Exploring Multiple Classification Systems for Online Time Series Anomaly Detection

1st Inmaculada Santamaria-Valenzuela
Computer Systems Department
Universidad Politécnica de Madrid,
Madrid, Spain
mi.santamaria@ upm.es
0000-0002-7497-8795

2nd Victor Rodriguez-Fernandez
Computer Systems Department
Universidad Politécnica de Madrid,
Madrid, Spain
victor.rfernandez@upm.es
0000-0002-8589-6621

3rd David Camacho
Computer Systems Department
Universidad Politécnica de Madrid,
Madrid, Spain
david.camacho@upm.es
0000-0002-5051-3475

Abstract—Online time series play a crucial role in the analysis and management of massive amounts of data. These series capture data points chronologically according to their acquisition time. Detecting anomalies (outliers) in these time series is crucial for understanding patterns and making informed decisions. This work exposes various techniques from the literature for online time series anomaly detection, and categorises them into statistical techniques. The paper shows several applications for statistical methods, machine learning, and hybrid methods that leverage the advantages of both statistical and deep learning techniques. Furthermore, ensembles are exposed as an efficient technique used with the mentioned models for anomaly detection in online time series. The paper discusses the challenges associated with temporal correlation, including the need for effective visualisation tools such as DeepVats. By providing an overview of the existing techniques and their applications, this work aims to contribute to the advancement of online time series anomaly detection and provide insight for future research in the field.

Keywords—Online Time Series, Anomaly detection, Visual Analytics, Ensembles

I INTRODUCTION

Online time series, used for massive data generation, capture data points arranged chronologically based on their acquisition time. These time-ordered data sequences, known as time series, often exhibit temporal correlations, making data modelling essential for understanding their behavior [1]. Detection of time series data anomalies or outliers is crucial in big data analysis because their patterns provide valuable information for decision making in a variety of applications such as fraud detection, healthcare monitoring, and performance analysis [2], [3]. Thus, with the growing availability of online time series data and the need for real-time insights, research on anomaly detection in online time series has gained significant interest in recent years.

Using real-time data streams, online time series data analysis provides valuable insight and enables proactive decision-making. However, online time series data analysis poses unique challenges and requires specialised techniques to identify unexpected patterns and anomalies effectively. The main issue is the inherent incompleteness of the datasets, as data are collected in near-real time (NRT). Thus, the correlation between data points can change over time, requiring incremental updates to the model [5]. Furthermore, the one-pass learning approach limits the analysis of each data point to a single database state, which presents unique considerations in online time series analysis[6]. Moreover,

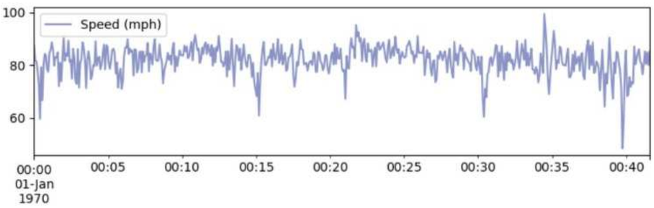

Fig. 1. Real-time traffic speed data obtained from speed_6005.csv of Numenta Anomaly Benchmark [4]. The data correspond to the Twin Cities metro area in Minnesota and was collected by the Minnesota Department of Transportation. The data has been rescaled by computing the mean in five-minute batches

the non-stationary nature of the data distribution and the occurrence of concept drift, where the underlying relationships shift over time, make the detection of anomalies in online time series more complicated [7].

The remainder of the paper is structured as follows. Section 2. Online time series anomaly detection challenges: presents the main challenges in the detection of anomalies for online time series. Section 3. Anomaly time series detection techniques: outlines different techniques and classifies them into statistical, Machine Learning, and hybrid methods. Then, it focusses on the Multiple Classification Systems as an efficient technique for anomaly detection that gets an advantage for those models. Section 4. DeepVats: Introduces in more detail the visualisation tool. Section 5. Application domains: Introduces different areas where anomaly detection is particularly relevant. Section 6. Research challenges and expectations.

II ONLINE TIME SERIES ANOMALY DETECTION CHALLENGES

An important portion of the data generated by digitalisation, the Internet of Things (IoT), and massive data generation is collected as time series. Thus, the data are stored as a sequence of data points ordered chronologically based on their acquisition time. Due to their nature, the data in the series are often correlated over time, making data modelling essential for understanding their behavior [1].

