

Detecção de Anomalias em Séries Temporais em Calorímetros Finamente Segmentados

Elias G. M. B. da Silva, Mariana F. da Silva, Eduardo F. de Simas Filho, Juan L. Marin e Edmar E. P. de Souza

Resumo— Calorímetros são utilizados em experimentos de física de partículas para medição de energia. Falhas na instrumentação eletrônica podem produzir leituras incorretas, comprometendo o desempenho do experimento. Este artigo avalia técnicas para detectar anomalias em séries temporais aplicadas ao monitoramento da qualidade de dados de calorímetros. Três modelos de aprendizado de máquina (TranAD, iTransformer e LSTM-AE) são comparados em diferentes cenários anômalos para módulos mortos e ruído eletrônico aumentado. O TranAD obteve Coeficiente de Correlação de Matthews (MCC) de 0,775 para módulos mortos; o LSTM-AE obteve MCC de 0,240 para ruído, valor 29% superior ao de trabalhos anteriores.

Palavras-Chave— Detecção de anomalias, séries temporais, calorímetros, monitoramento de qualidade de dados, aprendizado profundo.

Abstract— Calorimeters are used in particle physics experiments for energy measurement. Faults in the electronic instrumentation can produce incorrect readings, compromising experiment performance. This paper evaluates techniques for detecting anomalies in time series applied to data quality monitoring of calorimeters. Three machine learning models (TranAD, iTransformer, and LSTM-AE) are compared across different scenarios involving dead modules and increased electronic noise. TranAD achieved a Matthews Correlation Coefficient (MCC) of 0.775 for dead modules, while LSTM-AE achieved an MCC of 0.240 for noise, 29% higher than in previous works.

Keywords— Anomaly detection, time series, calorimeters, data quality monitoring, deep learning.

I. INTRODUÇÃO

A busca pela compreensão da composição do universo tem impulsionado pesquisas ao longo dos séculos, entre elas as da física de altas energias (HEP, do inglês *High Energy Physics*). Nesse contexto, destaca-se o LHC (do inglês *Large Hadron Collider*), construído no Centro Europeu para Pesquisa Nuclear (CERN, do francês *Conseil Européen pour la Recherche Nucléaire*), o maior colisor de partículas em atividade. Com aproximadamente 27 km de circunferência e operando a cerca de 100 m de profundidade na fronteira franco-suíça, o LHC acelera feixes de prótons em direções opostas com velocidade próxima à da luz, colidindo-os a cada 25 ns com energia de 13 TeV [1].

O experimento ATLAS (do inglês *A Toroidal LHC Apparatus*) é um dos maiores detectores do LHC, responsável por importantes descobertas científicas como a observação do bóson de Higgs em 2012, em conjunto com o experimento

CMS [2]. O detector utiliza calorímetros para medir a energia e a posição das partículas produzidas nas colisões, organizados em milhares de células com eletrônica de leitura associada. A alta taxa de colisões aumenta a probabilidade de ocorrência de partículas raras, mas também gera um grande volume de dados por evento, cerca de 70 TB/s, exigindo um sistema de filtragem online (*trigger*) para selecionar apenas os eventos de interesse [3].

O desgaste e defeitos nos módulos eletrônicos de calorímetros podem degradar a qualidade das informações coletadas, comprometendo a identificação das partículas e a confiabilidade dos resultados. O Monitoramento de Qualidade de Dados (DQM, do inglês *Data Quality Monitoring*) é, portanto, essencial para garantir a integridade das medidas em experimentos de HEP. Tradicionalmente realizado com grande intervenção humana, o DQM tem motivado a investigação de técnicas automatizadas baseadas em aprendizado de máquina [4].

Nesse contexto, técnicas computacionais para a detecção de anomalias em séries temporais (TSAD, do inglês *Time Series Anomaly Detection*) emergem como uma abordagem promissora para identificar automaticamente defeitos nos sensores do calorímetro. Trabalhos recentes [4], [5] demonstraram a aplicação de modelos baseados em transformers e autoencoders para DQM em calorímetros finamente segmentados, utilizando variáveis de forma do chuveiro (*shower shapes*) derivadas do calorímetro. No entanto, a detecção de anomalias sutis, como ruído eletrônico aumentado em células específicas, permanece um desafio em aberto.

