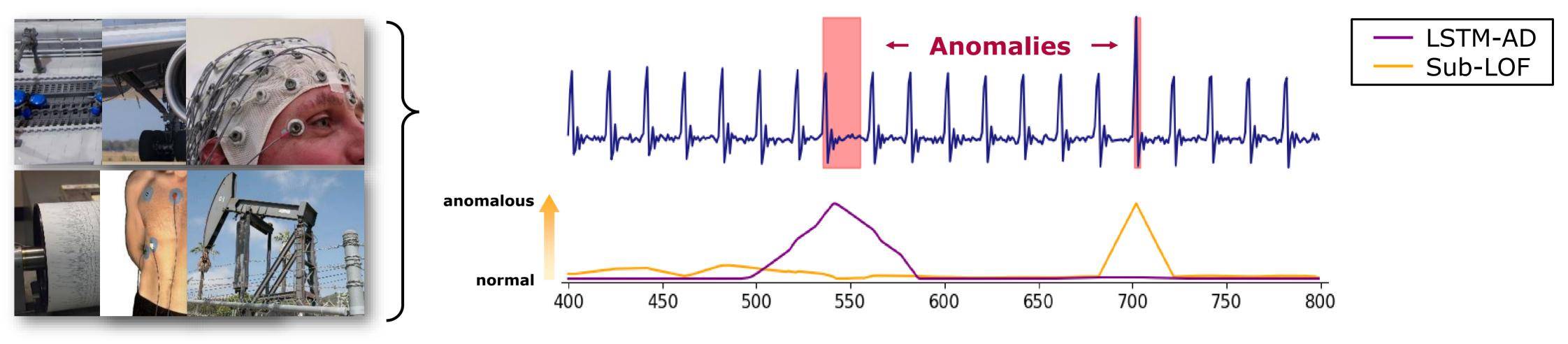
# **Anomaly Detection in Time Series: A Comprehensive Evaluation**

Detecting anomalies in time series is of central interest in many areas because anomalies can indicate important events, such as production faults or heart flicker. Data scientists have developed more 150 specialized algorithms for the automatic detection of anomalous subsequences to deal with the time series' size and complex patterns. However, choosing good algorithms for specific use cases is difficult because no comprehensive study that systematically evaluates the different approaches exists. In our comprehensive study, we carefully evaluate 71 state-of-the-art anomaly detection algorithms on 967 time series datasets.



LaserDBN[64]

### 71 out of 158 Algorithms

RobustPCA [65] Eros-SVMs [43] k-Means [107] XGBoosting [21] **KNN**[72] DWT-MLEAD [93] SR[74]**U-GMM-HMM** [38] I-HMM [88] NetworkSVM [116] MS-SVDD [105] sequenceMiner [14] AOSVM [27] SmartSifter [108]  ${\rm RUSBoost}\,[30]\,\, {\rm OC}\text{-}{\rm KFD}\,[75]$ Signal Analysis PhaseSpace-SVM[52]NoveltySVR [53] GLA [51] Stochastic Online DWT-MLEAD [92] **FFT** [73] Classic ML Random Black Forest [121] **S-SVM** [7] PCA [82] Hybrid K-Means [98] Random Forest Regressor [121] Normalizing Flow [77] SLADE-MTS [100] **PCC** [82] Hybrid KNN[85] LSTM-based EncDec-AD[54] HBOS [26] DeepLSTM [18] SSA [111] VAE-GAN [62] DAE [78] TCN-AE [94] **STOMP** [120] HMAD [28] DeepNAP [41] LSTM-VAE [69] MAD-GAN [45] OmniAnomaly [86] Series 2 Graph [11] $\operatorname{CoalESN}\left[63
ight]_{\emph{Torsk}}\left[34
ight]$ TwoFinger [56] GrammarViz[81] **AD-LTI**[104] Deep Learning PAD [20] DeepAnT [58]  $\mathbf{STORN}\left[84\right]$ KnorrSeq2 [66] Left STAMPi [112] **Donut** [106] MSCRED [115] Ocean WNN [101] MultiHTM [103] Telemanom [36] LSTM-AD [55] **TSBitmap** [102] DADS [80] *FAST-MCD* [76] HOT SAX [39] DissimilarityAlgo [4] RADM [23] SR-CNN [74] TAnoGAN [5] **VELC** [114] AE[78]MoteESN [17] Bagel[47]Norm [9] Data Mining MTAD-GAT [117] NumentaHTM [2] HealthESN [19] **ANODE** [60] Image-embedding-CAE [25] BoehmerGraph [8] VALMOD [49] PST [89]  $\mathrm{MGDD}\left[87
ight]$ MERLIN [61] **STAMP** [112] **ARMA** [12] TARZAN[40]MCOD[42]*CBLOF* [33] Isolation Forest [50] EIF [32] ILOF [71] DAD [110] NormA-SJ[10] LOCI/aLOCI [67] Subsequence IF [50] Subsequence LOF [13]

