*Review*

# A Systematic Review of Time Series Classification Techniques Used in Biomedical Applications

**Will Ke Wang[1], Ina Chen[2] , Leeor Hershkovich[1], Jiamu Yang[1], Ayush Shetty[1], Geetika Singh[1], Yihang Jiang[1], Aditya Kotla[1], Jason Shang[1], Rushil Yerrabelli[1], Ali R. Roghanizad[1], Md Mobashir Hasan Shandhi[1], and Jessilyn Dunn[1,*]**

[1] Duke University Biomedical Engineering
[2] Washington University at St. Louis Biomedical Engineering
* Correspondence: jessilyn.dunn@duke.edu

**Abstract:** Digital clinical measures are increasingly used by individuals and clinicians to monitor health outcomes or track behavioral and physiological characteristics of individuals. Digital clinical measures can be collected via various digital sensing technologies such as smartphones, smartwatches, wearables, implantables, etc. Time series classification is very commonly used in biomedical applications to discover digital biomarkers. While deep learning models for TSC are very common and powerful, there exist some fundamental challenges. This review presents non-deep learning models commonly used for time series classification in biomedical applications that achieve high performance. **Objective:** We performed a systematic review to characterize the techniques used in time series classification of digital clinical measures throughout all stages of data processing and model building. **Methods:** We conducted a literature search on PubMed, and the Institute of Electrical and Electronics Engineers (IEEE), Web of Science, and SCOPUS databases using a range of search terms to retrieve peer-reviewed articles reporting academic research on digital clinical measures in the five year period between June 2016 and June 2021. We identified and categorized research studies based on the types of classification algorithms and sensor input types. **Results:** We found 452 papers in total from four different databases: PubMed, IEEE, Web of Science Database, and SCOPUS. After removing duplicates and irrelevant papers, 135 articles remained for detailed review and data extraction. Among these, engineered features using time series methods that were subsequently fed into widely-used machine learning classifiers was the most commonly used technique and also most frequently achieved the best performance metrics (77 out of 135 articles). Statistical modeling (24 out of 135 articles) algorithms were the second most common and also second best classification technique. Wavelet-based classification models (8 out of 135 articles) were also common. Electroencephalogram (29 out of 135 articles) was the most common data type used as an input. Accuracy was the most commonly reported performance metric, with 67.65% of articles reporting on accuracy. In this review paper, we provide summaries of signal pre-processing methods, feature engineering and selection methods, time series models, as well as model interpretations. Importantly, we found that about 50% of the papers only report one performance metric, which may result in a skewed view of overall performance. **Conclusion:.** While high performance using time series classification models has been achieved in digital clinical, physiological, or biomedical measures, there are a lack of benchmark datasets, modeling methods, or reporting methodology. There is no single widely used method for time series model development or feature interpretation– many different methods have proven successful.

**Keywords:** time series classification; digital biomarkers; machine learning; algorithms; feature engineering

 doi:10.20944/preprints202208.0493.v1

## 1. Introduction

Time Series Classification (TSC) involves building predictive models that output a target variable or label from inputs of longitudinal or sequential observations across some time period **[1]**. These inputs could be from a single variable measured across time or multiple variables measured across time, where the measurements can be ordinal or numerical (discrete or continuous).

Time series data is a very common form of data, containing information of the (changing) state of any variable. Some common examples include stock market prices and temperature values across some period of time. We can perform many modeling tasks on time series data, including classification, regression, and forecasting. There are unique challenges that come with modeling time series, given that measurements in time obtained in real life settings are subject to random noise, and that any measurement at a particular point in time could be related to or influenced by measurements at other points in time **[1]**. Given this nature of time series data, we cannot simply utilize established machine learning algorithms such as logistic regression, support vector machine or random forest on the raw time series datasets because this data violates basic assumptions of those models. In recent times, there are two vastly different camps of time series classification techniques: deep-learning-based models vs non-deep learning-based models. While deep learning models are extremely powerful and show great promise in classification performance and generalizability, they also present challenges in the areas of hyperparameter tuning, training, and model complexity decisions. To enable evaluation of new models, a reasonable baseline is also needed for comparison. Further, there already exists a review on deep-learning time series classification methods **[2]** Therefore, the focus of this review is on non-deep learning based time series classification models.

This paper also focuses specifically on biomedical applications of time series classification because there has been a huge increase in the generation of biomedical time series datasets (such as data from wearable devices like Apple Watch and Fitbit) recently as well as research using such data. Examples include electrocardiogram (ECG, for cardiovascular dysfunction screening) **[3]**, electroencephalogram (EEG, for brainwave tracking) **[4]**, accelerometry (for activity recognition), and polysomnography (PSG for sleep tracking) **[5]**, etc. In addition, an increasing number of people use smart devices or wearables regularly **[6]** for general fitness tracking **[7]**, sleep tracking **[8]**, fall detection **[9]**, or arrhythmia detection **[10]**. There is a growing need to design better data mining and classification methods to discern important and useful information from biomedical time series data. This would lead to more reliable methods for screening, diagnosis, and monitoring, providing huge benefits for healthcare as a whole.

Biomedical time series data collected from human subjects often present challenges that impede the ability to leverage time series modeling techniques that are common in other fields. For example, biomedical time series datasets often include just a small number of human subjects due to the resources and effort needed for data collection and annotation (or labeling to produce ground truth), which makes applying deep learning models very difficult since they are extremely data hungry **[11]**. Another challenge is the non-ergodic nature of datasets collected from human subjects, meaning that human subjects have vast individual differences in mental and physical states, producing data that look very different from one subject to another **[12]**. This results in sample level observations or models that perform well on some individuals even while being completely useless for others.

 doi:10.20944/preprints202208.0493.v1

While both reviews and experimental evaluations of recent algorithmic advances have been done **[13]**, the usefulness and applicability of machine learning algorithms is also impacted by interpretability and simplicity, particularly for biomedical predictive or diagnostic tasks. This review systematically surveys papers published in recent years using time series classification machine learning algorithms on biomedical datasets to answer the following questions:

1) What are the most common time series classification algorithms used in biomedical data science in the past six years?

2) What are the best performing time series classification algorithms for common biomedical signals?

3) How is interpretability addressed in the scientific literature that describes applying TSC algorithms for specific biomedical tasks?

The motivation for this review came from the observation that the types of algorithms explored and the depth of analysis performed in time-series biomedical data science have not been well described. In general, there has been a strong emphasis on algorithmic performance and a lack of focus on interpretability and model simplicity. This review aims to provide a general and recent landscape of the types of time series classification algorithms on longitudinal biomedical data and these algorithms can be applied toward specific tasks and with more insightful analysis.

## 2. Materials and Methods

We developed a list of search terms specific for each of the following four databases: PubMed, IEEE, Scopus and Web of Science (Supplementary Table 1). We searched through four literature databases: PubMed, IEEE, Web of Science and Scopus. Among these databases, IEEE enforces a limit on the number of search terms to be a maximum of 20. Hence, the defined search terms on IEEE were different from those on PubMed, Web of Science, and Scopus. We limited our search to literature from the last six years to limit the search scope to recent work and for a manageable scope of review. The defined searches include general terms and variations of time series machine learning classification and the fields of biomedical data. While the approach limits the coverage of the review, more recent work is often built upon previous work and new time-series classification techniques are often compared to established techniques from previous work, therefore this method is expected to provide a sufficient representation of the field. We used Covidence for literature screening and data extraction. We defined a very clear focus on only non-deep-learning time-series classification techniques utilized on biomedical data. This boundary notably excludes time-series regression tasks and deep learning techniques.

There were two phases of screening before data extraction. The first phase was screening by titles and abstracts, which Covidence automatically extracted from the DOI URLs. This phase was completed by two reviews where each reviewer read through each title/abstract and labeled as "include" or "exclude". Conflicts were resolved by discussion among reviewers to reach consensus. 260 papers were found to be irrelevant in this phase, mainly due to the following reasons:

- Classification algorithm is not used on biomedical data or time series data
- The article does not focus on classification algorithms, but regression algorithms, clustering algorithms or other algorithms.
- The article focuses only on deep learning algorithms.

 doi:10.20944/preprints202208.0493.v1

The second phase was screening of the full texts, which were pulled automatically by Covidence if free full texts were readily available. The rest of the full texts were uploaded manually to the Covidence platform using university credentials for access. Again, two reviewers each went through all the papers and adjudicated inclusion of papers into the final data extraction. Conflicts were resolved by discussion to reach consensus. 40 papers were excluded in this phase, mainly due to the following reasons:

- No access to full paper
- Not enough information about classification algorithm performance included
- The data came from animals instead of humans
- The algorithms used are not classification algorithms

Data from each paper was extracted by one of five reviewers, and then verified, edited, and cleaned by the study lead. In Table 1 we detail the information extracted from each paper.

