ReLATE: Resilient Learner Selection for Multivariate Time-Series Classification Against Adversarial Attacks

Cagla Ipek Kocal[1], Onat Gungor[2], Aaron Tartz[1], Tajana Rosing[2], Baris Aksanli[1]

[1]San Diego State University
[2]University of California, San Diego
{ckocal0169, atartz0694, baksanli}@sdsu.edu, {ogungor, tajana}@ucsd.edu

Abstract

Minimizing computational overhead in time-series classification, particularly in deep learning models, presents a significant challenge. This challenge is further compounded by adversarial attacks, emphasizing the need for resilient methods that ensure robust performance and efficient model selection. We introduce ReLATE, a framework that identifies robust learners based on dataset similarity, reduces computational overhead, and enhances resilience. ReLATE maintains multiple deep learning models across well-known adversarial attack scenarios, capturing model performance. ReLATE identifies the most analogous dataset to a given target using a similarity metric, then applies the optimal model from the most similar dataset. ReLATE reduces computational time by an average of 80.25%, enhancing adversarial resilience and streamlining robust model selection, all without sacrificing accuracy, performing within 3.8% of the oracle.

Introduction

Various tasks rely on time-series data, i.e., sequences of observations collected over intervals (Adhikari and Agrawal 2013), including content-based querying (Yeh et al. 2023), anomaly detection (Schmidl et al. 2022), clustering (Holder, Middlehurst, and Bagnall 2024), and classification (Ismail Fawaz et al. 2019). Among these tasks, time-series classification with machine learning (ML) has crucial use cases, e.g., epileptic activity classification using EEG signals (Varlı and Yılmaz 2023), health prediction using biomedical data (Wang et al. 2022), and smart agriculture using multispectral satellite imagery (Simón Sánchez et al. 2022), requiring robust, resilient, and accurate ML-based solutions.

Time-series ML applications face significant challenges due to the dynamic nature of streaming data, which is often limited or incomplete in real-time environments, making it impractical to wait for sufficient data accumulation to retrain models (Suárez-Cetrulo et al. 2023). Moreover, training ML models on new data is both computationally expensive and time-consuming, further complicating the process (Dempster et al. 2020). In this context, deep learning (DL) models are often favored for multivariate time-series classification tasks due to their ability to automatically extract relevant features. However, these DL models could show significant variability in classification performance, as shown in Figure 1. These results highlight the substantial impact of model choice on classification outcomes, underscoring the critical need for careful and informed model selection in the context of DL-based time-series analysis. This motivates the need for efficient DL model selection methods that can adapt to new incoming data without requiring extensive retraining.

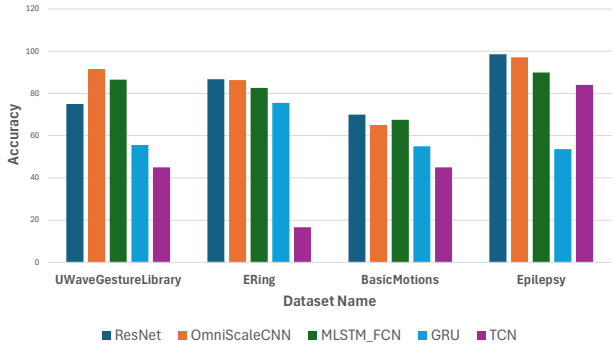

Figure 1: DL performance on multivariate time-series data

DL models are also vulnerable to adversarial attacks as models are susceptible to small, imperceptible changes in data, particularly when the data is limited or incomplete. These attacks introduce deliberate perturbations that obscure critical patterns, that might lead to misclassifications in critical applications (Goodfellow et al. 2014). For instance, small perturbations in medical sensor data could lead to incorrect diagnoses, potentially endangering lives, while adversarial attacks in security systems could result in unauthorized access or compromised safety (Fawaz et al. 2019; Gungor et al. 2023). Addressing these risks requires developing methods that not only adapt efficiently to new incoming data but also exhibit resilience against adversarial attacks.

