ML-based Plant Stress Detection from IoT-sensed Reduced Electromes

Marcos A. de Oliveira Jr.[*†], **Gregory Sedrez**[†], **Gerson Geraldo H. Cavalheiro**[†]
[*]Federal Institute of Education, Science and Technology Farroupilha, Santa Maria/RS, Brazil
[†]Technological Development Center, Federal University of Pelotas, Pelotas/RS, Brazil
{marcos.oliveira, gdbsedrez, gerson.cavalheiro}@inf.ufpel.edu.br

Abstract

The recognition of patterns in the electrical activities of plants (electromes, in time series format) has gained prominence in recent years. The use of Internet of Things (IoT) devices and Machine Learning (ML) techniques has automated and enhanced data collection and classification, helping researchers identify behaviors and classify them to detect plant stress. However, processing this information means dealing with large amounts of data, which is a major challenge from a computer science perspective. Thus, in this work, we propose an approach for reduction and classification of time series representing plant electromes to balance the trade-off between reduction and data quality, without compromising the classification task. We investigated the use of three time series approximation techniques (PAA, SAX, and MCB) in combination with ML algorithms, such as ANN, KNN, and SVM, in order to find the most suitable approach for this scope. The results validated the proposed approach, with the best performance obtained with the PAA+SAX techniques combined with the SVM algorithm, achieving good data reduction and improving stress detection, without compromising data quality. The main challenges in these tasks and future research directions are also discussed.

1 Introduction

Electrome stands for the electrical dimension of biological systems. It represents all ionic currents of any living entity, from the cellular level to the organismal level (Loof 2016). Through electrome analysis, especially variations in the electrical activity of plants, it is possible to identify behavioral changes in these organisms, such as growth stages, incidence of stress, and changes in climatic conditions. Obtaining and analyzing this information can help researchers and producers better understand plant growth conditions, the impacts of environmental variations on the crop, and reactions to treatments with agricultural input, providing a new dimension of biological understanding. Given the complexity of this context, the use of technology is very fruitful to automate and enhance the reliability of plant electrome analysis. With the spread of devices and systems on the Internet of Things (IoT) concept in recent years (Navarro, Costa, and Pereira 2020), agricultural sensors are presented as one of the most efficient and cost-effective ways to collect electrical elements from plants. In the same way, from sensor data, Machine Learning (ML) algorithms are widely used to identify patterns and perform predictive analysis for different tasks(Wallelign, Polceanu, and Buche 2018), such as the detection of stress in plants based on electrical activity.

Regarding the data format, one of the most used for recording electromes is Time Series (Jensen, Pedersen, and Thomsen 2017). This format is interesting because it aggregates electrical activity with temporal information. However, given the high frequency of data collection, sensing often results in very large time series. (Rosenberger, Kübel, and Rothfuß 2022). This poses a problem for applications in the IoT context, since IoT devices commonly have limited resources such as storage, processing, and connectivity. Techniques for reducing, compressing, approximating, or filtering time series are constant concerns in data compression research (Hu et al. 2021).

On the other hand, in terms of classification with ML algorithms, it is desirable to have good data quality. Quality does not necessarily mean quantity, but, in general, the larger the sample of data collected, the greater the coverage of cases for training an ML model. Arbitrarily reduced data can compromise the classification task and make the system unfeasible (Hamilton 2020). The challenge here is to balance the trade-off between data compression and data quality (Cruces, Seco, and Guitérrez 2019). Thus, it is desirable to investigate alternatives to enable the reduction of this volume of data without compromising quality, that is, those that enable good time series classification performance.

In this work, an approach to time series reduction and classification is proposed, specifically for plant electromes, through the use of time series approximation techniques combined with ML classification algorithms. The performance analysis performed here is based on metrics well known to the scientific community to demonstrate the feasibility of this approach, as well as its performance in relation to other existing techniques. This work extends previous efforts on this research topic (de Oliveira Jr. et al. 2021; 2022) and makes the following specific contributions:

1. Proposes a time series reduction and classification approach for plant electromes;

2. Analyze the electrome reduction, from the IoT perspective, and electrome stress detection performance, in terms of ML metrics;
3. Discuss the main challenges in electrome reduction and classification, as well as highlighting future research directions.

