organized operational neural network. *Biomed Signal Process Control* 113:109211. <https://doi.org/10.1016/J.BSPC.2025.109211>
31. Gabrielli D, Prenkaj B, Velardi P (2026) Seamless monitoring of stress levels leveraging a foundational model for time sequences. *Artif Intell Medicine* 173:103336. <https://doi.org/10.1016/J.ARTMED.2025.103336>
32. Ratanamahatana, C.A., Keogh, E.J (2005) Three myths about dynamic time warping data mining. In: Kargupta H, Srivastava J, Kamath C, Goodman A (eds) Proceedings of the 2005 SIAM International Conference on Data Mining, SDM 2005, Newport Beach, CA, USA, April 21–23, 2005, pp 506–510. <https://doi.org/10.1137/1.9781611972757.50>
33. Nestic S, Putina A, Bahri M, Huet A, Navarro JM, Rossi D, Sozio M (2022) STREAMRHF: Tree-Based unsupervised anomaly detection for data streams. 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA), pp 1–8. IEEE
34. Xu X, Liu M, Cheng X, Song L, Zhang J (2024) Online anomaly detection for streaming data in the presence of missing values. 2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp 3956–3961. IEEE
35. Lu T, Wang L, Zhao X (2023) Review of anomaly detection algorithms for data streams. *Appl Sci* 13(10):6353
36. Deng H, Runger G, Tuv E, Vladimir M (2013) A time series forest for classification and feature extraction. *Inf Sci* 239:142–153. <https://doi.org/10.1016/j.ins.2013.02.030>
37. Middlehurst M, Large J, Bagnall A (2020) The canonical interval forest (CIF) classifier for time series classification. In: Wu X, Jermaine C, Xiong L, Hu X, Kotevska O, Lu S, Xu W, Aluru S, Zhai C, Al-Masri E, Chen Z, Saltz J (eds) 2020 IEEE International Conference on Big Data (big Data), pp 188–195. IEEE. <https://doi.org/10.1109/bigdata50022.2020.9378424>, [arXiv:2008.09172](https://arxiv.org/abs/2008.09172)
38. Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and Regression Trees 40:874
39. Mensi A, Bicego M, Tax DMJ (2020) Proximity isolation forests. In: 25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10–15, 2021, pp 8021–8028. <https://doi.org/10.1109/ICPR48806.2021.9412322>
40. Preiss BR (1999) Data Structures and Algorithms with Object-Oriented Design Patterns in Java. <https://ci.nii.ac.jp/ncid/BA44302441>
41. Samariya D, Thakkar A (2023) A comprehensive survey of anomaly detection algorithms. *Annals of Data Science*
42. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research (JMLR)* 7(1):1–30
43. Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6):80. https://doi.org/10.1007/978-1-4612-4380-9_16
44. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC (2001) Estimating the support of a High-Dimensional distribution. *Neural Comput* 13(7):1443–1471
45. Kriegel H, Kröger P, Schubert E, Zimek A (2009) Outlier detection in Axis-Parallel subspaces of high dimensional data. In: Theeramunkong T, Kijssirikul B, Cercone N, Ho TB (eds) Advances in Knowledge Discovery and Data Mining, 13th Pacific-Asia Conference, PAKDD 2009, Bangkok, Thailand, April 27–30, 2009, Proceedings, vol 5476, pp 831–838. https://doi.org/10.1007/978-3-642-01307-2_86
46. Breunig MM, Kriegel H, Ng RT, Sander J (2000) LOF: identifying Density-Based local outliers. In: Chen W, Naughton JF, Bernstein PA (eds) Proceedings of the 2000 ACM SIGMOD International

- Conference on Management of Data, May 16–18, 2000, Dallas, Texas, USA, pp 93–104. <https://doi.org/10.1145/342009.335388>, <https://dl.acm.org/doi/pdf/10.1145/335191.335388>
47. Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, pp 359–370
 48. Goldstein M, Dengel A (2012) Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. KI-2012: poster and demo track
 49. Li Z, Zhao Y, Hu X, Botta N, Ionescu C, Chen GH (2023) ECOD: unsupervised outlier detection using empirical cumulative distribution functions. *IEEE Trans Knowl Data Eng* 35:12181–12193. <https://doi.org/10.1109/tkde.2022.3159580>
 50. Rousseeuw PJ, Driessen K (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41(3):212–223
 51. Azzari A, Bicego M, Combi C, Cracco A, Sala P (2025) TSRF-Dist: a novel time series distance based on extremely randomized canonical interval forests. *Data Min Knowl Disc* 39(3):27. <https://doi.org/10.1007/s10618-025-01098-3>
 52. Mietus JE (2002) The pNNx files: re-examining a widely used heart rate variability measure. *Heart* 88(4):378–380. <https://doi.org/10.1136/heart.88.4.378>
 53. Wang X, Wirth A, Wang L (2007) Structure-based statistical features and multivariate time series clustering. In: Proceedings—IEEE International Conference on Data Mining, ICDM, pp 351–360. <https://doi.org/10.1109/icdm.2007.103>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Alberto Azzari¹  · Alessandro Boaro² · Federica Basaldella³ · Sonia Nunes² · Francesco Sala² · Carlo Combi¹ · Pietro Sala¹ · Manuele Bicego¹

✉ Pietro Sala
pietro.sala@univr.it

Alberto Azzari
alberto.azzari@univr.it

Alessandro Boaro
alessandro.boaro@univr.it

Federica Basaldella
federica.basaldella@gmail.com

Sonia Nunes
sonia.nunes@univr.it

Francesco Sala
francesco.sala@univr.it

Carlo Combi
carlo.combi@univr.it

Manuele Bicego
manuele.bicego@univr.it

¹ Department of Computer Science, University of Verona, Verona 37134, VR, Italy

² Section of Neurosurgery, Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona 37134, VR, Italy

³ Division of Neurology, University Hospital at the University of Verona, Verona 37134, VR, Italy