We propose ReLATE to tackle the computational and retraining challenges in deep learning-based time-series classification with streaming data, particularly in the presence of adversarial attacks. ReLATE streamlines the selection of resilient algorithms using dataset similarity, eliminating the need for exhaustive testing and retraining across all models and datasets. This reduces overhead while ensuring robust classification performance. ReLATE achieves substan-

tial computational savings, averaging 80.25% reduction in overhead, while maintaining strong classification performance, within 3.8% of the oracle, providing a scalable and efficient solution for time-series classification under both adversarial and non-adversarial conditions.

Related Work

Multivariate Time-Series Classification (MTSC)

Numerous ML algorithms are designed to enhance the scalability and predictive capabilities of models for time-series classification (Faouzi 2022). Zheng et al. developed a framework with multi-channel deep convolutional neural networks in combination with Multilayer Perceptrons (MLPs) (Zheng et al. 2016). Grabocka et al. introduced a shapelet-based method with supervised selection and online clustering (Grabocka et al. 2016). Ruiz et al. proposed a method combining DL, shapelets, bag-of-words approaches, and independent ensembles (Ruiz et al. 2020). Baldán et al. employed feature-based methods with traditional classifiers (Baldán and Benítez 2021). Despite significant advancements in MTSC, the computational overhead associated with existing approaches remains a critical limitation. Most existing approaches rely on exhaustive training procedures, which demand significant computational resources.

Similarity Based Approaches

Similarity analysis plays a crucial role in ML and time-series analysis, offering a foundation for tasks such as dataset comparison, feature selection, and model evaluation. Marks examined three measures of similarity for comparing two sets of time-series vectors, including the Kullback-Leibler divergence, the State Similarity Measure, and the Generalized Hartley Metric (Marks 2013). Bounliphone et al. introduced a statistical test of relative similarity to address challenges in model selection for probabilistic generative frameworks (Bounliphone et al. 2015). Assegie et al. proposed a feature selection method using dataset similarity to improve the classification performance (Assegie et al. 2023). Existing methods are tailored for static datasets and do not account for the dynamic nature of time-series data; thus, they lack the adaptability required for evolving data and resilience against adversarial attacks.

Adversarial Attacks in MTSC

Several studies explored the vulnerabilities of time-series classification against adversarial attacks, subtle perturbations in data to reduce model performance, and proposed solutions to increase resiliency. Harford et al. focused on MTSC models, adapting Adversarial Transformation Network on a distilled model, and demonstrated that models like 1-NN Dynamic Time Warping and Fully Convolutional Networks are highly vulnerable (Harford et al. 2020). Galib et al. analyzed time-series regression and classification model performance under adversarial attacks, finding that Recurrent Neural Network (RNN) models were highly susceptible (Galib and Bashyal 2023). Siddiqui et al. proposed a regularization-based defense against adversarial attacks (Siddiqui et al. 2020). Gungor et al. developed a stacking ensemble learning-based framework that stays resilient against various adversarial attacks (Gungor et al. 2022). While these studies provide valuable insights, they often overlook the critical challenge of computational overhead and the need for extensive retraining or experimentation.

ReLATE addresses two challenges in existing methods: computational overhead and adversarial resilience. By leveraging a similarity-based technique, we eliminate the need for exhaustive DL model retraining when new data arrives. ReLATE not only adapts to the dynamic nature of new time-series data but also enhances robustness against adversarial attacks by prioritizing resilient model selection.

ReLATE: Proposed Framework

Our framework, ReLATE, enables resilient model selection for time-series classification while minimizing computational overhead. Instead of exhaustive DL model testing, ReLATE uses dataset similarity to identify robust models, selecting those most similar to the target dataset for improved performance. The framework includes a Performance Benchmark Database with pre-recorded metrics for various DL models and datasets under different adversarial attack scenarios. By matching new data with the most similar dataset in the repository, ReLATE selects the top-performing models tailored to the new dataset. The individual components of ReLATE are shown in Figure 2.

