

# Análise de Desempenho e Comparação de Técnicas de Explicabilidade em Modelos de Séries Temporais

Matheus Leandro de Melo Silva<sup>1,2</sup>, Lívia Almada Cruz<sup>2</sup>, Regis Pires Magalhães<sup>1,2</sup>, Emanuel Ferreira Coutinho<sup>1,2</sup>

<sup>1</sup>Programa de Pós-Graduação em Computação (PCOMP)

<sup>2</sup>Universidade Federal do Ceará (UFC) – Quixadá – CE – Brasil

matheusleandro@alu.ufc.br  
{livia.almada, regismagalhaes, emanuel.coutinho}@ufc.br

**Abstract.** *Time series are widely used because they represent the evolution of phenomena over time and reveal trends relevant to sensitive decision-making. Therefore, when machine learning models are applied to their classification, understanding their behavior becomes necessary. In this context, explainability is essential to validate model decisions and ensure transparency. This work analyzes and compares different explainability techniques applied to time-series classification, evaluating execution time and computational cost to highlight the advantages and limitations of the main approaches.*

**Resumo.** *Séries temporais são amplamente utilizadas por representarem a evolução de fenômenos ao longo do tempo e revelarem tendências relevantes para decisões sensíveis. Por isso, quando modelos de aprendizado de máquina são aplicados à sua classificação, torna-se necessário compreender seu comportamento. Nesse contexto, a explicabilidade é essencial para validar as decisões dos modelos e garantir transparência. Assim, este trabalho analisa e compara diferentes técnicas de explicabilidade aplicadas à classificação de séries temporais, avaliando tempo de execução e custo computacional para evidenciar as vantagens e limitações das principais abordagens.*

## 1. Introdução

O aprendizado de máquina tem se consolidado como uma ferramenta fundamental em diversas áreas, tanto para a automação de tarefas repetitivas quanto para o apoio à tomada de decisão em cenários complexos [Shinde and Shah 2018]. Esse avanço ocorre em paralelo ao crescimento expressivo da disponibilidade de dados, especialmente aqueles organizados na forma de séries temporais, resultado da ampla disseminação de sensores e sistemas de monitoramento capazes de coletar informações de maneira contínua e periódica [Ji et al. 2022]. Nesse contexto, decisões cada vez mais relevantes passam a depender diretamente de modelos preditivos, tornando indispensável que tais decisões sejam confiáveis, transparentes e passíveis de validação, sobretudo quando aplicadas em domínios críticos ou de alta responsabilidade.

Entre os diferentes tipos de dados disponíveis, as séries temporais se destacam por representarem fenômenos que evoluem ao longo do tempo, permitindo a identificação de tendências, sazonalidades e padrões complexos. Essas características tornam esse

tipo de dado especialmente relevante em áreas sensíveis, como saúde, finanças e sistemas industriais, nas quais decisões equivocadas podem acarretar consequências significativas [chung Fu et al. 2008]. Além disso, a natureza sequencial e a dependência temporal inerentes às séries temporais tornam sua modelagem particularmente desafiadora [Dongen et al. 2001], exigindo modelos capazes de capturar relações dinâmicas e, ao mesmo tempo, fornecer subsídios que permitam aos profissionais compreender e validar como essas decisões são produzidas.

Nesse cenário, a adoção de modelos computacionais cada vez mais sofisticados para o tratamento de séries temporais intensifica os desafios relacionados à interpretabilidade. Embora modelos profundos apresentem elevado desempenho na identificação e modelagem de padrões temporais complexos, eles frequentemente operam como verdadeiras “caixas-pretas”, oferecendo pouca ou nenhuma explicação sobre seus mecanismos internos e sobre os fatores que influenciam suas previsões [Veerappa et al. 2022]. Essa limitação torna-se ainda mais crítica em dados temporais, nos quais pequenos erros de previsão podem se propagar ao longo do tempo e resultar em impactos significativos em decisões futuras [Šimić et al. 2021]. Assim, compreender o porquê das previsões passa a ser um requisito essencial para o uso seguro e responsável desses modelos.

Com o avanço das pesquisas em explicabilidade de inteligência artificial (XAI), diversas técnicas de explicabilidade têm sido propostas visando tornar modelos complexos mais transparentes. No entanto, observa-se que grande parte dessas abordagens foi originalmente desenvolvida para dados estáticos, sendo posteriormente adaptada para o contexto de séries temporais. Além disso, apesar da ampla variedade de métodos disponíveis, ainda são escassos os estudos que realizam análises comparativas sistemáticas, considerando simultaneamente o custo computacional e a qualidade das explicações geradas. Essa ausência de diretrizes claras dificulta a escolha das técnicas mais adequadas para diferentes cenários de aplicação.

