

Técnicas de Aprendizado de Máquina Aplicadas à Identificação de Estresse em Plantas

Pedro Minoru Izumida de Almeida

Centro Universitário das Faculdades Metropolitanas
Unidas (FMU)
São Paulo, São Paulo, Brasil
pedroizumida@outlook.com

Orlando da Silva Junior

Universidade de São Paulo (USP)
São Paulo, São Paulo, Brasil
orlando.silvajr@usp.br

ABSTRACT

Plants can emit electrical signals under stress conditions, such as predator attacks and climatic changes. The capture and identification of these signals allows biologists and other researchers to identify the cause of the stress and contribute to the identification of external factors that may pose a risk to plant health. This research aims to classify the presence of stress through the electrical signals emitted by plants in response to different types of stimuli, generated in controlled environments, using Machine Learning techniques. Several algorithms are applied, and the results are compared. Tree-based models show better responses in identifying different types of stress in plants.

KEYWORDS

plant stress, machine learning, artificial neural networks, decision trees

1 INTRODUÇÃO

Em 1873, Burdon-Sanderson e John Scott [1] conduziram os primeiros experimentos envolvendo a identificação de sinais elétricos em plantas ao utilizar eletrodos nas folhas da espécie *Dioanaea muscipula* (vênus-caça-moscas, planta carnívora). Devido ao conceito presente na época sobre a emissão de sinais elétricos serem presentes apenas em seres capazes de uma rápida resposta ao ambiente, a planta carnívora foi selecionada, graças a sua capacidade de movimentação física.

Contudo, apesar de haver espécies sésseis, incapazes de se mover, são presentes indícios entre a similaridade nos sinais elétricos emitidos por seres humanos e plantas [2]. Há a redução no grau de entropia dos sinais emitidos pelo coração humano ao ser afetado pela fibrilação, um tipo de arritmia cardíaca. Analogamente, as plantas ao serem submetidas a condições prejudiciais ao seu bem-estar, elucidando o termo estresse, apresentam redução no grau de complexidade (entropia) de seus sinais [3].

Com o progresso do Aprendizado de Máquina (AM) nos últimos anos, estudos como [3] passaram a investigar o comportamento nos sinais elétricos de baixa tensão sobre o efeito de diferentes estímulos na dinâmica temporal da planta da soja. O estado-da-arte consiste em identificar os sinais elétricos emitidos

pelos plantas, tornando-as capazes de comunicarem o seu estado [2-3-4-5].

A análise envolvendo a saúde da planta pode levar em consideração diversos fatores, como os níveis de agentes químicos, temperatura, luminosidade etc. A planta pode ser utilizada como amostra viva para a monitoração de uma colheita, exemplificando o conceito de biossensor. O uso da ótica da própria planta, tem como origem a busca por um meio mais preciso para a classificar o seu status, o que pode impactar diretamente no aumento de produtividade da plantação. A adoção de biossensores implica ainda em uma solução sustentável e possivelmente capaz de reduzir custos atrelados a técnicas/equipamentos de maior complexidade.

O desafio nesta pesquisa, reside na melhor estratégia para modelar o problema envolvendo a classificação dos sinais elétricos emitidos pela planta da soja. São avaliados os impactos gerados ao subdividir os conjuntos de dados e a aplicação de técnicas de processamento de dados.

A sumarização dos dados, utilizada na etapa de processamento, é realizada para avaliação da hipótese de redução da complexidade dos conjuntos de dados ao resumir o comportamento que possuem. O principal objetivo deste trabalho consiste em classificar a presença de estresse em plantas através da emissão de seus sinais elétricos. São utilizadas e comparadas diferentes técnicas de AM.

2 REVISÃO BIBLIOGRÁFICA

Máquinas de Vetores de Suporte (SVM), Árvores de Decisão (DT) e *Multilayer Perceptron* (MLP) são alguns dos algoritmos comumente empregados para a classificação de sinais elétricos em plantas. A atividade elétrica em frutos de tomate é avaliada por meio da entropia aproximada através de DTs [6]. Florestas Aleatórias (RF) e DTs foram utilizadas para a classificação da condição de estresse hídrico ou do ritmo nictêmico (ciclo de um intervalo de 24 horas de processos biológicos) em plantas.