Module 1: DL Model Training

This module trains deep learning models on datasets from a specific application domain. The process begins with hyper-parameter tuning to identify the optimal settings for each model and dataset. With the best configurations, the models are trained, and their performance metrics (accuracy and F1-score) are recorded in the Performance Benchmark Database. The training dataset is used for model training, the validation dataset for hyper-parameter tuning, and the test dataset is reserved for evaluating final performance. Ultimately, this module processes the input datasets and generates the clean data (i.e., no adversarial attack) performance. The selected DL models span a diverse range, each deliberately selected to address specific challenges of time-series data, including capturing long-range temporal dependencies, identifying cyclical patterns within sequential data, and managing high dimensionality in multivariate settings. Overall, we select 14 different state-of-the-art DL models: LSTM (Graves and Graves 2012), GRU (Chung et al. 2014), MLP (Wang et al. 2017), FCN (Wang et al. 2017), ResNet (Wang et al. 2017), LSTM-FCN (Karim et al. 2017), GRU-FCN (Karim et al. 2017), Multi-Scale Weighted Dense Network (Elsayed et al. 2018), Temporal Convolutional Network (Bai et al. 2018), MLSTM-FCN (Karim et al. 2019), InceptionTime (Ismail Fawaz et al. 2020), Residual CNN (Zou et al. 2019), OmniScaleCNN (Tang et al. 2020), and Explainable Convolutional Model (Fauvel et al. 2021).

Module 2: Applying Adversarial Attacks

In this module, each dataset-model pairing is subjected to nine state-of-the-art adversarial attacks, including both

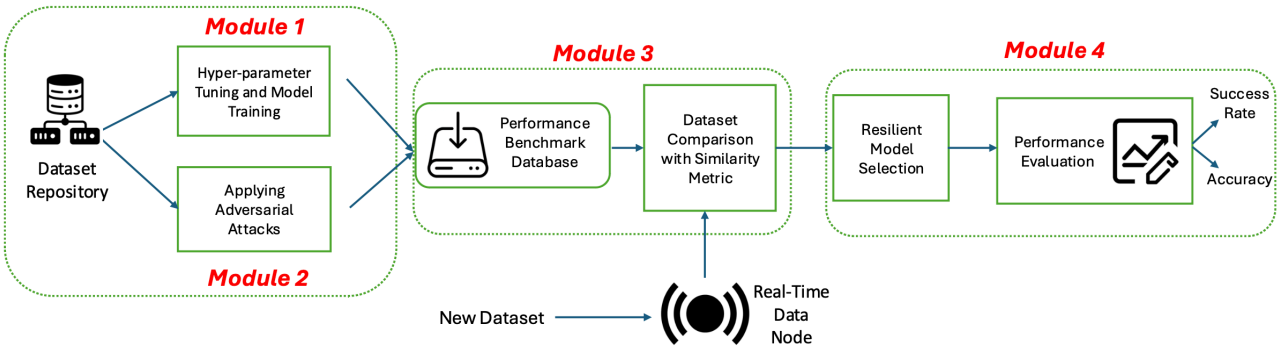

Figure 2: ReLATE framework components

white-box and black-box methods, to evaluate model resilience across diverse threat scenarios. White-box attacks exploit full access to the model's architecture, parameters, and gradients to generate adversarial examples, while black-box attacks depend solely on the model's outputs, without any knowledge of its internal structure (Chakraborty et al. 2021). This module generates adversarial attack versions of each dataset, testing the adversarial robustness of DL models against various types of attacks. The results of these evaluations, e.g., accuracy, F1-score, and attack success rate (ASR), are recorded in the Performance Benchmark Database. During this stage, adversarial attacks are applied to the test portion of the data using models trained in the first module with optimized hyperparameters. Each adversarial attack is chosen for its unique approach to disrupting data and exposing model weaknesses, ranging from simple gradient-based methods to complex iterative strategies. Overall, we select nine white-box and black-box adversarial attacks: Fast Gradient Sign Method (FGSM) (Goodfellow et al. 2014), DeepFool (Moosavi-Dezfooli et al. 2016), Carlini & Wagner (Carlini and Wagner 2017), Basic Iterative Method (BIM) (Kurakin et al. 2018), Momentum Iterative Method (MIM) (Dong et al. 2018), ElasticNet (Chen et al. 2018), Auto Projected Gradient Descent (Croce and Hein 2020), Zeroth order optimization (ZOO) (Chen et al. 2017), and Boundary attack (Brendel et al. 2017).

