Computer Science and Information Systems 23(2):757–774 https://doi.org/10.2298/CSIS250913020B

Enhanced ROCKET for the Automated Detection of Epileptic Tonic-Clonic Seizures Using Accelerometer Data*

Krisztian Buza[12], Alexandros Nanopoulos[3], and Noémi Ágnes Varga[4]

[1] Budapest University of Economics and Business
Budapest, Hungary
buza.krisztian@uni-bge.hu (corresponding author)
[2] BioIntelligence Group, Department of Mathematics-Informatics
Faculty of Technical and Human Sciences, Sapientia Hungarian University of Transylvania
Targu Mures, Romania
buza@biointelligence.hu
[3] Business Informatics, Baden-Wuerttemberg Cooperative State University
Mosbach, Germany
alexandros.nanopoulos@gmail.com
[4] Institute of Genomic Medicine and Rare Disorders, Semmelweis University
Budapest, Hungary
noemiagnesvarga@gmail.com

Abstract. The detection of epileptic tonic-clonic seizures during everyday life based on accelerometric data from wearable devices would enhance the diagnostic and the follow-up of the epileptic patients. We develop an algorithm which may contribute to recognition of tonic-clonic epileptic seizure based on accelerometer data that can be collected from mobile and wearable devices. We consider this task to be a multivariate time-series classification problem. State-of-the art solutions to this problem are based on machine learning techniques, such as Random Convolutional Kernel Transform (ROCKET). We enhance ROCKET by replacing standard convolution with dynamic convolution. Dynamic convolution was originally defined for univariate time series, therefore, we extend it to multivariate time series. We perform experiments on two publicly available real-world datasets related to tonic-clonic seizures. The experimental results show that the proposed enhancements of the ROCKET algorithm significantly reduce the average classification error. Moreover, our approach outperforms other time series classifiers, including several types of deep neural networks that are commonly used in the domain of time-series classification. An enhanced version of the ROCKET algorithm is proposed for the automated detection of epileptic tonic-clonic seizures using accelerometer data. To assist reproducibility and follow-up works, we made our implementation publicly available at https://github.com/kr7/seizure .

Keywords: Epilepsy, Multivariate Time Series Classification, Human Activity Recognition, Random Convolutional Kernel Transform, ROCKET, Dynamic Convolution

* This is an extended version of the conference paper [9].

1. Introduction

WHO reports that around 50 million people are diagnosed with epilepsy based on recurrent seizures worldwide.[5] A single unprovoked seizure is a common phenomenon in the general population, with at least 10% of the population experiencing a seizure during their lifetime [7]. The rate of seizure recurrence can vary widely, but it is estimated that up to 70% of people living with epilepsy could live seizure-free if properly diagnosed and treated.

The detection and monitoring of epileptic seizures require specialized devices such as electroencephalography (EEG) and well-trained experts to interpret the results. However, in the majority of cases, the first seizures happen outside a healthcare unit. Since the patients are unconscious or partially conscious during the seizure, the description of the seizure originates from family members or stranger eyewitnesses. Patient-reported outcomes (PROs) in the form of seizure diaries are commonly used methods for monitoring the disease activity, but diaries are limited due to variable adherence and periictal amnesia [21].

With the advent of clinically relevant data that can be collected by mobile and wearable devices, such as accelerometer data, the wish for reliable control during continuous monitoring, appropriate identification of seizures, and timely alerting became more pertinent and feasible [42].

Due to its evident importance, in this manuscript, we focus on the research topic of recognizing tonic-clonic epileptic seizures[6] based on accelerometer data. Our goal is a self-detection mechanism that is both: i) more accessible, due to the use of technology that is commonly available in mobile and wearable devices; ii) more acceptable, due to more accurate recognition with a limited amount of false alarms, which can be a considerable burden for both patients and medical staff [39], as it has been identified in other kinds of self-detection mechanisms [18], too. Both these factors can increase the detection coverage in broader sets of the population. Moreover, such a self-detection mechanism can become complementary to more precise EEG-based detection, effectively acting as a feasible form of pre-screening.

Within this context, we consider the automated recognition of epileptic seizure as a time-series classification task, for which state-of-the-art solutions are based on machine learning. The underlying time series are multivariate as the acceleration is usually measured along several axes. In particular, in our case, the acceleration is measured along three axes, therefore, we work with 3-dimensional time series. Our solution is based on "Random Convolutional Kernel Transform" or ROCKET for short [15]. ROCKET is a recent time-series classification algorithm that was shown to outperform other time-series classifiers. We extend ROCKET by replacing conventional convolution with *dynamic convolution* [10]. Dynamic convolution replace standard dot-product calculations of ROCKET by dynamic time warping (DTW) calculations. The resulting classifier is thereby more robust w.r.t. local shifts and elongations that are present in time-series data. Dynamic convolution was originally defined for univariate time series, nevertheless accelerometer data corresponds to multivariate time series. Thus, we propose a new variant of dynamic convolution that works with multivariate time series.

[5] https://www.who.int/news-room/fact-sheets/detail/epilepsy

[6] Tonic-clonic epileptic seizures involve both tonic (stiffening) and clonic (twitching) phases of muscle activity.

In our experimental evaluation, we used two publicly available real-world datasets. We performed experiments according to the 10×10-fold cross-validation protocol. The experimental results show that the proposed enhancements of the ROCKET algorithm significantly reduce the average error rate (misclassification ratio) of the detection of epileptic tonic-clonic seizures.

We compared our approach to other time-series classifiers too, including several prominent deep neural networks and we observed that our approach outperforms them as well. More importantly, the amount of false alarms is also reduced drastically, which, as mentioned above, is a critical requirement for the acceptance of the envisioned self-detection mechanism.

The remainder of the paper is organized as follows: Section 2 provides on overview of related works. Section 3 introduces the background that is necessary to understand our work. Section 4 presents our approach in detail which is followed by the experimental evaluation (Section 5) and conclusions (Section 6).

2. Related Work

Machine learning techniques have been widely applied to detection and classification tasks in the biomedical domain, including emotion recognition [44], the assessment of schizophrenia [30], and the detection as well as early prediction of epileptic seizures [1, 3, 12, 22, 26, 33, 50]. A comprehensive survey of epileptic seizure detection methods is provided in [19]. Although electroencephalography (EEG) has demonstrated strong performance in controlled laboratory environments, its applicability in everyday settings remains limited. In particular, Baumgartner and Koren [3] note that, in outpatient contexts, scalp-EEG–based seizure detection is constrained by poor patient acceptance, as individuals are unlikely to tolerate wearing electrode arrays during daily activities.