Diante desse contexto, o objetivo deste trabalho é realizar uma análise comparativa abrangente das principais técnicas de explicabilidade aplicadas a modelos de séries temporais, considerando tanto o desempenho computacional quanto a eficácia explicativa das abordagens avaliadas. Busca-se identificar os principais trade-offs entre custo computacional e qualidade das explicações, bem como fornecer subsídios que auxiliem pesquisadores e profissionais na seleção das técnicas mais apropriadas conforme diferentes necessidades e contextos de aplicação.

## 2. Trabalhos Relacionados

Diversos estudos têm investigado o uso de técnicas de Explainable Artificial Intelligence (XAI) em diferentes domínios, com o objetivo de aumentar a transparência e a confiabilidade de modelos complexos. Sahatova e Balabaeva (2022) realizaram uma análise comparativa de métodos de explicabilidade aplicados à detecção de objetos em imagens de tomografia computadorizada, utilizando o modelo MULAN e o conjunto de dados DeepLesion. Técnicas como EigenCAM, Gradient SHAP, Layer Activation, Occlusion e LIME foram avaliadas com base em métricas de infidelidade e consistência. Os resultados indicaram melhor desempenho de métodos baseados em gradiente, enquanto abordagens baseadas em perturbação apresentaram maior sensibilidade a variações nos dados. Embora esse trabalho evidencie a importância da escolha adequada das técnicas de expli-

cabilidade conforme o domínio e a arquitetura do modelo, sua análise concentra-se em dados visuais e não considera aspectos relacionados ao custo computacional, tampouco o contexto de séries temporais, que é o foco do presente estudo.

Duell et al. (2021) compararam métodos de XAI, como SHAP e LIME, aplicados à análise de registros eletrônicos de saúde para a predição de mortalidade por câncer de pulmão com um classificador XGBoost. O estudo avaliou explicações locais e globais e observou divergências significativas no ranking de importância das características entre as técnicas analisadas, apesar da identificação de atributos clinicamente relevantes por ambas. Destacou-se que tais discrepâncias podem impactar a confiança de profissionais de saúde nos modelos preditivos, reforçando a necessidade de validação das explicações. Diferentemente do presente trabalho, tal análise concentra-se em dados tabulares e não explora o comportamento das técnicas de explicabilidade em dados sequenciais, nem avalia sistematicamente o desempenho computacional das abordagens consideradas.

Ye et al. (2021) propuseram um classificador de COVID-19 baseado em imagens de tomografia computadorizada, integrando métodos de XAI como CAM, LIME e SHAP. O estudo introduz um módulo explicativo adaptativo que gera mapas de ativação durante o treinamento do modelo, buscando superar limitações de técnicas pós-processadas. Ademais, a integração de fatias de imagens para análises em nível de paciente foi abordada e comparações qualitativas e quantitativas das explicações foram geradas. Embora os resultados indiquem boa acurácia e explicações visualmente consistentes, o foco do trabalho está em imagens médicas e em avaliações qualitativas, sem uma análise comparativa aprofundada do custo computacional das técnicas. Em contraste, o presente trabalho direciona sua análise para modelos de classificação de séries temporais, enfatizando tanto a eficiência computacional quanto a qualidade explicativa das abordagens avaliadas.

### **3. Referencial Teórico**

Esta seção apresenta os principais conceitos que constituem a base teórica deste trabalho.

#### **3.1. Séries Temporais**

Séries temporais consistem em sequências de observações ordenadas no tempo, nas quais cada valor está associado a um instante ou intervalo temporal específico. Diferentemente de dados estáticos, esse tipo de dado apresenta uma estrutura sequencial intrínseca, na qual as observações não são independentes entre si [Gottman 1981]. Em geral, valores passados influenciam diretamente valores futuros, caracterizando a dependência temporal. Além disso, séries temporais frequentemente exibem padrões como tendência de longo prazo, sazonalidade associada a ciclos regulares e componentes aleatórias ou de ruído, que refletem variações não sistemáticas do fenômeno observado. Essas propriedades tornam as séries temporais especialmente adequadas para a análise de fenômenos dinâmicos e evolutivos [Cui et al. 2016].

Diversas tarefas podem ser formuladas a partir de dados temporais, uma delas é a classificação de séries temporais, que busca atribuir rótulos a séries completas ou a segmentos temporais, sendo essa abordagem recorrente em aplicações como reconhecimento de atividades humanas, análise biomédica e monitoramento de sistemas [Liu et al. 2018]. Onde, a detecção de anomalias visa identificar comportamentos atípicos ou desvios significativos em relação ao padrão esperado, desempenhando papel fundamental em cenários de diagnóstico, segurança e manutenção preditiva [Carden and Brownjohn 2008].