O trabalho de Boggia e Malaescu [5] avalia modelos TSAD baseados em *transformers*, incluindo iTransformer, TranAD e USAD, aplicados a dados sintéticos do Lorenzetti com anomalias injetadas. Os autores demonstram que módulos inativos são detectáveis com Coeficiente de Correlação de Matthews (MCC) [6] entre 0,5 e 0,77. A Colaboração do Experimento ATLAS tem investigado autoencoders para DQM do calorímetro com dados reais [4], demonstrando a viabilidade de abordagens não supervisionadas para monitoramento automatizado.

Este trabalho apresenta uma avaliação de métodos de TSAD aplicados ao monitoramento de qualidade de dados em calorímetros finamente segmentados, utilizando dados sintéticos gerados pelo simulador Lorenzetti [7]. São avaliados três modelos: TranAD [8], iTransformer [9] e LSTM-AE [10], em 13 cenários distintos, variando a intensidade de ruído, taxa de anomalia e presença de sobreposição de eventos em um mesmo canal de leitura (*pileup*). Adicionalmente, é realizada uma busca de hiperparâmetros para otimização do iTransformer, bem como a análise do poder discriminativo dos atributos de entrada utilizados (*shower shapes*). O texto está organizado

Elias Guimarães Miranda Barbosa da Silva, Eduardo Furtado de Simas Filho, Edmar Egídio Purcino de Souza e Mariana Farias da Silva, Laboratório de Sistemas Digitais, PPGEEC/UFBA, Salvador-BA, e-mail: {elias.barbosa, fariasmariana, eduardo.simas, edmar.egidio}@ufba.br; Juan Lieber Marin, Coordenação de Eletrônica, IFBA, Vitória da Conquista-BA, email: juan.marin@ifba.edu.br.

da seguinte forma: a Seção II descreve os conceitos teóricos importantes para o entendimento do trabalho. Já a Seção III descreve a metodologia utilizada. A Seção IV discute os resultados e impactos da seleção de características mais discriminativas na detecção de anomalias. As discussões são sumarizadas na Seção V¹.

II. FUNDAMENTAÇÃO TEÓRICA

A. Calorímetros Finamente Segmentados

Calorímetros finamente segmentados são detectores projetados para medir a energia de partículas com alta granularidade espacial [11]. Em experimentos de física de altas energias, como os associados ao LHC, esses detectores são tipicamente organizados em múltiplas camadas longitudinais e podem envolver um número muito grande de canais de leitura individuais [3]. Um exemplo da estrutura de um calorímetro de propósito geral é ilustrado na Figura 1, onde pode-se observar três camadas mais internas para detecção de partículas eletromagnéticas (ECAL 1, 2 e 3) e três mais externas, para detecção de partículas hadrônicas (HCAL 1, 2 e 3).

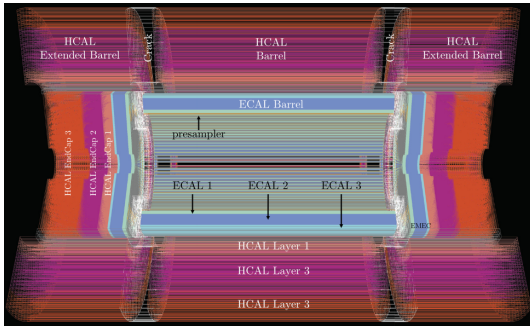


Fig. 1. Vista em corte transversal do calorímetro de uso geral utilizado para ilustrar um exemplo de simulação com o *framework* Lorenzetti Showers. [7]

Cada célula do calorímetro está associada a um canal de leitura composto por material absorvedor, meio ativo e eletrônica de leitura. Quando uma partícula atravessa o calorímetro, ela deposita energia no meio ativo, gerando um pulso elétrico amostrado pela eletrônica em múltiplos pontos temporais, tipicamente cinco amostras espaçadas de 25 ns em detectores do LHC. Esse pulso é então processado por um filtro de forma ótima (*optimal filter*) para estimar a amplitude e o tempo do sinal [12]. Um diagrama simplificado das etapas da eletrônica de leitura de um calorímetro é apresentado na Figura 2. Defeitos na eletrônica de leitura, como módulos inativos ou ruído excessivo, degradam essa estimativa e comprometem a qualidade dos dados coletados, motivando o desenvolvimento de técnicas automáticas de DQM.

