

# Supplementary Materials

## Search terms

**Supplementary Table S1.** Search terms for each database.

Database	Search Term
PubMed	("time-series"[tiab] OR "time series"[tiab] OR "temporal"[tiab] OR "sequential"[tiab] OR "continuous"[tiab] OR "longitudinal"[tiab]) AND ("classify"[tiab] OR "classification"[tiab] OR "algorithm"[tiab] OR "machine learning"[tiab]) AND ("*clinic"[tiab] OR "physiology"[tiab] OR "physiological"[tiab] OR "biological"[tiab] OR "*medical"[tiab] OR "health"[tiab] OR "sleep"[tiab] OR "activity"[tiab]) NOT ("gene"[tiab] OR "image"[tiab] OR "imaging"[tiab]) AND ("hidden markov model"[tiab] OR "state space model"[tiab] OR "state-space model"[tiab] OR "wavelet transform"[tiab] OR "time series classification"[tiab] OR "time series classifier"[tiab] OR "time series machine learning"[tiab] OR "dynamic time warping"[tiab] OR "DTW-kNN"[tiab] OR "Time Warp Edit"[tiab] OR "Move split merge"[tiab] OR "derivative distance transform"[tiab] OR "complex invariant distance"[tiab] OR "bag of patterns"[tiab] OR "bag-of-patterns"[tiab] OR "Symbolic Aggregate Approximation - Vector Space Model"[tiab] OR "SAXVSM"[tiab] OR "Bag of SFA Symbols"[tiab] OR "shapelet"[tiab] OR "time series forest"[tiab] OR "time series bag of features"[tiab] OR "learned pattern similarity"[tiab] OR "elastic ensemble"[tiab] OR "collective of transformation ensembles"[tiab] OR "rotation forest"[tiab] OR "kalman filters"[tiab] OR "ARMA model"[tiab]) AND English[lang] NOT ("deep learning"[tiab] OR "LSTM"[tiab] OR "temporal CNN"[tiab] OR "neural network"[tiab] OR "ResNet"[tiab] OR "LSTM"[tiab] OR "RNN"[tiab] OR "Long Short-Term Memory"[tiab])
IEEE	("Abstract": "time series") AND ("Abstract": "classification" OR "Abstract": "machine learning" OR "Abstract": "classifier" OR "Abstract": "algorithm") AND ("Abstract": "biomedical" OR "Abstract": "medical" OR "Abstract": "physiological" OR "Abstract": "biological") NOT ("Abstract": "deep learning" OR "Abstract": "lstm" OR "Abstract": "recurrent neural network" OR "Abstract": "neural network" OR "Abstract": "convolutional neural network" OR "Abstract": "rnn" OR "Abstract": "cnn" )
Scopus	(TITLE-ABS-KEY("*time-series classification") OR TITLE-ABS-KEY("*time-series classifier")) AND (TITLE-ABS-KEY("*clinical") OR TITLE-ABS-KEY("*biomedical") OR TITLE-ABS-KEY("*physiological")) AND NOT (TITLE-ABS-KEY("*gene") OR TITLE-ABS-KEY("*imaging"))
Web of Science	KP=((("time series" OR time-series OR temporal) AND (classification OR "machine learning" OR classifier OR algorithm OR automatic OR "hidden markov model" OR "state space model" OR "state-space model" OR "wavelet transform" OR "time series classification" OR "time series classifier" OR "time series machine learning" OR "dynamic time warping" OR "DTW-kNN" OR "Time Warp Edit" OR "Move split merge" OR "derivative distance transform" OR "complex invariant distance" OR "bag of patterns" OR "bag-of-patterns" OR "Symbolic