Finding unexpected patterns in data is crucial for many situations, such as the detection of credit card fraud, the monitoring of variations of physiological variables over time for observation of healthcare care, or the proper performance of applications [2]. The problem of identifying patterns in data that deviate from the expected behaviour is known as anomaly detection (also referred to as outliers).

There exist many occasions where processing information in real time is necessary, such as to analyse daily physical activity using activity monitors [8], [9]. These data are collected in what is known as streaming or on-line time series. The data of this series is collected and updated (near) in real-time (NRT). Unlike conventional time series that rely on historical data, online time series are used to monitor and analyse events as they occur. When data arrive in NRT, the order or arrival time of the data is crucial, as it provides valuable additional information to improve the results.

The analysis of NRT data faces three main challenges. First, the data set is always incomplete. The correlation between the data points can vary as the data arrive, rendering the previously established model obsolete. This requires incremental model training[5], where the model parameters must be adjusted to accommodate the patterns arriving at regular intervals. Secondly, due to continuous updates in the data, each collected data point can only be analysed once with each state of the database, a concept known as one-pass learning[6]. Lastly, and related to the previous points, there is the issue of nonstationary data distribution, which changes over time. This is known as concept drift, where the underlying concept that the model is trying to learn fluctuates over time, representing the relationship between the input data and the target variable. The model adaptation in each iteration, combined with the data updates, mitigates the negative impact of concept drift on the accuracy of the model's results.

These challenges make anomaly detection in online time series a greater challenge compared to traditional approaches, and it represents a relatively new and emerging research field that has gained increasing interest in recent years[7].

III ANOMALY TIME SERIES DETECTION TECHNIQUES

These challenges make anomaly detection in online time Several methods have been developed to detect anomalies in time series. From statistical approaches such as moving average and standard deviation-based techniques to more advanced algorithms such as autoregressive integrated moving average (ARIMA)[10], exponential smoothing [11], and machine learning-based approaches such as Support Vector Machines (SVM) [12], random forests[13], and DL architectures such as recurrent neural networks[14]. These techniques aim to capture temporal dependencies, identify abnormal patterns, and provide valuable insights for the detection of anomalies and subsequent decision-making processes. This section briefly introduces these techniques and their applications.

A Statistical techniques

These methods use statistical techniques to model the data and detect anomalies. Examples of such methods include ARIMA models (AutoRegressive Integrated Moving Average) [10] and exponential smoothing models [15].

The ARIMA (Autoregressive Integrated Moving Average) method is a time series analysis technique that combines autoregressive (AR), differencing (I), and moving average (MA) components to model patterns in data and predict future values. This technique has been used in the current year in different areas. For example, it has been used

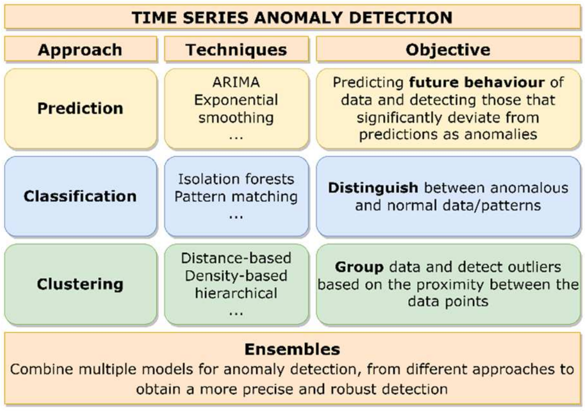

Fig. 2. Time series anomaly detection approaches and techniques

to analyse the consumption of fossil fuels in Slovakia [16]. The authors fit multiple ARIMA models with different parameter combinations to ensure that the selected ARIMA models explain all the autocorrelation in the series and are adequate for modelling the actual time series and predicting future values getting reliable predictions. Another example is the use of ARIMA to predict and control the impact of economic activities related to globalisation, industry, and urbanisation on human health and the environment. The objective is to better understand the negative effects of these activities and implement appropriate control measures.

