

CONCATPIPELINE: UM TOOLBOX PARA ANÁLISE DE DATASETS DE IMAGEAMENTO DE CÁLCIO COM MINI MICROSCÓPIOS

Daniel Almeida Filho^{1,2}; Ludmila Anjos^{1,3}; Ammar Marvi²; Ying Cai²; Megha Sehgal²; Yang Shen²; Alcino J. Silva²

¹ Centro Universitário SENAI CIMATEC; Salvador-BA;

² Neurobiology, Psychiatry and Psychology Departments and Integrative Center for Learning and Memory, University of California Los Angeles, Los Angeles, CA, USA

³ Graduanda em Engenharia Elétrica; Iniciação Científica PIBIC-CIMATEC/CNPq

contato: daniel.almeidaf@ieeb.org.br

RESUMO

O estudo se propõe a aprimorar a análise de registros de imageamento de cálcio em roedores, mediante o desenvolvimento de um conjunto de ferramentas para detecção e acompanhamento de células neuronais em múltiplas sessões. Após a otimização do código e a realização de testes extensivos, embora a literatura indique que a concatenação de sessões aprimora a extração de células, a análise prática ainda carece de uma conclusão estatística definitiva, devido à falta de finalização na avaliação dos resultados do software desenvolvido para a detecção de células pós-concatenação de sessões. Ao final de seu desenvolvimento, o estudo almeja que o algoritmo se estabeleça como uma referência na análise de imageamento de cálcio, ressaltando a importância de sua adoção por instituições de pesquisa interessadas em avanços na compreensão das interfaces cérebro-máquina.

PALAVRAS-CHAVE: miniscope; concatpipeline; imageamento de cálcio; matlab.

1. INTRODUÇÃO

Atualmente, a captura de imagens de cálcio representa o estado mais avançado para monitorar centenas de neurônios individuais em diferentes regiões cerebrais simultaneamente, ao longo de períodos que podem variar de dias a meses, em roedores em comportamento livre. No entanto, a análise desses dados após a gravação ainda está em estágios iniciais de desenvolvimento. Isso se deve, em parte, ao fato de que as técnicas de microscopia de fóton único utilizadas para esses registros apresentam limitações, como baixa resolução espacial e alto nível de ruído. Além disso, há uma escassez de avaliações sistemáticas dos diferentes métodos para acompanhar as células ao longo do tempo ou em várias sessões consecutivas. Por exemplo, ainda não está claro se é mais eficaz detectar células correspondentes em sessões independentes ou concatenar as sessões antes de realizar a detecção das células.

Com o intuito de aprimorar o registro simultâneo de centenas de neurônios em diversas áreas cerebrais e monitorar sua atividade ao longo de períodos prolongados em roedores em livre comportamento, este estudo tem como objetivo principal desenvolver um conjunto de ferramentas para análise de dados obtidos por meio de imageamento de cálcio, utilizando mini microscópios (miniscopes) de fóton único. Nesse contexto, será analisada a eficácia das técnicas atualmente empregadas na análise desses dados, mediante a leitura de trabalhos antecessores, e será realizada uma análise direta dos dados brutos provenientes dos mini microscópios.

2. METODOLOGIA

Durante a condução deste estudo, uma metodologia detalhada foi seguida para garantir a qualidade e eficácia do algoritmo desenvolvido para análise de registros de imageamento de cálcio por mini microscópios.

Primeiramente, foi dedicado um tempo para analisar e otimizar o código já existente, disponível em <https://github.com/Almeida-FilhoDG/ConcatMiniscope>, com o objetivo de tornar os processos mais eficientes computacionalmente, sem comprometer a qualidade dos resultados. Além disso, o código foi revisado para torná-lo mais acessível e utilizável pela comunidade de pesquisa, considerando que a maioria dos usuários do toolbox possui baixa experiência em programação.

Posteriormente, testes abrangentes foram conduzidos para validar a qualidade e funcionalidade do algoritmo. Esses testes foram realizados em diferentes conjuntos de dados, provenientes de várias espécies, registrados em diferentes equipamentos e regiões cerebrais. As capacidades do algoritmo em

detectar células e acompanhar sua atividade ao longo de múltiplas sessões foram avaliadas, considerando registros de diferentes tamanhos, formatos e qualidades.

Utilizando os dados provenientes de diferentes regiões cerebrais foram realizadas análises da veracidade das células detectadas, por meio do método da matriz de confusão, que consiste em uma forma de comparação de dados analisados por duas pessoas distintas, com o intuito de identificar quais células são em consenso consideradas verdadeiras ou falsas.

Para complementar a análise, o método proposto, que envolve a concatenação de sessões, foi comparado com outros métodos de análise de sessões individuais e definição de correspondência entre células, utilizando como referência trabalhos anteriores.

3. RESULTADOS E DISCUSSÃO

A tarefa desafiadora atual na área da neurociência consiste na detecção e extração automática de células e sua atividade a partir de registros de indicadores fluorescentes de cálcio obtidos através de dois fótons ou um fóton.¹ Além disso, para analisar dinamicamente a plasticidade em redes neurais, que são capazes de armazenar novas informações codificadas, são necessárias gravações de longo prazo, que se estendem por várias sessões de horas a semanas. Nesse contexto, a habilidade de rastrear a atividade de células específicas em diferentes sessões torna-se essencial.² Algumas evidências recentes indicam que, ao invés de extrair células de uma única sessão e tentar alinhá-las para comparar entre sessões, a abordagem de concatenar sessões em um único vídeo de longa duração poderia ser uma alternativa eficaz para melhorar a extração de células e monitorar sua atividade ao longo do tempo.