	Aggregate Approximation - Vector Space Model" OR "SAXVSM" OR "Bag of SFA Symbols" OR "shapelet" OR "time series forest" OR "time series bag of features" OR "learned pattern similarity" OR "elastic ensemble" OR "collective of transformation ensembles" OR "rotation forest" OR "kalman filters" OR "ARMA model") NOT ("RNN" OR "deep learning*" OR "LSTM" OR "temporal CNN*" OR "Neural Network*")) AND TS= ( biomedical OR medical OR physiological OR biological NOT gene* NOT image*)
--	---

## Algorithm Types and Example Papers

**Supplementary Table S2.** Description for each algorithm type and example papers.

Best Algorithm Type	Description	Example papers
Feature extraction	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.	See Supplementary Table S4a,b
Distance-based methods	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.	[79]
Shapelet based	Shapelet-based methods are similar to significant pattern mining. Time series shapelets are subsequences that maximize classification performance.	[1]
Statistical Modeling	This technique uses statistical modeling (such as Kalman filters or 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.	[37]
Wavelet Transform	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.	[26]
Ensemble-based	Ensemble-based classification algorithms utilize multiple algorithms to make predictions and then aggregate the results coming from these different algorithms	[54]

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.	[62]
-----------------------------------	---	------

## Preprocessing Methods

**Supplementary Table S3.** Summary of time series signal preprocessing methods and example papers.

Preprocessing Steps	Preprocessing Method	Example Papers
Filtering or filters	<ul style="list-style-type: none"> <li>• High Pass Filter</li> <li>• Low Pass Filter</li> <li>• Butterworth Filter</li> <li>• Notch Filter</li> <li>• Moving Average Filter</li> <li>• Median Filter</li> <li>• Wiener Filter</li> <li>• Wavelet FilterSoftware</li> <li>• Chebyshev Type II Band-pass Filter</li> </ul>	[55,80,81]
Artifact removal	<ul style="list-style-type: none"> <li>• Component Analysis</li> <li>• Noise <ul style="list-style-type: none"> <li>○ Discrete Wavelet Transform</li> </ul> </li> <li>• Anonymous Amplitude</li> <li>• Detrending <ul style="list-style-type: none"> <li>○ Locally Estimated Scatterplot Smoothing (LOESS)</li> </ul> </li> <li>• Outlier Detection</li> </ul>	[33,70]
Segmentation	<ul style="list-style-type: none"> <li>• Sliding Window</li> <li>• Interval Proportion</li> <li>• Data Augmentation</li> <li>• Pan-Tompkins Algorithm</li> <li>• Thresholding</li> <li>• Linear interpolation of ectopic beats</li> <li>• Window Discretization</li> </ul>	[71,82]
Resampling	<ul style="list-style-type: none"> <li>• Downsampling</li> <li>• Downscaling</li> </ul>	[41]
Smoothing	<ul style="list-style-type: none"> <li>• Exponential Smoothing <ul style="list-style-type: none"> <li>○ Single</li> <li>○ Double</li> </ul> </li> </ul>	[56,81]

	<ul style="list-style-type: none"> <li>■ Fuzzy Inference Double Exponential Smoothing (FIDES) <ul style="list-style-type: none"> <li>○ Tripple</li> </ul> </li> <li>● Smoothing Filter <ul style="list-style-type: none"> <li>○ Mean Filter</li> <li>○ Moving average filtering</li> <li>○ Median filtering</li> <li>○ Three point linear smoothing filtering</li> <li>○ Five point linear smoothing filtering.</li> </ul> </li> </ul>	
Dimension reduction	<ul style="list-style-type: none"> <li>● Autoregressive (AR) Model</li> <li>● Principal Component Analysis (PCA)</li> </ul>	[83]
Synthetic Data Generation using statistical explosion	<ul style="list-style-type: none"> <li>● Data explosion model</li> </ul>	[29]
Data cleaning	<ul style="list-style-type: none"> <li>● Ensemble of various methods like <ul style="list-style-type: none"> <li>○ outlier detection</li> <li>○ data balancing(undersampling)</li> <li>○ selective removal of missing data.</li> </ul> </li> </ul>	[29,54]
Other	<ul style="list-style-type: none"> <li>● Normalization</li> <li>● Squared</li> <li>● Multivariate feature discretization</li> <li>● Sub-sequence Enumeration</li> <li>● Data Transformation</li> <li>● Data imputation</li> </ul>	[24,84]

## Feature Engineering Techniques and Signal Specific Feature Engineering

**Supplementary Table S4.** (a) Example of feature engineering techniques and papers. (b) Features for common signals and example papers.