O Lorenzetti Showers é um *framework* de simulação de calorímetro desenvolvido para geração de dados sintéticos com controle sobre as condições de simulação [7]. Durante a geração dos dados, na etapa de digitalização, é possível injetar anomalias controlando o parâmetro *noiseFactor*, que multiplica o desvio padrão do ruído padrão por um fator

escalar. O Lorenzetti tem sido utilizado para estudos como a calibração de energia em experimentos de HEP [13], bem como a mitigação do *crosstalk* em calorímetros utilizando redes neurais artificiais [14].

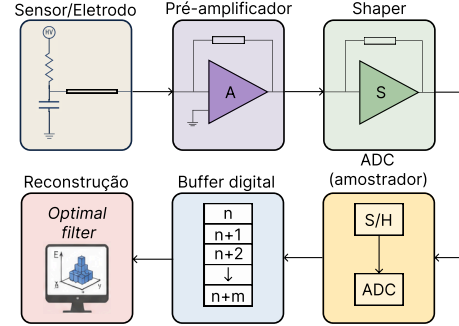


Fig. 2. Diagrama simplificado das etapas da eletrônica de leitura de um calorímetro

B. Variáveis analisadas

A descrição dos candidatos a partículas reconstruídos no calorímetro envolve duas categorias de variáveis. A primeira reúne grandezas cinemáticas, tais como energia total (E), energia transversa (E_T), pseudorapidez (η) e ângulo azimutal (ϕ). A segunda compreende as *shower shapes*, que capturam o perfil espacial e energético do chuveiro a partir da distribuição de energia nas células vizinhas ao candidato [3]:

- R_η : razão entre a energia depositada em uma janela de 3×7 células e a energia em 7×7 células na segunda camada do calorímetro eletromagnético, centrada na posição do *cluster*;
- R_ϕ : razão entre a energia em 3×3 células e a energia em 3×7 células na mesma camada, sensível ao espalhamento lateral na direção ϕ ;
- R_{had} : razão entre a energia transversa no calorímetro hadrônico e a energia transversa do *cluster* eletromagnético, dada por $R_{had} = E_{T,had}/E_T$;
- E_{ratio} : razão entre a diferença e a soma das energias do depósito máximo (E_{maxs1}) e do segundo máximo (E_{2tsts1}) na primeira camada do calorímetro EM, dada por $E_{ratio} = (E_{maxs1} - E_{2tsts1}) / (E_{maxs1} + E_{2tsts1})$;
- $w_{\eta2}$: largura lateral do chuveiro na direção η , calculada em uma janela de 3×5 células por meio da soma ponderada em energia:

$$w_{\eta2} = \sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left(\frac{\sum_i E_i \eta_i}{\sum_i E_i} \right)^2}, \quad (1)$$

onde E_i é a energia da i -ésima célula.

C. Pileup

O *pileup* refere-se à sobreposição de múltiplas interações próton-próton no mesmo cruzamento de feixes. À medida que a luminosidade do LHC aumenta, o número médio de interações por cruzamento $\langle \mu \rangle$ cresce, resultando em sinais adicionais que se sobrepõem ao evento de interesse. O *pileup* degrada as variáveis de calorimetria, pois a energia de interações secundárias contamina a medida do evento principal [15]. Neste trabalho, experimentos com $\langle \mu \rangle = 40$ são avaliados para quantificar o impacto do *pileup* na detecção de anomalias.

¹Conforme Portaria CNPq n° 2.664/2026, utilizou-se IA generativa (Claude, Anthropic) para revisão textual. (Conteúdo e resultados são responsabilidade dos autores)

III. METODOLOGIA

A Figura 3 apresenta o fluxo adotado para o desenvolvimento do presente trabalho. As etapas são detalhadas ao longo desta Seção.