Para avaliar de maneira mais precisa o comportamento dos dados, é possível que os conjuntos sejam particionados em janelas no formato $M \times N$, em que M corresponde a quantidade de subconjuntos e N o volume de amostras em cada um [8, 9]. Por sua vez, as características envolvidas em sinais biológicos são geralmente avaliadas através de medidas estatísticas de dispersão, como média, variância, curtose e intervalo interquartil [9].

3 METODOLOGIA

Os dados de entrada para os modelos foram compostos pela diferença entre o sinal captado por dois canais/eletrodos fixados em regiões diferentes da planta da soja [3]. O volume dos conjuntos foi delimitado em 160 mil registros para cada estímulo, como o frio, baixa luminosidade e o agente químico manitol induzindo a planta a um estado de baixo potencial osmótico. A **Figura 1** ilustra o fluxo desenvolvido.



Figura 1. Fluxo 1ª parte - divisão dos dados

Os dados são normalizados pelo desvio padrão e divididos na metade entre o estado normal/sob estresse e para cada fase de aplicação do estímulo. A planta da fortuna possui dados durante a aplicação do estímulo, neste caso o conjunto de dados foi separado também entre as fases antes/durante e durante/depós.

Quatro experimentos foram conduzidos. A janela com tamanho 1600×50 foi adotada nos três experimentos finais. Os dados foram particionados reservando 80% para treinamento e 20% para teste. No quarto experimento os dados foram sumarizados por meio das medidas estatísticas de dispersão, incluindo média, intervalo interquartil, variância, desvio padrão, assimetria e curtose. Os modelos SVM, DT, MLP e RF foram utilizados para a classificação de forma binária. O modelo XGBoost (XGB) também foi avaliado e comparado junto aos demais. Os experimentos foram programados em Python com auxílio das bibliotecas Scikit-Learn e XGboost.

A medida F1Score (F1) foi adotada para a avaliação do desempenho dos modelos de AM, sendo formada pela média harmônica entre a precisão e a revocação. A precisão avalia o grau de resultados positivos [10], ou seja, as amostras da planta com status de estresse classificadas de fato como estresse, e possui como intuito medir os acertos dos modelos. A métrica de revocação mede a assertividade do classificador [10], por exemplo, entre a classe da planta com estresse e o total avaliado, isto é, dados em estado normal somados aos de estresse. O conceito da medida F1 também envolve uma escala de zero e um, sendo um (1) reservado aos melhores resultados e zero (0) atribuído aos de baixo desempenho.

$$F1 = \frac{2}{Precisão^{-1} + Revocação^{-1}} = 2 * \frac{Precisão * Revocação}{Precisão + Revocação} \quad (3)$$

Ao consolidar os dados em sua forma bruta temos o 1º experimento, consumindo o volume total dos dados para a classificação. O 2º experimento, divisão das bases brutas, terá ocorrido caso os dados não sejam consolidados. O formato $M \times N$ também foi utilizado no 2º experimento, contudo gerou M arquivos apartados contendo N amostras, classificados de forma isolada. Ao evitar o processo de consolidar as bases, classificar e processar os dados o 3º experimento adota o conceito de janelas, formando os

subconjuntos com os dados em sua forma bruta. Em contrapartida no 4º experimento, ao não classificar, mas optar por adotar as janelas estatísticas as amostras são processadas através das medidas estatísticas de dispersão em cada subconjunto. Por fim, os dados são submetidos aos modelos para a classificação da presença do estresse. A **Figura 2** descreve a segunda parte do processo, apresentando o fluxo para cada experimento.

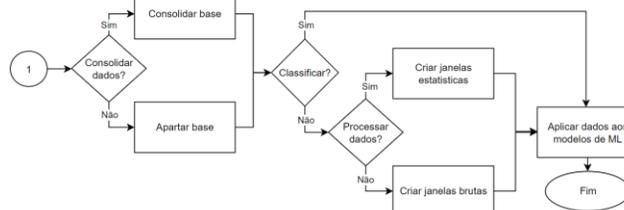


Figura 2. Fluxo 2ª parte - experimentos

4 RESULTADO E DISCUSSÕES

Para cada experimento foi avaliado a frequência de classificações (eixo vertical) obtidos ao analisar a medida de F1Score (eixo horizontal), dando ênfase aos resultados obtidos entre a faixa de 70-100 pontos.