The exponential smoothing model is a forecasting technique to predict future values in time-series data. It is based on the concept that more decent data carry more weight in the prediction than older data. Exponential smoothing is widely used for time series forecasting. It can be used to improve the estimation of seasonal coefficients in demand forecasting for homogeneous groups of products or series[11]. This technique has been used for different purposes, such as forecasting energy demand[17], prediction of hotel daily room demand[18] or the study of surface water temperature.

B Machine learning

Machine learning methods use algorithms to automatically learn and improve data without explicit programming. The goal is to complete tasks and make decisions intelligently and quickly. They can be subdivided into two main types.

On the one hand, supervised learning. These methods use a labelled data set to train a model that can detect anomalies in future data sets. Examples of such methods include decision trees, neural networks, and k-nearest neighbours. One possible application of this method is the classification of animal behaviour state based on environmental characteristics [19].

On the other hand, unsupervised learning. These methods do not require labelled data and are used to identify anomalous patterns in the data. Examples of such methods include Principal Component Analysis (PCA), clustering methods, and density-based clustering models. These models can be used, for example, for text classification [20] or High-Performance Computer unsupervised anomaly detection [21].

C Hybrid methods

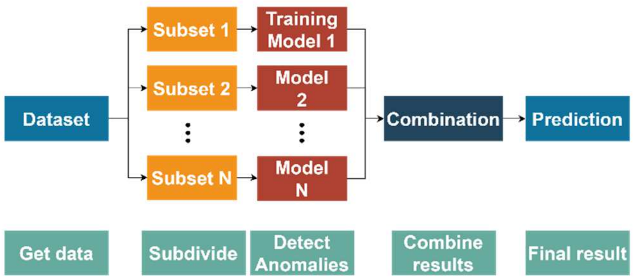

Fig. 3. Multiple Classification System workflow sketch

Hybrid methods integrate both statistical techniques and machine learning methods to detect anomalies in time series [22]. Thus, a combination of the strengths of both approaches is offered, allowing for more comprehensive and accurate time series modelling and greater adaptability to different situations and missing data. An application of hybrid methods is the improvement of efficiency and precision in the detection of electrical theft in smart grids[23].

D Advanced approaches in anomaly detection

There exist other techniques used in conjunction with the aforementioned algorithms and usually combining them (Fig. 2), such as distance-based, pattern matching or Multiple Classification Systems (MCS). Distance-based techniques are used to classify anomaly points according to the distance of new data from old data, clustering to embed data in a multidimensional space, and then using unsupervised machine learning to classify data into anomalous and non-anomalous points. Pattern matching detects anomalous subsequences in time series by modelling them in a supervised way using known characteristics for expected anomalous subsequences, and then new observations are compared with a database of labelled anomaly event flags to determine if they are anomalous or not. If there is a lack of labelled anomalies, common historical patterns are used instead of explicit characteristics. Finally, Multiple Classification Systems (MCS) [24]–[27]. MCS is one of the most effective techniques in other areas of machine learning, such as classification [28] or clustering [29].

As Fig. 3 shows, these systems start with an initial database, which is divided into several subsets to apply a different anomaly detection model to each subset. Subsequently, the results are combined to provide a prediction.

Recent studies have shown that MCS is among the most promising research directions for obtaining more robust and accurate anomaly detectors[30] It is a useful technique for enhancing the performance of detectors by reducing the model's dependency on the dataset and complementing the weaknesses of individual detectors while refining their strengths.

E Online time series anomaly detection

The detection of anomalies in online time series poses unique challenges due to their nature. The absence of complete datasets and the requirement for (near) real-time anomaly detection are specific challenges that need to be addressed in models designed for this type of series [27].

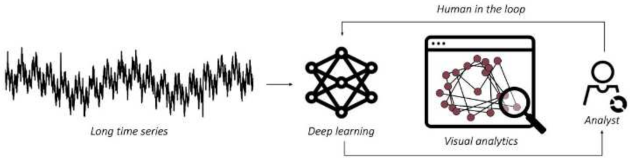

Fig. 4. General outline of DeepVats. Figure obtained from [31].

In the field of anomaly analysis in online time series, there is a recognised need to develop MCS [7] in two ways: taking into account the temporal correlation of the data and making heterogeneous systems. The following lines expose different open research lines that will be studied in future work.