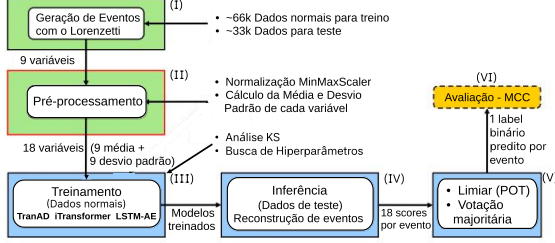


Fig. 3. Fluxograma aplicado no desenvolvimento da pesquisa.

Os dados utilizados neste trabalho foram gerados pelo simulador Lorenzetti na etapa (I). O *dataset* é composto por aproximadamente 100k eventos de jatos QCD por cenário, sendo 33.496 reservados para teste e os demais (~66k eventos normais) para treino. Todos os experimentos são realizados em duas condições de *pileup*: sem *pileup* e com *pileup* médio ($\mu = 40$), totalizando 13 cenários.

Dois tipos de anomalias são injetadas na digitalização dos dados de teste:

- Módulos mortos: células com *noiseFactor* = 0, simulando módulos eletrônicos completamente inativos que zeram o sinal de saída para todas as células afetadas, com ~29% de anomalias no conjunto de teste. Essa anomalia extrema é sinalizada de forma confiável no DQM, tornando-a um caso de teste adequado para validar algoritmos TSAD.
- Ruído aumentado: *noiseFactor* $\in \{5, 10, 20, 50\}$, multiplicando o desvio padrão do ruído eletrônico gaussiano. Três regimes de taxa de anomalia são avaliados: ~0,32%, 1,03% e 22%.

Na etapa (II) são calculadas a média (*mean*) e desvio padrão (std) de cada variável, totalizando 18 variáveis que compõem o *dataset* de treino e teste. Os dados são normalizados com MinMaxScaler antes do treinamento.

Na etapa (III), três modelos de TSAD não supervisionados são avaliados:

- TranAD [8]: modelo baseado em *transformer* com treinamento adversarial e auto-condicionamento, projetado para amplificar o erro de reconstrução em anomalias sutis.
- iTransformer [9]: *transformer* com *embedding* invertido que aplica o mecanismo de atenção sobre as variáveis em vez da dimensão temporal, capturando correlações multivariadas.
- LSTM-AE [10]: autoencoder baseado em LSTM que combina a capacidade de captura de dependências temporais das redes recorrentes com a compressão de representações dos autoencoders.

Os modelos são treinados de forma não supervisionada para reconstruir eventos normais, utilizando apenas dados sem anomalias no treinamento. As configurações de treinamento de cada modelo são detalhadas na Tabela I. Todos os modelos utilizam janela deslizante de tamanho $W = 10$ com passo

$S = 1$, otimizador AdamW com *weight decay* de 10^{-5} , *scheduler* StepLR (passo = 5, $\gamma = 0,9$) e são treinados por 10 épocas. Uma busca de hiperparâmetros é realizada para o iTransformer, por ser o modelo mais sensível aos mesmos, com $W \in \{10, 50, 100\}$ e tamanho interno do modelo $D \in \{2, 10, 20\}$ com o objetivo de investigar configurações que possam otimizar os resultados para todas ou alguma classe de anomalias em específico.

TABELA I
CONFIGURAÇÕES DE TREINAMENTO DOS MODELOS.

Parâmetro	TranAD	iTransf.	LSTM-AE
<i>Learning rate</i>	0,0001	0,0001	0,002
Ativação	GELU	GELU	Tanh/Sigmoid
<i>Batch size</i>	128	32	100
Função de perda	MSE adv.	MSE	MSE

Na inferência, etapa (IV), cada modelo treinado é aplicado aos dados de teste, produzindo um *score* de anomalia para cada uma das 18 características por evento. Para o iTransformer e o LSTM-AE, esse *score* corresponde ao MSE entre a entrada e a saída reconstruída por dimensão. Para o TranAD, corresponde ao erro adversarial com amplificação progressiva [8].