O 1º experimento, **Figura 3**, apresentou a maior concentração em 100 pontos, mas com leves distribuições entre as faixas de F1 para 60-70 e 80-90.

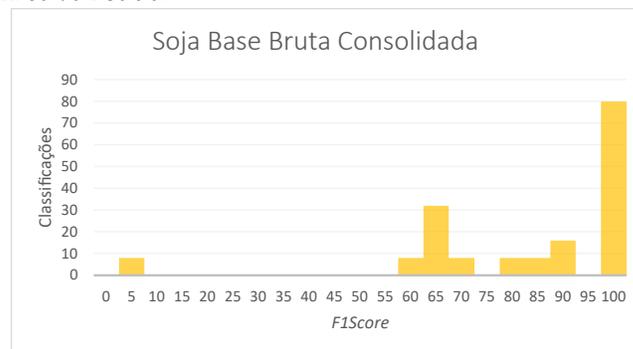


Figura 3. Primeiro Experimento

A variação entre os picos de menor expressividade, presentes no 2º experimento, indicou a existência de trechos mais relevantes na identificação do estresse. O surgimento de picos discretos, preenchendo lacunas no experimento anterior, sugeriu melhora dos modelos ao classificar a presença do estresse apenas realizando a divisão dos dados. Contudo, a dispersão ainda presente sugeriu impacto negativo nos modelos ao perder do contexto geral dos dados. A combinação destes comportamentos evidenciou o benefício gerado ao realizar a divisão dos dados. A frequência de classificações com os resultados agrupados em cinco unidades para o eixo x (F1Score) são presentes na **Figura 4**.

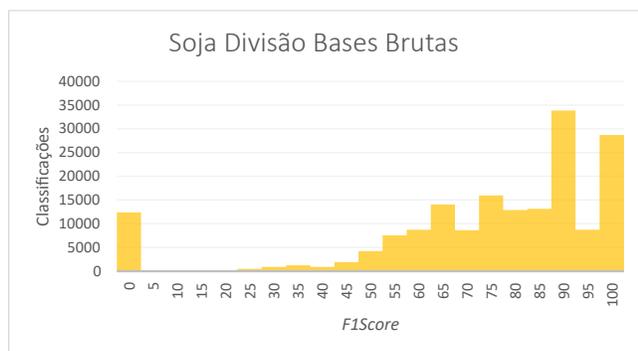


Figura 4. Segundo Experimento

No 3º experimento, **Figura 5**, o aumento no volume das concentrações para F1, 90 e 95, demonstrou melhora no desempenho dos modelos ao consumir o volume total dos dados, porém através do conceito de janelas e com os dados brutos. A dispersão entre os resultados obtidos instigou a adesão de técnicas de processamentos de dados realizada no 4º experimento.

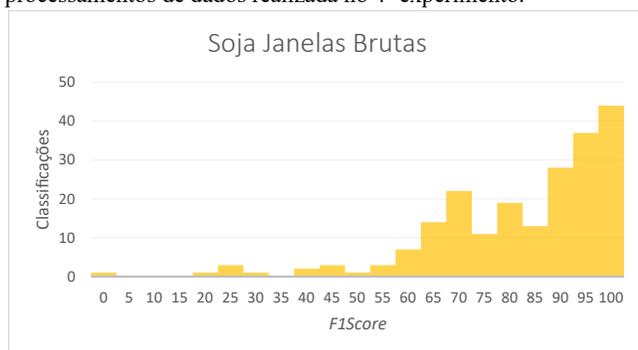


Figura 5. Terceiro Experimento

O 4º e último experimento, **Figura 6**, composto pelas janelas estatísticas apresentou o melhor desempenho ao processar os dados. Houve a redução na distribuição dos resultados.