The increasing availability of low-cost sensors and the widespread adoption of smart devices, such as smartphones and smartwatches, have stimulated growing interest in the automated detection of tonic-clonic seizures using accelerometer data. Several studies have explored this direction [8, 39, 6, 35, 34]. Bruno et al. [8] evaluated both medically certified and non-certified devices, with particular emphasis on the requirements of patients and caregivers. Regalia et al. [39] introduced what they described as the first commercially available multimodal wrist-worn devices designed to capture the physiological signatures of ongoing generalized tonic-clonic seizures. Beniczky et al. [6] investigated the clinical reliability of a wireless, wrist-worn accelerometer for detecting generalized tonic-clonic seizures, while Onorati et al. [35] examined classifiers based on accelerometer signals. More recently, Lupión et al. [34] analyzed seizure detection using low-cost IoT devices in combination with federated machine learning techniques.

The majority of these studies emphasize practical considerations related to seizure detection in daily life, such as device usability, patient comfort, and user experience. While these aspects are undoubtedly important, the present work focuses on the underlying machine learning methodologies used for classification. In prior research, classification models were often treated as black boxes, with existing algorithms applied without detailed justification, or classifiers constructed using manually engineered features [35], frequently without in-depth discussion of the employed learning techniques. In contrast, we investigate advanced classification approaches, including deep learning models and state-of-the-art methods such as ROCKET.

As we formulate the automated recognition of epileptic tonic-clonic seizures as a time series classification problem, we briefly review relevant work in this area. Time series classification constitutes a unifying framework [23] for a wide range of detection and recognition tasks and has given rise to numerous methodological approaches. These include neural network–based models [20, 25, 48], Bayesian networks [36], hidden Markov models [17], decision trees [27], and methods based on frequent pattern discovery, commonly referred to as motifs or shapelets [47, 24, 32]. An additional line of research includes hubness-aware classifiers [38].

An influential study by Xi et al. [46] demonstrated that a simple k-nearest neighbor classifier combined with dynamic time warping (DTW) distance was highly competitive with, and in some cases superior to, many alternative classifiers proposed prior to that work. The effectiveness of DTW arises from its elastic alignment mechanism, which accommodates temporal shifts and local variations in length between time series.

Subsequent advances in time series classification increasingly leveraged deep learning techniques [48, 14, 49]. Among these, fully convolutional networks (FCNs) have been identified as a particularly strong baseline [45]. Wang et al. [45] reported that FCNs outperformed all other evaluated time series classifiers in terms of overall accuracy, with residual networks (ResNets) achieving comparable performance.

Despite their widespread adoption, convolutional neural networks (CNNs) exhibit inherent limitations in time series analysis. As discussed in [10], CNNs are primarily suited to rigid pattern matching. Even when convolutional layers are combined with pooling operations, the resulting architectures mainly handle translational invariance and fail to adequately accommodate temporal elongations of local patterns. Furthermore, their ability to address translations remains limited and inconsistent. These shortcomings motivated the development of dynamic convolution [10], which seeks to overcome such constraints.

More recently, the Random Convolutional Kernel Transform (ROCKET) was proposed as a highly effective approach to time series classification [15]. ROCKET and its deterministic variant, MiniROCKET [16], have demonstrated performance that surpasses many deep learning–based models, including FCNs. Nevertheless, both methods rely on standard convolution operations, and therefore inherit their associated limitations. In this work, we address these constraints and propose an approach designed to overcome them.

Preliminary investigations of the integration of dynamic convolution with ROCKET were presented in [11]. However, that work focused exclusively on univariate time series classification. In contrast, the present study addresses the detection of epileptic tonic-clonic seizures using multivariate accelerometer data. For completeness, we note that an earlier version of this work was reported in [9], outlining the core idea. The current paper substantially extends this preliminary study by broadening the experimental evaluation to additional datasets, comparing the proposed approach with a wider range of state-of-the-art methods, and providing a detailed analysis of false alarm rates.

3. Background

Next, we present a brief overview of the key concepts and techniques essential for understanding our work. Specifically, we begin with a formal definition of the time series classification task, followed by a review of related works (Section 2), dynamic time warping (Section 3.2), dynamic convolution (Section 3.3), and ROCKET (Section 3.4). Finally, we present our approach in Section 4 and the datasets used in the experimental evaluation (Section 5.1).

3.1. Problem Formulation (Multivariate Time Series Classification)

Given a set $\mathcal{C}$ of class labels, and a set $\mathcal{D}$ (called training set) of time series together with their class labels

$$\mathcal{D} = \left\{ \left(x^{(i)}, y^{(i)}\right)_{i=1}^{n} \right\}, \;\; y^{(i)} \in \mathcal{C} \tag{1}$$

where $x^{(i)}$ is a multivariate time series. As in our case, the acceleration is measured along three axis, therefore

$$x^{(i)} = \Big((x_{1,1}^{(i)}, x_{1,2}^{(i)}, x_{1,3}^{(i)}), \ldots, (x_{l,1}^{(i)}, x_{l,2}^{(i)}, x_{l,3}^{(i)}) \Big),$$

where each $x_{j,k}^{(i)}$ is a real number, $j \in \{1, 2, ..., l\}$ and $k \in \{1, 2, 3\}$. We aim at finding a model $\mathcal{M}$ that is able to determine the class label $y' \in \mathcal{C}$ of any new (test) time series x'.

In this study, in line with the above problem formulation, we assume that tonic-clonic seizures are detected based on an acceleration time series of length l. In the case of a real-world application, this could be achieved by considering both the current observation and the previous $l-1$ observations as the input of $\mathcal{M}$.

3.2. Dynamic Time Warping

Dynamic Time Warping (DTW) is an elastic distance measure for time series, built on the principles of dynamic programming [41]. It enables shifts and elongations when aligning two time series, allowing for a more flexible comparison.

The DTW distance of two time series, $x = (x_1, \ldots, x_l)$ and $x' = (x'_1, \ldots, x'_{l'})$, is computed by populating an $l \times l'$ matrix D. Each entry $d_{i,j} \in D$ represents the distance between two prefixes—one from x and the other from x'. Specifically, $d_{i,j}$ is the distance between the prefixes $(x_1, \ldots, x_i)$ and $(x'_1, \ldots, x'_j)$. It is computed as follows:

$$d_{i,j} = |x_i - x'_j| + \min \{d_{i,j-1}, d_{i-1,j}, d_{i-1,j-1}\} \tag{2}$$

where the terms of the minimum correspond to the cases of elongation in x, elongation in x' or matching the next elements in both time series.[7]

The matrix entries $d_{i,j}$ are computed in a column-wise manner, following the sequence: $d_{1,1}$, $d_{2,1}$, $\ldots$, $d_{l,1}$, $d_{1,2}$, $d_{2,2}$, $\ldots$, $d_{l,2}$, and so forth until $d_{l,l'}$. The initialization begins with the first entry, defined as $d_{1,1} = |x_1 - x'_1|$. When certain terms, such as $d_{i,j-1}$, $d_{i-1,j}$, or $d_{i-1,j-1}$, are undefined (i.e., if $i-1=0$ or $j-1=0$), they are ignored. In such cases, the minimum in Eq. (2) is computed only over the defined terms. Finally, the DTW distance between the two time series is given by $d_{l,l'}$.

