

# mHealth: da coleta de dados à regressão de parâmetros clínicos por séries temporais e imagens obtidas por smartphone

Aluna: Isabela Guarnier De Mitri (isademitri@usp.br)

Orientador: Diego Furtado Silva (diegofsilva@icmc.usp.br)

Colaborador: André Luís Mendes Fakhoury (andrefakhoury@usp.br)

Universidade de São Paulo

## Motivação e Objetivos

Aprimorando o uso de Aprendizado de Máquina em mHealth[1], com foco na melhoria do desempenho do algoritmo ROCKET (Random Convolutional Kernel Transform)[2] na análise de séries temporais em saúde, o objetivo deste projeto é reduzir a instabilidade do algoritmo por meio de técnicas de customização e otimização de kernels.

## Métodos e Procedimentos

Neste projeto, utilizamos o ROCKET para transformar séries temporais em um conjunto de características[3], explorando formas de aprimorar a precisão e estabilidade do algoritmo. Avaliamos três abordagens principais: (1) modificamos a distribuição dos kernels, comparando Normal e T-Student para avaliar impacto na acurácia; (2) investigamos como a quantidade e customização de bias e dos kernels influenciam o desempenho; (3) visualização do UMAP (Uniform Manifold Approximation and Projection)[4] para identificar padrões e redundâncias nas features e kernels;

## Resultados e Aplicações

Dataset	NORMAL		TSTUDENT	
	Acurácia	Desvio-padrão	Acurácia	Desvio-padrão
ArrowHead	0.7699428571	0.0219092064	0.7672	0.024045216
DistalPhalanxTW	0.679856115	0.012439587	0.6735251798	0.0139327319
FaceAll	0.7556449704	0.010280119	0.7790532544	0.025914187
FiftyWords	0.721098901098	0.0090263463	0.72632967032	0.00871507739
GunPoint	0.97986666	0.0054693539	0.975466666	0.004937707
Ham	0.7540952380	0.029224160	0.762857142	0.0268858133

Figura 1: Tabela comparativa de acurácia para distribuições Normal e T-Student.

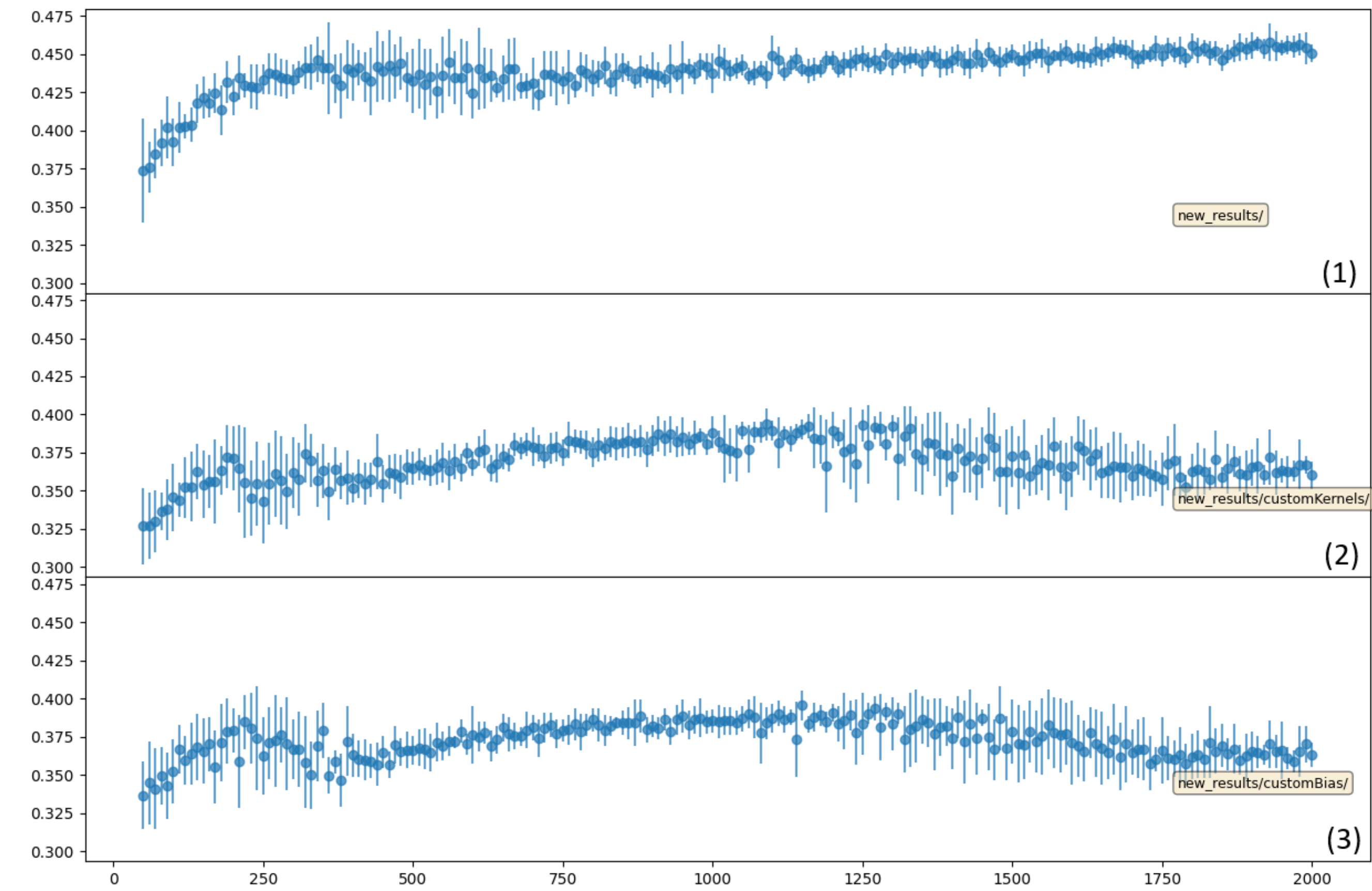


Figura 2: Comparação dos desempenhos em três cenários distintos utilizando o dataset In-lineSkate: (1) com bias customizado, (2) com kernels customizados e (3) com configuração padrão (sem customizações).