Na etapa (V), os 18 *scores* por evento são convertidos em *labels* binários. O limiar de detecção é determinado automaticamente para cada dimensão pelo método *Peak-over-Threshold* (POT) [16], baseado na Teoria dos Valores Extremos, com nível de risco $q = 10^{-5}$. Os 18 *labels* resultantes são combinados por votação majoritária (*local majority voting*): um evento é declarado anômalo se pelo menos $N/2$ dimensões forem classificadas como anômalas.

O desempenho é avaliado na etapa (VI) pelo Coeficiente de Correlação de Matthews (MCC), definido como:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2)$$

onde TP , TN , FP e FN são os verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, respectivamente. O MCC é preferível ao F1-score em problemas com classes desbalanceadas, pois pondera igualmente positivos e negativos. Cada experimento é repetido com 3 sementes aleatórias distintas, reportando-se a média $\pm$ desvio padrão.

Foi realizada uma investigação do poder discriminativo de cada *shower shape* por meio do *score* Kolmogorov-Smirnov (KS) para analisar a separabilidade de cada característica entre as classes anômala e normal. Em seguida, as 9 características de maior KS entre as 18 foram repassadas à etapa (III), sendo os três modelos de TSAD reavaliados com as mesmas configurações no mesmo *dataset*, para verificar se haveria manutenção ou melhoria dos resultados, justificando uma simplificação dos métodos aplicados.

O custo computacional também foi considerado na seleção dos modelos. Em uma GPU NVIDIA RTX 4060, o treinamento por cenário (10 épocas) levou 1,9 min para o TranAD, 2,0 min para o iTransformer e 0,6 min para o LSTM-AE. O modelo USAD, avaliado por Boggia e Malaescu [5], não

explora eficientemente o paralelismo da GPU, demandando 26 min por cenário, e foi por isso descartado. A implementação está disponível publicamente².

IV. RESULTADOS E DISCUSSÃO

A Tabela II apresenta os resultados do MCC para os 13 cenários avaliados, o melhor resultado por linha está destacado em negrito. A nomenclatura dos cenários obedece o seguinte padrão: wa representa dados com anomalia, sigla do inglês *with anomaly*; wp40 indica a presença de *pileup* ($\mu = 40$); o valor após o símbolo $\times$ representa o nível do *noiseFactor*; o termo percentual apresentado entre parênteses representa a taxa de anomalias presente no cenário. O iTransformer otimizado foi avaliado nos cenários de ruído com maior taxa de anomalia, onde a configuração padrão apresentou maior margem para melhoria.

TABELA II
MCC OBTIDO PARA OS 13 CENÁRIOS AVALIADOS.

Tipo	Cenário	TranAD	iTransf. (pad.)	iTransf. (otim.)	LSTM-AE
Módulos	<i>Dead cells</i> (~29%)	0,775	0,566	0,566	0,646
Ruído sem <i>pileup</i>	wa $\times$ 5 (0,32%)	0,008	0,000	—	0,000
	wa $\times$ 10 (0,32%)	0,010	0,000	—	0,000
	wa $\times$ 20 (0,32%)	0,012	0,000	—	0,013
	wa $\times$ 50 (1,03%)	0,076	0,037	—	0,050
	wa $\times$ 20 (1,03%)	0,008	0,014	0,030	0,006
	wa $\times$ 20 (22%)	0,226	0,041	0,105	0,240
Ruído com <i>pileup</i> ($\mu = 40$)	wp40 wa $\times$ 5 (0,32%)	0,006	0,025	—	0,007
	wp40 wa $\times$ 10 (0,32%)	0,006	0,000	—	0,000
	wp40 wa $\times$ 20 (0,32%)	0,006	0,000	—	0,000
	wp40 wa $\times$ 50 (1,03%)	0,004	0,008	—	0,002
	wp40 wa $\times$ 20 (1,03%)	0,002	0,012	0,009	0,009
	wp40 wa $\times$ 20 (22%)	0,204	0,034	0,157	0,205