The results obtained in this work validated the proposed approach for reducing and classifying electromes in time series format. It was possible to achieve a satisfactory data reduction, as demonstrated by the compression ratio and space saving metrics, and a good classification performance, improving the accuracy of stress detection with compressed data, that is, the reduction did not compromise its quality.

This article is structured as follows. Section 2 presents Related Works found in the literature; Section 3 presents the Methodology used in this work and the proposed approach; Section 4 presents the Experiments and Results; Section 5 presents a Discussion and Future Research Directions; and, Section 6 presents the Conclusion.

2 Related Works

Some published work has already investigated the reduction and classification of time series that represent plant electromes. From the works found in the literature, the ones that most resemble and inspire the approach proposed here are described below.

In (Souza et al. 2017) and (Pereira et al. 2018) alternatives are investigated for time series classification using data collected from soybean plants, before and after environmental stimuli, such as cold, low light, and osmotic stress. Both use the same data set and seek to recognize the behavior patterns of plants in relation to variations in the environment. In this approach, the authors use a data reduction strategy called Interval Arithmetic (IA), which reduces a time series interval by an average, maximum, and minimum value of that interval. For classification, algorithms such as Artificial Neural Networks, Convolutional Neural Network, Optimum-Path Forest, K-Nearest Neighbors, and Support Vector Machine were explored. Experiments indicated that the models achieved better results with larger test sets (approximately 80% of the data sets) and better classification performance (accuracy greater than 80%) for cold and osmotic stresses.

In (Reissig et al. 2021), an approach for the classification of electrical activities of tomatoes was presented, to identify different stages of maturation: mature green, breaker, and light red. Fast Fourier Transform, Wavelet Transform, Power Spectral Density, and Approximate Entropy techniques were used to analyze the time series. Decision Tree, Support Vector Classification, Gaussian Process, K-NN, Bayes, Random Forest, and a Dummy classifier models were used for classification. The results showed that it is possible to classify the stages of maturation using the electrical activity of the fruit. It was also observed, using precision, sensitivity, and F1 score techniques, that the breaker stage was the most classifiable in all stages (accuracy around 80%).

In (Toledo 2019), an approach for collecting and classifying bean electromes is proposed. The authors explore different stimuli applied to bean plants and develop ML models to recognize plant behavior patterns before and after a stimulus. In this work, IA was also used as a data processing strategy. ML Bayes, SVM (Support Vector Machine), K-NN (K-Nearest Neighbors), ANN (Artificial Neural Network), DT (Decision Trees), and RF (Random Forest) techniques were used. The results showed that the larger the size of the test set, in relation to the entire data set, the better the accuracy achieved, but this accuracy varies for each stimulus, with results in the range of 40 to 60%. The important contribution here is the feasibility of this classification, at first, and then the improvement of this accuracy index.

From these previous works, it is possible to observe that the problem of pattern recognition in electrical activities of plants (electromes) is relatively new, as is the application of ML techniques in this context. For this reason, the performance of existing classifiers does not yet reach perfection levels, but even so, the accuracy around 60% is already acceptable and profitable. Although there are some initial studies, there is still room for research, mainly from the perspective of data reduction combined with classification, since improving this process will enhance the implementation of monitoring systems in IoT contexts. Thus, this research aims to advance the state of the art in the reduction and classification of electromes in time series format to assist in future applications in biology, agriculture, and other related areas.

3 Methodology

In this section, the methodology used in the work is detailed. Initially, an application of stress detection in plants is presented using data collected from sensors as input to train ML models. It is also explained how the performance of ML algorithms is measured, which metrics are used, and how they are interpreted. So, an approach for electrome reduction and classification in time series format is proposed. For this study, the same data set presented in (Toledo 2019) was used and the same methodology was followed to detect stress in plants, as detailed below.