1 Temporal correlation ensembles

Despite the inherent correlation in time series data, existing MCS techniques do not adequately consider it. Therefore, it is interesting to investigate whether the combination of time series methods should take into account the temporal correlation of the results when combining them to understand the atypical nature of the incoming data. Thus, we should be able to use prediction algorithms to estimate the value of the current point x_t and predict the next n points: $x_{t+1}, \dots, x_{t+n}$. Therefore, when we need to predict the value of x_{t_n}, we will have n predictions from previous steps, which can be combined using the average or weighted average to minimise the impact of previous anomalies.

2 Heterogeneous ensembles

It is interesting to investigate the combination of algorithms of different types to compare the accuracy of heterogeneous MCS (HMCS) with that of individual algorithms based on HMC models that have been previously developed in the area [7].

3 Visual exploration for anomaly detection

Furthermore, the importance of visual exploration of the algorithms developed should be noted. From this need arises the idea of developing interactive visual systems based on Deep Learning techniques. Specifically, DeepVATS, an open source tool, is introduced, which brings the visual tools of Deep Visual Analytics to time series data. DeepVATS trains an unsupervised time series masked auto-encoder that reconstructs segments of a time series and projects the acquired knowledge from the model onto an interactive chart, making it easy to observe patterns and anomalies in time series data [31].

The following section provides a more detailed introduction to DeepVATS as an interesting tool for visually detecting anomalies in time series.

IV DEEPVATS

DeepVATS is a tool for Visual Analytics of Time Series using Deep Learning. It consists of various modules, including a deep learning (DL) module to train neural network models, a data management module, and a visualisation (VA) module to explore analysis results. Figure

4 shows how an analyst can analyse long time series using visual analytics to improve deep learning models and analyse its results graphically. Additionally, DeepVATS includes several algorithms and data analysis techniques, such as the Masked Time Series AutoEncoder (MTSAE) for anomaly detection and dimensionality reduction algorithms such as UMAP, t-SNE, and PCA for projecting data into a lower-dimensional space. The tool also allows the integration of custom algorithms to suit specific user needs. In summary, DeepVATS is a comprehensive and flexible tool for visual analysis of time series data that combines deep learning techniques and data analysis to detect patterns and anomalies in the data. In addition, DeepVATS uses a recurrent neural network (RNN) to predict future values of a time series.

One of the key components is the Masked Time Series AutoEncoder (MTSAE). MTSAE is a variation of the Masked AutoEncoder (MAE) architecture that is specifically designed for time series data. Like MAE, MTSAE is a self-supervised scalable learner of large visual representations. MTSAE can be used for anomaly detection in time-series data by training it on datasets specifically designed for detecting repetitive patterns and anomalies. The knowledge contained in the model's latent space can be projected into an interactive graphical user interface to extract high-level information about the structure of the series, such as repetitive patterns or outliers. This model has been trained by randomly masking steps of the input series and making the model reconstruct them when training.

In summary, the main tools of DeepVATS include:

- Interactive user interface. DeepVATS provides an interactive user interface that allows users to load, visualise, and analyse time series.
- Time series visualisation, providing an interactive visualisation of time series that allows users to explore and analyse patterns in the data.
- Integration with TensorFlow, thus allowing users to take advantage of the deep learning capabilities of TensorFlow for time series analysis. TensorFlow is an open source software library for data flow and differentiable programming on a range of tasks. It is a popular platform for building and training machine learning models, particularly deep neural networks.

DeepVATS is especially useful for combining the ensemble algorithms for anomaly detections with de visual tools, allowing for the visual analysis of online time series outliers.

V APPLICATION DOMAINS

Multiple classification systems for online time series anomaly detection have proven to be highly effective across a range of domains, including Social Network analysis, industry, space, and misinformation. In social network analysis, these systems enable the identification of anomalous behaviour patterns, such as bot-generated content or suspicious network connections, thereby ensuring a safer and more trustworthy online environment. In industrial settings, the ability to detect anomalies in time series data allows early identification of equipment failures, leading to improved maintenance and reduced downtime. Similarly, in the domain of space exploration, these classification systems play a crucial role in monitoring spacecraft telemetry data and promptly detecting anomalies that could indicate potential system malfunctions or deviations from expected trajectories.