## 3. Results

After removing duplicates and irrelevant papers, 135 articles remained for review and data extraction (Figure 1.a). Time series classification modeling typically consists of 3 main steps, signal preprocessing/transformation, modeling, and classification (Figure 1.b). The classification step is basically the process of model tuning, training and validation. The different types of algorithms used in the modeling steps are adapted from the categories summarized in the "The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances" by Ruiz et al. [13] (Figure 1.b). We illustrate the concept of each algorithm type in Figure 4. The most common techniques found in our search are feature engineering and selection, statistical modeling, distance-based, index development and shape-based methods (Table 1, Supplementary Table 2).

### *3.1. Summary of Algorithms*

Among the articles reviewed, electroencephalogram (EEG) signals were the most common biosignals investigated. For detailed information about the types of signals investigated in these papers (Figure 2.a). Engineered time series features fed into widely used machine learning classifier models is the most commonly used technique and is most often found to achieve the best performance (77 out of 135 articles). Statistical modeling (24 out of 135 articles) algorithms were the second most common. Wavelet-based classification models (8 out of 135 articles) are also common (Figure 2.b). Of papers which reported accuracy, 64% achieved accuracy higher than 90%. Of those reported F1-score, 70% achieved F1-score higher than 0.90. Of those reported AUC-ROC values, 24% achieved AUC-ROC higher than 0.90. Of those that reported sensitivity and/or specificity, 54% and 57%, respectively, achieved scores higher than 0.90 Of those that reported Cohen's Kappa, 43% achieved Cohen's Kappa higher than 0.90.

We recorded the classification performance metrics of all the articles included in this review, including accuracy, F1-score, Area Under Curve of Receiver Operating Characteristics (AUCROC), sensitivity, specificity, and Cohen's Kappa . The accuracy score is the most commonly reported performance metric, with 68% of the articles reporting accuracy scores (Figure 3.a). All other performance metrics are seldom reported: 30% reported f1-score, 19% reported AUC-ROC, 35% reported specificity, 43% reported sensitivity, and 6% reported Cohen's Kappa. This is concerning because oftentimes using only one or two performance metrics to evaluate a classifier is unreliable and does not tell the whole story of performance [14], [15].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0493.v1

In Figure 3.b), we see the performance metrics reported by each article colored by white (reported) or black (not reported). 86.7% reported one or more performance metrics. 50.4% reported two or more performance metrics. 37.8% reported three or more performance metrics. Only 2 papers out of 135 reported more than 6 performance metrics.

#### 3.1.1. Preprocessing

In all 135 papers, 50% specifically mentioned the preprocessing methods used. The most common preprocessing method is filtering, which was used mainly for artifact removal or noise reduction. Some other common preprocessing methods include resampling (downsampling for lower frequency or upsampling for higher frequency), segmentation, and smoothing. Also common are the use of discrete wavelet transform to decompose the original signal into different frequency bands [16]–[18], the use of continuous wavelet transform to expand the feature space [19], and the use of Fourier transform for signal decomposition and feature extraction [20], [21]. There are also intelligent upsampling techniques such as the use of synthetic data generation for a larger sample during preprocessing [22]. We present a summary of commonly used pre-processing methods in Supplementary Table 3.

### *3.2. Feature Engineering*

Feature engineering was the most commonly used method of time series classification. The feature engineering pipeline (Fig 1b) usually consists of the following steps:

1. Preprocessing: this step takes raw data as the input and performs some manipulation of the data to return cleaner signals. Common steps include artifact removal, filtering, and segmentation.

2. Signal transformation: this step can be used in preprocessing and also as a precursor to feature extraction. Some manipulation is performed on the signal to represent it in a different space. Common choices are Fourier Transform and Wavelet transforms.

3. Feature extraction: in this step features are extracted from the time series data as a new representation of the original time series.

4. Feature selection: this step selects the features that are the most descriptive, or have the most explanation power. Feature selection is also frequently performed in conjunction with model building.

5. Model selection: the best model is found through hyperparameter tuning and/or comparisons between different types of algorithms.

6. Model validation: performance metrics are calculated for all of the final models. This is frequently done in conjunction with model selection and often using some form of cross-validation.

An example feature engineering technique for time series is shown in Figure 5. We present summary tables of extracted features for general time series data as well as specific signals (HRV, EEG, etc.) and feature selection methods in Supplementary Table 4.a) and 4.b). We also present a summary table for all the found feature selection methods, shown in Supplementary Table 5. Feature engineering is used for all signal types across many different applications.

 doi:10.20944/preprints202208.0493.v1

### *3.3. Other Methods*

3.3.1 Ensemble Methods:

Ensemble-based methods are characterized by the connection of multiple algorithmic models together that join forces to make the final prediction. These methods may or may not need an additional feature engineering step. Some algorithms that do not necessitate feature engineering in this category are Hierarchical Vote Collective of Transformation-based Ensembles and Bag of Symbolic Fourier Approximation Symbols ensemble algorithms (BOSS) [13]. Newman el al. [23] describe a novel 3-classifier ensemble algorithm which achieves an accuracy of 0.9877, F1-score of 0.98, sensitivity of 0.9911 and specificity of 0.9863. The algorithm detects short periods of artificially induced nystagmus, a vision condition in which the eyes make repetitive, uncontrolled movements, from continuous eye movement data. The ensemble of classifiers include a linear discriminant analysis (LDA), a support vector machine (SVM), and boosted trees. The final classification decision is made by the majority vote. Elsayed et al. [24] found the optimal univariant ECG signal classifier after testing 8 different state-of-the-art time series classification methods. These models are: the fully convolutional network (FCN), long short-term memory and fully convolutional network (LSTM-FCN) and its attention-based LSTM model (ALSTM-FCN), the residual network mode (ResNet), the deep gated recurrent and convolutional network hybrid model (GRU-FCN), dynamic time warping model (DTW), multilayered perceptrons model (MLP), and the noise reduction based model, BOSS. GRU-FCN has the best performance with the top accuracy on five out of six tested datasets, with a reported accuracy score of 0.92.

3.3.2. State-space Models:

State-space models are characterized by the construction of a state and transition model where the transitions are modeled by probabilities. Often, state-space models are most intuitively used for sequence-to-sequence or point-wise classification. For example, She et al. [25] introduced an adaptive transfer learning algorithm to classify and segment events from non-stationary, multi-channel temporal data recorded by Empatica E4 wristband, including heart rate (HR), 3-axis accelerometry (ACC), electrodermal activity (EDA), and skin temperature (TEMP). The algorithm adaptively adjusts to shifts in distribution over time by using Fisher's linear discriminant analysis (FLDA) and a multivariate hidden Markov model (HMM), and achieves an accuracy of 0.9981 and F1-score of 0.9987. Garcia et al. [26] proposed a method based on dynamic affect (or emotional state) recognition from multimodal physiological signals such as EEG, Electromyography (EMG) and Electrooculography (EOG). Gaussian process latent variable models (GP-LVM) to learn a latent space is used to map high dimensional data (multimodal physiological signals) to a low dimensional latent space. In the affect recognition process, a support vector classifier is implemented to evaluate the relevance of the latent space features. The model achives an accuracy of 0.90556.

3.3.3. Shape/Pattern-based:

These models are characterized by mining or comparing shapes or patterns in a time or sequence vector. For example, Zhou et al. [27] published an algorithm that can consider the interaction among signals collected at spatiotemporally distinct points, where different classes of multichannel EEG data is characterized and differentiated by fuzzy temporal patterns. This algorithm achieves an accuracy of 0.9318, and a F1-score of 0.931 classifying positive vs negative emotion states.

 doi:10.20944/preprints202208.0493.v1

3.3.4. Distance-based:

These models calculate the distance (or differences) of time series data vectors. For example, Forestier et al. [28] propose an algorithm to detect the optimal partial alignment (optimal subsequence matching), and a system to predict multivariate signals by maximum a posteriori probability estimation and filtering. This scoring function is based on dynamic time warping. They were able to achieve an accuracy of 0.95, F1-score of 0.926, and a sensitivity of 0.896.