A Figura 4 apresenta o MCC médio $\pm$ desvio padrão das 3 sementes para os quatro cenários principais. O TranAD obteve o melhor desempenho para módulos mortos ($0,775 \pm 0,009$), seguido pelo LSTM-AE ($0,646 \pm 0,020$) e pelo iTransformer ($0,566 \pm 0,010$). Esses resultados indicam que todos os modelos testados conseguem detectar corretamente parte das anomalias de módulos mortos. Para anomalias de ruído aumentado com taxa de 22%, o LSTM-AE supera trabalhos anteriores [5] em 29% ($0,240 \pm 0,008$ vs. $0,186 \pm 0,026$), enquanto o iTransformer com configuração padrão obtém apenas $0,041 \pm 0,001$.

Para o cenário de 1% de anomalias de ruído aumentado, todos os modelos apresentam MCC próximo de zero, confirmando que essa configuração é indetectável com *shower shapes*. Esse resultado condiz com os resultados obtidos por [5] e reforça a motivação para abordagens baseadas em pulsos eletrônicos brutos como trabalho futuro.

A busca de hiperparâmetros do iTransformer revelou que a configuração padrão ($W = 10$, $D = 2$) já é ótima para módulos mortos. No entanto, para anomalias de ruído aumentado, configurações maiores melhoram significativamente o desempenho: $W = 50$, $D = 20$ para o cenário sem *pileup* eleva o MCC de 0,041 para 0,105 (+157%); $W = 10$,

$D = 20$ com *pileup* eleva de 0,034 para 0,157 (+362%). Esses resultados indicam que o iTransformer com configuração padrão não aproveita totalmente seu potencial para anomalias de ruído, sendo sensível à escolha dos hiperparâmetros.

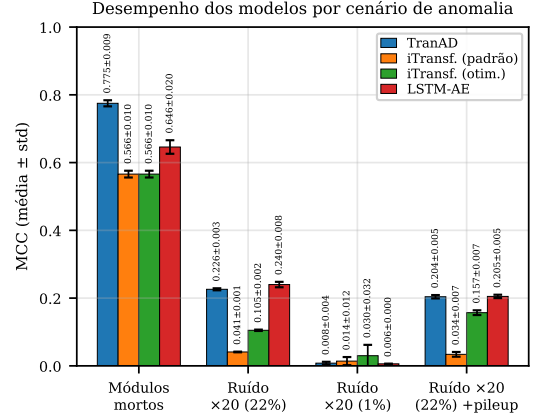


Fig. 4. MCC médio ($\pm$ std, 3 sementes) por modelo e cenário. iTransf. (otim.) refere-se ao iTransformer com hiperparâmetros otimizados por busca exaustiva.

Comparando os cenários com e sem *pileup* ($\langle \mu \rangle = 40$) para o cenário de ruído aumentado com 22% de anomalias, observa-se uma degradação moderada: TranAD de 0,226 para 0,204 (−9,7%) e LSTM-AE de 0,240 para 0,205 (−14,6%), conforme apresentado na Figura 5. O *pileup* não altera o ranking dos modelos, mas reduz consistentemente o MCC em todos os cenários de ruído aumentado, sendo o iTransformer o mais afetado. Esse resultado é fisicamente esperado, pois o *pileup* adiciona sobreposição de interações próton-próton que degrada as variáveis de calorimetria [15].

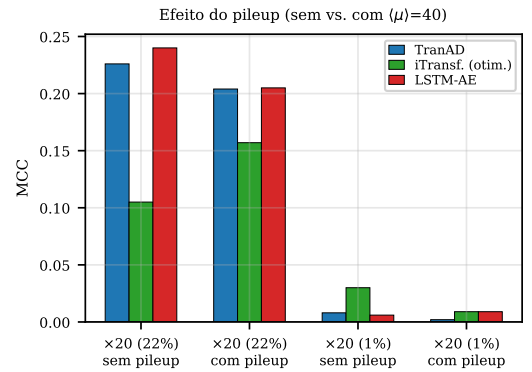


Fig. 5. Efeito do *pileup* no MCC médio por modelo para cenários com taxas de anomalia 22% e 1% com *noiseFactor* $\times 20$.