Furthermore, in the battle against misinformation, these systems aid in identifying anomalous content propagation patterns or sudden spikes in user engagement, helping to mitigate the spread of false information and protect the integrity of online discourse. Overall, the application of multiple classification systems for online time series anomaly detection has proven invaluable in diverse domains, enhancing security, efficiency, and trustworthiness.

A Industry 4.0

Anomaly detection plays a crucial role in Industry 4.0 by identifying rare events so accurate detections are taken. For instance, in solar radio flux forecasting, anomaly detection is crucial, as rare events can significantly impact the model's precision [32]. Another example is the analysis of the resources needed for firefighting, where the analysis over time according to the fight status evolution could be a big improvement in optimising the number of assets needed for correct performance. Analysing anomalies would become vital for the correct use of resources as a high degree of precision is needed to ensure the correct extinguishment of the fire [33].

1 Internet Of Things (IoT)

Time series are also used to store the evolution of variables throughout the execution of business processes such as coal mining [34]. Anomaly analysis in those series is crucial for the conformance between the business process model and the data provided by its execution in the cases where the data are given as a set of time series containing the evolution of the variables involved in the process rather than as an event log. The use of time series allows the modelling of the business industry for problems and risks forecasting. They can also be used for historical data analysis to detect rare (potentially riskful) patterns

2 Aerospace engineering

Time series have also been used for the analysis of the behaviour of Unnamed Aerial Vehicles (UAV) operators over time using a profile-based model where the evolution of the operator's performance during a mission is the main measure unit [35]. This information can be useful for the improvement of the quality of simulation-based training systems, as it can help exploit general patterns of behaviour between simulations, detect out-of-normal performance by anomaly detection, and study whether the behaviour of a specific operator changes when in dangerous situations.

.

B Social Network Analysis

Understanding the dynamics of social networks and their impact on various aspects of society is of great importance. This section shows two applications of Anomaly Detection in the area of social network analysis: behaviour analysis and the detection of misinformation and fake news.

1 Behavioural Analysis

The behavioural analysis in Social Networks may provide great insights for commerce strategies [36]. For instance, we could assume a special interest in a product by detecting an anomaly in the data that shows much more hashtags in a concrete time than in the rest. In addition, the product owner could change strategies if society does not talk about them as much as expected. Furthermore, game strategies can be improved by analysing the behaviour of professional games [37]. Applying anomaly detection to those strategies can help to detect possible unwanted actions that could refer to a bad player performance, may be due to fatigue. We could then analyse both the best strategy and the most appropriate one for each player based on his/her accuracy in the game.

2 Missinformation and fake news

Time series have also been used to detect fake news on Twitter using time series and information from users involved in spreading the news. Anomaly analysis could potentially be used to detect unusual patterns in news distribution, which could show more probabilities of its fakeness. For example, if a piece of news spreads much faster or more widely than usual, this could be considered an anomaly and warrant further investigation to determine whether the new is true or false [38].

VI RESEARCH CHALLENGES AND EXPECTATIONS

Detecting anomalies in real time is crucial to timely decision-making and intervention. However, performing anomaly detection on large-scale online time series data in real-time poses significant computational and scalability challenges. Future research should focus on developing efficient algorithms and frameworks that can handle the high volume and velocity of data while maintaining real-time detection accuracy. This will enable the deployment of effective real-time anomaly detection systems in various domains. The unique characteristics of online time series data make anomaly detection in this context a complex task. As a result, there are different research challenges that need to be addressed. This section highlights the anticipated outcomes and potential advances that can be achieved by tackling these challenges.

1 Multiple Classification Systems designed considering data temporal correlation

One of the primary challenges in time series anomaly detection is effectively handling the temporal correlation of data points. The presence of temporal data dependencies and patterns in online time series data requires the development of robust techniques that can accurately capture and model these correlations. Future research should explore advanced algorithms and methodologies that can effectively handle the complex temporal patterns and dependencies present in classical time series data to apply them to online time series.

2 Heterogeneous Multiple Classification Systems

Heterogeneous ensembles, where individual classifiers may have different characteristics and performances, may deal to most robust models with better anomaly detection efficience. These techniques can reduce the dependence of the model on the dataset, complement the debilities of each individual detector, and improve their force points. Thus, the next step is to combine individual models into heterogeneous ones to improve the models.