### 3.2. Aprendizado de Máquina

Aprendizado de máquina compreende um conjunto de técnicas que permitem sistemas computacionais aprenderem padrões a partir de dados, sem a necessidade de regras explicitamente programadas [El Naqa and Murphy 2015]. Em abordagens supervisionadas, modelos são treinados a partir de exemplos rotulados para aprender uma função capaz de associar corretamente entradas a classes, sendo amplamente utilizados em tarefas de classificação. Esse paradigma tem se mostrado eficaz em problemas complexos, ou com necessidade de análise de muitos dados [Jiang et al. 2020].

Aplicado ao contexto de séries temporais, esta técnica surge como uma alternativa flexível aos métodos estatísticos clássicos [Wang et al. 2022]. Onde, na classificação de séries temporais, abordagens baseadas em extração de características temporais permitem a aplicação de modelos clássicos, como máquinas de vetores de suporte ou florestas aleatórias, enquanto modelos modernos, como redes neurais recorrentes, convolucionais e baseadas em atenção, aprendem representações diretamente a partir dos dados brutos. Porém, apesar do desempenho alcançado, essas abordagens frequentemente operam como caixas-pretas, o que dificulta a compreensão dos padrões temporais que influenciam suas decisões [Azodi et al. 2020].

Paralelamente, métodos baseados em projeções convolucionais aleatórias destacaram-se por combinar alto desempenho preditivo com baixo custo computacional na classificação de séries temporais [Liu et al. 2023]. Essas técnicas utilizam classificadores simples sobre representações transformadas, tornando-se especialmente atrativas em cenários com grandes volumes de dados. No entanto, embora eficientes, seus mecanismos internos de transformação permanecem pouco transparentes, evidenciando uma troca recorrente entre desempenho e interpretabilidade [Uribarri et al. 2024].

### 3.3. Explicabilidade

A explicabilidade em inteligência artificial (XAI), busca fornecer métodos capazes de esclarecer como e por que um modelo produz determinada saída. Em aplicações envolvendo dados temporais, essa necessidade é ainda mais pronunciada, uma vez que decisões baseadas em padrões sequenciais podem impactar diretamente análises futuras e processos de tomada de decisão contínuos.

Abordagens de explicabilidade podem ser classificadas inicialmente conforme o momento em que a explicação é incorporada ao modelo. Técnicas ad hoc referem-se a modelos projetados desde sua concepção com estruturas inerentemente interpretáveis, de modo que suas decisões possam ser compreendidas diretamente a partir de seus componentes internos. Exemplos incluem modelos lineares, árvores de decisão e algumas arquiteturas restritas que mantêm relações explícitas entre entradas e saídas. Em contrapartida, métodos post-hoc são aplicados após o treinamento de modelos já existentes, visando explicar suas previsões sem modificar sua estrutura interna. Essa categoria é particularmente relevante para modelos de alto desempenho, como redes neurais profundas e métodos baseados em transformações complexas, que normalmente operam como caixas-pretas.

Outra distinção fundamental no campo da XAI diz respeito ao grau de dependência da técnica em relação ao modelo a ser explicado. Técnicas agnósticas ao modelo são projetadas para operar independentemente da arquitetura ou do algoritmo

subjacente, tratando o modelo como uma função de mapeamento entre entradas e saídas. Essas abordagens permitem uma aplicação ampla e flexível, sendo especialmente úteis em cenários comparativos ou quando o acesso aos parâmetros internos do modelo é limitado. Porém, técnicas específicas ao modelo exploram informações internas, como gradientes, pesos ou mecanismos de atenção, para gerar explicações mais detalhadas e potencialmente mais fiéis ao funcionamento do modelo. Embora essas técnicas possam oferecer maior precisão explicativa, sua aplicabilidade é restrita a classes específicas de modelos.

No contexto de séries temporais, grande parte das técnicas de explicabilidade foi originalmente desenvolvida para dados estáticos e posteriormente adaptada para lidar com a natureza sequencial dos dados. Essa adaptação envolve desafios adicionais, como a preservação da dependência temporal e a interpretação da importância relativa de diferentes instantes ou segmentos da série. Explicações podem assumir a forma de mapas de relevância temporal, atribuindo pesos a janelas ou pontos específicos no tempo, ou destacar padrões temporais recorrentes que influenciam previsões do modelo. Entretanto, a correlação entre instantes temporais pode dificultar a interpretação direta dessas explicações, uma vez que alterações locais podem afetar globalmente o comportamento do modelo.

## 4. Metodologia

Esta seção descreve os procedimentos metodológicos adotados no trabalho, abrangendo coleta dos dados, treinamento do modelo de classificação de séries temporais e seleção das técnicas de explicabilidade utilizadas para a interpretação das previsões.