3.3.5. Other Methods:

There are other methodologies that are difficult to characterize. One common method is using statistical modeling to learn the characteristics of the time series sequences, and then the tuned models (the parameters or probablistic descriptors) can be used for classification purposes. For example, İşcan et al. [29] published a high performance model called LLGMN, composed of a log-linear model and a Gaussian mixture model (GMM). The model is used to classify ECG patterns (to discover QRS complexes), and it returns a posterior probability for the training data. This model was able to achieve a best accuracy of 0.9924.

Another common method is designing a composite metric or index based on domain knowledge or data-driven metrics. For example, Zhou et al. [30] proposed a new algorithm to detect gait events on three walking terrains. The detection uses acceleration jerk signals by obtaining gait paramets and detecing peaks of jerk signals. The indices are built using time-frequency methods and heuristically determining the peaks of accelerometry signals. The mean F1-score was above 0.98 for HS (heel-strike) event detection and 0.95 for TO (toe-off) event detection on the three terrains, tested on eight healthy subjects walking on level ground, upstairs, and downstairs.

Some articles focus specifically on investigating the wavelet transform and increasing its usefulness for specific use cases. For example, Ji et al. [31] attempted to find out which mother wavelet was the best choice for classifying gait events when using Continuous Wavelet Transform (CWT) by systematic investigations. Different choices of mother wavelet used produced significantly different algorithm performances. “Daubecgues 6 (Db6)” has the highest detection accuracy with the lowest detection time-error, achieving a final accuracy of 1.0. Lu et al. [32] proposed two methods: Discrete Wavelet Transform (DWT) and Extra Trees Classifier, and a personal identification method based on Convolutional Neural Networks (CNN) and Continuous Wavelet Transform (CWT). Nested five-fold cross-validation was used for model selection and model assessment, achieving accuracies of 0.99206 and 0.99203 respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0493.v1

### *3.4. Figures, Tables and Schemes*

**Table 1.** Data fields extracted from identified academic research.

| Categories | Choices/Sub-fields | Definitions/Descriptions |
|---|---|---|
| General (Relevant) Information | Article Type | The type of article for a particular paper being reviewed, such as Journal Article/Conference Article/Review Paper |
| | Area of application | Describes the area of biomedical signal and application this paper is about. |
| | Aim of study | Defines the specific challenge or question this paper is aimed at tackling |
| | Name of Publisher/Journal/Conference | Site of article publication. |
| | Classification Task | Defines the kind of classification task performed in this article. (Pointwise classification, window classification, or whole sequence classification.) |
| | Input data (X) | The type of input biomedical time series data.* |
| | Label (Y) | The output label or variable. Example: sleep vs wake, healthy vs diseased. |
| | Data source or open dataset name | States if the data is open source and where is the dataset hosted. |
| | Population Size | The number of subjects are included in this dataset. |
| | Data exclusion criteria | States the criteria considered to exclude subjects or specific parts of the data. |
| | All algorithms tested | List (or examples) of all the algorithms tested. |
| | Best algorithm name | The name of the best algorithm. |
| Classification Task | Whole-Series Classification | For whole time series classification (WSC), an entire time series sample (of length n) has just one associated label. |
| | Sequence-to-sequence (point-wise) | Each point in the entire time series sample (of length n) has one associated label, producing n labels. |
| | Window-based Classification or Onset Detection | For each entire time series sample (of length n), we have associated labels of where an event started and ended, producing a number of labels smaller than n but potentiallty bigger than 1. |

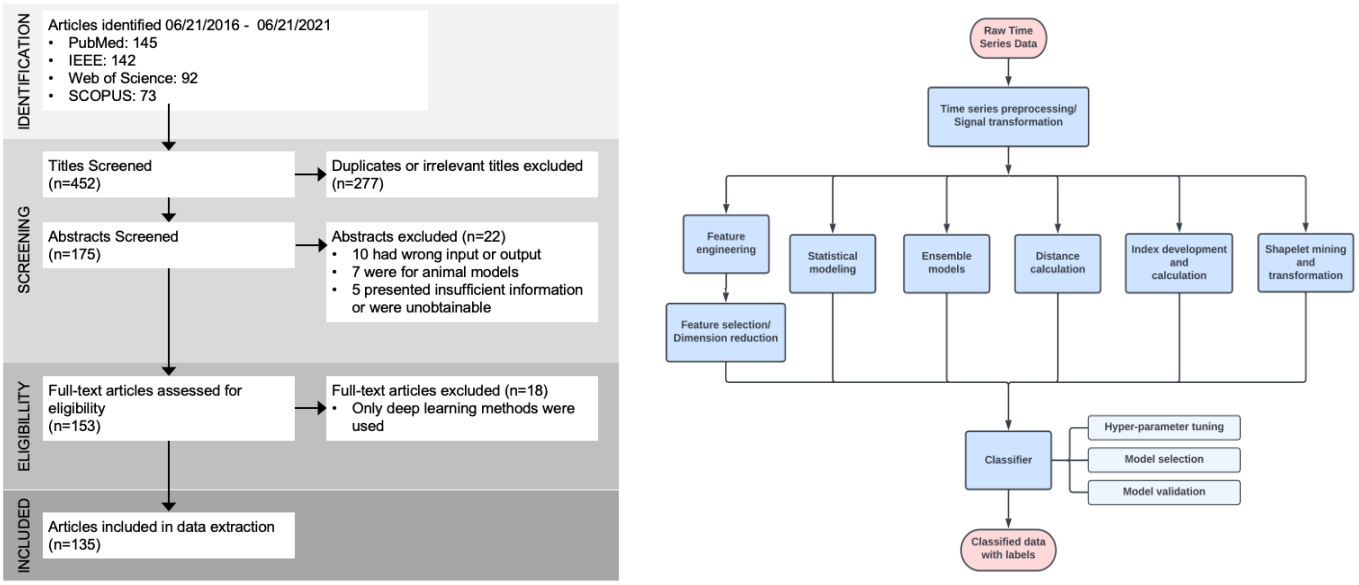
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0493.v1

| | | |
|---|---|---|
| | | We understand this to be a compromise between time pointwise classification and whole time series classification. |
| Best Algorithm Class | Feature Engineering | The type of time series classification technique where features are extracted to describe a particular time series sample and the features are fed into traditional machine learning algorithms as inputs of the predictive modeling. |
| | Statistical Modeling | This technique uses statistical modeling (such as Kalman filters or state-space models like Hidden Markov Models) to describe or fit to the time series observed. Using the information obtained from statistical models, we can make decisions or extract features to be used as inputs to machine learning algorithms. |
| | Wavelet Transform[8] | Wavelet Transform can be used for signal cleaning (preprocessing), signal decomposition (preprocessing) as well as feature extraction. This technique is widely used and can be considered an integral part of time series machine learning. |
| | Distance-based methods[7] | This method is based on defining or quantifying the difference between 2 or more time series data samples. The first class of classification methods rely on some notion of distance between two time series or time series subsequences. Intuitively, a distance can be seen as a proxy for dissimilarity. Two time series that are in close proximity (i.e., they have a small distance) under some distance measure, are likely to come from the same class. |
| | Ensemble-based | Ensemble-based classification algorithms utilize multiple algorithms to make predictions and then aggregate the results coming from these different algorithms |
| | Shapelet/Shape-based | Shapelet-based methods are similar to significant pattern mining. Time series shapelets are subsequences that maximize classification performance. |
| | Non-linear index and thresholding | This time series classification method is based on defining indices based on domain- and data-driven time series features. The thresholds for these indices can be predefined or found through statistical learning. The thresholds are then used to make predictions of classes. |
| | Other | Any other methods of time series classification that cannot be easily categorized. |
| Best Algorithm | Accuracy | The degree of correctness of a calculation of the best algorithm reported. |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0493.v1