Module 3: Similarity Comparison

When a new dataset is introduced to the system, mimicking real-time data conditions, ReLATE initiates a similarity comparison process to aid in model selection. The process begins by training a lightweight DNN-based similarity function, built on a simple CNN architecture, using the training portion of each dataset in the Performance Benchmark Database. Once trained, these CNNs are used to extract feature embeddings from the validation portions of their respective datasets. For the incoming dataset, a separate similarity function is trained using the training portion of its data, and embeddings are extracted from the remaining validation portion. These embeddings are then normalized to ensure efficient quantification of average similarity between datasets based on their feature distributions. The embeddings extracted from the incoming dataset are compared to those in the Performance Benchmark Database using a predefined Embedding Similarity Metric, such as Cosine similarity. The dataset with the highest similarity score is identified, and its top three performing DL models (ranked by test performance recorded in Modules 1 and 2) are selected. This approach minimizes computational overhead by eliminating the need to retrain or test all models on the incoming dataset.

Table 1: Selected datasets from the UEA repository

Dataset	Train	Test	Dim.	Len.	Classes
RacketSports	151	152	6	30	4
NATOPS	180	180	24	51	6
UWaveGestureLibrary	120	320	3	315	8
Cricket	108	72	6	1197	12
ERing	30	270	4	65	6
BasicMotions	40	40	6	100	4
Epilepsy	137	138	3	206	4

Module 4: Resilient Model Selection

In this module, we evaluate the performance of the top three DL models selected using the similarity-based approach from Module 3 on the new incoming dataset. The models are trained on the training portion of the new dataset and evaluated on the test portion to assess accuracy and resilience. For clean data, we measure performance with accuracy. For the adversarial data, we assess resilience using attack success rate (ASR). This selection is directly informed by the similarity analysis, ensuring that the chosen model is well-suited to handle the unique characteristics and potential adversarial challenges of the new data. The best model is then deployed on the new dataset for real-time use.

Experimental Setup

Dataset Description: We use seven datasets from the UEA multivariate time-series classification repository (Bagnall et al. 2018). We focus on Human Activity Recognition (HAR) due to its inherently multivariate and complex motion data. However, our method can be applied to any domain that involves time-series data. The chosen datasets were guided by UEA's type categorization criteria to ensure relevance and diversity. Variations in training size, test size,

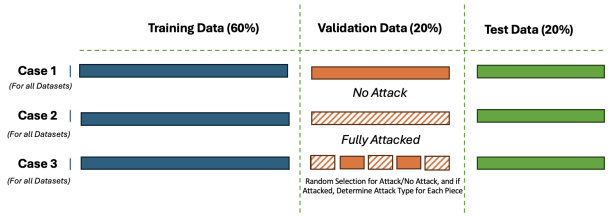

Figure 3: Dataset partitioning for each case

dimensions, sequence length, and class count, as shown in Table 1, ensure diverse dataset characteristics.

Hardware Setup: We use a PC equipped with an Intel Core i7-9700K CPU (8 cores), 32 GB of RAM, and a 16 GB NVIDIA GeForce RTX 2080 dedicated GPU.

Evaluation Metrics: We use three metrics: accuracy, F1-score, and attack success rate (ASR). Accuracy measures the proportion of correctly classified instances out of the total number of samples. F1-score evaluates the balance between precision and recall, making it well-suited for datasets with class imbalance. ASR measures the effectiveness of adversarial attacks by calculating the percentage of instances where model predictions are successfully altered.