Plant Stress Detection from Electromes

Plants are subject to multiple environmental variations, some of which generate stress, such as lack of water and soil salinity. These stressful situations lead to financial losses, and the manifestation of stress can be observed by reactions involving changes in plant electrical activity. The approach proposed here follows the methodology to detect plant stress presented in (Toledo 2019), which uses bean plants for a laboratory experiment. In the mentioned work, experiments were carried out to monitor seedlings subjected to induction of saline stress. Each experiment monitored the electrical activity of bean plants for 4 consecutive hours, two hours before application of the solution and two hours after. The time series produced (electromes) for each two-hour range has around 450,000 readings. For classification purposes, values 0 and 1 were assigned for the measured time series, with 0 for no stress and 1 for stress. In addition to the original time series, complementary features were also extracted, such as Fast Fourier Transform, Approximate Entropy, Variance, Skewness, and Kurtosis. Then, ML models

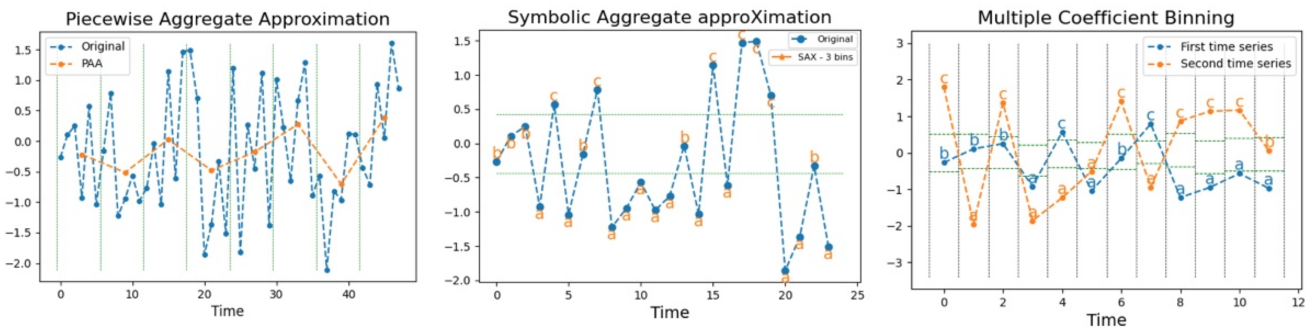

Figure 1: Operation of time series approximation algorithms: *PAA*, *SAX* and *MCB*.

were trained to recognize time series patterns with and without stress.

Electrome Approximation

Initially, in (Toledo 2019), to reduce time series, a strategy called Interval Arithmetic (IA) was used. This method replaces each sequence of 30,000 points, equivalent to 8 minutes of monitoring, with three values: the average, maximum, and minimum of the interval. The interval size of 30 thousand points was adopted, according to the authors, because it presents a better result in terms of accuracy. To explore other possibilities to reduce the volume of data and analyze the impact of these alternatives on model training, three approximation algorithms were chosen and implemented, as detailed below. Fig. 1 illustrates the operation of the algorithms graphically.

Piecewise Aggregate Approximation (PAA): Create an alternative representation of a time series, segmenting it into intervals and replacing them with their average value (Keogh et al. 2002). The objective of the technique is to decrease the number of points and reduce the noise of a time series, preserving the trend.

Symbolic Aggregate approXimation (SAX): Categorize the data, reduce noise, and capture the time series trend (Lin et al. 2007). This method transforms each time series independently into a sequence of symbols. The parameters of this technique are the number of *bins* (horizontal cuts), the strategy to determine the width of the bins and the alphabet of characters.

Multiple Coefficient Binning (MCB): Also, categorize the data, transform them into strings, reduce noise, and capture the time series trend (Schäfer and Högqvist 2012). However, this method stores each time period independently, setting the bins according to the interval, unlike *SAX*, which categorizes the intervals according to the distribution of the entire series.

The implementation of the time series approximation algorithms, detailed above, was done in the Python language, in the Google Colab tool environment. The PYTS library (Faouzi and Janati 2020) was used. The techniques were placed sequentially, with algorithm *PAA* being the first to be executed, which would later be combined with *SAX* and *MCB*. The implementation of the algorithms is available on the GitHub repository[1].

Electrome Classification

Electrome classification was performed using trained ML models, with the aim of identifying stress patterns (0 without stress and 1 with stress) in the time series of the input data set. For model training and performance verification, the data sets were always separated into train (80%) and test (20%) sets. Model training was also implemented in the same programming environment as the approximation techniques, using the open source library *Scikit-learn* (Pedregosa et al. 2011) to have access to tools for predictive data analysis. The ML algorithms were implemented as follows.

K-Nearest Neighbors (KNN): Uses the Euclidean distance between points to classify the data according to their k nearest neighbors (Guo et al. 2004). For the *KNN* algorithm, the default parameters were used, with $n = 3$.