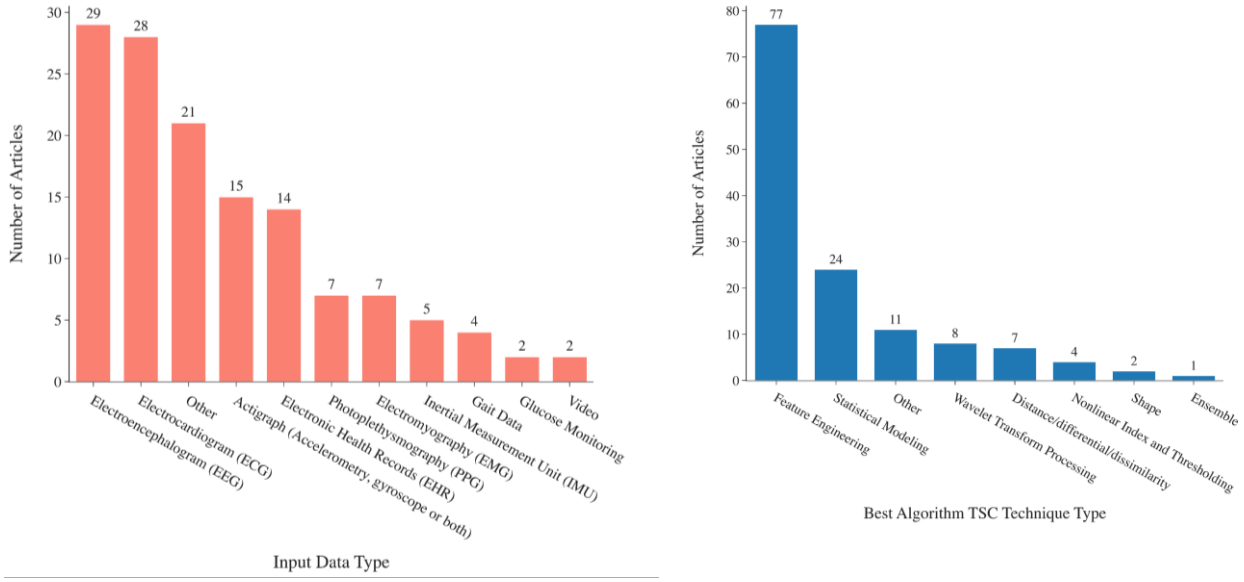
| | | |
|---|---|---|
| Performances [33], [34] | | $\frac{TP + TN}{TP + FN + TN + FP}$ |
| | F1-score | The harmonic mean of precision and recall of the best algorithm reported. $2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$ |
| | Area Under Curve of Receiver-Operating Characteristic | The measure of the usefulness of a test in general of the best algorithm reported. |
| | Sensitivity | The percentage of true positives of the best algorithm reported. $\frac{TP}{TP + FN}$ |
| | Specificity | The percentage of true negatives of the best algorithm reported. $\frac{TN}{TN + FP}$ |
| | Cohen's Kappa | A statistical measure of inter-rater reliability for categorical variables of the best algorithm reported. $\frac{p_0 - p_e}{1 - p_e}$ |
| | Positive Predictive Value | The percentage of positive test results is a true positive. $\frac{TP}{TP + FP}$ |
| | Negative Predictive Value | The percentage of negative test results is a true negative. $\frac{TN}{TN + FN}$ |
| | False Positive Rate | The percentage of false alarm of the best algorithm reported $\frac{FP}{FP + TN}$ |
| | Area Under Precision-Recall Curve | A model performance metric for binary responses that is appropriate for rare events and not dependent on model specificity |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0493.v1




(**a**) (**b**)

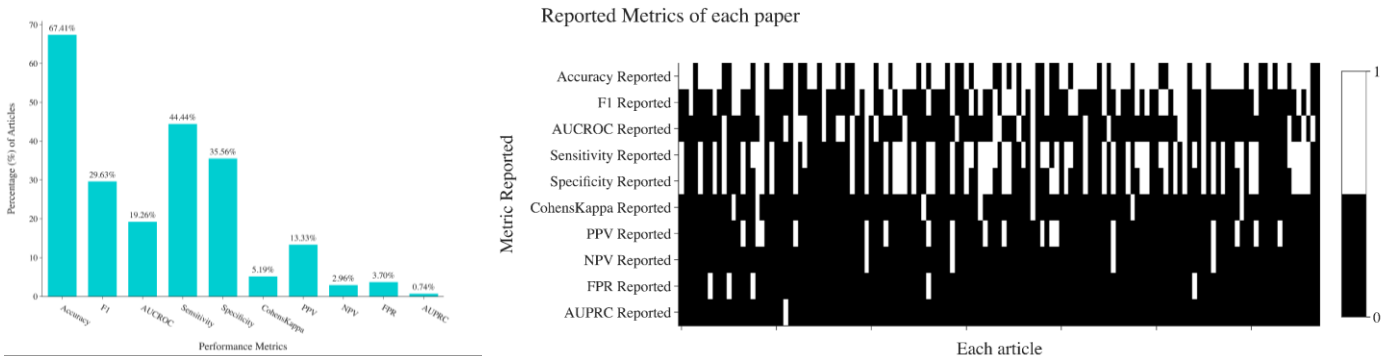
**Figure 1.** (**a**) Review results and the number of papers through each selection process; (**b**) Flow chart of the common steps in time series classification techniques found in this review. Raw time series signals usually go through some steps of preprocessing for artifact removal or noise reduction, and then passed through the modeling stage. The modeling stage can use many different types of algorithms, such as feature engineering and selection, statistical modeling and distance calculation (Table 1). Classifiers are then tuned, trained, validated and compared to find the best model for a specific task.




(**a**) (**b**)

**Figure 2.** (**a**) Numbers of papers found in this review focusing on each different biosignal type specified on the horizontal axis; (**b**) Number of articles found for the categories of time series classification methods (horizontal axis) used in biomedical applications.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0493.v1




(**a**) (**b**)

**Figure 3.** (**a**) Percentages of performance metrics reported in studies reviewed; (**b**) Reported metrics of each paper are represented horizontally. Each article's reported metrics are represented across one vertical line. White block means metric reported (=1), while black block means metrics not reported (=0).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0493.v1