Dataset Similarity Calculation: We quantify dataset similarity using a custom CNN with two 1D convolutional layers, adaptive max-pooling, and dropout to extract features from both clean and attacked data. The final fully connected layer maps these features to class predictions. The resulting embeddings are then normalized using L2 normalization to ensure consistency and scale invariance. We use cosine similarity between the embeddings, which measures the angular similarity between two vectors in a multi-dimensional space:

$$\text{Cosine Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the embedding vectors. Cosine similarity ranges from 0 (orthogonal) to 1 (identical).

Baselines: We compare ReLATE against three baseline approaches: random model selection, Oracle (best model) and worst model performance. Random model selection uses a Monte Carlo approach, where an ML model is randomly chosen for each dataset, and its performance is recorded. This process is repeated 1,000 times, with the average performance calculated as the baseline score for random selection. Oracle represents the maximum accuracy recorded on the test data for each dataset among all evaluated models. This score reflects the upper performance bound achievable by selecting the most optimal model with exhaustive search. While the Oracle is computationally intensive and impractical in real-world scenarios, it serves as a reference point for the best model performance. The worst model performance represents the performance of the least effective model among all the possible DL models in the database, reflecting the lower performance bound.

New Incoming Data Setup

We conduct a series of experiments where each of the seven datasets in our Performance Benchmark Database is treated as a new, unseen arrival in rotation. In each round, one dataset is designated as the "new arrival", while the remaining datasets are treated as the pre-existing "drive" datasets. For each drive dataset, we record key performance metrics, accuracy, F1-score, and ASR, across all models, including adversarially attacked versions. Figure 3 visually illustrates each case, explicitly showing the training-validation splits. In all cases, the training dataset is split into an 80% training set and a 20% validation set to facilitate model training and validation for comparisons. This split is determined after conducting several trials to identify the most effective training-validation ratio. The custom CNN is trained using the training portion of the dataset, while the validation portion is utilized for similarity measurement during case implementation. Overall, we design three cases to assess model selection under different adversarial conditions:

Case 1: No Adversarial Attack – We assume that the new incoming dataset is clean and compare it against the clean versions of each dataset in the drive. We find the most similar dataset in the drive to the new dataset, and select its three best-performing models. We evaluate the selected models on the newly arriving dataset to identify the best among them.

Case 2: Fully Attacked – We assume the new incoming dataset is subject to a single type of adversarial attack and perform dataset similarity analysis under similarly attacked conditions. We identify the most similar attacked dataset and select its three best models, based on ASR, where lower ASR indicates better model resilience. The selected models are then evaluated on the arriving dataset.

Case 3: Partially Attacked – We create a randomized partial attack scenario to simulate unknown and intermittent attacks. We divide each dataset into five segments and determine what action applied to each segment with a two step randomization procedure. The first step decides whether a segment would be attacked. If attacked, the second step selects the specific adversarial attack type. This attack pattern is applied identically across all datasets for consistency, to both the incoming and all drive datasets.

Results

We evaluate ReLATE's performance based on previously introduced cases. Table 2 shows ReLATE's performance on Case 1. Oracle is based on the highest accuracy achieved for each dataset, averaged over all datasets. Random selection reports average accuracy across multiple random trials for all datasets. The worst-performing model performance has the lowest accuracy values for each dataset, averaged across all datasets. We can observe that ReLATE achieves accuracy close to Oracle, with only a 3.3% difference, significantly outperforming the random selection by 12.4%. This reflects ReLATE's ability to effectively match datasets with suitable models, leveraging dataset similarity to guide accurate model selection without exhaustively testing all of them.

