

ESTIMADOR KERNEL DA DENSIDADE NA ANÁLISE DE AMOSTRAS DE UM FLUORÍMETRO PORTÁTIL

Randson Davi Mendes Santos¹; Lucca Coelho Muniz²; Bruno Arvan³; Marcos Vinicius Sousa da Silva⁴; Valmara Silveira Ponte⁵; Vinicius Pinto Rocha⁶; Cleber Vinicius Ribeiro de Almeida⁷; Milena Botelho Pereira Soares⁸; Valéria Loureiro da Silva⁹.

¹Graduado em Física; Bolsista INOVA-FIOCRUZ; davissantosii@gmail.com

²Graduando em Engenharia Elétrica; Bolsista INOVA-FIOCRUZ; lmunizcoelho@gmail.com

³Graduando em Engenharia Elétrica; Bolsista INOVA-FIOCRUZ; arvanbruno@gmail.com

⁴Graduado em Engenharia Elétrica; Bolsista INOVA-FIOCRUZ; marvin.sou@gmail.com

⁵Mestre em Mecatrônica; Centro Universitário SENAI CIMATEC, Salvador-BA; valmara.ponte@fieb.org.br

⁶Doutor em Patologia Humana e Experimental; Centro Universitário SENAI CIMATEC, Salvador-BA; vinicius.rocha@fieb.org.br

⁷Mestre em Engenharia Elétrica; Centro Universitário SENAI CIMATEC, Salvador-BA; cleber.vinicius@fieb.org.br

⁸Doutora em Biofísica; Centro Universitário SENAI CIMATEC, Salvador-BA; milenabpsoares@gmail.com

⁹Doutora em Física; Centro Universitário SENAI CIMATEC, Salvador-BA; valeria.dasilva@fieb.org.br

RESUMO

A análise de fluorescência possui inúmeras aplicações em dispositivos da área de saúde, é o caso do teste RT-qPCR que detecta o SARS-COV-2. Buscando sanar o problema da mobilidade de aparelhos que realizam esse teste, foi desenvolvido um fluorímetro portátil para detectar e mensurar a fluorescência da reação PCR. Para otimizar o aparelho de bancada, foram realizados testes para quatro grupos de amostras: com a presença do gene humano, com água em substituição do material genético, com SARS-COV-2 e na ausência do SARS-COV-2 buscando reproduzir os resultados encontrados no teste padrão ouro. Os valores permitiram a construção de curvas de densidade de probabilidade através do estimador kernel da densidade. A análise permitiu selecionar a melhor corrente de operação do aparelho de bancada e do futuro protótipo, encontrando o valor ideal para 0,26 A. Nas próximas etapas será construída a matriz de confusão, buscando estabelecer os limites de operação do projeto.

PALAVRAS-CHAVE: SARS-COV-2; Estimador Kernel da Densidade; PCR.

1. INTRODUÇÃO

A luminescência é a emissão de luz por qualquer substância, ocorrendo a partir de estados eletrônicos excitados. Ela pode ser dividida em duas categorias, a fluorescência e a fosforescência¹. Essencialmente, na fluorescência, a fonte de luz de um determinado comprimento de onda ($\lambda_{excitação}$) excita o material fluorescente, os átomos do mesmo, por sua vez, emitem luz num comprimento de onda maior ($\lambda_{emissão}$). Esse fenômeno permite a utilização dos materiais fluorescentes, também chamados de fluoróforos, como marcadores para a presença de determinadas substâncias nas áreas das ciências biológicas¹.

Na pandemia de SARS-COV-2 uma dessas aplicações acabou assumindo um papel de destaque, o uso para diagnóstico do vírus através do método RT-qPCR (Reverse Transcription – quantitative Polymerase Chain Reaction). Na RT-qPCR ocorre uma reação na qual o material genético alvo é amplificado e a detecção é feita através da fluorescência de um fluoróforo ligado ao material genético². Nesse caso, uma fonte de luz é utilizada para excitar o fluoróforo e com a detecção da fluorescência emitida é possível determinar a presença do vírus numa amostra. O principal problema desse tipo de análise é seu alto custo, necessidade de equipamentos especializados e pessoal treinado³.

Buscando resolver essa problemática foi proposto em parceria com a Fiocruz o desenvolvimento de um aparelho que pudesse mensurar a fluorescência e que fosse portátil, além de possuir um baixo custo, para leitura de técnicas de PCR alternativas mais rápidas e fáceis de executar, como a técnica RT-RPA (Transcriptase Reverse-recombinase Polymerase Amplification), facilitando o acesso a esse tipo de teste mais confiável. Durante a construção do experimento de bancada foram coletados dados que pudessem validá-lo, e esses dados precisaram ser tratados com ferramentas estatísticas. Uma das ferramentas adotadas foi o Estimador Kernel da Densidade ou KDE.

O KDE permite através de um banco de dados com amostras aleatórias construir uma função de probabilidade de densidade⁴. Com essa função é possível construir uma curva e assim identificar melhor como os valores mensurados estão distribuídos e até mesmo comparar os valores obtidos para tipos diferentes de amostras como é o caso das analisadas no projeto. Foi utilizado em outras situações como, por exemplo, no diagnóstico de câncer⁵.