### 4.1. Coleta dos dados

Os dados utilizados neste estudo foram obtidos a partir do repositório *Time Series Classification*<sup>1</sup>, amplamente reconhecido na literatura por disponibilizar conjuntos de dados de referência, resultados experimentais e links para códigos relacionados a tarefas de classificação, clusterização e regressão de séries temporais. Os conjuntos de dados disponíveis nesse repositório encontram-se em formato adequado para processamento computacional, abrangendo séries temporais univariadas e multivariadas, acompanhadas de seus respectivos rótulos de classe.

Para este trabalho, foi selecionado um conjunto de dados pertencente ao domínio agrícola, composto por séries temporais rotuladas que representam diferentes tipos de plantações. Essas séries foram geradas a partir de imagens de satélite, nas quais cada instância temporal descreve a evolução espectral das áreas monitoradas ao longo do tempo, permitindo a distinção entre diferentes classes de cultivo. O conjunto de dados utilizado é composto por um conjunto de treinamento com 7.200 séries temporais e um conjunto de teste com 16.800 séries. Cada série possui comprimento fixo de 46 instantes temporais e está associada a uma dentre 24 classes distintas.

### 4.2. Treinamento do modelo

O modelo de classificação das séries temporais foi treinado utilizando a linguagem Python e a biblioteca *aeon*<sup>2</sup>, um framework especializado em aprendizado de máquina para dados

<sup>1</sup><https://timeseriesclassification.com/description.php?Dataset=Crop>

<sup>2</sup><https://www.aeon-toolkit.org/en/stable/>

temporais. Dentre os algoritmos disponíveis, foi selecionado o método dos  $k$ -vizinhos mais próximos ( $k$ -Nearest Neighbors – KNN), conforme estudos anteriores que utilizaram esse conjunto de dados como referência, conforme apresentado em [Tan et al. 2017].

Durante o treinamento, foram utilizados os parâmetros padrão recomendados pela biblioteca *aeon*, para garantir simplicidade, reprodutibilidade e alinhamento com práticas consolidadas na literatura. Essa escolha também visa evitar vieses decorrentes de ajustes manuais excessivos de hiperparâmetros, permitindo que a análise posterior de explicabilidade seja conduzida sobre um modelo representativo e amplamente utilizado. Por fim, o desempenho do modelo foi avaliado por meio da acurácia de classificação no conjunto de teste, assegurando que o classificador apresentasse resultados satisfatórios antes da aplicação das técnicas de explicabilidade.

### 4.3. Seleção de técnicas XAI

As técnicas de explicabilidade adotadas neste trabalho foram selecionadas com base no levantamento da literatura apresentado por [Theissler et al. 2022] sobre métodos de *Explainable Artificial Intelligence* aplicáveis à classificação de séries temporais. A partir desse levantamento, definiu-se como critério de seleção a escolha de técnicas que fossem simultaneamente desenvolvidas especificamente para dados temporais, operassem de forma *post-hoc* e fossem agnósticas ao modelo. Técnicas *post-hoc* permitem a geração de explicações após o treinamento do classificador, sem alterações em sua arquitetura, enquanto métodos agnósticos baseiam-se exclusivamente nas entradas e saídas do modelo, possibilitando sua aplicação a diferentes algoritmos.

Com base nesses critérios, foram selecionadas três técnicas representativas de diferentes paradigmas explicativos. A técnica *Native Guide* emprega perturbações em janelas temporais para identificar regiões da série mais relevantes para a decisão do classificador. A abordagem *Conceptual Explanations* fornece explicações baseadas em características conceituais, como tendência, variabilidade e sazonalidade, facilitando a interpretação por especialistas do domínio. Por fim, a técnica *LASTS (Local Agnostic Subsequence-based Time Series Explanations)* utiliza subsequências discriminativas (*shapelets*) para gerar explicações locais e contrafactuais. A escolha dessas técnicas permite uma análise comparativa abrangente entre diferentes perspectivas explicativas, considerando aspectos como custo computacional, fidelidade e interpretabilidade.

### 4.4. Projeto do experimento

Este experimento pretende investigar e comparar o desempenho computacional e as características das explicações geradas por diferentes técnicas de explicabilidade aplicadas à classificação de séries temporais. A análise busca fornecer subsídios práticos para a escolha de métodos de *Explainable Artificial Intelligence* em cenários reais, considerando a influência da complexidade temporal sobre o custo computacional e a interpretabilidade das explicações.

**Fatores e tratamentos:** Fator 1 – Técnica de explicabilidade - *Native Guide*, *Conceptual Explanations*, *LASTS*; Fator 2 – Complexidade da janela temporal - w1: curta (ex.: 5 *timesteps*), w2: média (ex.: 15 *timesteps*), w3: longa (ex.: 30 *timesteps*); Fator 3 – Dataset - D1: Plantações (ex.: classificação do tipo de plantas).