A análise de separabilidade via fator KS revelou que o poder discriminativo das características depende do tipo de anomalia, conforme ilustrado na Figura 6.

Para módulos mortos, R_{had} , ϕ_{std} e e_{mean} são as mais discriminativas ($KS \approx 0,26$), refletindo o impacto físico do módulo morto na fração hadrônica e na energia total do *cluster*. Para ruído aumentado, ϕ_{std} , ϕ_{mean} e $R_{\eta, std}$ são mais significativos, com valores KS menores ($\approx 0,17$). Em ambos os casos, os desvios padrão são consistentemente mais discriminativos que as médias correspondentes.

²Repositório: github.com/EliasGMBS/tsad-calorimetro

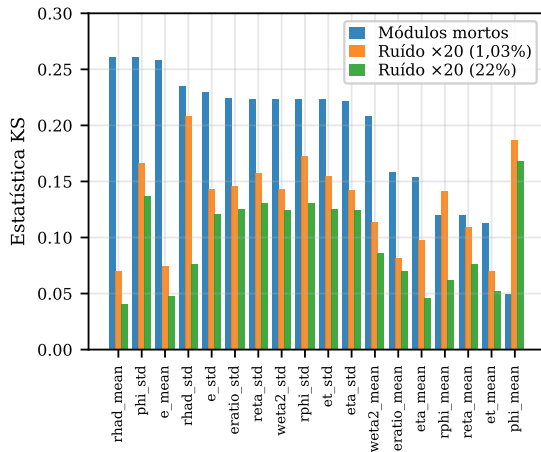


Fig. 6. Separabilidade KS das 18 características por tipo de anomalia. O ranking muda conforme o tipo de defeito.

A substituição das 18 características pelas 9 características de maior KS produziu resultados mistos. Para módulos mortos, o MCC deteriora em todos os modelos (TranAD: $-0,144$; LSTM-AE: $-0,227$), indicando que as características de menor KS ainda carregam informação complementar útil para a detecção de anomalias. Para ruído aumentado, o TranAD melhora com seleção ($+0,060$ para 22%), enquanto o LSTM-AE piora ($-0,115$). Esses resultados, apresentados na Figura 7, demonstram que não existe uma estratégia universal de seleção de características para TSAD em calorímetros; os resultados dependem tanto do modelo quanto do tipo de anomalia.

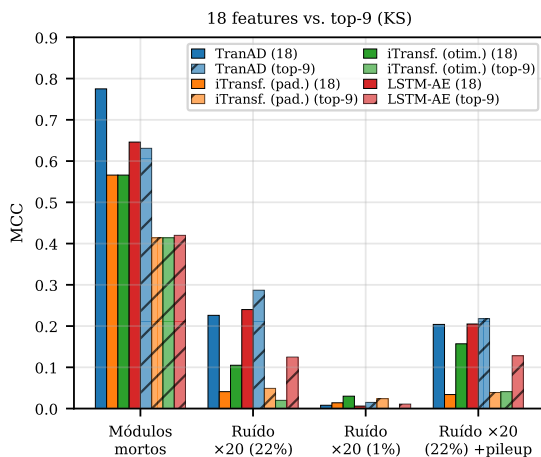


Fig. 7. Efeito da seleção das características no MCC para quatro cenários.

V. CONCLUSÕES

Este trabalho apresentou uma avaliação de três modelos de TSAD: TranAD, iTransformer e LSTM-AE, para monitoramento de qualidade de dados de calorímetros finamente segmentados, em 13 cenários com variação de tipo de anomalia, intensidade de ruído, taxa de anomalias e condições de *pileup*.

O TranAD mostrou-se o modelo mais adequado para detecção de módulos mortos (MCC $0,775 \pm 0,009$). O LSTM-AE, não avaliado em trabalhos anteriores nesse contexto, revelou-se competitivo para anomalias de ruído aumentado. O iTransformer com hiperparâmetros otimizados melhora sub-

tancialmente para anomalias de ruído aumentado, mas ainda é inferior ao TranAD e ao LSTM-AE nos cenários mais difíceis.