*COPOD* [48] *IF-LOF* [22] Outlier Detection GeckoFSM [79] Hybrid Isolation COF [91] BLOF [33] DBStream [31] LOF [13] DILOF [59] *Forest* [57]

**EDBN** [70] EM-HMM [68] Learning CxDBN [96] *MultiHMM* [46] HSMM [90] FuzzyDNBC [95] ConInd [3] S-H-ESD[35] SH-ESD+[97]MA [12] EWMA [37] SARIMA [29] Kalman Filter [29]  $\mathrm{AR}\left[12
ight]$ Statistics **PCI**[113] pEWMA [15] MedianMethod [6] EWMA-STR [118] Holt-Winter's [1] **ARIMA** [37] **DSPOT** [83] RePAD [44]

AMD Segmentation [109] Holt's [37]

#### 6 Method Families

Encoding

Distance

**†** Forecasting Reconstruction Distribution + Isolation Tree

#### 967 Datasets

Collection	Datasets
CalIt2	1
Daphnet	3
Exathlon	2
Genesis	1
IOPS	4
KDD-TSAD	249
MGAB	10
MITDB	4
NAB	56
NASA-MSL	16
NASA-SMAP	35
SMD	23
SVDB	16
WebscopeS5	360
GutenTAG	187

## **Evaluation Results**

SurpriseEncoding [16]

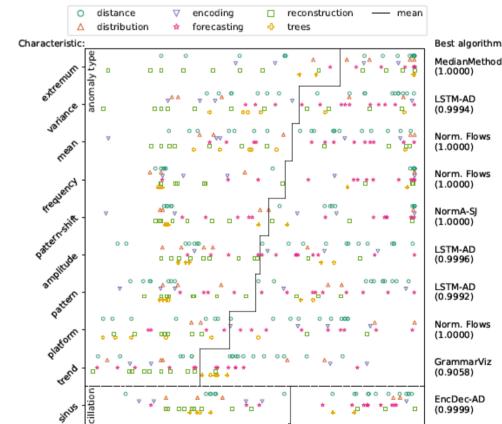
Ensemble GI[24]

NormA-smpl<sup>[10]</sup>

SCRIMP++[119]

- Judgement heavily depends on metric, dataset, and use case. Thus, there is no clear winner!
- Overall high error rates despite our strong investment.
- Hyperparameter optimization is important, but also very difficult.
- Algorithms have their individual strengths and weaknesses.