**Variáveis Independentes:** Técnica de explicabilidade, Tamanho da janela temporal.

**Variáveis Dependentes:** Latência média por explicação (ms), Custo computacional (uso médio de CPU e RAM), Estabilidade das explicações frente a diferentes tamanhos de janela temporal.

#### 4.5. Formulação das Hipóteses

Espera-se que técnicas de explicabilidade baseadas em perturbação e amostragem local, como o *Native Guide* e o *LASTS*, apresentem maior latência média e custo computacional à medida que o tamanho da janela temporal aumenta, em função do maior número de avaliações necessárias sobre o modelo. Em contrapartida, técnicas baseadas em conceitos agregados, como *Conceptual Explanations*, tendem a apresentar menor custo computacional e maior estabilidade das explicações, especialmente em janelas temporais longas. Além disso, pressupõe-se que o aumento da complexidade temporal impacte negativamente a clareza das explicações locais, tornando mais desafiadora a identificação precisa dos segmentos temporais mais relevantes para a decisão do classificador.

### 5. Resultados

Esta seção apresenta a avaliação experimental das técnicas de explicabilidade aplicadas à classificação de séries temporais, considerando três abordagens distintas: *Native Guide*, *Conceptual* e *LASTS*. As técnicas foram comparadas sob diferentes perspectivas, incluindo custo computacional, uso de recursos, granularidade das explicações e sensibilidade a perturbações no sinal de entrada. Os resultados são apresentados com métricas quantitativas e visualizações gráficas, permitindo uma análise comparativa objetiva.

#### 5.1. Ambiente Experimental

Os experimentos foram realizados em um computador equipado com processador Intel Core i5-11400H, 32 GB de memória RAM e GPU NVIDIA RTX 3050, utilizando o sistema operacional Windows 11.

#### 5.2. Tempo Médio de Execução

A Figura 1 apresenta o tempo médio de execução das técnicas avaliadas para diferentes tamanhos de janela. Observa-se uma diferença entre os métodos, com a abordagem *Conceptual* apresentando desempenho computacional significativamente superior, com tempos médios da ordem de milissegundos, variando aproximadamente entre 0,0024 s e 0,0030 s. Em contrapartida, as técnicas *Native Guide* e *LASTS* demandam tempos substancialmente maiores, situando-se na faixa de segundos, com médias entre 1,66 s e 3,98 s para o *Native Guide* e entre 2,92 s e 3,48 s para o *LASTS*. Além disso, nota-se que o *Native Guide* apresenta redução do tempo de execução à medida que o tamanho da janela aumenta, enquanto o *LASTS* mantém custo computacional relativamente estável. Esses resultados podem evidenciar que métodos baseados em conceitos de alto nível são mais eficientes computacionalmente, ao passo que abordagens que operam diretamente sobre regiões temporais impõem maior custo, o que pode limitar sua aplicabilidade em cenários interativos ou de larga escala.

#### 5.3. Uso médio do hardware

A Figura 2 apresenta o uso médio de CPU observado durante o período de geração das explicações para cada técnica avaliada. Nota-se que as barras permanecem próximas