(a)		
Feature Type	Feature Names	Example Papers
Amplitude	<ul style="list-style-type: none"> <li>● Mean absolute value</li> <li>● Arithmetic mean</li> <li>● Harmonic mean</li> <li>● Geometric mean</li> <li>● Root mean square</li> </ul>	[23,85]

	<ul style="list-style-type: none"> <li>• Skewness</li> <li>• Kurtosis</li> <li>• Minimum</li> <li>• Maximum</li> <li>• Mode</li> <li>• First Higher-Order Mean Absolute Value</li> <li>• Normalized 1st Higher-Order Mean Absolute Value</li> <li>• Second Higher-Order Mean Absolute Value</li> <li>• Normalized 2nd Higher-Order Mean Absolute Value</li> <li>• Peak-to-peak Amplitude</li> <li>• Peak Amplitude</li> <li>• Mean Relative Time of the Peaks</li> <li>• Mean Relative Time of the Valleys</li> </ul>	
Frequency	<ul style="list-style-type: none"> <li>• Trend</li> <li>• Seasonality</li> <li>• Total power in frequency band</li> <li>• Power spectral density</li> <li>• Bandwidth</li> <li>• Center Frequency</li> <li>• Median Frequency</li> <li>• Mean Frequency</li> <li>• Mode Frequency</li> <li>• Root Mean Square Successive Interval Differences</li> <li>• Ratio between Low Frequency and High Frequency (LF/HF)</li> <li>• Mean normal-to-normal interval (NNI)</li> <li>• Standard deviation of normal-to-normal interval (SDNNI)</li> <li>• Root mean square of successive differences between NNIs (RMSSD)</li> <li>• Number of NNIs &gt; 20 ms (NN20)</li> <li>• Proportion of NN20 among total number of NNIs (pNN20)</li> <li>• Zero-crossings</li> </ul>	[85,86]
Stationarity	<ul style="list-style-type: none"> <li>• Internal Degree of Stationarity</li> <li>• Modified Integral Degree of Stationarity</li> <li>• Modified Mean Degree of Stationarity</li> <li>• Median</li> <li>• Standard Deviation of Mean Vector</li> <li>• Standard Deviation of Standard Deviation Vector</li> </ul>	[85]
Entropy	<ul style="list-style-type: none"> <li>• Fuzzy entropy</li> <li>• Approximate entropy</li> <li>• Sample entropy</li> <li>• Shannon Entropy</li> <li>• Permutation entropy</li> <li>• Slope Entropy</li> <li>• correct conditional entropy</li> <li>• Improved multiscale permutation entropy</li> </ul>	[80,87]