3 Visualization tools

Furthermore, in the battle against misinformation, these Finally, it is essential to understand and analyse the results obtained in real time. DeepVATS stands out as an excellent tool for anomaly visualisation, as it effectively combines ensemble algorithms with visual tools. To further enhance anomaly detection analysis, we shall incorporate the algorithms developed into the library. This integration will significantly improve anomaly analysis, improving DeepVATS as a resource for users seeking comprehensive and accurate anomaly detection capabilities.

In conclusion, addressing these research challenges will contribute to the advancement of time series anomaly detection techniques. By developing robust techniques to handle temporal correlation, address heterogeneity in ensembles, and develop visualisation tools, we can expect significant improvements in the accuracy and effectiveness of anomaly detection systems for online time series data.

ACKNOWLEDGMENT

This work has been funded by PCI2022-134990-2 (MARTINI) of the CHISTERA IV Cofund 2021 program, funded by MCIN/AEI/10.13039/501100011033 and by the “European Union NextGenerationEU/PRTR; by Grant PLEC2021-007681 (XAI-DisInfodemics) and PID2020-117263GB-100 (FightDIS) funded by MCIN/AEI/ 10.13039/501100011033 and, as appropriate, by “ERDF A way of making Europe”, by the “European Union” or by the “European Union NextGenerationEU/PRTR”, by the research project CIVIC: Intelligent characterisation of the veracity of the information related to COVID-19, granted by BBVA FOUNDATION GRANTS FOR SCIENTIFIC RESEARCH TEAMS SARS-CoV-2 and COVID-19, by European Comission under IBERIFIER - Iberian Digital Media Research and Fact-Checking Hub (2020-EU-IA-0252), and by "Convenio Plurianual with the Universidad Politécnica de Madrid in the actuation line of Programa de Excelencia para el Profesorado Universitario".

REFERENCES

[1] S. Shahriari, S. A. Sisson, y T. Rashidi, «Copula ARMA-GARCH modelling of spatially and temporally correlated time series data for transportation planning use», *Transp. Res. Part C Emerg. Technol.*, vol. 146, p. 103969, ene. 2023, doi: 10.1016/j.trc.2022.103969.

[2] J. Soldani y A. Brogi, «Anomaly Detection and Failure Root Cause Analysis in (Micro) Service-Based Cloud Applications: A Survey», *ACM Comput Surv*, vol. 55, n.º 3, feb. 2022, doi: 10.1145/3501297.

[3] X. Zhang *et al.*, «LSHiForest: A Generic Framework for Fast Tree Isolation Based Ensemble Anomaly Analysis», en *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 2017, pp. 983-994. doi: 10.1109/ICDE.2017.145.

[4] A. Lavin y S. Ahmad, «Evaluating Real-Time Anomaly Detection Algorithms – The Numenta Anomaly Benchmark», en *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015, pp. 38-44. doi: 10.1109/ICMLA.2015.141.

[5] Y. Zhang, Y. Liu, G. Yang, y J. Song, «SSIT: a sample selection-based incremental model training method for image

recognition», *Neural Comput. Appl.*, vol. 34, n.º 4, pp. 3117-3134, feb. 2022, doi: 10.1007/s00521-021-06515-4.

[6] D. Li y Z. Zeng, «CRNet: A Fast Continual Learning Framework With Random Theory», *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1-14, 2023, doi: 10.1109/TPAMI.2023.3262853.

[7] A. Iturria Aguinaga, «Reduction of False Positives in Online Outlier Detection over Time Series using Ensemble Learning», PhD Thesis, Ciencias de la Computación e Inteligencia Artificial (UGR), Francisco Herrera Triguero, Javier Del Ser Lorente, 2023.

[8] S. Gurz, N. G. Temel, A. T. Sengul, Y. Buyukkarabacak, M. G. Pirzirenli, y A. Basoglu, «Learning Curve for Uniportal VATS Anatomical Pulmonary Resections: the Activity Monitor Operating Characteristic Method», *Indian J. Surg.*, ene. 2023, doi: 10.1007/s12262-023-03667-6.