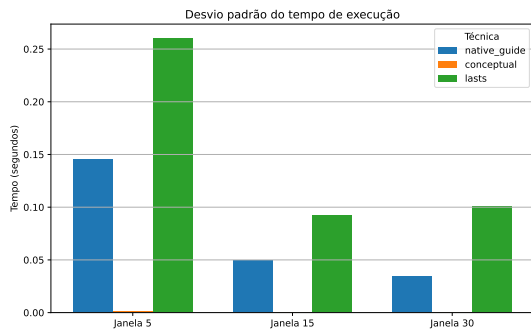


Figura 1. Tempo médio de execução

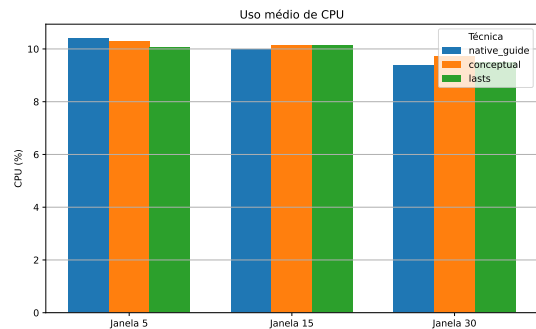


Figura 2. Consumo de CPU

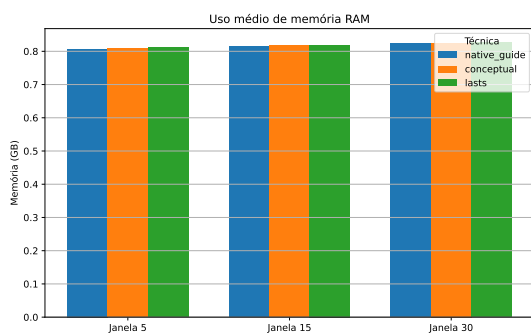


Figura 3. Consumo de Memória RAM

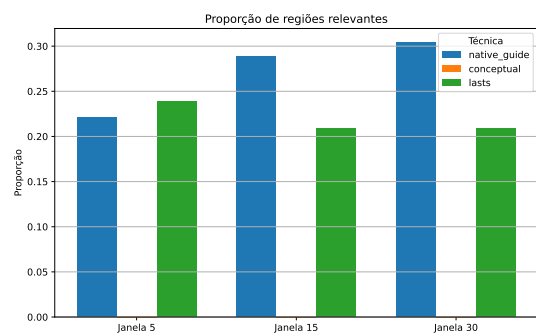


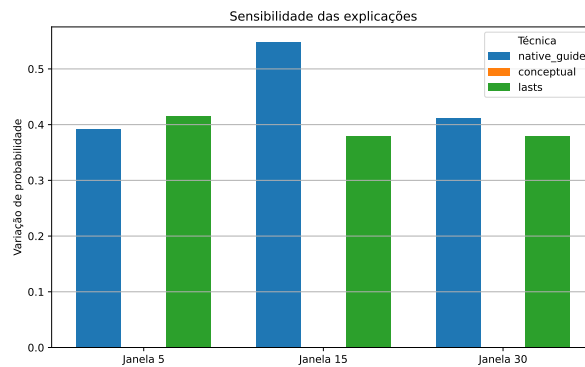
Figura 4. Proporção de pontos relevantes

entre si para todas as abordagens e tamanhos de janela, com valores médios situados aproximadamente entre 9% e 10%. Esse comportamento pode indicar que, no intervalo de execução analisado, não há diferenças expressivas no consumo de processamento entre as técnicas. O uso observado reflete majoritariamente a carga associada à execução do modelo de classificação e ao gerenciamento dos dados durante o processo de explicação, sugerindo que, apesar das diferenças nos tempos de execução, o impacto médio sobre a CPU do sistema mantém-se semelhante entre as abordagens.

A Figura 3 ilustra o consumo médio de memória RAM observado durante a execução das técnicas de explicação. Nota-se que as barras permanecem visualmente muito próximas entre si para todas as abordagens e tamanhos de janela, com valores médios variando aproximadamente entre 0,80 GB e 0,83 GB. Semelhante ao consumo do processador, esse resultado pode indicar que não há diferença expressiva no uso de memória entre as técnicas, uma vez que o consumo observado corresponde majoritariamente à memória necessária para manter o modelo de classificação e os dados carregados durante o processo de explicação. Assim, embora Native Guide e LASTS imponham maior custo em termos de processamento, elas não introduzem sobrecarga relevante de memória, mantendo um perfil de consumo semelhante ao da abordagem Conceptual.

#### 5.4. Análise das Regiões Temporais Relevantes e Sensibilidade

A Figura 4 apresenta a proporção média de regiões temporais identificadas como relevantes pelas técnicas Native Guide e LASTS. A abordagem Conceptual não é considerada



**Figura 5. Sensibilidade das explicações**

nessa métrica, uma vez que sua explicação é formulada exclusivamente no nível de conceitos abstratos, sem a identificação explícita de pontos ou segmentos do sinal temporal, o que justifica seu valor nulo no gráfico. Observa-se que o Native Guide seleciona, em média, aproximadamente 27,1% da série temporal como relevante, enquanto o LASTS identifica cerca de 21,9%. Esse resultado indica que o Native Guide tende a produzir explicações mais abrangentes, ao passo que o LASTS adota uma estratégia mais seletiva na delimitação das regiões explicativas.

Complementarmente, a sensibilidade das explicações às perturbações é avaliada por meio da taxa média de mudança na predição do modelo, apresentada na Figura 5. Verifica-se que o Native Guide apresenta maior sensibilidade, com valor médio em torno de 0,53, enquanto o LASTS atinge aproximadamente 0,45. Esse comportamento sugere que as regiões destacadas pelo Native Guide exercem influência mais pronunciada sobre as decisões do modelo. Entretanto, essa maior sensibilidade pode estar associada à menor parcimônia da técnica, uma vez que uma fração mais extensa da série é considerada relevante. Em contrapartida, o LASTS concentra-se em regiões mais restritas, resultando em explicações mais concisas, ainda que com impacto ligeiramente inferior na predição.