	<ul style="list-style-type: none"> <li>• Multiscale sign series entropy</li> </ul>	
Variability	<ul style="list-style-type: none"> <li>• Interquartile Range</li> <li>• Variability Range</li> <li>• Standard Deviation</li> <li>• Variance</li> <li>• Average amplitude change</li> <li>• Mean Resting Rate</li> <li>• Slope Resting Rate</li> </ul>	[80]
Similarity	<ul style="list-style-type: none"> <li>• Median Coherence</li> <li>• Mutual Information</li> <li>• Modified Integral of Coherence</li> <li>• Mean Coherence</li> <li>• Modified Mean Coherence</li> </ul>	[85]
Linearity	<ul style="list-style-type: none"> <li>• Lag dependence function</li> <li>• Population lag dependence function</li> <li>• Linear fit coefficients</li> <li>• Quadratic fit coefficients</li> <li>• Autoregressive modeling</li> <li>• ARIMA-fit coefficients</li> </ul>	[64,85]
Correlation	<ul style="list-style-type: none"> <li>• Cross-correlation</li> <li>• Auto-correlation</li> <li>• Pearson correlation</li> <li>• Autocovariance</li> </ul>	[48]
Plot-based	<ul style="list-style-type: none"> <li>• Explanation: defined specifically for ECG and each RR interval is plotted as a function of the previous RR interval.</li> <li>• VAI: mean of all the absolute value of angular differences between the lines plotted from the original point to every scatter point and the diagonal line, measuring the angular dispersion of all the points</li> <li>• VLI: standard deviation of all distances of scatter points from the original point, measuring the distance dispersion of all the points</li> <li>• SD1: measuring the width of the ellipse and indicating the short-term variability</li> <li>• Probability density distribution of delta RR intervals: <ul style="list-style-type: none"> <li>○ Blank pane ratio</li> <li>○ Entropy</li> </ul> </li> <li>• Activity (variance of time function)</li> <li>• Poincare plot</li> <li>• Recurrence plot</li> </ul>	[3,80,88]
Transformation Techniques	<ul style="list-style-type: none"> <li>• Temporal Pattern (Shapelet) Extraction</li> <li>• Fast Fourier Transform (FFT)</li> <li>• Short-time fourier transform (STFT)</li> <li>• Waveform decomposition:</li> </ul>	[64,71,89,90]

	<ul style="list-style-type: none"> <li>○ EEG decomposition to 5 levels: (43.4-86.8Hz, 21.7-43.4Hz, 10.8-21.7Hz, 5.4-10.8Hz, 2.7-5.4Hz, 0-2.7Hz)</li> <li>● Dual-tree Complex (Discrete) Wavelet Transform: PPG decomposed into six levels to calculate detail coefficients (high frequency) and approximation coefficients (low frequency).</li> <li>● The Daubechies wavelet of order 4 is the most suitable one for automated epileptic seizure detection.</li> <li>● Continuous wavelet transform: ECG signals can be decomposed up to 5 levels using Daubechies-6 (db6) mother wavelet function.</li> <li>● Principal component analysis</li> <li>● Empirical Mode Decomposition</li> <li>● Recurrence quantification analysis</li> <li>● Singular value decomposition</li> <li>● Dynamic time warping time series distances calculation</li> <li>● Derivatives</li> </ul>	
Undefined	<ul style="list-style-type: none"> <li>● Global volatility Index</li> <li>● Hurst Exponents</li> <li>● Fractal Dimension</li> <li>● Higuchi fractal dimension</li> <li>● Principal components Hjorth parameters</li> <li>● Time Series motif features (Symbolic aggregate approximation): used in the Genetic algorithm</li> <li>● Signal-to-noise ratio</li> <li>● Serial Correlation (TSA)</li> <li>● Self similarity (raw)</li> <li>● Periodicity (raw)</li> <li>● Average Maharaj distance (raw)</li> <li>● Signal-to-noise ratio</li> <li>● Pattern variability</li> <li>● DWT-RWE (Discrete Wavelet Transform and Relative Wavelet Energy)</li> <li>● Relative Wavelet Energy</li> <li>● Average of the first 10 scales of multi-scale entropy</li> <li>● Quadratic sample entropy in TQ intervals (TQEn): specific for ECG or PPG</li> <li>● Lempel-ziv complexity (LZC)</li> <li>● Kolmogorov complexity (KC)</li> </ul>	[25,29]
(b)		
Signal Name	Features	Example
HRV	Frequency-domain: <ul style="list-style-type: none"> <li>● Spectral power in the range <math>f \leq 0.4</math> Hz (Total PWR)</li> <li>● Very low frequency (VLF): 0.0033 - 0.04 Hz</li> <li>● Low frequency (LF): 0.04 - 0.15 Hz</li> <li>● High frequency (HF): 0.15 - 0.4 Hz</li> </ul>	[86,91]