Support Vector Machine (SVM): Searches, through the distances between points in the same set, for a linear classifier that better classifies these data (Evgeniou and Pontil 2001). The *SVM* algorithm was also implemented with the default parameters.

Artificial Neural Network (ANN): A mathematical model inspired by the neural structure (Jain, Mao, and Mohiuddin 1996). The neural network was implemented with two layers, 3 and 5 nodes, with the activation function *relu*.

Performance Evaluation Metrics

For the performance analysis in relation to the volume of data, metrics *Compression Ratio (CR)* and *Space Saving (SS)* (Lelewer and Hirschberg 1987) were observed. Both were chosen because they address the issue of data size. The *CR* gives the reduction ratio achieved, dividing the original file size by the reduced file size, and, in addition, the *SS* indicates the amount of space reduced by the reduction, relatively, in percentage. Regarding the evaluation of the ML models, metrics *Accuracy* and *Confusion Matrix* were used,

[1]GitHub - https://github.com/gdbsedrez/TSRCElectromes

Table 1: Performance of approximation methods in terms of Size Reduction, Compression Ratio and Space Saving.

Approximation	Size Red.	Comp. Ratio	Sp. Saving
IA (Toledo 2019)	315 KB	-	-
PAA	277 KB	1.1372	12.06%
PAA+SAX	259 KB	1.2162	17.77%
PAA+MCB	259 KB	1.2162	17.77%

showing the false and true positive and negative cases of the classification.

4 Experiments and Performance Analysis

To evaluate the proposed approach, experiments were carried out by applying it to the available electrome data set, comparing the original solution with the IA approach. For performance comparison purposes, the same data set collected in (Toledo 2019) was used and the IA technique was implemented. The original data collected are not public because of a confidentiality agreement with a partner company, but the processed data used in this work are also available on the GitHub repository.

First, the reduction of electromes was implemented and evaluated. Table 1 presents the results found for CR and SS. The data set was collected from 37 measurements taken 2 hours before and 2 hours after a stress stimulus, stored in 74 files in text format, containing a floating point time series by file. Each file, with these settings, is about 6 MB, with the total size of the data set available for the experiment being 442 MB. The IA technique is able to reduce this volume of data to a total of 315 KB. The values presented in the *Reduced Size* column consider the storage consumption of the raw time series plus the other features extracted from them, such as approximate entropy and transforms, among others, as detailed in (Toledo 2019).

Regarding the reduction of data volume, it was observed that the implemented techniques achieve better performance, compared to the IA technique. The PAA technique allows for a good initial reduction in the size of the input data set, around 12%, and when applied in combination with SAX or MCB, due to categorization of the data, achieves even higher space savings, over 17%. The improvements, compared to IA, imply a significant reduction in the space in absolute values. In terms of storage, for instance, the level of reduction provided by our approach means an average of one hour more of measurement in relation to the initial 4 hours. That is, if the reduction were performed at the edge (embedded in the sensor), it would allow the sensor to monitor the electrical activity of the plants for an extra hour, generating the same volume of data, or to monitor the same four hours, reducing the need for storage. This high storage gain makes the use of these techniques highly viable in IoT applications for plant stress detection, in agricultural scenarios with limited online technological resources and/or limited hardware. In addition, a lower storage cost can also impact lower energy costs, increasing the battery cycle of IoT devices, as well as a lower communication cost, sending smaller amounts of data, for example.

Table 2: Performance of classification algorithms in terms of Accuracy.

Approximation	Accuracy		
	KNN	SVM	ANN
IA (Toledo 2019)	53.36%	68.26%	50.96%
PAA	57.21%	68.75%	49.51%
PAA+SAX	**62.01%**	**69.23%**	51.92%
PAA+MCB	60.09%	68.72%	**54.32%**

Regarding electrome classification, Table 2 presents classification performance, from the data set reduced by IA and by the other approximation algorithms, where each column represents the accuracy achieved by each ML algorithm. The best accuracies for each ML algorithm are highlighted. The first concern, from a time series classification point of view, is to ensure that data reduction does not compromise classification. This can be verified and ensured by the results indicated in the table, where all approximation techniques maintained or increased performance on average. Compared to classification after data reduction with IA, the PAA technique increases the performance of the KNN and SVM algorithms and has a small decrease for ANN, while the combinations of PAA+SAX and PAA+MCB increased the accuracy in all ML algorithms. The greatest emphasis should be given to the PAA+SAX technique in combination with the SVM algorithm, which reached the highest level of accuracy and therefore is the most indicated, according to this study, for the reduction and classification of electromes.