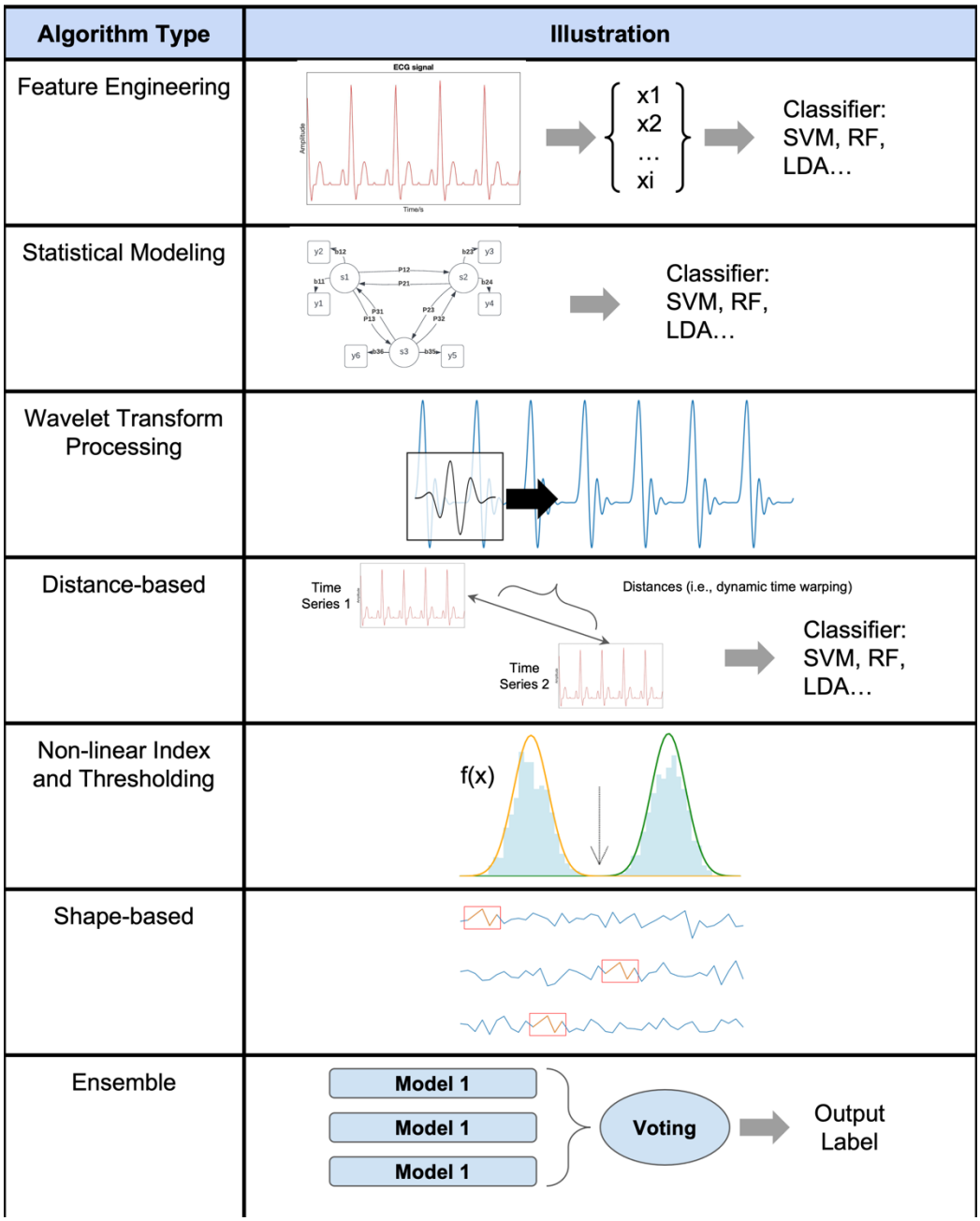



**Figure 4.** Conceptual representation of non-deep learning time series classification modeling types, adapted from [1], [35]–[39]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0493.v1

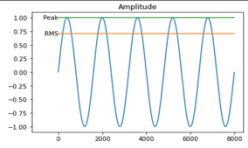
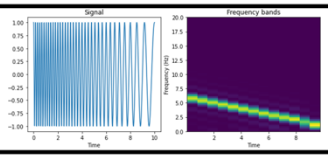
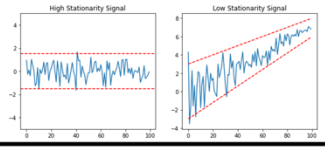
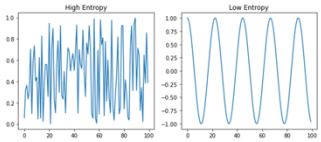
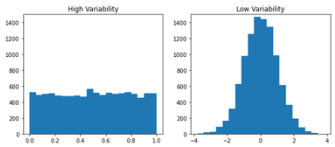
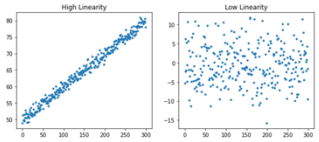
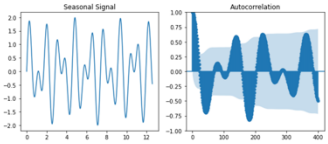
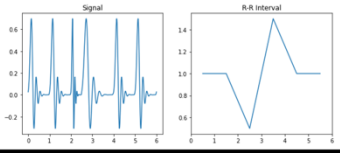
| Feature Type | Description | Illustration |
|---|---|---|
| Amplitude | Amplitude features describe how distant a signal's values are from 0 |  |
| Frequency | Frequency features represent the properties of the Fourier transform of a signal |  |
| Stationarity | Stationarity describes the consistency of signal properties over time, such as mean and variance |  |
| Entropy | Entropy measures the number of states of a system, or the ability to probabilistically determine the next state of a system. |  |
| Variability | Variability measures how similar in value each of the measurements in a signal are |  |
| Linearity | Linearity features quantify how much a system changes with a constant rate |  |
| Correlation | Correlation features describe how dependent a signal is on a previous state |  |
| Plot-based | Features related to properties of an ECG graph |  |

**Figure 5.** Illustration of different types of feature engineering techniques, adapted from [40]

## 4. Discussion

While deep learning methods have seen wide usage and high performance in health informatics in recent years, this review demonstrates the utility and power of non-deep learning machine learning algorithms. Many papers reviewed here with a focus on conventional machine learning algorithms achieved almost perfect performance in classification metrics (i.e. 0.999 in classification accuracy). Compared to deep learning approaches, many conventional machine learning algorithms can be used off-the-shelf, without the researcher needing to rebuild the model architecture and tune a large number of hyperparameters. Conventional machine learning algorithms are also generally easier to train, optimize, and deploy, due to their light-weight model (not necessarily needing a large number of parameters like in deep neural networks). This review also serves to identify the non-deep learning time series classification techniques that can serve as a competitive

baseline comparator for researchers to understand whether newly designed deep learning networks are truly performing well or not.

Among all of the papers reviewed, feature engineering methods followed by off-the-shelf machine learning techniques such as Support Vector Machine and Random Forests are by far the most common. To aid future researchers building and testing feature engineering algorithms, we have provided an almost exhaustive list of features and transformation techniques that can be applied on longitudinal data in health informatics. We also provided a summary of the most common preprocessing methods, but do not claim the summary to be exhaustive since preprocessing methods are very frequently domain dependent, and often decisions about preprocessing are made with the researchers' own experience and discretion with considerations about different characteristics of each unique dataset.

Out of all of the papers reviewed, there is a lack of standard in terms of the classification metrics reported, which goes against best practices of reporting multiple metrics to fully describe the performances of algorithms tested. Nearly half of the papers only report one metric, typically the accuracy score, which is prone to bias [15], [41].

Interpretation of models is very useful, especially in understanding clinical digital measures, and this practice should be widely adopted. Fortunately, the non-deep learning models are generally very interpretable. We see many different methods for interpreting models and features in our review (Supplementary Table 6). However, there is no set standard or commonly used method for interpretation for most signal types and/or algorithms. We do see that EEG digital measures very commonly use visualization of statistical significance over a representation of the cranium as a method of interpretation, and similar common practices should be adopted for other signal types.

While rigorous, our paper selection method would have benefited from a third reviewer to break ties and resolve discrepancies. Also, we were not aware of any existing classification system for categorizing the time-series classification algorithms, and thus we developed our own, which may be sub-optimal. It is evident that many papers are difficult to categorize or assign a single category because studies often incorporate multiple different approaches, for example, using Dynamic Time Warping to calculate distances between time series motifs, and subsequently using those distances as input features into a Support Vector Machine. Additionally, although we sought to exclude deep learning approaches through our search term design, some papers examined both deep-learning and non-deep-learning classification algorithms and we felt compelled to include these papers in our review, both to not exclude the non-deep learning methods, as well as to gain insight on the direct comparison between these two approaches.

## 5. Conclusions

In conclusion, our group performed this systematic review to survey the landscape of non-deep learning based time series classification methods used in biomedical applications. Here, we review and summarize these methods and present our findings. We found that non-deep learning time series classification techniques can be extremely powerful, giving great algorithm performances, while also allowing great interpretability. However, this field still lacks standardization about model testing and validation procedure and reporting metrics, which should be addressed to allow better reproducibility and understanding of the presented algorithms presented by researchers in this field.

**Supplementary Materials:** The following supporting information can be downloaded at: www.mdpi.com/xxx/s1,

Supplementary Table 1. Search terms for each database

Supplementary Table 2. Description for each algorithm type and example papers.

Supplementary Table 3. Summary of time series signal preprocessing methods and example papers.

Supplementary Table 4.a. Example of feature engineering techniques and papers.

Supplementary Table 4.b. Features for common signals and example papers.

Supplementary Table 5. Summary of feature selection methods and example papers

Supplementary Table 6. Summary the different types of model interpretation methods discussed or used in each article, with example usages.

**Author Contributions:** Conceptualization: Will Ke Wang, Jessilyn Dunn; preparation of draft and writing: Ina Chen, Leeor Hershkovich, Jiamu Yang; methodology and validation: Will Ke Wang, Geetika Singh, Yihang Jiang, Aditya Kotla, Jason Shang, Rushil Yerrabelli; review and editing: Md Mobashir Hasam Shandhi, Ali R. Roghanizad, Jessilyn Dunn. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** In this section, you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

**Conflicts of Interest:** The authors declare no conflict of interest

## References

[1] C. Bock, M. Moor, C. R. Jutzeler, and K. Borgwardt, "Machine Learning for Biomedical Time Series Classification: From Shapelets to Deep Learning," in *Artificial Neural Networks*, H. Cartwright, Ed. New York, NY: Springer US, 2021, pp. 33–71. doi: 10.1007/978-1-0716-0826-5_2.