[9] S.-M. Lee, D.-Y. Kim, y J. Woo, «Glucose Transformer: Forecasting Glucose Level and Events of Hyperglycemia and Hypoglycemia», *IEEE J. Biomed. Health Inform.*, vol. 27, n.º 3, pp. 1600-1611, 2023, doi: 10.1109/JBHI.2023.3236822.

[10] M. H. Mgammal, E. M. Al-Matari, y T. F. Alruwaili, «Value-added-tax rate increases: A comparative study using difference-in-difference with an ARIMA modeling approach», *Humanit. Soc. Sci. Commun.*, vol. 10, p. 121, 2023, doi: 10.1057/s41599-023-01608-y.

[11] I. Svetunkov, H. Chen, y J. E. Boylan, «A new taxonomy for vector exponential smoothing and its application to seasonal time series», *Eur. J. Oper. Res.*, vol. 304, n.º 3, pp. 964-980, 2023.

[12] Z. Shang, Z. Zhao, R. Yan, y X. Chen, «Core loss: Mining core samples efficiently for robust machine anomaly detection against data pollution», *Mech. Syst. Signal Process.*, vol. 189, p. 110046, 2023.

[13] F. A. Morales, J. M. Ramírez, y E. A. Ramos, «A mathematical assessment of the isolation random forest method for anomaly detection in big data», *Math. Methods Appl. Sci.*, vol. 46, n.º 1, pp. 1156-1177, 2023.

[14] J. Schmitt, I. Horstkötter, y B. Bäker, «Electrical lithium-ion battery models based on recurrent neural networks: A holistic approach», *J. Energy Storage*, vol. 58, p. 106461, 2023.

[15] Z. Chen, D. Zhou, E. Zio, T. Xia, y E. Pan, «Adaptive transfer learning for multimode process monitoring and unsupervised anomaly detection in steam turbines», *Reliab. Eng. Syst. Saf.*, vol. 234, p. 109162, 2023, doi: 10.1016/j.ress.2023.109162.

[16] I. Sardar, M. A. Akbar, V. Leiva, A. Alsanad, y P. Mishra, «Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: methodology, evaluation, and case study in SAARC countries», *Stoch. Environ. Res. Risk Assess.*, vol. 37, n.º 1, pp. 345-359, 2023.

[17] C. Rao, Y. Zhang, J. Wen, X. Xiao, y M. Goh, «Energy demand forecasting in China: A support vector regression-compositional data second exponential smoothing model», *Energy*, vol. 263, p. 125955, 2023.

[18] E. J. Jamshidi, Y. Yusup, J. S. Kayode, y M. A. Kamaruddin, «Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: A case study on surface water temperature», *Ecol. Inform.*, vol. 69, p. 101672, 2022, doi: https://doi.org/10.1016/j.ecoinf.2022.101672.

[19] S. Bergen *et al.*, «A review of supervised learning methods for classifying animal behavioural states from environmental features», *Methods Ecol. Evol.*, vol. 14, n.º 1, pp. 189-202, 2023, doi: 10.1111/2041-210X.14019.

[20] L. S. da Costa, I. L. Oliveira, y R. Fileto, «Text Classification Using Embeddings: A Survey», *Knowl. Inf. Syst.*, vol. 65, n.º 7, pp. 2761-2803, mar. 2023, doi: 10.1007/s10115-023-01856-z.

[21] M. Molan, A. Borghesi, D. Cesarini, L. Benini, y A. Bartolini, «RUAD: Unsupervised Anomaly Detection in HPC Systems», *Future Gener. Comput. Syst.*, vol. 141, n.º C, pp. 542-554, abr. 2023, doi: 10.1016/j.future.2022.12.001.

[22] H. Mubarak *et al.*, «A hybrid machine learning method with explicit time encoding for improved Malaysian photovoltaic power prediction», *J. Clean. Prod.*, vol. 382, p. 134979, 2023, doi: https://doi.org/10.1016/j.jclepro.2022.134979.

[23] A. Ullah, N. Javaid, M. Asif, M. U. Javed, y A. S. Yahaya, «Alexnet, adaboost and artificial bee colony based hybrid model for electricity theft detection in smart grids», *IEEE Access*, vol. 10, pp. 18681-18694, 2022.