To complement the analysis, through the algorithms' confusion matrices, it is possible to identify those that are most relevant in the detection of stress in electromes. The confusion matrix indicates the true and false positives and negatives, that is, between the hits and misses, which classes were the most difficult to classify. From a plant monitoring point of view, it is more useful for an algorithm to find all existing stresses, so false positive cases are less relevant than false negatives. Failure to detect a stress can lead to the death of the plant, whereas the detection of non-existent stress can be dealt with in other ways. Therefore, the lower the index in the lower left quadrant, the more appropriate the algorithm for this task. Figure 2 presents the confusion matrices after training each ML algorithm with the data sets reduced by the time series reduction techniques discussed here.

It is possible to observe that the ANN algorithm is the one that presents the greatest difficulty in dealing with false negatives, while KNN and SVM have practically the same performance in the classification of all cases, and the approximation techniques explored here improve slightly, in a balanced way between false and true negatives and positives, especially the PAA+SAX technique. Therefore, considering both analysis of reduction and classification of electomes, the greater suitability of the use of the PAA+SAX approximation techniques, in a combined way, to reduce the volume of data in electromes and the use of these data as input for the SVM classification algorithm, to achieve better performance in stress detection in plants.

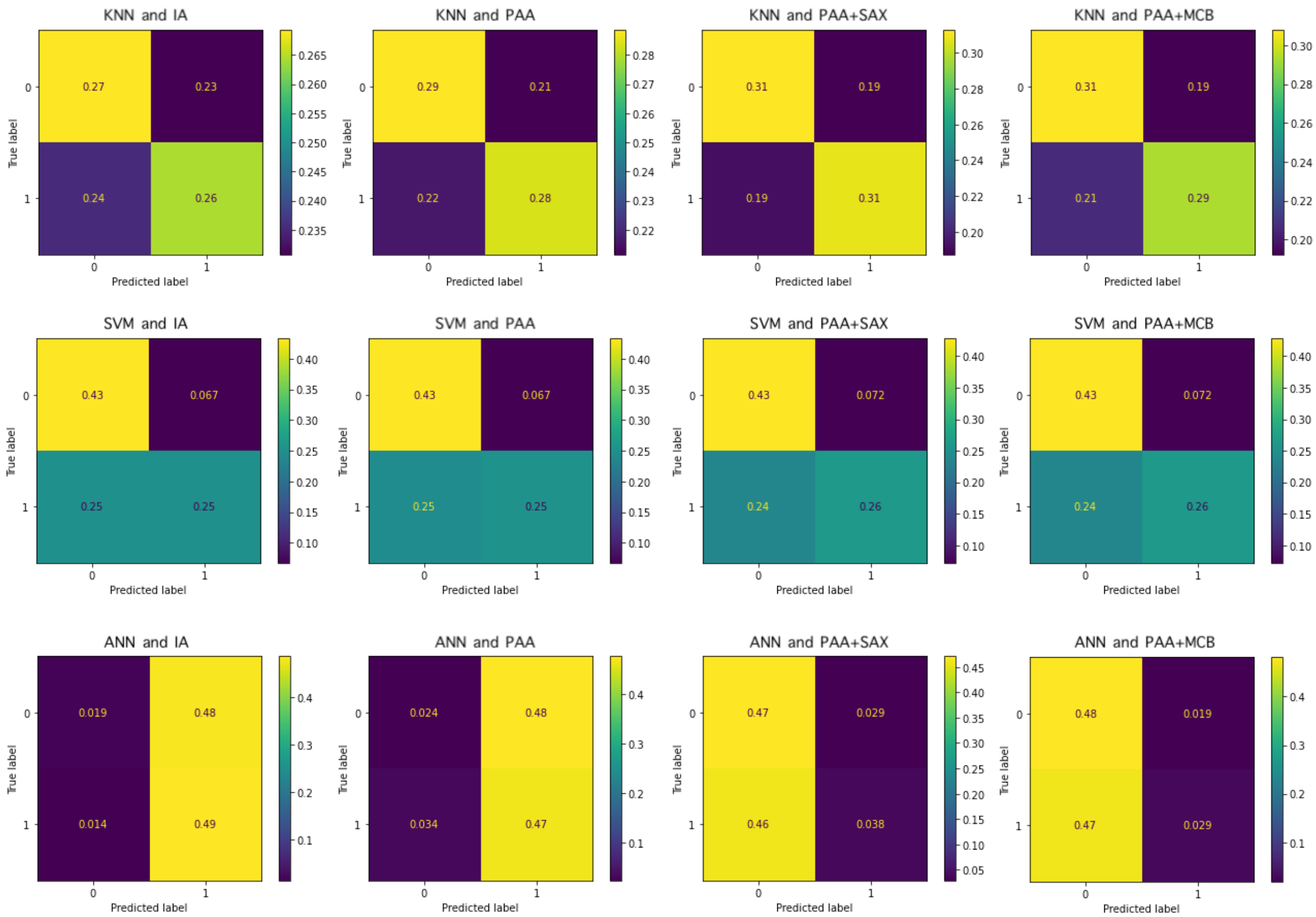

Figure 2: Confusion matrix of the ML algorithms combined with the approximations techniques.

5 Discussion

Reducing the data volume can be done in several ways, such as by decreasing the sampling rate. However, in the scope of electromes, from the perspective of plant electrophysiology, electrical variations can occur at very small intervals (milliseconds), so a high sampling rate is important. Therefore, a reduction strategy that does not affect data quality is needed. Thus, the main scientific contribution of this work is to demonstrate precisely the possibilities of data reduction without compromising the quality of the stored data. Through the experiments carried out, it was possible to design and validate an approach for the reduction and classification of electromes in a time series format. This approach uses the PAA time series approximation technique, in combination with SAX, to reduce the data volume of the time series, together with the SVM classification algorithm, which showed better performance in detecting plant stress from eletromes.

Challenges and Research Directions