[7] Instead of $|x_i - x'_j|$, one can calculate $(x_i - x'_j)^2$ in Eq. (2). In our study, we used the variant with $|x_i - x'_j|$.

3.3. Dynamic convolution

In time series classifiers like ROCKET, convolution serves as a local pattern detector. As discussed in [10], the combination of convolution and max pooling provides limited flexibility in pattern matching by reducing sensitivity to the exact position of a pattern within the time series. Specifically, max pooling can only maintain consistent outputs when a pattern shifts within its pooling window. If a pattern shifts beyond this range, the detection may be affected. Additionally, conventional convolution is unable to handle more complex temporal distortions, such as elongations of local patterns. These limitations motivated the development of dynamic convolution, which enhances pattern matching by incorporating time-warping mechanisms.

The core concept of dynamic convolution is to replace the standard dot product (or inner product) used in conventional convolution with the computation of Dynamic Time Warping (DTW) distances between the convolutional kernel and segments of the time series. This approach allows for more flexible pattern matching by accounting for temporal distortions, such as shifts and elongations within local patterns.

3.4. ROCKET

The "ROCKET" time series classifier begins by generating a predefined number (F) of random convolutional kernels (filters). By default, ROCKET utilizes $F = 10,000$ convolutional kernels. The parameters of each kernel, including its length, weights, bias, and dilation, are randomly sampled from appropriate distributions. Specifically:

- *The window size* (w) of the convolutional kernel is chosen uniformly from $\{7, 9, 11\}$.
- *Weights* are drawn from a standard normal distribution (mean = 0, standard deviation = 1).
- *Bias* is sampled uniformly from the interval $(-1, 1)$.
- *Dilation* is set to $\lfloor 2^d \rfloor$, where d is selected uniformly from the range $(0, \lfloor \log_2((l - 1)/(w - 1)) - 1 \rfloor)$, with l representing the length of the input time series.
- *Zero padding* is applied with a probability of 0.5 for each convolution.

This randomized kernel generation process allows ROCKET to efficiently capture diverse temporal features while maintaining computational efficiency.

Given a *single* input time series (whether from the training or test set), ROCKET applies convolution with all F randomly generated kernels, producing F convoluted time series. For each of these F convoluted time series, ROCKET performs two global pooling operations:

- *Global Max Pooling:* Extracts the maximum value from the convoluted time series.
- *Global PPV Pooling:* Computes the *Proportion of Positive Values* (PPV), which represents the fraction of values in the convoluted time series that are positive.

As a result, each input time series is transformed into a feature vector containing $2F$ real-valued features (by default, 20,000 features when $F = 10,000$).

For classification, Dempster et al. [15] apply *ridge regression* when the dataset contains fewer than $2F$ time series or *logistic regression* when the dataset is larger.

Despite its simplicity, ROCKET has been shown to outperform various deep learning-based methods, including fully convolutional networks (FCNs). This strong performance may be attributed to its ability to efficiently capture diverse temporal patterns through a large number of random convolutional kernels.

In their follow-up work, Dempster et al. [16] found that the features resulting from "PPV pooling" are more important than the ones resulting from max pooling. In this study, we use ROCKET-PPV and ROCKET-MAX to denote those versions of ROCKET that use only the features resulting from "PPV pooling" and max pooling, respectively.

Ruiz et al. [40] noted that the multivariate extension of ROCKET, at least the one included in the sktime software library, assigns kernels to channels randomly. Although this may be useful in certain cases, this approach inherently assumes that channels are independent. Therefore, we consider an alternative version of ROCKET for multivariate time series. In particular, we generate multivariate kernels. Whenever the opposite is not explicitly stated, we use ROCKET with multivariate kernels. In the experimental evaluation, we use "ROCKET-UNI" to denote the variant that generates univariate kernels and assigns those univariate kernels to channels randomly.

4. Our Approach

In this section, we describe in detail how the proposed approach enhances ROCKET.

Given ROCKET's F convolutional kernels, we propose to calculate dynamic convolution instead of conventional convolution and apply global max pooling to the convoluted time series.

Originally, dynamic convolution was defined for univariate time series. We handle multivariate time series as follows: we generate multivariate convolutional kernels and use them when calculating convolutions. In particular, for each segment of the time series, we calculate the DTW distances for each channel separately and sum the DTW distances of the different channels. This is illustrated in Fig. 1.

As mentioned previously, in case of the original ROCKET approach, global max pooling and PPV ("portion of positive values") pooling are applied to the convoluted time series. As DTW distances are nonnegative, we omit PPV pooling. In case of conventional convolution, the role of max pooling is to select the similarity (particularly, the value of the dot product) of the time series segment that is most similar to the pattern associated with the convolutional kernel. In contrast to dot product, in case of DTW, low values mean high similarity, and high values correspond to substantial difference between the segment and the pattern associated with the kernel. Therefore, we use *min pooling* instead of max pooling.

Calculating multivariate dynamic convolution with F kernels and applying global max pooling results in F features.

Using this representation of time series, according to which each time series is represented as an F-dimensional feature vector, similarly to ROCKET, we train ridge regression or logistic regression to classify time series. Ridge regression is used in case if the number of time series in the dataset is less than the number of features, while logistic regression is used otherwise.