Figure 4 compares Oracle (red), ReLATE (green), random model selection (blue), and the worst model (purple) performance under diverse adversarial attack scenarios. These scenarios include the results of Case 2 for the attacks DeepFool, Boundary Attack, and Carlini & Wagner. The purpose of evaluating these diverse attack types is to assess ReLATE's

Table 2: ReLATE results for Case 1 using accuracy

Case	Oracle	ReLATE	Random Model Selection	Worst Model Performance
Case 1	91.8	88.5	76.1	32.51

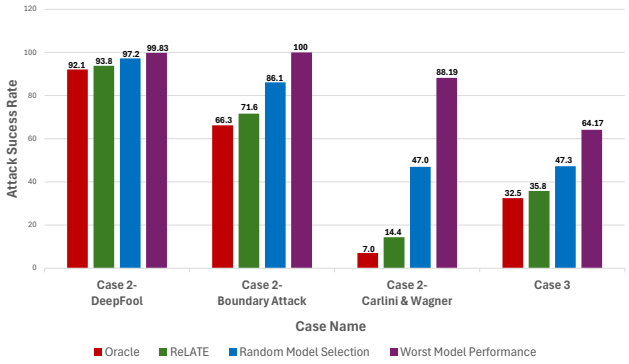

Figure 4: ReLATE ASR results for Case 2 and 3

performance across a range of adversarial conditions. Also, case 3 is repeated ten times to account for the randomization inherent in their scenarios and to evaluate model consistency across distinct random attack scenarios. The evaluation also considers their average performance across these ten random adversarial attack scenarios. For these cases, Oracle is defined as the lowest ASR for each dataset, averaged across all datasets. Random selection performance is calculated with a Monte Carlo approach, while the worst model performance corresponds to the highest ASR values averaged across all datasets. In these cases, ReLATE achieves an ASR that is 4.1% higher than Oracle's average and 14.2% lower than random model selection's average across the two cases: Case 2 and 3. This highlights ReLATE's ability to leverage dataset specific similarities to recommend models with higher adversarial robustness. By aligning model recommendations with the feature distributions of datasets, ReLATE mitigates the impact of attacks.

ReLATE's performance consistently aligns closer to Oracle across all cases, even when the range between the minimum and maximum values is broad. This demonstrates ReLATE's ability to achieve near-optimal performance and maintain robustness across diverse scenarios, all with computational savings by avoiding exhaustive model testing.

Overhead Analysis

To achieve the best possible model performance (Oracle), all DL models have to be trained. Thus, Oracle overhead is determined by evaluating all models for each case. This process involves training each model and applying all adversarial attacks, ensuring a comprehensive assessment of their performances. Since ReLATE focuses on choosing the most similar dataset and evaluating only the top three models, it significantly reduces computational overhead. In Case 1, ReLATE reduces model training and evaluation overhead by 85%, demonstrating its efficiency in scenarios without adversarial attacks. In Case 2, ReLATE reduces model training and evaluation overhead by 78.16% for the DeepFool attack, 73.0% for the Carlini & Wagner attack, and 80.50% for the Boundary Attack, with an average improvement of 77.22%. In Cases 3, ReLATE achieves an average overhead reduction of 81.19% respectively. Overall, ReLATE reduces overhead by an average of 81.14% across all cases. Considering that the similarity metric calculation overhead accounts for 1% in Case 1, 1.08% in Case 2, and 0.58% in Case 3 of the Oracle overhead on average across datasets, ReLATE achieves an 84% reduction in Case 1, 76.14% in Case 2, and 80.61% in Case 3 when this similarity metric overhead is included. These case results represent the average reduction scores across all datasets, culminating in an average overall reduction of 80.25% across all cases.

Conclusion

Time-series data presents challenges due to its dynamic and often unpredictable nature. This variability complicates the task of anticipating the type of data to be encountered, whether adversarially attacked, incomplete, or limited. In such scenarios, traditional model retraining is impractical due to the substantial computational overhead required, particularly in real-time environments where data accumulation may be insufficient. Moreover, the vulnerability of deep learning models to adversarial attacks exacerbates these challenges, as even small perturbations in data can lead to significant misclassifications. These factors underscore the critical need for efficient model selection methods that minimize retraining overhead while maintaining resilience against adversarial attacks. To address these issues, we propose ReLATE, a resilient learner selection mechanism against adversarial attacks. ReLATE leverages dataset similarity to efficiently select resilient models for multivariate time-series classification, minimizing the need for exhaustive model testing. Experimental results show that ReLATE reduces computational overhead by an average of 80.25%, performs within 3.8% of the Oracle, and outperforms random model selection by an average of 13.6%.