### 5.5. Número Médio de Conceitos

A Figura 6 apresenta o número médio de conceitos extraídos por cada técnica ao longo das explicações geradas. Observa-se que a abordagem Conceptual produz, em média, aproximadamente três conceitos por explicação, enquanto LASTS e Native Guide extraem cerca de dois e um conceito, respectivamente. Esse comportamento indica que a técnica baseada em conceitos tende a fornecer uma decomposição mais rica e interpretável do processo decisório do modelo, ao representar padrões de alto nível de forma explícita. Por outro lado, as abordagens que operam diretamente sobre regiões temporais relevantes concentram-se em um número reduzido de elementos explicativos, privilegiando explicações mais pontuais e localizadas, o que pode limitar a abrangência semântica da interpretação.

### 5.6. Discussão

Observa-se que a abordagem Conceptual ocupa uma região de baixo custo computacional e alta densidade conceitual, tornando-se particularmente adequada para aplicações que exigem explicações rápidas e interpretáveis em nível semântico. Em contraste, Native Guide e LASTS apresentam maior custo computacional, porém fornecem explicações

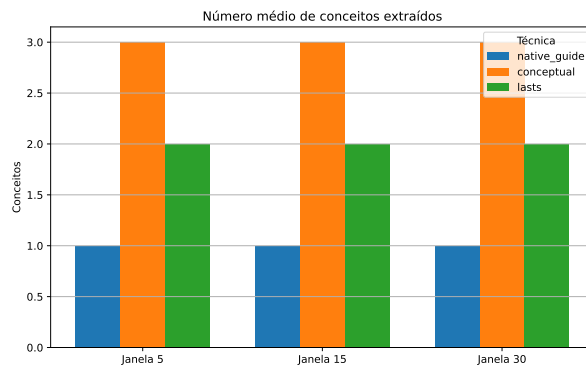


Figura 6. Número de conceitos extraídos

mais diretamente associadas a regiões específicas da série temporal, o que pode ser desejável em contextos que demandam maior fidelidade local. De forma geral, os resultados evidenciam que não existe uma abordagem universalmente superior, mas sim diferentes compromissos entre eficiência, granularidade e fidelidade explicativa. Técnicas conceituais mostram-se altamente eficientes e interpretáveis em nível abstrato, enquanto métodos baseados em regiões temporais oferecem maior detalhamento local, ao custo de maior consumo de recursos computacionais. Essas observações reforçam a importância da escolha da técnica de explicabilidade conforme os requisitos específicos da aplicação, como restrições de tempo, disponibilidade de recursos e nível desejado de interpretação.

## 6. Conclusão

Este trabalho apresentou uma análise comparativa de técnicas de explicabilidade aplicadas à classificação de séries temporais, utilizando um classificador do tipo K-Nearest Neighbors treinado sobre o conjunto de dados Crop, composto por séries de comprimento fixo igual a 46. As técnicas avaliadas, native guide, conceptual e lasts, foram analisadas sob diferentes tamanhos de janela, considerando métricas de desempenho computacional, fidelidade e consumo de recursos. Os resultados experimentais evidenciaram diferenças significativas entre as abordagens avaliadas. A técnica conceptual destacou-se pelo reduzido tempo de execução, apresentando desempenho superior às demais, o que a torna particularmente adequada para cenários que exigem baixa latência e alta eficiência computacional. Além disso, essa abordagem demonstrou estabilidade na extração de conceitos relevantes, com número médio de conceitos consistente e valores de importância bem definidos. Por outro lado, as técnicas native guide e lasts apresentaram maior custo computacional, porém forneceram explicações mais detalhadas no nível temporal, evidenciadas pela identificação de pontos relevantes e pelas variações observadas nas previsões e probabilidades do modelo. Esses resultados indicam que tais métodos podem ser mais apropriados em contextos nos quais a interpretabilidade local e a análise detalhada do comportamento do modelo ao longo da série temporal sejam prioritárias.

De forma geral, os experimentos demonstram que não existe uma técnica de explicabilidade universalmente superior, sendo a escolha da abordagem fortemente dependente dos requisitos da aplicação, como tempo de resposta, nível de detalhamento da explicação e disponibilidade de recursos computacionais. Assim, este estudo contribui ao fornecer uma avaliação dessas técnicas em um cenário de classificação de séries tem-

porais curtas. Entretanto, uma limitação presente se deve como a análise foi conduzida exclusivamente utilizando o classificador KNN, que é um modelo tradicionalmente interpretável e baseado em instâncias. Embora adequado para a avaliação inicial das técnicas de explicabilidade, a ausência de modelos considerados caixa-preta limita a generalização dos resultados para cenários de deep learning frequentemente associados às aplicações de *Explainable Artificial Intelligence*. Além disso, os experimentos foram realizados apenas sobre o conjunto de dados de agricultura, por conta da falta de dados de qualidade sobre a área da saúde, composto por séries temporais curtas e de comprimento fixo. Dessa forma, não foi possível avaliar de maneira aprofundada o impacto do aumento do tamanho das séries temporais sobre a escalabilidade das técnicas, especialmente em relação ao custo computacional e ao consumo de memória.