De forma mais específica, a concatenação de sessões pode aumentar tanto a quantidade quanto a qualidade na extração de células,³ proporcionando também a capacidade de rastrear essas mesmas células em sessões distintas. Isso é fundamental para analisar a atividade da rede neural e a plasticidade associada a novas experiências comportamentais, aprendizado e memória.⁴ É válido destacar que recentemente foi observado que algumas células podem ser identificadas apenas em um subconjunto das sessões gravadas, sugerindo que essas células permanecem silenciosas durante as sessões em que não são detectadas.⁵ Este aspecto é especialmente relevante ao analisar células que são compartilhadas entre diferentes sessões de codificação, pois essa observação é crucial para entender o padrão de conexão entre memórias ao explorar diversos contextos em intervalos de tempo curtos e para a teoria da alocação neuronal.⁶

No presente estudo, o software Matlab foi empregado para desenvolver e aprimorar um pipeline dedicado à detecção de células neuronais em um mamífero específico. Durante a condução do projeto, foram identificadas novas oportunidades de otimização por meio da análise dos códigos, além da realização de testes para validar a qualidade da detecção das células identificadas pela ferramenta. Esta última foi desenvolvida utilizando o método de detecção após a concatenação das sessões.

Na fase de processamento inteligente conduzida pelo software desenvolvido, áreas contendo células autênticas foram identificadas, bem como regiões contendo ruídos que, apesar de apresentarem características semelhantes às células, não correspondiam a células reais. Estas regiões foram computacionalmente classificadas como células verdadeiras e falsas, respectivamente. Para avaliar a qualidade da detecção, alguns pesquisadores deste experimento realizaram uma inspeção manual das células classificadas como falsas e verdadeiras, utilizando critérios pré-estabelecidos e objetivos.

Com o objetivo de avaliar o desempenho do modelo de classificação, utilizou-se a matriz de confusão como ferramenta para realizar uma comparação estatística da concordância nas classificações de células verdadeiras e falsas entre dois pesquisadores. Essa abordagem visa preparar o terreno para estabelecer um ground truth no futuro, ao introduzir um terceiro pesquisador. O método do *ground truth* envolve a determinação de um consenso sobre a veracidade de um conjunto de células através da análise de três pesquisadores. Esta abordagem possibilitará estimar o desempenho do modelo de detecção na identificação de células verdadeiras.

A matriz de confusão proporciona uma comparação entre a classificação real de uma célula e a sua classificação predita.⁷ Para utilizar este método, é necessário que um dos pesquisadores seja definido como o detentor do conjunto real de dados contendo células falsas e verdadeiras, enquanto o outro realiza as predições. A partir dessas informações, foram extraídos os resultados apresentados na figura 1.

Matrizes de confusão desenvolvidas pela comparação do conjunto de dados biológicos celulares de dois pesquisadores

Matriz de Confusão Animal 1



Matriz de Confusão Animal 2



A diagonal principal, da esquerda para a direita, representa as células que foram consensualmente classificadas como verdadeiramente positivas e negativas, respectivamente. Para o animal 1, houve 100% de concordância entre os pesquisadores quanto à veracidade das células analisadas; no entanto, para o animal 2, houve uma variação mínima de 0,3226% na classificação. Em relação às células classificadas como verdadeiramente negativas, houve uma disparidade maior entre ambos, com uma proximidade percentual de 72,2222% e 67,033% para os animais 1 e 2, respectivamente. A análise dos resultados sugere uma relevante concordância na categorização feita pelos pesquisadores, o que representa um bom resultado preliminar a ser explorado futuramente para a obtenção do *ground truth*.

4. CONSIDERAÇÕES FINAIS

Antecipa-se que o algoritmo apresentado neste estudo se estabeleça como uma referência na análise de registros de imageamento de cálcio obtidos por mini microscópios. A transferência dessa tecnologia da Universidade da Califórnia Los Angeles (UCLA) para o SENAI-CIMATEC está em curso como uma das potenciais plataformas para futuras Interfaces Cérebro-Máquina a médio/longo prazo, visando torná-la acessível aos laboratórios de pesquisa no Brasil e América Latina a curto prazo. É fundamental que o SENAI-CIMATEC assuma um papel de liderança nesse campo tecnológico com grande potencial para o futuro.

5. REFERÊNCIAS

- Stringer et al (2019). Computational processing of neural recordings from calcium imaging data This review comes from a themed issue on Machine Learning, Big Data, and Neuroscience. *Current Opinion in Neurobiology*, 55, 22–31. <https://doi.org/10.1016/j.conb.2018.11.005>
- Sheintuch et al (2017). Tracking the Same Neurons across Multiple Days in Ca²⁺ Imaging Data. *Cell Reports*, 21(4), 1102–1115. <https://doi.org/10.1016/j.celrep.2017.10.013>
- Gonzalez et al (2019). Persistence of neuronal representations through time and damage in the hippocampus. *Science*, 365(6455), 821–825. <https://doi.org/10.1126/SCIENCE.AAV9199>
- Ghandour et al (2019). Orchestrated ensemble activities constitute a hippocampal memory engram. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-10683-2>
- Kumar et al (2020). Sparse Activity of Hippocampal Adult-Born Neurons during REM Sleep Is Necessary for Memory Consolidation. *Neuron*. <https://doi.org/10.1016/j.neuron.2020.05.008>
- Sehgal et al (2018). Memory allocation mechanisms underlie memory linking across time. *Neurobiology of Learning and Memory*, 153, 21–25. <https://doi.org/10.1016/j.nlm.2018.02.021>
- Monard, M. C.; Baranauskas, J. A. Conceitos sobre Aprendizado de Máquina. Disponível em: <https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>. Acesso em: 17 de Mar. 2018