Anomalias de ruído aumentado com baixa taxa (1%) permanecem indetectáveis com *shower shapes* para todos os modelos, evidenciando a limitação fundamental da representação baseada em variáveis agregadas por evento. A análise KS confirmou que o poder discriminativo das variáveis depende do tipo de anomalia, e que os desvios padrão são mais informativos que as médias correspondentes.

Como trabalho futuro, pretende-se investigar abordagens baseadas nos pulsos eletrônicos brutos, compostos por 5 amostras por célula, armazenados nos arquivos ESD do Lorenzetti, hipótese central da dissertação em desenvolvimento. Espera-se que a representação em nível de pulso supere as *shower shapes* na detecção de anomalias de ruído aumentado, em especial para baixas taxas de anomalias.

AGRADECIMENTOS

Os autores agradecem ao CNPq e à CAPES (Programa CAPES-COFECUB) pelo apoio financeiro.

REFERÊNCIAS

- [1] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation*, vol. 3, p. S08001, Aug. 2008.
- [2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012.
- [3] ATLAS Collaboration, "The ATLAS experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, p. S08003, 2008.
- [4] ATLAS Collaboration, "Autoencoder-based Time Series Anomaly Detection for ATLAS Liquid Argon Calorimeter Data Quality Monitoring," *SciPost Physics Proceedings*, 2025, arXiv:2512.05977.
- [5] L. Boggia and B. Malaescu, "Synthetic Data Generation with Lorenzetti for Time Series Anomaly Detection in High-Energy Physics Calorimeters," in *SciPost Physics Proceedings*, 2025.
- [6] D. Chicco e G. Jurman, "The advantages of the Matthews Correlation Coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, n. 1, p. 6, 2020.
- [7] M. V. de Araujo et al., "Lorenzetti: A general purpose calorimeter simulation framework," *Computer Physics Communications*, vol. 286, p. 108671, 2023.
- [8] S. Tuli, G. Casale e N. R. Jennings, "TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data," *Proc. VLDB Endow.*, vol. 15, n. 6, pp. 1201–1214, 2022.
- [9] Y. Liu et al., "iTransformer: Inverted Transformers Are Effective for Time Series Forecasting," *ICLR*, 2024.
- [10] R.-J. Hsieh, J. Chou e C.-H. Ho, "Unsupervised Online Anomaly Detection on Multivariate Sensing Time Series Data for Smart Manufacturing," *IEEE SOCA*, pp. 90–97, 2019.
- [11] G. I. Gonçalves, "Estimação de energia para calorímetros finamente segmentados operando em condições severas de empilhamento de sinais," Dissertação de Mestrado, UERJ, 2020.
- [12] M. Delmastro, "Quality factor analysis and optimization of digital filtering signal reconstruction for liquid ionization calorimeters," *NIM-A*, vol. 600, pp. 545–554, 2009.
- [13] P. R. A. da Silva, E. F. Simas Filho, P. C. M. A. Farias, E. E. P. de Souza, J. L. Marin, and J. M. de Seixas, "Machine learning for energy calibration in high energy physics experiments using the Lorenzetti Showers Framework," *IEEE Transactions on Nuclear Science*, 2026, doi: 10.1109/TNS.2026.3686269.
- [14] M. S. Santos, E. F. Simas Filho, P. C. M. A. Farias, B. Laforge, and J. M. Seixas, "Crosstalk mitigation in an ATLAS-like high energy liquid argon calorimeter using artificial neural networks," *NIM-A*, vol. 1087, p. 171382, 2026, doi: 10.1016/j.nima.2026.171382.
- [15] J. L. Marin, E. F. de Simas Filho, E. E. P. de Souza et al., "Conjunto de Redes Neurais Artificiais para Detecção Online de Fótons de Altas Energias," *CBIC*, 2021.
- [16] A. Siffer et al., "Anomaly Detection in Streams with Extreme Value Theory," *Proc. ACM KDD*, pp. 1067–1075, 2017.