	<ul style="list-style-type: none"> <li>• Total frequency (TF): <math>\leq 0.4</math> Hz</li> <li>• Ratio between Low Frequency and High Frequency (LF/HF)</li> <li>• Spectral Shannon's entropy (SpectEn)</li> </ul> <p>Linear, time-domain:</p> <ul style="list-style-type: none"> <li>• Average of all normal-to-normal intervals (AVNNI)</li> <li>• Mean normal-to-normal interval (NNI)</li> <li>• Standard deviation of normal-to-normal interval (SDNNI)</li> <li>• Root mean square of successive differences between NNIs (RMSSD)</li> <li>• Number of NNIs <math>&gt; 20/50</math> ms (NN20/NN50)</li> <li>• Proportion of NN20/NN50 among total number of NNIs (pNN20/pNN50)</li> <li>• Standard deviation of differences between NNIs (SDSD)</li> <li>• Total number of NNIs divided by height of NNI histogram (HRV_TI)</li> <li>• Baseline width of the minimum square difference triangular interpolation of the highest peak of the histogram of NNIs (TINN)</li> <li>• Average of all R peaks amplitude (mean_R)</li> <li>• Standard deviation of all R peaks amplitude (std_R)</li> <li>• Average of all RRs (mean_RR)</li> <li>• Minimum of all RRs (min_RR)</li> <li>• Median of all RRs (median_RR)</li> <li>• Coefficient of variation of RR intervals (CVrr)</li> <li>• Difference between maximum and minimum RR intervals (<math>\Delta RR_{\text{Imax}}</math>)</li> <li>• Normalized absolute deviation (NADev)</li> <li>• Normalized absolute difference (NADiff)</li> </ul> <p>Time-frequency:</p> <ul style="list-style-type: none"> <li>• Standard deviation of Haar 's wavelet (HaarWavSD)</li> </ul> <p>Nonlinear, phase space:</p> <ul style="list-style-type: none"> <li>• Poincare plot standard deviations ratio (SD1/SD2)</li> <li>• Spatial filling index, <math>d = 2</math>, bins = 5 (SFI)</li> <li>• Central tendency measure (CTM)</li> <li>• Sequential trend analysis (STA)</li> <li>• Recurrence plot features, <math>d = 2</math>, lag = 1 (REC, LMean, DET, RecShEn, Lam)</li> </ul> <p>Nonlinear, fractal:</p> <ul style="list-style-type: none"> <li>• Detrended fluctuation analysis short-term (<math>\leq 11</math> beats) and mid-term (<math>&gt; 11</math> beats) complexity (<math>DFA_{\alpha 1}</math>, <math>DFA_{\alpha 2}</math>)</li> </ul> <p>Nonlinear, entropy:</p> <ul style="list-style-type: none"> <li>• Corrected conditional Shannon's entropy (CCE)</li> <li>• Renyi's entropy (RenyiEn)</li> <li>• Sample entropy (SampEn)</li> <li>• Maximum Sample entropy (MaxSampEn)</li> <li>• Fuzzy approximate entropy (FuzzyApEn1)</li> </ul>	
--	---	--