Acknowledgements

This work has been funded in part by NSF, with award numbers #1826967, #1911095, #2003279, #2052809, #2100237, #2112167, #2112665, and in part by PRISM and CoCoSys, centers in JUMP 2.0, an SRC program sponsored by DARPA.

References

Adhikari, R.; and Agrawal, R. K. 2013. An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.

Assegie, T. A.; Salau, A. O.; Omeje, C. O.; and Braide, S. L. 2023. Multivariate sample similarity measure for feature selection with a resemblance model. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(4): 4512–4520.

Bagnall, A.; Lines, J.; Hills, J.; Mapp, J.; and Keogh, E. 2018. UEA Multivariate Time Series Classification Repos-

itory. https://www.timeseriesclassification.com/dataset.php. Accessed: 2024-11-29.

Bai, S.; et al. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Baldán, F. J.; and Benítez, J. M. 2021. Multivariate times series classification through an interpretable representation. *Information Sciences*, 569: 596–614.

Bounliphone, W.; Belilovsky, E.; Blaschko, M. B.; Antonoglou, I.; and Gretton, A. 2015. A test of relative similarity for model selection in generative models. *arXiv preprint arXiv:1511.04581*.

Brendel, W.; et al. 2017. Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. *arXiv preprint arXiv:1712.04248*.

Carlini, N.; and Wagner, D. 2017. Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy (sp)*, 39–57. Ieee.

Chakraborty, A.; et al. 2021. A survey on adversarial attacks and defences. *CAAI Transactions on Intelligence Technology*, 6(1): 25–45.

Chen, P.-Y.; Sharma, Y.; Zhang, H.; Yi, J.; and Hsieh, C.-J. 2018. Ead: elastic-net attacks to deep neural networks via adversarial examples. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Chen, P.-Y.; Zhang, H.; Sharma, Y.; Yi, J.; and Hsieh, C.-J. 2017. Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. In *Proceedings of the 10th ACM workshop on artificial intelligence and security*, 15–26.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Croce, F.; and Hein, M. 2020. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. In *International conference on machine learning*, 2206–2216. PMLR.

Dempster, A.; et al. 2020. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5): 1454–1495.

Dong, Y.; Liao, F.; Pang, T.; Su, H.; Zhu, J.; Hu, X.; and Li, J. 2018. Boosting adversarial attacks with momentum. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 9185–9193.

Elsayed, N.; et al. 2018. Deep gated recurrent and convolutional network hybrid model for univariate time series classification. *arXiv preprint arXiv:1812.07683*.

Faouzi, J. 2022. Time series classification: A review of algorithms and implementations. *Machine Learning (Emerging Trends and Applications)*.

Fauvel, K.; Lin, T.; Masson, V.; Fromont, É.; and Termier, A. 2021. Xcm: An explainable convolutional neural network for multivariate time series classification. *Mathematics*, 9(23): 3137.

Fawaz, H. I.; Forestier, G.; Weber, J.; Idoumghar, L.; and Muller, P.-A. 2019. Adversarial attacks on deep neural networks for time series classification. In *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE.

Galib, A. H.; and Bashyal, B. 2023. On the susceptibility and robustness of time series models through adversarial attack and defense. *arXiv preprint arXiv:2301.03703*.

Goodfellow, I. J.; et al. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Grabocka, J.; et al. 2016. Fast classification of univariate and multivariate time series through shapelet discovery. *Knowledge and information systems*, 49: 429–454.

Graves, A.; and Graves, A. 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, 37–45.