The computational complexity of the calculation of dynamic convolution with window size w for a time series of constant length is $\mathcal{O}(w^2)$, while the complexity of "conventional" convolution is $\mathcal{O}(w)$. This is due to the fact that DTW calculations are quadratic in the length of the input time series whereas the scalar product can be calculated in linear

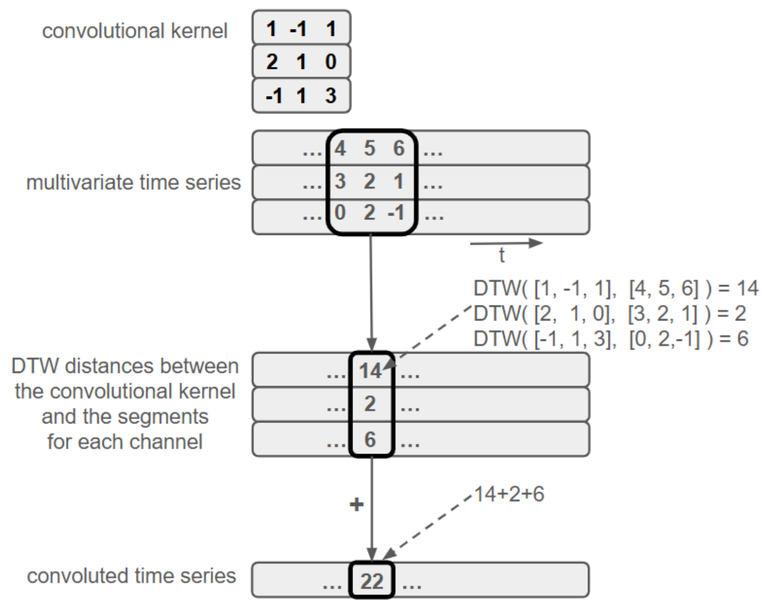

Fig. 1. Illustration of multivariate dynamic convolution

time. However, as detailed in the previous section, the window size w of the convolutional kernel is usually set to a small value, such as $w = 7, 9$ or 11, thus the actual runtime of the approach only increases moderately. On the other hand, for the case of large convolutional windows, we point out that there are various approaches to speed up DTW calculations, such as limiting the size of the warping window. (Please note that the warping window of DTW is different from the aforementioned convolutional window.) If setting the size of the warping window to a constant value, the theoretical complexity of DTW calculations becomes $\mathcal{O}(w)$ which is the same as the theoretical complexity of scalar product calculations in the "conventional" convolution.

5. Experimental Evaluation

The goal of our experiments is to assess the accuracy of the proposed approach for the recognition of epileptic tonic-clonic seizures and to compare it to relevant variants of ROCKET as well as to other prominent approaches from the literature.

5.1. Data

We performed experiments on two datasets[8], in particular: (i) the "Epilepsy" dataset [43] and (ii) the "OpenSeizure" dataset [28].

[8] See also https://timeseriesclassification.com/description.php?Dataset=Epilepsy and https://ieee-dataport.org/documents/open-seizure-database-v100 for more information about the datasets.

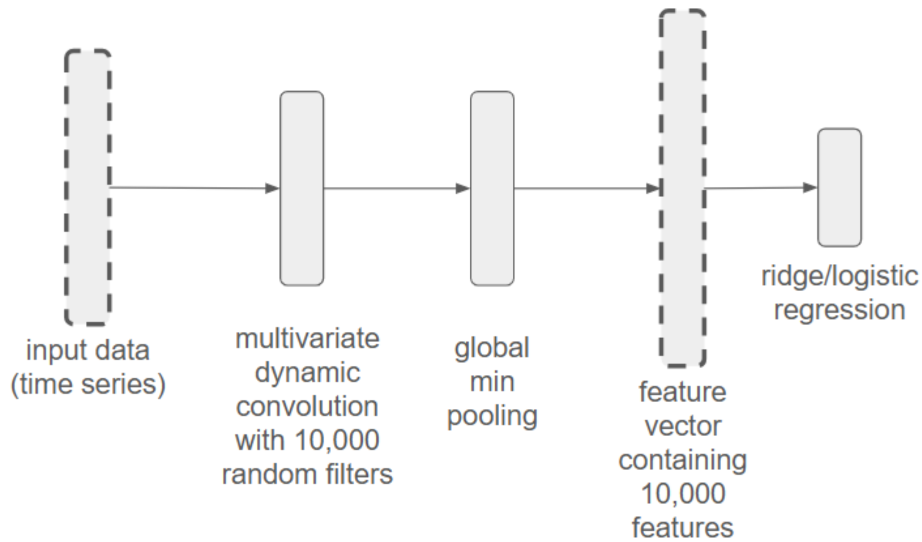

Fig. 2. Our approach, EROCKET, for the classification of multivariate time series. It takes a time series as input and performs multivariate dynamic convolution using F random kernels and global min pooling. This results in a vector of F features in total. This representation is used in ridge regression or logistic regression.

Epilepsy According to the description at timeseriesclassification.com, "the data contains multivariate time series measured by an accelerometer." It was collected from six healthy participants using a tri-axial accelerometer on the dominant wrist while they were conducting four different activities.

These activities are walking, running, sawing, and seizure mimicking. *Walking* includes different paces and gestures: walking slowly while gesturing, walking slowly, walking normally, and walking fast, each of which lasts for 30 seconds. *Running* refers to running a 40-meter-long corridor. *Sawing* was performed with a saw for 30 seconds. In the case of *seizure mimicking*, the participants were seated, and the seizure itself lasted for 30 seconds, but an additional 5-6 seconds before and 30 seconds after the mimicked seizure were recorded. Each participant performed each activity 10 times at least. The mimicked seizures were trained and controlled, following the protocol defined by a medical expert. All the activities were carried out indoors, either inside an office or in the corridor around it. The sampling frequency was 16 Hz. Some activities lasted about 30 seconds, others up to 2 minutes.

The data was truncated to the length of the shortest time series. Prefixes and suffixes of flat series were truncated to the shortest series ($\approx$13 seconds), taking a random interval of activity for series longer than the minimum. A single case from the original data (ID002 Running 16) was removed because the data was not collected correctly. After preprocessing, the data contains a total of 275 accelerometer time series, each of which has a length of $l = 206$. Out of the 275 time series, 68 belong to the class of seizure mimicking (or epilepsy), while 74, 73, and 60 belong to the classes of walking, running, and sawing, respectively. One of the accelerometer time series from each of the classes is shown in Fig. 3.