Given the relative novelty in electrome reduction and classification and the constant technological advancement, there are still many open challenges in this topic, especially from the perspective of IoT and ML. In general, from studies already carried out, it is noted that few authors consider context peculiarities in the development of time series reduction solutions. The scope of plant electrophysiology is an example of a very specific context, since the collection of data from electrical activities produces significantly large time series with specific characteristics, which generic compressors do not consider and do not deal with. Additionally, other characteristics of the context, such as infrastructure issues, can influence the development of solutions. Energy limitations, connectivity and communication restrictions, or sensor reliability also deserve space in the discussion. Also, ML methods have been little explored in the reduction step, rather than just in the classification. Given the potential and suitability of ML models in different contexts, exploring the use of ML to reduce time series can help increase the performance of applications as a whole.

6 Conclusion and Future Work

This work proposes an approach for Time Series reduction and classification and presents experiments to validate its functioning and performance. The time series analyzed and studied here are those that represent plant electromes, characterized by storing information about electrical activities of plants and having a large volume, due to the high frequency of collection by IoT sensors. This approach explores mathematical methods for time series approximation and ML-based techniques for time series classification, to detect stress in plant electromes. From real data collected by sensors, a first step was taken to reduce the volume of data, where three time series approximation techniques were implemented: *Piecewise Aggregate Approximation - PAA*, *Symbolic Aggregate approXimation - SAX* and *Multiple Coefficient Binning - MCB*. The reduced time series were then used as input for the training of artificial intelligence models. For training, algorithms *KNN, SVM* and *ANN* were used. The results of the experiment were analyzed from *Compression Rate* and *Space Saving* metrics, for data reduction, and with *Accuracy* and *Confusion Matrix*, for classification models. The results obtained validated the proposed approach of using the PAA+SAX technique to reduce time series in combination with the SVM algorithm for plant stress classification. The main challenges and open issues for future work are: IoT context analysis, which can vary greatly; analysis of infrastructure limitations, such as energy and communication costs; and, the use of ML in the data reduction stage, in addition to classification.