[2] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Min. Knowl. Discov.*, vol. 33, no. 4, pp. 917–963, Jul. 2019, doi: 10.1007/s10618-019-00619-1.

[3] Y. Li, X. Tang, A. Wang, and H. Tang, "Probability density distribution of delta RR intervals: a novel method for the detection of atrial fibrillation," *Australas. Phys. Eng. Sci. Med.*, vol. 40, no. 3, pp. 707–716, Sep. 2017, doi: 10.1007/s13246-017-0554-2.

[4] B. Garcia-Martinez, A. Martinez-Rodrigo, A. Fernandez-Caballero, J. Moncho-Bogani, and R. Alcaraz, "Nonlinear predictability analysis of brain dynamics for automatic recognition of negative stress," *Neural Comput. Appl.*, vol. 32, no. 17, pp. 13221–13231, Sep. 2020, doi: 10.1007/s00521-018-3620-0.

[5] Y. R. Tabar, K. B. Mikkelsen, M. L. Rank, M. C. Hemmsen, and P. Kidmose, "Investigation of low dimensional feature spaces for automatic sleep staging," *Comput. Methods Programs Biomed.*, vol. 205, p. 106091, Jun. 2021, doi: 10.1016/j.cmpb.2021.106091.

[6] "Smartwatch penetration 2020," *Statista*. https://www.statista.com/statistics/1107874/access-to-smartwatch-in-households-worldwide/ (accessed Jul. 25, 2022).

[7] J. Xie, D. Wen, L. Liang, Y. Jia, L. Gao, and J. Lei, "Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study," *JMIR MHealth UHealth*, vol. 6, no. 4, Apr. 2018, doi: 10.2196/mhealth.9754.

[8] E. Guillodo *et al.*, "Clinical Applications of Mobile Health Wearable–Based Sleep Monitoring: Systematic Review," *JMIR MHealth UHealth*, vol. 8, no. 4, p. e10733, Apr. 2020, doi: 10.2196/10733.

[9] P. Bet, P. C. Castro, and M. A. Ponti, "Fall detection and fall risk assessment in older person using wearable sensors: A systematic review," *Int. J. Med. Inf.*, vol. 130, p. 103946, Oct. 2019, doi: 10.1016/j.ijmedinf.2019.08.006.

[10] M. P. Turakhia *et al.*, "Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study," *Am. Heart J.*, vol. 207, pp. 66–75, Jan. 2019, doi: 10.1016/j.ahj.2018.09.002.

[11] G. Marcus, "Deep Learning: A Critical Appraisal." arXiv, Jan. 02, 2018. doi: 10.48550/arXiv.1801.00631.

[12] A. J. Fisher, J. D. Medaglia, and B. F. Jeronimus, "Lack of group-to-individual generalizability is a threat to human subjects research," *Proc. Natl. Acad. Sci.*, vol. 115, no. 27, pp. E6106–E6115, Jul. 2018, doi: 10.1073/pnas.1711978115.

[13] A. P. Ruiz, M. Flynn, J. Large, M. Middlehurst, and A. Bagnall, "The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Min. Knowl. Discov.*, vol. 35, no. 2, pp. 401–449, Mar. 2021, doi: 10.1007/s10618-020-00727-3.

[14] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation," in *AI 2006: Advances in Artificial Intelligence*, Berlin, Heidelberg, 2006, pp. 1015–1021. doi: 10.1007/11941439_114.

[15] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216–231, Jul. 2019, doi: 10.1016/j.patcog.2019.02.023.

[16] S. S. S. Mole and K. Sujatha, "An efficient Gait Dynamics classification method for Neurodegenerative Diseases using Brain signals," *J. Med. Syst.*, vol. 43, no. 8, p. 245, Jun. 2019, doi: 10.1007/s10916-019-1384-4.

[17] D. Joshi, A. Khajuria, and P. Joshi, "An automatic non-invasive method for Parkinson's disease classification," *Comput. Methods Programs Biomed.*, vol. 145, pp. 135–145, Jul. 2017, doi: 10.1016/j.cmpb.2017.04.007.

[18] H. T. Tor *et al.*, "Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals," *Comput. Methods Programs Biomed.*, vol. 200, p. 105941, Mar. 2021, doi: 10.1016/j.cmpb.2021.105941.

[19] S. Mesbah, F. Gonnelli, C. A. Angeli, A. El-baz, S. J. Harkema, and E. Rejc, "Neurophysiological markers predicting recovery of standing in humans with chronic motor complete spinal cord injury," *Sci. Rep.*, vol. 9, no. 1, p. 14474, Oct. 2019, doi: 10.1038/s41598-019-50938-y.

[20] N. X. Anh, R. M. Nataraja, and S. Chauhan, "Towards near real-time assessment of surgical skills: A comparison of feature extraction techniques," *Comput. Methods Programs Biomed.*, vol. 187, 2020, doi: 10.1016/j.cmpb.2019.105234.

[21] P. Durongbhan *et al.*, "A Dementia Classification Framework Using Frequency and Time-Frequency Features Based on EEG Signals," *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 27, no. 5, pp. 826–835, May 2019, doi: 10.1109/TNSRE.2019.2909100.

[22] S. Bhattacharya, O. Mazumder, D. Roy, A. Sinha, and A. Ghose, "Synthetic Data Generation Through Statistical Explosion: Improving Classification Accuracy of Coronary Artery Disease Using PPG," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 1165–1169. doi: 10.1109/ICASSP40776.2020.9054570.

[23] J. L. Newman, J. S. Phillips, S. J. Cox, J. FitzGerald, and A. Bath, "Automatic nystagmus detection and quantification in long-term continuous eye-movement data," *Comput. Biol. Med.*, vol. 114, 2019, doi: 10.1016/j.compbiomed.2019.103448.

[24] N. Elsayed, A. S. Maida, and M. Bayoumi, "An analysis of univariate and multivariate electrocardiography signal classification," 2019, pp. 396–399. doi: 10.1109/ICMLA.2019.00074.

[25] X. She *et al.*, "Adaptive multi-channel event segmentation and feature extraction for monitoring health outcomes," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 8, pp. 2377–2388, Aug. 2021, doi: 10.1109/TBME.2020.3038652.

[26] H. F. García, M. A. Álvarez, and Á. A. Orozco, "Gaussian process dynamical models for multimodal affect recognition," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2016, pp. 850–853. doi: 10.1109/EMBC.2016.7590834.

[27] P.-Y. Zhou and K. C. C. Chan, "Fuzzy feature extraction for multichannel EEG classification," *IEEE Trans. Cogn. Dev. Syst.*, vol. 10, no. 2, pp. 267–279, 2018, doi: 10.1109/TCDS.2016.2632130.

[28] G. Forestier, F. Petitjean, L. Riffaud, and P. Jannin, "Automatic matching of surgeries to predict surgeons' next actions," *Artif. Intell. Med.*, vol. 81, pp. 3–11, Sep. 2017, doi: 10.1016/j.artmed.2017.03.007.

[29] M. İşcan, F. Yığıt, and C. Yilmaz, "Heartbeat pattern classification algorithm based on Gaussian mixture model," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2016, pp. 1–6. doi: 10.1109/MeMeA.2016.7533715.

[30] H. Zhou *et al.*, "Towards Real-Time Detection of Gait Events on Different Terrains Using Time-Frequency Analysis and Peak Heuristics Algorithm," *Sensors*, vol. 16, no. 10, Art. no. 10, Oct. 2016, doi: 10.3390/s16101634.

[31] N. Ji *et al.*, "Appropriate Mother Wavelets for Continuous Gait Event Detection Based on Time-Frequency Analysis for Hemiplegic and Healthy Individuals," *Sensors*, vol. 19, no. 16, Art. no. 16, Jan. 2019, doi: 10.3390/s19163462.

[32] L. Lu, J. Mao, W. Wang, G. Ding, and Z. Zhang, "A Study of Personal Recognition Method Based on EMG Signal," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 4, pp. 681–691, Aug. 2020, doi: 10.1109/TBCAS.2020.3005148.