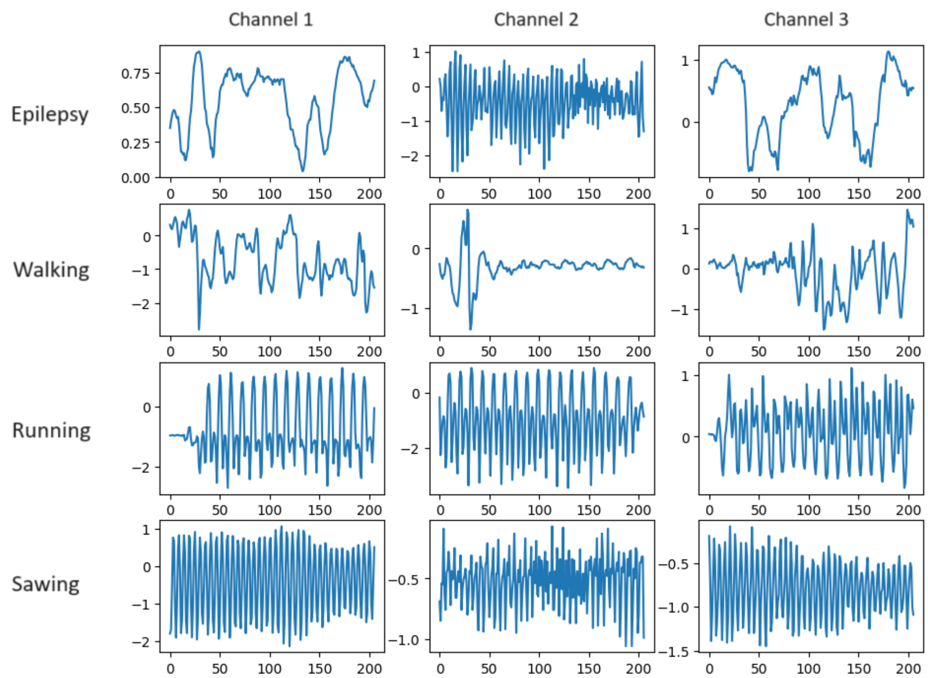

Fig. 3. An excerpt of the considered dataset: one of the accelerometer time series from each of the classes. Please note that we consider multivariate time series with three channels, each of which is shown separately.

OpenSeizure The OpenSeizure dataset is a "publicly accessible resource designed to advance non-electroencephalogram seizure detection research".[9] It contains multimodal sensor data observed in real-world, in-home environments. From the dataset, we considered tonic-clonic seizures that had associated 3-dimensional time series data. As the dataset contains labeled false alarms as well, we selected the same amount of false alarms with 3-dimensional time series. For each tonic-clonic seizure and for each of the selected false alarms, we selected 500 time points from the middle of the 3-dimensional time series. Each time series was normalized to have a mean of zero and a standard deviation of one. The script we used to preprocess the data is available in the same repository where we published the implementation of our approach, see Section 5.5.

[9] https://ieee-dataport.org/documents/open-seizure-database-v100

As we consider the classification task of distinguishing tonic-clonic seizures from false alarms, the OpenSeizure dataset is inherently more challenging compared to the Epilepsy dataset.

5.2. Baselines

Our approach is based on ROCKET, therefore, we compare its accuracy to ROCKET and its variants ROCKET-PPV, ROCKET-MAX and ROCKET-UNI (see Section 3.4). To obtain a more complete evaluation, we additionally examine other standard baseline time-series classifiers: (i) multilayer perceptron (MLP), (ii) a simple convolutional neural network (CNN), (iii) fully convolutional network (FCN), (iv) a residual network (ResNet), and (v) a classifier based on the transformer architecture (Transformer).

Although recent time series classifiers outperform multilayer perceptrons, we decided to include MLP in the experimental evaluation because feed-forward neural networks with at least one hidden layer and non-linear activation are known to be universal function approximators [5]. The MLP considered in our experiments contains a single hidden layer with 100 units and ReLU activation and an output layer with softmax activation.

The simple convolutional neural network (CNN) has a single convolutional layer, followed by a fully connected layer with 32 units and the output layer in which each unit corresponds to one of the classes. The convolutional layer contains 16 kernels with size of 32. We used softmax activation in the output layer and ReLU activation in the convolutional and fully connected layers.

The architectures of FCN and ResNet models are based on the study of Wang et al. [45], nevertheless, we omitted batch normalization, because we observed substantially higher accuracy without batch normalization both in case of FCN and ResNet.

Inspired by the success of transformer-based architectures for time series classification, see e.g. ShapeFormer [31], we implemented a computationally efficient transformer for time series classification. For simplicity, we refer to this model as Transformer. Our implementation of Transformer, as well as the implementations of other baselines are included in our code base, which we have made publicly available, see Section 5.5.

In case of the Epilepsy dataset, we trained the aforementioned neural networks (MLP, CNN, FCN, ResNet and Transformer) for 1000 epochs with a learning rate of 10^{-5} and used the Adam optimizer [29] with cross-entropy loss and set the batch size to 16. In case of the OpenSeizure dataset, we trained the neural networks for 100 epochs with a learning rate of 10^{-4} and set the batch size to 32. According to our observations, these settings allowed the neural networks to converge in all the examined cases.

5.3. Experimental Protocol

We performed experiments according to the 10×10-fold cross-validation protocol. With 10-fold cross-validation, we mean that we partition the data into 10 disjoint splits, one of these splits is used as test data, while the other 9 splits are used as training data. In the case of 10-fold cross-validation, the experiment (i.e., training and evaluation of the classifier) is repeated 10 times with a different split being used as test data each time. In case of 10×10-fold cross-validation, the original random splitting of the data is performed 10 times, therefore 10 instances of 10-fold cross-validation are performed in total.

5.4. Evaluation metrics

We used average classification error, i.e., the ratio of incorrectly classified time series, to assess the performance of our approach and the baselines. Lower values indicate better performance.

We report classification error averaged over the $10 \times 10 = 100$ folds of the cross-validation together with its standard deviation both for our approach and the baselines.

Additionally, we used paired t-test to assess whether the differences between our approach and the baselines – in terms of classification error – are statistically significant or not. We used the Nadeau and Bengio correction[10] to account for the overlapping among the repeated folds.

As low false alarm rates are of paramount importance, see e.g. [39], we consider the epilepsy class (seizure mimicking) as the positive class and report False Positive Rate (FPR), i.e., the fraction of false positives among all negative instances (time series belonging to other classes).

5.5. Implementation

We implemented our approach and the baselines in Python using the numpy and pytorch software libraries. As the dataset contains less than 10,000 time series, we used ridge regression – particularly the "RidgeClassifier" from the scikit-learn machine learning library – throughout the experiments, both in the case of our approach and other variants of ROCKET.

To calculate DTW distances quickly, we implemented DTW in Cython, which combines the efficiency of C with Python's rapid prototyping [4]. We executed the experiments in Google Colab.[11] To assist reproduction of our work, independent validation of the results and to facilitate follow-up works, we published the implementation of our approach (source codes) in our GitHub repository

https://github.com/kr7/seizure

in form of an IPython notebooks.