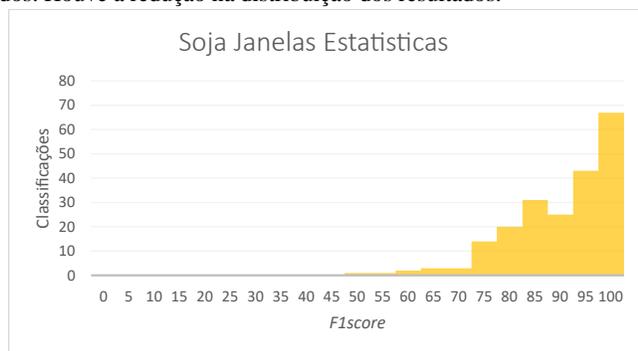


Figura 6. Quarto Experimento

A **Tabela 1**, a seguir, apresenta os resultados dos modelos com melhor desempenho ao classificar a presença do estresse,

evidenciando a influência positiva ao adotar a combinação do conceito de janelas e medidas estatísticas.

Estímulo	1ª (Posição)	2ª (Posição)	3ª (Posição)
Frio	96.63 (XGB)	96.62 (MLP)	96.44 (RF)
Luminosidade	87.42 (RF)	86.62 (SVM)	84.75 (XGB)
Manitol	99.26 (DT)	99.12 (SVM)	99.11 (XGB)

Tabela 1. Janela 1600x50 ranking dos melhores classificadores

5 CONCLUSÕES

Este trabalho apresenta resultados prévios para a classificação da presença de estresse em plantas a partir dos sinais elétricos que emitem. Modelos baseados em árvores (XGB e RF) apresentam os melhores resultados em relação a outros paradigmas. A viabilidade ao adotar técnicas de AM é reforçada ao resumir o comportamento dos dados, uma vez que este processo se mostrou capaz de potencializar a eficácia dos classificadores ao identificarem a presença de estresse em plantas de soja.

REFERÊNCIAS

- [1] John Scott Burdon-Sanderson. 1873. Note on the electrical phenomena which accompany irritation of the leaf of *Dionæa muscipula*. Proceedings of the Royal Society of London 21, 139-147 (1873), 495-496.
- [2] F. Z. Simmi, L. J. Dallagnol, A. S. Ferreira, D. R. Pereira, and G. M. Souza. 2020. Electrome alterations in a plant-pathogen system: Toward early diagnosis. *Bioelectrochemistry* 133 (2020), 107493.
- [3] Gustavo M. Souza, Arlan S. Ferreira, Gustavo F. R. Saraiva, and Gabriel R. A. Toledo. 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 (2017), e1290040.
- [4] Segundo-Ortin, M., & Calvo, P. 2022. Consciousness and cognition in plants. *Wiley Interdisciplinary Reviews: Cognitive Science*, 13(2), e1578.
- [5] Segundo-Ortin, M., & Calvo, P. 2023. Plant sentience? Between romanticism and denial: *Science. Animal Sentience*, 8(33), 1.
- [6] G. N. Reissig, T. F. de Carvalho Oliveira, A. G. Parise, Á. V. L. Costa, D. A. Posso, C. V. Rombaldi, and G. M. Souza. 2023. Approximate entropy: a promising tool to understand the hidden electrical activity of fruit. *Communicative & Integrative Biology* 16, 1 (2023), 2195236.
- [7] D. Tran, F. Dutoit, E. Najdenovska, N. Wallbridge, C. Plummer, M. Mazza, Laura E. Raileanu, and C. Camps. 2019. Electrophysiological assessment of plant status outside a Faraday cage using supervised machine learning. *Scientific reports* 9, 1 (2019), 17073.
- [8] E. Buss, T. Aust, M. Wahby, T. L. Rabbel, S. Kernbach, and H. Hamann. 2023. Stimulus classification with electrical potential and impedance of living plants: comparing discriminant analysis and deep-learning methods. *Bioinspiration & Biomimetics* 18, 2 (2023), 025003.
- [9] S. K. Chatterjee, S. Das, K. Maharatna, E. Masi, L. Santopolo, S. Mancuso, and A. Vitaletti. 2015. Exploring strategies for classification of external stimuli using statistical features of the plant electrical response. *Journal of the Royal Society Interface* 12, 104 (2015), 20141225.
- [10] Bishop, C. 2006. Pattern recognition and machine learning. Springer google schola, 2, 5-43.