	<ul style="list-style-type: none"> <li>• Maximum Fuzzy approximate entropy (MaxFuzzyApEn)</li> <li>• Multiscale sample entropy (MultiSampEn)</li> </ul> <p>Nonlinear, other:</p> <ul style="list-style-type: none"> <li>• Lempel-Ziv complexity (LZComp)</li> <li>• Multiscale asymmetry index mean (MultiAiMean)</li> <li>• Standard deviation of Multiscale asymmetry index (MultiAiStDev)</li> <li>• Allan factor</li> </ul> <p>Symbolic dynamics:</p> <ul style="list-style-type: none"> <li>• Alphabet entropy average (AlphEn)</li> <li>• AlphEn average (AlphEnAver)</li> <li>• Maximum AlphEn (MaxAlphEn)</li> <li>• AlphEn variance (AlphEnVar)</li> <li>• Rate of occurrence of AlphEn (Rate_A - Rate_AA)</li> <li>• alphabet AlphEn averages (AverAlphEn B - AverAlphEn AA)</li> </ul>	
EEG	<p>Frequency bands:</p> <ul style="list-style-type: none"> <li>• Delta (&lt; 5 Hz)</li> <li>• Theta (4–7.99 Hz)</li> <li>• Alpha (8–12.99 Hz)</li> <li>• Beta (13–20 Hz)</li> </ul> <p>Waveform decomposition:</p> <ul style="list-style-type: none"> <li>• 43.4-86.8Hz</li> <li>• 21.7-43.4Hz</li> <li>• 10.8-21.7Hz</li> <li>• 5.4-10.8Hz</li> <li>• 2.7-5.4Hz</li> <li>• 0-2.7Hz</li> </ul>	[55]
Photoplethysmography (PPG)	<p>Entropy measures</p> <ul style="list-style-type: none"> <li>• ApEn</li> <li>• SaEn</li> <li>• PE</li> <li>• IMPE</li> </ul>	[92]
Electromyography (EMG)	<p>Amplitude</p> <ul style="list-style-type: none"> <li>• First Higher-Order Mean Absolute Value (HOMAV1)</li> <li>• Normalized 1st Higher-Order Mean Absolute Value (HOMAV1n)</li> <li>• Second Higher-Order Mean Absolute Value (HOMAV2)</li> <li>• Normalized 2nd Higher-Order Mean Absolute Value (HOMAV2n)</li> <li>• Mean Absolute Value (MAV)</li> <li>• Peak to Peak Amplitude (P2P)</li> <li>• Peak Amplitude (PK)</li> <li>• Root Mean Square (RMS)</li> <li>• Mean Relative Time of the Peaks (TMNP)</li> <li>• Mean Relative Time of the Valleys (TMNV)</li> </ul> <p>Variability</p>	[85]

	<ul style="list-style-type: none"> <li>• Interquartile Range (IQR)</li> <li>• Range (R)</li> <li>• Standard Deviation (SD)</li> <li>• Variance (VAR)</li> <li>• Mean Resting Rate (MNRR)</li> <li>• Root Mean Square Successive Interval Differences (RMSSD)</li> <li>• Slope Resting Rate (slopeRR)</li> </ul> <p>Stationarity</p> <ul style="list-style-type: none"> <li>• Internal Degree of Stationarity (IDS)</li> <li>• Median (MD)</li> <li>• Modified Integral Degree of Stationarity (MIDS)</li> <li>• Modified Mean Degree of Stationarity (MMNDS)</li> <li>• Standard Deviation of Mean Vector (SDMN)</li> <li>• Standard Deviation of Standard Deviation Vector (SDSD)</li> </ul> <p>Entropy</p> <ul style="list-style-type: none"> <li>• Approximate Entropy (ADEn)</li> <li>• Fuzzy Entropy (FuzzyEn)</li> <li>• Sample Entropy (SampEn)</li> <li>• Shannon Entropy (ShannonEn)</li> <li>• Spectral Entropy (SpectralEn)</li> </ul> <p>Linearity</p> <ul style="list-style-type: none"> <li>• Lag Dependence Function (LDF)</li> <li>• Population Lag Dependence Function (PLDF)</li> </ul> <p>Similarity</p> <ul style="list-style-type: none"> <li>• Correlation Coefficient (CC)</li> <li>• Median Coherence (MDCOH)</li> <li>• Mutual Information (MI)</li> <li>• Modified Integral of Coherence (MICOH)</li> <li>• Mean Coherence (MNCOH)</li> <li>• Modified Mean Coherence (MMNCOH)</li> </ul> <p>Frequency</p> <ul style="list-style-type: none"> <li>• Bandwidth (BW)</li> <li>• Center Frequency (CF)</li> <li>• Median Frequency (MDF)</li> <li>• Mean Frequency (MNF)</li> <li>• Mode Frequency (MOF)</li> <li>• Zero Crossings (ZC)</li> </ul>	
Electro-Corticography (ECOG)	<p>Low pass and Band pass filtering</p> <ul style="list-style-type: none"> <li>• Count of spikes/poly spikes</li> <li>• Amplitude of spikes/poly spikes</li> <li>• Mean amplitude of epileptic discharge (MAG)</li> </ul>	[93]
Electronic Health Records (EHR)	<ul style="list-style-type: none"> <li>• Date of death is between the ICU admission and discharge time</li> <li>• 25 Phenotype labels</li> </ul>	[94]