5.6. Experimental Results

Our results on the classification error are shown in Table 1. Our approach, EROCKET, outperforms all examined (deep) neural networks and other variants of ROCKET. We also report the significance levels (p-values) of the t-tests that measure whether the differences (in terms of classification error) between our approach (EROCKET) and its competitors are significant or not.

Furthermore, we consider false alarms, which can critically burden the acceptance of self-detection by a wider set of users, as mentioned in Section 1. Figure 4 shows the false positive rate (FPR) as a percentage in case of the Epilepsy dataset. As one can see, the proposed approach attains the smallest amount of false alarms, which makes it suitable for the examined application context.

[10] https://gist.github.com/jensdebruijn/13e8eeda85eb8644ac2a4ac4c3b8e732
[11] https://colab.research.google.com

Table 1. Average classification error, its standard deviation (after the $\pm$ sign, calculated over the 10×10 folds) and the significance level (p-value) of the t-test that measures whether the difference (in terms of classification error) between our approach (EROCKET) and its competitor is significant or not.

Approach	**Epilepsy dataset**		**OpenSeizure dataset**	
	classification error	p-value	classification error	p-value
MLP	$24.11\% \pm 13.36$	≈ 0	$43.61\% \pm 12.64$	0.0019
CNN	$10.52\% \pm 5.07$	4.30×10^{-6}	$41.30\% \pm 13.99$	0.0050
FCN	$2.16\% \pm 2.59$	0.0820	$26.02\% \pm 12.53$	0.5069
ResNet	$5.60\% \pm 4.32$	0.0012	$25.05\% \pm 12.36$	0.5362
Transformer	$3.82\ \% \pm 3.36$	0.0073	$36.04\ \% \pm 14.93$	0.0232
ROCKET	$1.74\% \pm 2.20$	0.1326	$30.66\% \pm 12.51$	0.1271
ROCKET-UNI	$2.77\% \pm 2.92$	0.0324	$27.67\% \pm 12.81$	0.4053
ROCKET-PPV	$1.54\% \pm 2.54$	0.2484	$31.22\% \pm 13.42$	0.1102
ROCKET-MAX	$2.61\% \pm 3.18$	0.0596	$30.92\% \pm 12.65$	0.1218
EROCKET (our approach)	**0.40%** $\pm$ **1.14**		**22.84%** $\pm$ **12.16**	

We also examined the performance of our approach, EROCKET, and two of the strongest baselines, ROCKET and FCN in the presence of noise in the data. In particular, we added zero-mean Gaussian noise with various standard deviations $\sigma \in \{0, 0.1, 0.25\}$ to the time series of the Epilepsy dataset and repeated the experiments with noisy data. Noise was added to both training and test data, so that our experiments simulate the real-world scenario when the sensor provides noisy data. The results are shown in Fig. 5. As one can see, with increasing noise level, the performance of all the approaches decrease, nevertheless, our approach, EROCKET, consistently outperforms its competitors in case of noisy data as well.

5.7. Discussion

The primary goal of our study was to develop a classification algorithm that outperforms state-of-the-art time series classifiers for the examined research problem. Our results are encouraging and show that the proposed classifier can be effective in the context of a self-detection mechanism of epileptic tonic-clonic seizures. Nevertheless, it cannot be claimed that similar accuracy could be observed in a clinical setting: the Epilepsy dataset contains only four types of activities and it might be more challenging to distinguish tonic-clonic seizure from *any* other activity types. On the other hand, distinguishing seizures from false alarms in case of the OpenSeizure dataset is a rather challenging task. Thus, we should interpret the results in this paper as a feasibility study regarding the use of advanced machine learning algorithms, whereas a more thorough clinical application may be subject of future works. Moreover, aspects related to privacy-preserving training of EROCKET, such as training according to the federated learning paradigm [34], comprise another possible extension.

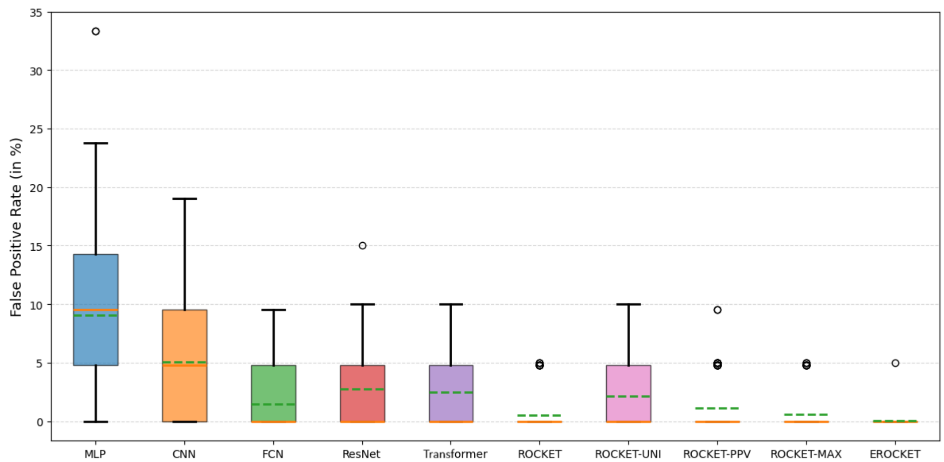

Fig. 4. Boxplots of false positive rates (FPR), measured as a percentage, for the proposed approach (EROCKET) and its competitors in case of the Epilepsy dataset. The mean FPR (over the 10×10 folds) is shown by the dashed line.

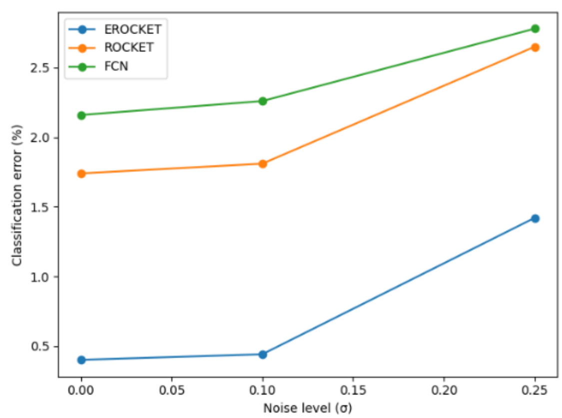

Fig. 5. Classification error of our approach (EROCKET), ROCKET and FCN in case of noisy data.