7 Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

Cruces, N.; Seco, D.; and Guitérrez, G. 2019. A Compact Representation of Raster Time Series. In *2019 Data Compression Conference (DCC)*, 103–111.

de Oliveira Jr., M.; Sedrez, G.; Monteiro, A.; Puntel, F.; and Cavalheiro, G. 2021. Effects of agro-sensor time series approximation on plant stress detection: an experimental study. In *XIII SBIAGRO*. RS, Brasil: SBC.

de Oliveira Jr., M.; Sedrez., G.; de Souza., G.; and Cavalheiro, G. 2022. An Application with Jetson Nano for Plant Stress Detection and On-field Spray Decision. In *Proceedings of the 11th International Conf. on Sensor Networks*.

Evgeniou, T., and Pontil, M. 2001. Support Vector Machines: Theory and Applications. volume 2049, 249–257.

Faouzi, J., and Janati, H. 2020. pyts: A Python Package for Time Series Classification. *Journal of Machine Learning Research* 21(46):1–6.

Guo, G.; Wang, H.; Bell, D.; and Bi, Y. 2004. KNN Model-Based Approach in Classification.

Hamilton, J. D. 2020. *Time series analysis*. Princeton university press.

Hu, S.; Xi, Z.; McGowin, G.; Sukthankar, G.; Fiore, S. M.; and Oden, K. 2021. Representing time series data in intelligent training systems. *The International FLAIRS Conference Proceedings* 34.

Jain, A. K.; Mao, J.; and Mohiuddin, K. M. 1996. Artificial neural networks: a tutorial. *Computer* 29(3):31–44.

Jensen, S. K.; Pedersen, T. B.; and Thomsen, C. 2017. Time Series Management Systems: A Survey. *IEEE Trans. on Knowledge and Data Engineering* 29(11).

Keogh, E.; Chakrabarti, K.; Pazzani, M.; and Mehrotra, S. 2002. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. *Kn. and Inf. Syst.*

Lelewer, D. A., and Hirschberg, D. S. 1987. Data Compression. *ACM Comput. Surv.* 19(3):261–296.

Lin, J.; Keogh, E.; Wei, L.; and Lonardi, S. 2007. Experiencing SAX: A Novel Symbolic Representation of Time Series. *Data Min. Knowl. Discov.* 15.

Loof, A. 2016. The cell's self-generated 'electrome': The biophysical essence of the immaterial dimension of Life? *Communicative & Integrative Biology* 9.

Navarro, E.; Costa, N.; and Pereira, A. 2020. A Systematic Review of IoT Solutions for Smart Farming. *Sensors* 20(15).

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Pereira, D. P.; Papa, J. P.; Saraiva, G. F. R.; and Souza, G. M. 2018. Automatic classification of plant electrophysiological responses to environmental stimuli using machine learning and interval arithmetic. *Computers and Electronics in Agriculture* 145.

Reissig, G. N.; Oliveira, T. F. C.; Costa, A. V. L.; Parise, A. G.; Pereira, D. R.; and Souza, G. M. 2021. Machine Learning for Automatic Classification of Tomato Ripening Stages Using Electrophysiological Recordings. *Front. in Sustainable Food Syst.* 5.

Rosenberger, J.; Kübel, A.; and Rothfuß, F. 2022. Comparison and extension of autoencoder models for uni- and multivariate signal compression in IIoT. In *2022 Data Compression Conference (DCC)*, 481–481.

Schäfer, P., and Högqvist, M. 2012. SFA: A symbolic fourier approximation and index for similarity search in high dimensional datasets. 516 – 527.

Souza, G. M.; Ferreira, A. S.; Saraiva, G. F. R.; and Toledo, G. R. A. 2017. Plant "electrome" can be pushed toward a self-organized critical state by external cues: Evidences from a study with soybean seedlings subject to different environmental conditions. *Plant Signaling & Behavior* 12(3):e1290040. PMID: 28277967.

Toledo, G. R. A. 2019. *Caracterização eletrofisiológica do feijão (*Phaseolus vulgaris *L.) cv. BRS-Expedito sob diferentes disponibilidades hídricas (in Portuguese)*. Ph.D. Dissertation, UFPel, Pelotas.

Walleligh, S.; Polceanu, M.; and Buche, C. 2018. Soybean plant disease identification using convolutional neural network. In Brawner, K., and Rus, V., eds., *Proceedings of the 31st International FLAIRS Conference, Florida, USA*.