[33] M. Hossin and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, pp. 01–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.

[34] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, Jun. 2006, pp. 233–240. doi: 10.1145/1143844.1143874.

[35] *The Elements of Statistical Learning*. Accessed: Aug. 14, 2022. [Online]. Available: http://link.springer.com/book/10.1007/978-0-387-84858-7

[36] "Hidden Markov model," *Wikipedia*. Jul. 18, 2022. Accessed: Aug. 14, 2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Hidden_Markov_model&oldid=1098931761

[37] S. Talebi, "The Wavelet Transform," *Medium*, Jan. 08, 2021. https://towardsdatascience.com/the-wavelet-transform-e9cfa85d7b34 (accessed Feb. 13, 2022).

[38] M. Regan, "K Nearest Neighbors & Dynamic Time Warping." Aug. 03, 2022. Accessed: Aug. 14, 2022. [Online]. Available: https://github.com/markdregan/K-Nearest-Neighbors-with-Dynamic-Time-Warping

[39] "Figure 10: Using the Gaussian Mixture Model to Estimate the Threshold.," *ResearchGate*. https://www.researchgate.net/figure/Using-the-Gaussian-Mixture-Model-to-Estimate-the-Threshold_fig4_307984756 (accessed Aug. 14, 2022).

[40] S. Walter *et al.*, "Automatic pain quantification using autonomic parameters," *Psychol. Neurosci.*, vol. 7, pp. 363–380, Dec. 2014, doi: 10.3922/j.psns.2014.041.

[41] G. Canbek, T. Taskaya Temizel, and S. Sagiroglu, "BenchMetrics: a systematic benchmarking method for binary classification performance metrics," *Neural Comput. Appl.*, vol. 33, no. 21, pp. 14623–14650, Nov. 2021, doi: 10.1007/s00521-021-06103-6.

[42] Y. Shi, F. Li, T. Liu, F. R. Beyette, and W. Song, "Dynamic Time-frequency Feature Extraction for Brain Activity Recognition," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2018, pp. 3104–3107, Jul. 2018, doi: 10.1109/EMBC.2018.8512914.

[43] N. El-Rashidy, S. El-Sappagh, T. Abuhmed, S. Abdelrazek, and H. M. El-Bakry, "Intensive Care Unit Mortality Prediction: An Improved Patient-Specific Stacking Ensemble Model," *IEEE Access*, vol. 8, pp. 133541–133564, 2020, doi: 10.1109/ACCESS.2020.3010556.

[44] J. Rodenas, M. Garcia, R. Alcaraz, and J. J. Rieta, "Combined Nonlinear Analysis of Atrial and Ventricular Series for Automated Screening of Atrial Fibrillation," *Complexity*, p. 2163610, 2017, doi: 10.1155/2017/2163610.

[45] S. Hong, H. Kwon, S. H. Choi, and K. S. Park, "Intelligent system for drowsiness recognition based on ear canal electroencephalography with photoplethysmography and electrocardiography," *Inf. Sci.*, vol. 453, pp. 302–322, Jul. 2018, doi: 10.1016/j.ins.2018.04.003.

[46] A. Zarei and B. M. Asl, "Automatic Detection of Obstructive Sleep Apnea Using Wavelet Transform and Entropy-Based Features From Single-Lead ECG Signal," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1011–1021, May 2019, doi: 10.1109/JBHI.2018.2842919.

[47] L. Liu *et al.*, "Ambulatory Human Gait Phase Detection Using Wearable Inertial Sensors and Hidden Markov Model," *Sensors*, vol. 21, no. 4, Art. no. 4, Jan. 2021, doi: 10.3390/s21041347.

[48] M. Adam *et al.*, "Automated characterization of cardiovascular diseases using relative wavelet nonlinear features extracted from ECG signals," *Comput. Methods Programs Biomed.*, vol. 161, pp. 133–143, Jul. 2018, doi: 10.1016/j.cmpb.2018.04.018.

[49] C. K. Zhang, Y. Y. Chen, A. Yin, and X. Wang, "Anomaly detection in ECG based on trend symbolic aggregate approximation," *Math. Biosci. Eng. MBE*, vol. 16, no. 4, pp. 2154–2167, Mar. 2019, doi: 10.3934/mbe.2019105.

[50] R. He *et al.*, "Automatic Detection of Atrial Fibrillation Based on Continuous Wavelet Transform and 2D Convolutional Neural Networks," *Front. Physiol.*, vol. 9, p. 1206, 2018, doi: 10.3389/fphys.2018.01206.

[51] J. Liu, Y. Zhao, B. Lai, H. Wang, and K. L. Tsui, "Wearable Device Heart Rate and Activity Data in an Unsupervised Approach to Personalized Sleep Monitoring: Algorithm Validation," *JMIR MHealth UHealth*, vol. 8, no. 8, Aug. 2020, doi: 10.2196/18370.

[52] J. Y. Nancy, N. H. Khanna, and A. Kannan, "A bio-statistical mining approach for classifying multivariate clinical time series data observed at irregular intervals," *Expert Syst. Appl.*, vol. 78, pp. 283–300, Jul. 2017, doi: 10.1016/j.eswa.2017.01.056.

[53] S. X. Lee and S. Y. Leemaqz, "Automated Wrist Pulse diagnosis of Pancreatitis via Autoregressive Discriminant models," 2017, pp. 1262–1266. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85080915306&partnerID=40&md5=f92bd168c60aae1a6ffbda888f5663f0

[54] F. Bagattini, I. Karlsson, J. Rebane, and P. Papapetrou, "A classification framework for exploiting sparse multivariate temporal features with application to adverse drug event detection in medical records," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 7, Jan. 2019, doi: 10.1186/s12911-018-0717-4.

[55] E. Campbell, A. Phinyomark, and E. Scheme, "Feature Extraction and Selection for Pain Recognition Using Peripheral Physiological Signals," *Front. Neurosci.*, vol. 13, p. 437, May 2019, doi: 10.3389/fnins.2019.00437.

[56] A. Jovic, K. Brkic, and G. Krstacic, "Detection of congestive heart failure from short-term heart rate variability segments using hybrid feature selection approach," *Biomed. Signal Process. Control*, vol. 53, p. 101583, Aug. 2019, doi: 10.1016/j.bspc.2019.101583.

[57] R. Nawaz, K. H. Cheah, H. Nisar, and V. V. Yap, "Comparison of different feature extraction methods for EEG-based emotion recognition," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 910–926, Sep. 2020, doi: 10.1016/j.bbe.2020.04.005.

[58] E. Zdravevski *et al.*, "Improving Activity Recognition Accuracy in Ambient-Assisted Living Systems by Automated Feature Engineering," *IEEE Access*, vol. 5, pp. 5262–5280, 2017, doi: 10.1109/ACCESS.2017.2684913.

[59] N. Reamaroon, M. W. Sjoding, K. Lin, T. J. Iwashyna, and K. Najarian, "Accounting for Label Uncertainty in Machine Learning for Detection of Acute Respiratory Distress Syndrome," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 407–415, Jan. 2019, doi: 10.1109/JBHI.2018.2810820.

[60] Z. Haddi *et al.*, "Relevance Vector Machine as Data-Driven Method for Medical Decision Making," in *2019 18th European Control Conference (ECC)*, Jun. 2019, pp. 1011–1016. doi: 10.23919/ECC.2019.8796141.

[61] D. S. Wickramasuriya, M. K. Tessmer, and R. T. Faghih, "Facial Expression-Based Emotion Classification using Electrocardiogram and Respiration Signals," in *2019 IEEE Healthcare Innovations and Point of Care Technologies, (HI-POCT)*, Nov. 2019, pp. 9–12. doi: 10.1109/HI-POCT45284.2019.8962891.

[62] A. R. Malik and J. Boger, "Zero-Effort Ambient Heart Rate Monitoring Using Ballistocardiography Detected Through a Seat Cushion: Prototype Development and Preliminary Study," *JMIR Rehabil. Assist. Technol.*, vol. 8, no. 2, p. e25996, May 2021, doi: 10.2196/25996.