6. Conclusions

In this study, we focused on the recognition of epileptic tonic-clonic seizures based on accelerometer data. We cast this task as a time series classification problem for which recent approaches are based on "Random Convolutional Kernel Transform" (ROCKET). Considering two publicly available time series datasets containing sensor data, we observed that ROCKET achieves relatively accurate classification on its own, especially in case of the Epilepsy dataset. Nevertheless, in medical and healthcare applications, high accuracy and low FPR are essential, therefore, we enhanced ROCKET by incorporating multivariate dynamic convolution into ROCKET. This way, our approach, EROCKET, achieved an significantly lower classification error compared to the original ROCKET algorithm. Our approach not only outperforms ROCKET, but other time series classifiers as well, both in terms of classification error and FPR. Furthermore, we point out that EROCKET may be used in other applications of multivariate time-series classification, such as the classification of electrocardiograph signals [13], user verification based on mouse dynamics [2] or person authentication based on keystroke dynamics [37].

References

1. Ahmad, I., Wang, X., Javeed, D., Kumar, P., Samuel, O.W., Chen, S.: A hybrid deep learning approach for epileptic seizure detection in EEG signals. IEEE Journal of Biomedical and Health Informatics (2023)
2. Antal, M., Denes-Fazakas, L.: User verification based on mouse dynamics: a comparison of public data sets. In: 2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI). pp. 143–148. IEEE (2019)
3. Baumgartner, C., Koren, J.P.: Seizure detection using scalp-EEG. Epilepsia 59, 14–22 (2018)
4. Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D.S., Smith, K.: Cython: The best of both worlds. Computing in Science & Engineering 13(2), 31–39 (2010)
5. Bengio, Y., Goodfellow, I., Courville, A.: Deep learning, vol. 1. MIT press Cambridge, MA, USA (2017)
6. Beniczky, S., Polster, T., Kjaer, T.W., Hjalgrim, H.: Detection of generalized tonic–clonic seizures by a wireless wrist accelerometer: a prospective, multicenter study. Epilepsia 54(4), e58–e61 (2013)
7. Berg, A.T., Shinnar, S.: The risk of seizure recurrence following a first unprovoked seizure: a quantitative review. Neurology 41(7), 965–965 (1991)
8. Bruno, E., Viana, P.F., Sperling, M.R., Richardson, M.P.: Seizure detection at home: Do devices on the market match the needs of people living with epilepsy and their caregivers? Epilepsia 61, S11–S24 (2020)
9. Buza, K.: Activity recognition based on accelerometer data with enhanced rocket algorithm. In: 18th International Symposium on Applied Computational Intelligence and Informatics. pp. 321–326. IEEE (2024)
10. Buza, K., Antal, M.: Convolutional neural networks with dynamic convolution for time series classification. In: Advances in Computational Collective Intelligence: 13th International Conference, ICCCI 2021, Kallithea, Rhodes, Greece, September 29–October 1, 2021, Proceedings 13. pp. 304–312. Springer (2021)
11. Buza, K., Antal, M.: Rocket with dynamic convolution for time series classification. In: International Conference on Computational Collective Intelligence. pp. 271–282. Springer (2024)
12. Buza, K., Koller, J., Marussy, K.: PROCESS: projection-based classification of electroencephalograph signals. In: Artificial Intelligence and Soft Computing: 14th International Conference, ICAISC 2015, Zakopane, Poland, June 14-18, 2015, Proceedings, Part II. pp. 91–100. Springer (2015)
13. Buza, K., Nanopoulos, A., Schmidt-Thieme, L., Koller, J.: Fast classification of electrocardiograph signals via instance selection. In: 2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology. pp. 9–16. IEEE (2011)
14. Chen, W., Shi, K.: A deep learning framework for time series classification using relative position matrix and convolutional neural network. Neurocomputing 359, 384–394 (2019)

15. Dempster, A., Petitjean, F., Webb, G.I.: Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. Data Mining and Knowledge Discovery 34(5), 1454–1495 (2020)
16. Dempster, A., Schmidt, D.F., Webb, G.I.: Minirocket: A very fast (almost) deterministic transform for time series classification. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. pp. 248–257 (2021)
17. Esmael, B., Arnaout, A., Fruhwirth, R.K., Thonhauser, G.: Improving time series classification using hidden markov models. In: 2012 12th International Conference on Hybrid Intelligent Systems (HIS). pp. 502–507. IEEE (2012)
18. Faes, L., Islam, M., Bachmann, L.M., Lienhard, K.R., Schmid, M.K., Sim, D.A.: False alarms and the positive predictive value of smartphone-based hyperacuity home monitoring for the progression of macular disease: a prospective cohort study. Eye 35, 3035–3040 (2021)
19. Farooq, M.S., Zulfiqar, A., Riaz, S.: Epileptic seizure detection using machine learning: Taxonomy, opportunities, and challenges. Diagnostics 13(6), 1058 (2023)
20. Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.: Deep learning for time series classification: a review. Data Mining and Knowledge Discovery 33(4), 917–963 (2019)
21. Fisher, R.S., Blum, D.E., DiVentura, B., Vannest, J., Hixson, J.D., Moss, R., Herman, S.T., Fureman, B.E., French, J.A.: Seizure diaries for clinical research and practice: limitations and future prospects. Epilepsy & Behavior 24(3), 304–310 (2012)
22. Gabeff, V., Teijeiro, T., Zapater, M., Cammoun, L., Rheims, S., Ryvlin, P., Atienza, D.: Interpreting deep learning models for epileptic seizure detection on EEG signals. Artificial intelligence in medicine 117, 102084 (2021)
23. Geller, Z., Kurbalija, V., Ivanović, M.: FAP: A time series analysis and mining framework for scientific and practical applications. Computer Science and Information Systems 22(4), 1379–1403 (2025)
24. Grabocka, J., Wistuba, M., Schmidt-Thieme, L.: Fast classification of univariate and multivariate time series through shapelet discovery. Knowledge and information systems 49, 429–454 (2016)
25. Hüsken, M., Stagge, P.: Recurrent neural networks for time series classification. Neurocomputing 50, 223–235 (2003)
26. Jaiswal, A.K., Banka, H.: Epileptic seizure detection in EEG signal using machine learning techniques. Australasian physical & engineering sciences in medicine 41, 81–94 (2018)
27. Jankowski, D., Jackowski, K., Cyganek, B.: Learning decision trees from data streams with concept drift. Procedia Computer Science 80, 1682–1691 (2016)
28. Jones, G., Pordoy, J.: Open seizure database v1.0.0. IEEE Dataport (2023), https://dx.doi.org/10.21227/jftq-3e97
29. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
30. Ksiazek, K., Masarczyk, W., Glomb, P., Romaszewski, M., Buza, K., Sekula, P., Cholewa, M., Kolodziej, K., Gorczyca, P., Piegza, M.: Deep learning approach for automatic assessment of schizophrenia and bipolar disorder in patients using rr intervals. PLOS Computational Biology 21(9), e1012983 (2025)
31. Le, X.M., Luo, L., Aickelin, U., Tran, M.T.: Shapeformer: Shapelet transformer for multivariate time series classification. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 1484–1494 (2024)
32. Lines, J., Davis, L.M., Hills, J., Bagnall, A.: A shapelet transform for time series classification. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 289–297 (2012)
33. Liu, T., Truong, N.D., Nikpour, A., Zhou, L., Kavehei, O.: Epileptic seizure classification with symmetric and hybrid bilinear models. IEEE Journal of Biomedical and Health Informatics 24(10), 2844–2851 (2020)