Gungor, O.; et al. 2022. Stewart: Stacking ensemble for white-box adversarial attacks towards more resilient data-driven predictive maintenance. *Computers in Industry*, 140: 103660.

Gungor, O.; et al. 2023. Adversarial-hd: Hyperdimensional computing adversarial attack design for secure industrial internet of things. In *Proceedings of Cyber-Physical Systems and Internet of Things Week 2023*, 1–6.

Harford, S.; et al. 2020. Adversarial attacks on multivariate time series. *arXiv preprint arXiv:2004.00410*.

Holder, C.; Middlehurst, M.; and Bagnall, A. 2024. A review and evaluation of elastic distance functions for time series clustering. *Knowledge and Information Systems*, 66(2): 765–809.

Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; and Muller, P.-A. 2019. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4): 917–963.

Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D. F.; Weber, J.; Webb, G. I.; Idoumghar, L.; Muller, P.-A.; and Petitjean, F. 2020. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6): 1936–1962.

Karim, F.; Majumdar, S.; Darabi, H.; and Chen, S. 2017. LSTM fully convolutional networks for time series classification. *IEEE access*, 6: 1662–1669.

Karim, F.; Majumdar, S.; Darabi, H.; and Harford, S. 2019. Multivariate LSTM-FCNs for time series classification. *Neural networks*, 116: 237–245.

Kurakin, A.; et al. 2018. Adversarial examples in the physical world. In *Artificial intelligence safety and security*, 99–112. Chapman and Hall/CRC.

Marks, R. E. 2013. Validation and model selection: Three similarity measures compared. *Complexity Economics*, 2: 41–61.

Moosavi-Dezfooli, S.-M.; et al. 2016. Deepfool: a simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2574–2582.

Ruiz, A. P.; et al. 2020. Benchmarking multivariate time series classification algorithms. *arXiv preprint arXiv:2007.13156*.

Schmidl, S.; et al. 2022. Anomaly detection in time series: a comprehensive evaluation. *Proceedings of the VLDB Endowment*, 15(9): 1779–1797.

Siddiqui, S. A.; et al. 2020. Benchmarking adversarial attacks and defenses for time-series data. In *International Conference on Neural Information Processing*, 544–554. Springer.

Simón Sánchez, A.-M.; González-Piqueras, J.; de la Ossa, L.; and Calera, A. 2022. Convolutional neural networks for agricultural land use classification from Sentinel-2 image time series. *Remote Sensing*, 14(21): 5373.

Suárez-Cetrulo, A. L.; et al. 2023. A survey on machine learning for recurring concept drifting data streams. *Expert Systems with Applications*, 213: 118934.

Tang, W.; Long, G.; Liu, L.; Zhou, T.; Blumenstein, M.; and Jiang, J. 2020. Omni-scale cnns: a simple and effective kernel size configuration for time series classification. *arXiv preprint arXiv:2002.10061*.

Varlı, M.; and Yılmaz, H. 2023. Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning. *Journal of Computational Science*, 67: 101943.

Wang, W. K.; Chen, I.; Hershkovich, L.; Yang, J.; Shetty, A.; Singh, G.; Jiang, Y.; Kotla, A.; Shang, J. Z.; Yerrabelli, R.; et al. 2022. A systematic review of time series classification techniques used in biomedical applications. *Sensors*, 22(20): 8016.

Wang, Z.; et al. 2017. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, 1578–1585. IEEE.

Yeh, C.-C. M.; Chen, H.; Dai, X.; Zheng, Y.; Wang, J.; Lai, V.; Fan, Y.; Der, A.; Zhuang, Z.; Wang, L.; et al. 2023. An efficient content-based time series retrieval system. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 4909–4915.

Zheng, Y.; Liu, Q.; Chen, E.; Ge, Y.; and Zhao, J. L. 2016. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science*, 10: 96–112.

Zou, X.; Wang, Z.; Li, Q.; and Sheng, W. 2019. Integration of residual network and convolutional neural network along with various activation functions and global pooling for time series classification. *Neurocomputing*, 367: 39–45.