[63] N. Liu, M. Sun, L. Wang, W. Zhou, H. Dang, and X. Zhou, "A support vector machine approach for AF classification from a short single-lead ECG recording," *Physiol. Meas.*, vol. 39, no. 6, p. 064004, Jun. 2018, doi: 10.1088/1361-6579/aac7aa.

[64] A. Goshvarpour and A. Goshvarpour, "Evaluation of Novel Entropy-Based Complex Wavelet Sub-bands Measures of PPG in an Emotion Recognition System," *J. Med. Biol. Eng.*, vol. 40, no. 3, pp. 451–461, Jun. 2020, doi: 10.1007/s40846-020-00526-7.

[65] M. Kolodziej, A. Majkowski, R. J. Rak, A. Rysz, and A. Marchel, "DECISION SUPPORT SYSTEM FOR EPILEPTOGENIC ZONE LOCATION DURING BRAIN RESECTION," *Metrol. Meas. Syst.*, vol. 25, no. 1, pp. 15–32, 2018, doi: 10.24425/118167.

[66] Y. Hu, W. An, R. Subramanian, N. Zhao, Y. Gu, and W. Wu, *Faster Clinical Time Series Classification with Filter Based Feature Engineering Tree Boosting Methods*, vol. 914. 2021, p. 260. doi: 10.1007/978-3-030-53352-6_23.

[67] B. Miao, J. Guan, L. Zhang, Q. Meng, and Y. Zhang, "Automated Epileptic Seizure Detection Method Based on the Multi-attribute EEG Feature Pool and mRMR Feature Selection Method," in *Computational Science - Iccs 2019, Pt Iii*, vol. 11538, J. M. F. Rodrigues, P. J. S. Cardoso, J. Monteiro, R. Lam, V. V. Krzhizhanovskaya, M. H. Lees, J. J. Dongarra, and P. M. A. Sloot, Eds. 2019, pp. 45–59. doi: 10.1007/978-3-030-22744-9_4.

[68] W. Mumtaz, L. Xia, M. A. M. Yasin, S. S. A. Ali, and A. S. Malik, "A wavelet-based technique to predict treatment outcome for Major Depressive Disorder," *PLOS ONE*, vol. 12, no. 2, p. e0171409, Feb. 2017, doi: 10.1371/journal.pone.0171409.

[69] S. Lotfan, S. Shahyad, R. Khosrowabadi, A. Mohammadi, and B. Hatef, "Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures," *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 199–213, Mar. 2019, doi: 10.1016/j.bbe.2018.10.008.

[70] J. Cimbalnik *et al.*, "Physiological and pathological high frequency oscillations in focal epilepsy," *Ann. Clin. Transl. Neurol.*, vol. 5, no. 9, pp. 1062–1076, Sep. 2018, doi: 10.1002/acn3.618.

[71] H. Ren, Z. Ye, and Z. Li, "Anomaly detection based on a dynamic Markov model," *Inf. Sci.*, vol. 411, pp. 52–65, Oct. 2017, doi: 10.1016/j.ins.2017.05.021.

[72] R. H. Elden, V. F. Ghoneim, and W. Al-Atabany, "A computer aided diagnosis system for the early detection of neurodegenerative diseases using linear and non-linear analysis," in *2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME)*, Mar. 2018, pp. 116–121. doi: 10.1109/MECBME.2018.8402417.

[73] "Heartbeat Classification Using Abstract Features From the Abductive Interpretation of the ECG | IEEE Journals & Magazine | IEEE Xplore." https://ieeexplore-ieee-org.proxy.lib.duke.edu/document/7750556 (accessed Feb. 06, 2022).

[74] R. Gupta and P. Kundu, "Dissimilarity factor based classification of inferior myocardial infarction ECG," in *2016 IEEE First International Conference on Control, Measurement and Instrumentation (CMI)*, Jan. 2016, pp. 229–233. doi: 10.1109/CMI.2016.7413745.

[75] R. Mohammadi-Ghazi, Y. M. Marzouk, and O. Buyukozturk, "Conditional classifiers and boosted conditional Gaussian mixture model for novelty detection," *Pattern Recognit.*, vol. 81, pp. 601–614, Sep. 2018, doi: 10.1016/j.patcog.2018.03.022.

[76] G. H. Park, S. J. Kim, and Y. S. Cho, "Development of a voiding diary using urination recognition technology in mobile environment," *J. Exerc. Rehabil.*, vol. 16, no. 6, pp. 529–533, Dec. 2020, doi: 10.12965/jer.2040790.395.

[77] S. A. David, J. A. T. Machado, C. M. C. Inacio, and C. A. Valentim, "A combined measure to differentiate EEG signals using fractal dimension and MFDFA-Hurst," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 84, p. 105170, May 2020, doi: 10.1016/j.cnsns.2020.105170.

[78] M. Li, S. Tian, L. Sun, and X. Chen, "Gait Analysis for Post-Stroke Hemiparetic Patient by Multi-Features Fusion Method," *Sensors*, vol. 19, no. 7, p. E1737, Apr. 2019, doi: 10.3390/s19071737.

[79] K. Gunnarsdottir, V. Sadashivaiah, M. Kerr, S. Santaniello, and S. V. Sarma, "Using demographic and time series physiological features to classify sepsis in the intensive care unit," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2016, pp. 778–782, Aug. 2016, doi: 10.1109/EMBC.2016.7590817.

[80] L. Mertzanis, A. Panotonoulou, M. Skoularidou, and I. Kontoyiannis, "Deep Tree Models for 'Big' Biological Data," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2018, pp. 1–5. doi: 10.1109/SPAWC.2018.8445994.

[81] K. Orphanou, A. Dagliati, L. Sacchi, A. Stassopoulou, E. Keravnou, and R. Bellazzi, "Incorporating repeating temporal association rules in Naïve Bayes classifiers for coronary heart disease diagnosis," *J. Biomed. Inform.*, vol. 81, pp. 74–82, May 2018, doi: 10.1016/j.jbi.2018.03.002.

[82] R. C. Lacson, B. Baker, H. Suresh, K. Andriole, P. Szolovits, and E. Lacson, "Use of machine-learning algorithms to determine features of systolic blood pressure variability that predict poor outcomes in hypertensive patients," *Clin. Kidney J.*, vol. 12, no. 2, pp. 206–212, Jul. 2018, doi: 10.1093/ckj/sfy049.

[83] O. Özdenizci *et al.*, "Time-Series Prediction of Proximal Aggression Onset in Minimally-Verbal Youth with Autism Spectrum Disorder Using Physiological Biosignals," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2018, pp. 5745–5748. doi: 10.1109/EMBC.2018.8513524.

[84] D. Cuesta-Frau *et al.*, "Characterization of artifact influence on the classification of glucose time series using sample entropy statistics," *Entropy*, vol. 20, no. 11, 2018, doi: 10.3390/e20110871.

[85] P. Peng, H. Wei, L. Xie, and Y. Song, "Epileptic Seizure Prediction in Scalp EEG Using an Improved HIVE-COTE Model," 2020, vol. 2020-July, pp. 6450–6457. doi: 10.23919/CCC50068.2020.9188930.

[86] X. Li, Y. Zhang, F. Jiang, and H. Zhao, "A novel machine learning unsupervised algorithm for sleep/wake identification using actigraphy," *Chronobiol. Int.*, vol. 37, no. 7, pp. 1002–1015, Jul. 2020, doi: 10.1080/07420528.2020.1754848.

[87] H. Bragança, J. G. Colonna, W. S. Lima, and E. Souto, "A Smartphone Lightweight Method for Human Activity Recognition Based on Information Theory," *Sensors*, vol. 20, no. 7, Art. no. 7, Jan. 2020, doi: 10.3390/s20071856.

[88] M. H. Pham *et al.*, "Validation of a Step Detection Algorithm during Straight Walking and Turning in Patients with Parkinson's Disease and Older Adults Using an Inertial Measurement Unit at the Lower Back," *Front. Neurol.*, vol. 8, p. 457, Sep. 2017, doi: 10.3389/fneur.2017.00457.

[89] P. Ivaturi, M. Gadaleta, A. C. Pandey, M. Pazzani, S. R. Steinhubl, and G. Quer, "A Comprehensive Explanation Framework for Biomedical Time Series Classification," *IEEE J. Biomed. Health Inform.*, 2021, doi: 10.1109/JBHI.2021.3060997.