34. Lupión, M., Sanjuan, J.F., Medina-Quero, J., Ortigosa, P.M.: Epilepsy seizure detection using low-cost iot devices and a federated machine learning algorithm. In: International Symposium on Ambient Intelligence. pp. 229–238. Springer (2022)
35. Onorati, F., Regalia, G., Caborni, C., Migliorini, M., Bender, D., Poh, M.Z., Frazier, C., Kovitch Thropp, E., Mynatt, E.D., Bidwell, J., et al.: Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors. Epilepsia 58(11), 1870–1879 (2017)
36. Pavlovic, V., Frey, B.J., Huang, T.S.: Time-series classification using mixed-state dynamic bayesian networks. In: Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). vol. 2, pp. 609–615. IEEE (1999)
37. Peška, L., Veselỳ, P., Skopal, T., Buza, K.: Person authentication using visual representations of keyboard typing dynamics. In: 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS). pp. 1–6. IEEE (2022)
38. Radovanović, M., Nanopoulos, A., Ivanović, M.: Time-series classification in many intrinsic dimensions. In: Proceedings of the 2010 SIAM International Conference on Data Mining. pp. 677–688. SIAM (2010)
39. Regalia, G., Onorati, F., Lai, M., Caborni, C., Picard, R.W.: Multimodal wrist-worn devices for seizure detection and advancing research: focus on the empatica wristbands. Epilepsy research 153, 79–82 (2019)
40. Ruiz, A.P., Flynn, M., Large, J., Middlehurst, M., Bagnall, A.: The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining and Knowledge Discovery 35(2), 401–449 (2021)
41. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing 26(1), 43–49 (1978)
42. Shah, S., Gutierrez, E.G., Hopp, J.L., Wheless, J., Gil-Nagel, A., Krauss, G.L., Crone, N.E.: Prospective multicenter study of continuous tonic-clonic seizure monitoring on apple watch in epilepsy monitoring units and ambulatory environments. Epilepsy & Behavior 158, 109908 (2024)
43. Villar, J.R., Vergara, P., Menéndez, M., de la Cal, E., González, V.M., Sedano, J.: Generalized models for the classification of abnormal movements in daily life and its applicability to epilepsy convulsion recognition. International journal of neural systems 26(06), 1650037 (2016)
44. Wang, X., Chuncao, L., Liu, Y., Liang, W., Kuanching, L., Poniszewska-Maranda, A.: A spatiotemporal graph neural network for EEG emotion recognition based on regional and global brain. Computer Science and Information Systems 22(3), 971–989 (2025)
45. Wang, Z., Yan, W., Oates, T.: Time series classification from scratch with deep neural networks: A strong baseline. In: 2017 International joint conference on neural networks (IJCNN). pp. 1578–1585. IEEE (2017)
46. Xi, X., Keogh, E., Shelton, C., Wei, L., Ratanamahatana, C.A.: Fast time series classification using numerosity reduction. In: Proceedings of the 23rd international conference on Machine learning. pp. 1033–1040 (2006)
47. Ye, L., Keogh, E.: Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. Data mining and knowledge discovery 22, 149–182 (2011)
48. Zhao, B., Lu, H., Chen, S., Liu, J., Wu, D.: Convolutional neural networks for time series classification. Journal of Systems Engineering and Electronics 28(1), 162–169 (2017)
49. Zheng, Y., Liu, Q., Chen, E., Ge, Y., Zhao, J.L.: Time series classification using multi-channels deep convolutional neural networks. In: International conference on web-age information management. pp. 298–310. Springer (2014)
50. Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., Guo, H., Xiang, J.: Epileptic seizure detection based on EEG signals and cnn. Frontiers in neuroinformatics 12, 95 (2018)

Krisztian Buza is an Associate Professor at the Budapest University of Economics and Business. He received his diploma degree in computer science from the Budapest University of Technology and Economics in 2007 and earned his Ph.D. from the University of Hildesheim in 2011. In addition to his academic career, he has gained industrial experience in the automotive sector, where he worked as a Senior Deep Learning Engineer, contributing to the development and application of machine learning and deep learning solutions in industry. Dr. Buza is the co-author of more than 50 scientific publications and received the Best Paper Award at the IEEE Conference on Computational Science and Engineering in 2010. He regularly serves as a reviewer for leading journals and conferences, including Neurocomputing, Knowledge-Based Systems, Knowledge and Information Systems, the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, and the Pacific-Asia Conference on Knowledge Discovery and Data Mining. His research interests include machine learning and data mining, with a particular focus on time series classification and biomedical applications.

Alexandros Nanopoulos is a Professor of Business Informatics at Baden-Württemberg Cooperative State University, with extensive experience in data science and AI. He holds a PhD in Data Mining from Aristotle University of Thessaloniki and a Diploma in Computer Science from the same institution. With a career spanning over two decades, Dr. Nanopoulos has held senior roles at leading organizations in the industry, where he developed AI-driven solutions and data products focused on predictive modeling, price optimization, demand forecasting, and more. He has also worked as consultant on projects related to the automotive industry, and has contributed to academic research as a junior professor at the University of Hildesheim and Catholic University Eichstätt- Ingolstadt. Dr. Nanopoulos has authored more than 130 research papers and has served on the program committees of top-tier data mining conferences such as ACM KDD and ECML. His work includes the development of innovative machine learning algorithms and contributions to areas like recommendation systems, social media mining, and data mining for large-scale applications.

Noemi Agnes Varga is a clinical geneticist and a neurology fellow in the final year of training. Since completing medical school, Dr. Varga has focused on neuroscience, conducting significant research on the genetic background of autism spectrum disorders during a PhD at the Semmelweis University Szentagothai János Doctoral School. With a keen interest in understanding common neurological disorders, such as epilepsy and recognizing rare neurogenetic diseases that present with common symptoms, Dr. Varga is dedicated to advancing deeper understanding of common pathways in the neurology.

Received: September 13, 2025; Accepted: March 5, 2026.