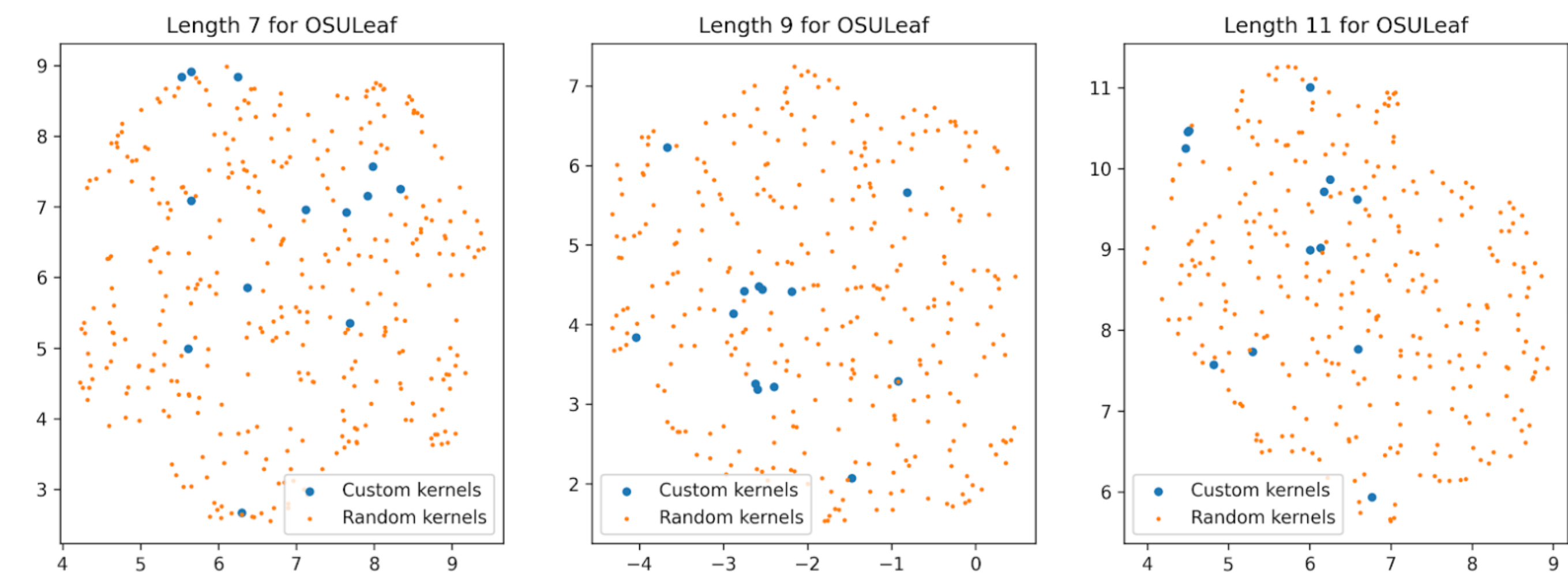


Figura 3: Visualização UMAP dos kernels com comprimentos 7, 9 e 11 aplicados ao dataset OSULeaf.

A modificação da distribuição dos kernels mostrou ter um impacto limitado na melhoria da acurácia geral, o que sugere que a distribuição padrão já era suficientemente eficaz para a maioria dos datasets testados. Observou-se que o aumento do número de kernels atinge um ponto de saturação, além do qual a adição de mais kernels não apenas deixa de melhorar a acurácia, mas também pode prejudicá-la devido à redundância. A customização dos kernels, por sua vez, não resultou em melhorias significativas e, em alguns casos, reduziu a acurácia.

## Conclusão

O algoritmo ROCKET apresenta um grande potencial para a análise de séries temporais em saúde, mas sua natureza aleatória representa um desafio significativo para garantir a estabilidade e a robustez dos resultados em aplicações práticas. Nossos experimentos revelaram que, apesar da customização e otimização dos kernels, as melhorias na acurácia foram limitadas e, em algumas situações, até prejudiciais. A saturação observada ao aumentar o número de kernels indica que, além de um certo ponto, a redundância pode comprometer o desempenho do algoritmo. A investigação de novas abordagens e a exploração de diferentes técnicas de aprendizado de máquina podem ser caminhos promissores para melhorar a aplicação do ROCKET na área da saúde.

## Referências

[1] World Health Organization. mhealth. use of appropriate digital technologies for public health: Report by director-general, 2018. 71st World Health Assembly Provisional Agenda Item, Vol. 12, p. A71.

[2] A. Dempster, F. Petitjean, and G. I. Webb. Rocket: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495, 2020.

[3] M. Middlehurst, J. Large, M. Flynn, J. Lines, A. Bostrom, and A. Bagnall. Hivemote 2.0: A new meta ensemble for time series classification. *Machine Learning*, 110(11):3211–3243, 2021.

[4] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. Unsupervised scalable representation learning for multivariate time series. *arXiv preprint arXiv:1802.03426*, 2018.