Como trabalhos futuros, sugere-se a ampliação da análise para diferentes classificadores de séries temporais, incluindo modelos de aprendizado profundo, como CNNs, RNNs e *Transformers*, permitindo investigar como as técnicas de explicabilidade se comportam em arquiteturas mais complexas. Também se recomenda a utilização de conjuntos de dados com séries temporais mais longas e características distintas, possibilitando uma análise mais robusta da escalabilidade, estabilidade do consumo de recursos e granularidade das explicações geradas. Além disso, a investigação de métricas adicionais de qualidade das explicações pode contribuir para uma compreensão mais aprofundada da relação entre fidelidade, interpretabilidade e eficiência computacional.

## Referências

- Azodi, C. B., Tang, J., and Shiu, S.-H. (2020). Opening the black box: interpretable machine learning for geneticists. *Trends in genetics*, 36(6):442–455.
- Carden, E. P. and Brownjohn, J. M. (2008). Arma modelled time-series classification for structural health monitoring of civil infrastructure. *Mechanical systems and signal processing*, 22(2):295–314.
- chung Fu, T., lai Chung, F., Luk, R., and man Ng, C. (2008). Representing financial time series based on data point importance. *Engineering Applications of Artificial Intelligence*, 21(2):277–300.
- Cui, Z., Chen, W., and Chen, Y. (2016). Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995*.
- Dongen, H. P. V., Ruf, T., Olofsen, E., VanHartevelt, J. H., and Kruyt, E. W. (2001). Letter to the editor: Analysis of problematic time series with the lombscargle method, a reply to ‘emphasizing difficulties in the detection of rhythms with lombscargle periodograms’. *Biological Rhythm Research*, 32(3):347–354.
- Duell, J., Fan, X., Burnett, B., Aarts, G., and Zhou, S.-M. (2021). A comparison of explanations given by explainable artificial intelligence methods on analysing electronic health records. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–4. IEEE.
- El Naqa, I. and Murphy, M. J. (2015). What is machine learning? In *Machine learning in radiation oncology: theory and applications*, pages 3–11. Springer.
- Gottman, J. M. (1981). Time-series analysis. *A comprehensive introduction for social scientists*.

- Ji, C., Du, M., Hu, Y., Liu, S., Pan, L., and Zheng, X. (2022). Time series classification based on temporal features. *Applied Soft Computing*, 128:109494.
- Jiang, T., Gradus, J. L., and Rosellini, A. J. (2020). Supervised machine learning: a brief primer. *Behavior therapy*, 51(5):675–687.
- Liu, C.-L., Hsaio, W.-H., and Tu, Y.-C. (2018). Time series classification with multivariate convolutional neural network. *IEEE Transactions on industrial electronics*, 66(6):4788–4797.
- Liu, Q., Liu, B., and Du, Y. (2023). An algorithm to improve the performance of convolutional neural networks for tsc tasks. *Engineering Reports*, 5(4):e12589.
- Sahatova, K. and Balabaeva, K. (2022). An overview and comparison of xai methods for object detection in computer tomography. *Procedia Computer Science*, 212:209–219.
- Shinde, P. P. and Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA)*, pages 1–6.
- Šimić, I., Sabol, V., and Veas, E. (2021). Xai methods for neural time series classification: A brief review. *arXiv preprint arXiv:2108.08009*.
- Tan, C. W., Webb, G. I., and Petitjean, F. (2017). Indexing and classifying gigabytes of time series under time warping. In *Proceedings of the 2017 SIAM international conference on data mining*, pages 282–290. SIAM.
- Theissler, A., Spinnato, F., Schlegel, U., and Guidotti, R. (2022). Explainable ai for time series classification: a review, taxonomy and research directions. *Ieee Access*, 10:100700–100724.
- Uribarri, G., Barone, F., Ansuini, A., and Fransén, E. (2024). Detach-rocket: sequential feature selection for time series classification with random convolutional kernels. *Data Mining and Knowledge Discovery*, 38(6):3922–3947.
- Veerappa, M., Anneken, M., Burkart, N., and Huber, M. F. (2022). Validation of xai explanations for multivariate time series classification in the maritime domain. *Journal of Computational Science*, 58:101539.
- 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.
- Ye, Q., Xia, J., and Yang, G. (2021). Explainable ai for covid-19 ct classifiers: an initial comparison study. In *2021 IEEE 34th international symposium on computer-based medical systems (CBMS)*, pages 521–526. IEEE.