## Feature Selection Methods and Examples

**Supplementary Table S5.** Summary of feature selection methods and example papers.

Feature Selection Method Name	Feature Selection Method Brief Description	Example Papers
ANOVA	Uses the ANOVA statistical test to rank contribution of variance in the output to the variance in each feature.	[70]
Chi-Squared	Computes the Chi-squared statistic for each feature to test for independence between each feature and the output	[87]
PCA	Exploits the eigenvectors of the covariance matrix to select the subset of features that contributed most to the output	[57,87]
Relief-F	Compares instances within the training set with the most similar (nearest) instance that is labeled the same, and the most similar (nearest) instance that is labeled differently. Uses these “near-hit” and “near-miss” instances to select which features contributed to the comparison/distinction, respectively.	[55]
Mutual-I	Incrementally selects the feature with the highest effect on the joint probability density function between the input and output.	[55]
Fisher-S	Each feature’s Fisher-score is calculated by computing the distance between each class’s mean of the feature, divided by the variances of the feature in each.	[55]
Minimum Redundancy Maximum Relevance	Incrementally add features to the selected set based on mutual information until the set reaches termination condition. Average mutual information between selected features and its category, and between unselected features and its category are Relevance and Redundancy.	[57,95]
Compensation Distance Evaluation Technique (CDET)	Find the feature sets with the highest difference between two data sets by calculating a value that describes the quality of a feature parameter to separate the condition.	[61]
Temporal interpreted tolerance rough set attribute selection	The proposed attribute selection method determines the attribute relevance and significance based on the temporal pattern mean and trend for each attribute. A temporal similarity measure computes the similarity for each attribute based on its temporal pattern. This temporal based similarity measure defines the lower approximations to construct positive regions. The significance of the attribute is measured using temporal tolerance based degree of dependency. The computed significance attributes are used to predict the effectiveness of the attribute in the classification process.	[56]
Rank-based feature	Descending order of the features’ z-value of area under its receiver	[95]

selection according to their relevance with the class labels	operating characteristics curve. Top-ranked features are removed to eliminate their relevance that might be redundant during classification	
Recursive Feature Elimination (RFE) wrapper approach	Searching for a subset of features by fitting the model, ranking features by importance, discarding the least important features, and refitting the mode until a specified number of features remains.	[54]
Dynamic Time Warping	Calculate an optimal distance between two time series with lengths respectively	[79]
Data drift sensitivity estimate	The proposed framework attempts to identify drift sensitive features by generating an artificial dataset containing all rows of the training and validation datasets. All feature importance estimations are then performed by training an ERT classifier with 1000 trees and using its feature importance estimates Likewise, an artificial target label (i.e. class) is generated, which denotes from which dataset the corresponding row originates.	[64]
Diversified Forward-Backward (DFB) (Wrapper method)	It is a wrapper feature selection method, which trains a model with the training set and evaluates it with the validation set, aiming to evaluate the quality of a feature set. It is parallelized and uses a greedy heuristic to narrow down the search space.	[64]