Dim.	Learn. Algorithm	TL OOM ERR	AUC-ROC all datasets	AUC-PR all datasets	AUC- $\mathbf{P}_T \mathbf{R}_T$ all datasets	AUC-ROC GutenTAG only	O distance	
UNIVARIATE	<ul> <li>Sub-LOF [22]</li> </ul>	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	AUC-ROC all datasets	AUC-PR all datasets	AUC-P <sub>T</sub> R <sub>T</sub> all datasets	AUC-ROC GutenTAG only	△ distribution ★ forecasting ♣ trees Characteristic:	Best alg Median (1.0000 LSTM-A (0.9994 Norm. F (1.0000 Norm.F (1.0000 NormA- (1.0000 LSTM-A (0.9996 LSTM-A
	TSBitmap [144]     DSPOT [122]     FFT [111]     S-H-ESD [62]      Donut [150]     ★ RForest [21]     IE-CAE [44]     ★ XGBoosting [34]     ★ OceanWNN [143]     Bagel [79]     SR-CNN [112]     ▼ TARZAN [71]	$\begin{array}{ccccccc} 0 & & 0 & & 0 & & 0 & \\ 6 & & & 0 & & 0 & & 0 & \\ 0 & & & 0 & & 0 & & 0 & \\ 0 & & & 0 & & 0 & & 0 & \\ 1 & & & 1 & & 2 & & \\ 12 & & & 0 & & 0 & & 0 & \\ 12 & & & 0 & & 0 & & 0 & \\ 0 & & & 0 & & 1 & & \\ 0 & & & 0 & & 0 & & 1 & \\ 0 & & & 0 & & 0 & & 1 & \\ 0 & & & 0 & & & 1 & & \\ 19 & & & 0 & & & 2 & & \\ 22 & & & 0 & & & 1 & & \\ 0 & & & 0 & & & 1 & & \\ 0 & & & 0 & & & 1 & & \\ 0 & & & 0 & & & 1 & & \\ 0 & & & 0 & & & 1 & & \\ \end{array}$						Norm. 1 (1.0000 Gramm (0.9058 EncDec (0.9999 Sub-LO (0.9916 EncDec (0.9898
MULTIVARIATE	<ul> <li>k-Means [151]</li> <li>KNN [110]</li> <li>Torsk [60]</li> <li>EIF [58]</li> <li>iForest [83]</li> <li>HBOS [47]</li> <li>DBStream [55]</li> <li>CBLOF [59]</li> <li>COPOD [80]</li> <li>IF-LOF [36]</li> <li>LOF [22]</li> <li>COF [130]</li> <li>PCC [121]</li> </ul> <li>LSTM-AD [89]</li> <ul> <li>HealthESN [32]</li> <li>Telemanom [64]</li> <li>RBForest [165]</li> <li>EncDec-AD [88]</li> <li>DeepAnT [94]</li> <li>OmniAnomaly [125]</li> <li>LaserDBN [100]</li> </ul> <ul> <li>LaserDBN [100]</li> </ul>	$\begin{array}{cccccccccccccccccccccccccccccccccccc$						Sub-LO (0.9854 EncDec (0.9324 DBStrei (0.9997 LSTM-A (0.9975 LSTM-A (0.9700 Sub-LO (0.9989 EncDec (0.9948 Sub-LO (0.9280
	Image: Second State       RobustPCA [101]         Image: TAnoGan [8]       Image: Hybrid KNN [124]         Image: Hybrid KNN [124]       Image: Hybrid KNN [124]	0% 0% 0% 65% 0% 1%						LSTM-A (0.9748 Norm. F (0.9738



- Simple and fast algorithms are very competitive.
- Robust and effective algorithms: Sub-LOF, GrammarViz, DWT-MLEAD, k-Means, and Telemanom.
- Deep learning algorithms are not superior to classic algorithms.
- Supervised algorithms are not superior to semi-supervised or unsupervised algorithms.

Sebastian Schmidl 💿
Phillip Wenig 💿
Thorsten Papenbrock 💿

Information Systems Group Hasso Plattner Institute, University of Potsdam Potsdam, Germany h E-Mail: {firstname.lastname}@hpi.de

TimeEval





	ttps://github.com/HPI-Information-Sy	<pre>ystems/TimeEval</pre>	doi:10.14778/3538598.3538602
--	--------------------------------------	----------------------------	------------------------------