[24] J. Del Ser, M. N. Bilbao, I. Laña, K. Muhammad, y D. Camacho, «Efficient Fake News Detection using Bagging Ensembles of Bidirectional Echo State Networks», en *IEEE 2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1-7. doi: 10.1109/IJCNN55064.2022.9892331.

[25] H. Liz, J. Huertas-Tato, M. Sánchez-Montañés, J. Del Ser, y D. Camacho, «Deep Learning for Understanding Multilabel Imbalanced Chest X-Ray Datasets», *Future Gener. Comput. Syst.*, vol. 144, n.º C, pp. 291-306, abr. 2023, doi: 10.1016/j.future.2023.03.005.

[26] R. Xin, H. Liu, P. Chen, y et al, «Robust and accurate performance anomaly detection and prediction for cloud applications: a novel ensemble learning-based framework», *J. Cloud Comput.*, vol. 12, n.º 7, p. 7, 2023, doi: 10.1186/s13677-022-00383-6.

[27] A. A. Cook, G. Mısırlı, y Z. Fan, «Anomaly detection for IoT time-series data: A survey», *IEEE Internet Things J.*, vol. 7, n.º 7, pp. 6481-6494, 2019, doi: 10.1109/JIOT.2019.2958185.

[28] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, 1st ed. Chapman & Hall, 2012.

[29] T. Alqurashi y W. Wang, «Clustering ensemble method», *Int. J. Mach. Learn. Cybern.*, vol. 10, pp. 1227-1246, 2019, doi: 10.1007/s13042-017-0756-7.

[30] H. Wang, M. J. Bah, y M. Hammad, «Progress in Outlier Detection Techniques: A Survey», *IEEE Access*, vol. 7, pp. 107964-108000, 2019, doi: 10.1109/ACCESS.2019.2932769.

[31] V. Rodriguez-Fernandez, D. Montalvo, F. Piccialli, G. J. Nalepa, y D. Camacho, «DeepVATS: Deep Visual Analytics for Time Series». 2023.

[32] E. Stevenson, V. Rodriguez-Fernandez, E. Minisci, y D. Camacho, «A deep learning approach to solar radio flux forecasting», *Acta Astronaut.*, vol. 193, pp. 595-606, 2022, doi: https://doi.org/10.1016/j.actaastro.2021.08.004.

[33] H. Liz-López, J. Huertas-Tato, J. Pérez-Aracil, C. Casanova-Mateo, J. Sanz-Justo, y D. Camacho, «Spain on Fire: A novel wildfire risk assessment model based on image satellite processing and atmospheric information», *ArXiv Prepr. ArXiv230605045*, 2023.

[34] V. Rodriguez-Fernandez, A. Trzcionkowska, A. Gonzalez-Pardo, E. Brzychczy, G. J. Nalepa, y D. Camacho, «Conformance checking for time-series-aware processes», *IEEE Trans. Ind. Inform.*, vol. 17, n.º 2, pp. 871-881, 2020.

[35] V. Rodríguez-Fernández, H. D. Menéndez, y D. Camacho, «Analysing temporal performance profiles of UAV operators using time series clustering», *Expert Syst. Appl.*, vol. 70, pp. 103-118, 2017.

[36] G. Bello-Orgaz, R. M. Mesas, C. Zarco, V. Rodriguez, O. Cordón, y D. Camacho, «Marketing analysis of wineries using social collective behavior from users' temporal activity on Twitter», *Inf. Process. Manag.*, vol. 57, n.º 5, p. 102220, 2020.

[37] A. M. Rama, V. Rodriguez-Fernandez, y D. Camacho, «Finding behavioural patterns among league of legends players through hidden markov models», en *Applications of Evolutionary Computation: 23rd European Conference, EvoApplications 2020, Held as Part of EvoStar 2020, Seville, Spain, April 15–17, 2020, Proceedings 23*, Springer, 2020, pp. 419-430.

[38] M. Previti, V. Rodriguez-Fernandez, D. Camacho, V. Carchiolo, y M. Malgeri, «Fake news detection using time series and user features classification», en *Applications of Evolutionary Computation: 23rd European Conference, EvoApplications 2020, Held as Part of EvoStar 2020, Seville, Spain, April 15–17, 2020, Proceedings 23*, Springer, 2020, pp. 339-353.