2. METODOLOGIA

A otimização do fluorímetro desenvolvido foi feita através da comparação do resultado medido no fluorímetro desenvolvido com o obtido num equipamento de RT-qPCR, modelo QuantStudio 1™ da Thermo Fisher, localizado no ISIS - SAS do CIMATEC, permitindo ter um valor de referência para comparação.

As análises ocorreram no período de um mês. Foram medidos os valores de luz produzidas pelas amostras em ADC Counts, unidade de medida adotada pelo sensor óptico. Detectaram-se três comprimentos de ondas diferentes, o ciano, que representa boa parte da fluorescência produzida, o verde e o amarelo. Cada uma dessas análises foi feita para um valor de corrente diferente (0,13 A; 0,26 A; 0,39 A; 0,52 A; 0,65 A) de alimentação de um LED de bancada (435 nm). Ao todo foram 120 amostras, sendo 47 positivas de controle (presença do material genético humano), 25 negativas de controle (são os brancos, não possuíam material genético, e sim água no lugar) e 25 positivas virais e 23 negativas virais, amostras que no teste padrão ouro RT-qPCR deram positiva ou negativa para a presença de SARS-Cov-2.

A análise KDE foi realizada para o conjunto de amostras restantes após remoção dos outliers (106 amostras) identificados por análise de Boxplot⁶. Para essa análise foi utilizado o Python como linguagem principal de programação⁷ já que possui algumas bibliotecas interessantes para a tarefa. Foram utilizadas três bibliotecas diferentes, a Pandas⁸, para manipulação e importação de dados, a Seaborn⁹, para estatística e visualização e, por fim, a Matplotlib¹⁰, para visualização e plotagem dos dados.

Para utilizar essas bibliotecas foi necessário do preâmbulo importa-las:

```
import pandas as pds
import matplotlib.pyplot as plt
import seaborn as sns
```

A leitura dos dados através do Pandas:

```
posit = pds.read_excel(r"D:\GD\Projeto Fiocruz\Bancos de Dados\Corrente_Positivo.xlsx")
negat = pds.read_excel(r"D:\GD\Projeto Fiocruz\Bancos de Dados\Corrente_Negativo.xlsx")
virus = pds.read_excel(r"D:\GD\Projeto Fiocruz\Bancos de Dados\Corrente_Viral.xlsx")
nvirus = pds.read_excel(r"D:\GD\Projeto Fiocruz\Bancos de Dados\Corrente_NViral.xlsx")
```

Foram plotados os gráficos de KDE confrontando amostras positivas e negativas de controle e de vírus, para as três cores, como no trecho:

```
fig, controle_ciano = plt.subplots()
controle_ciano = sns.kdeplot(posit["Ciano"], color = "green", shade = True)
controle_ciano = sns.kdeplot(negat["Ciano"], color = "red", shade = True)
controle_ciano.set(title = "Positivo e Negativo de Controle (KDE) - Ciano", xlabel = "Intensidade Luminosa (ADC Counts)", ylabel = "Densidade")
```

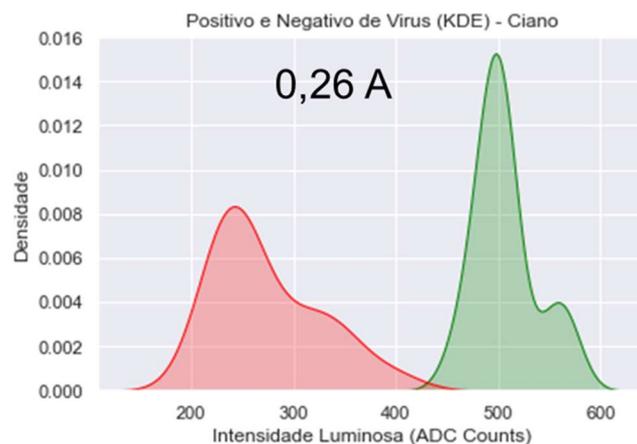
Por fim, validamos a curva do KDE sobrepondo-a ao histograma das amostras:

```
fig, ax1 = plt.subplots()
ax1.set(title = "Histograma e KDE - Ciano - Controle Positivas", xlabel = "Intensidade Luminosa (ADC Counts)", ylabel = "Densidade")
sns.kdeplot(posit["Ciano"], color = "cyan", ax = ax1)
ax2 = ax1.twinx()
sns.histplot(posit["Ciano"], color = "black")
ax2.set(ylabel = "Quantidade")
```

3. RESULTADOS E DISCUSSÃO

Através das análises das curvas de densidade obtidas pelo KDE foi possível comparar os resultados encontrados para as correntes diferentes que alimentavam o LED. A melhor corrente seria aquela que apresentasse melhor separação entre as curvas de densidade para as amostras positivas e negativas de controle, e positivas e negativas de vírus.

A imagem ao lado é um exemplo de resultado obtido para o ciano no caso em que comparamos amostras positivas e negativas para a presença de Sars-Cov-2. Esse resultado se repete nos outros casos e, por isso, a melhor corrente foi a de 0,26 A. Através dos gráficos foi possível determinar que apenas o ciano e o verde são eficientes para análise que será realizada no protótipo final.



4. CONSIDERAÇÕES FINAIS

Através dos dados avaliados foi possível determinar a corrente a ser utilizada na fonte de excitação, ou seja, no LED, que foi de 0,26 A. Foi possível ainda, determinar as cores com as quais é possível distinguir as amostras de tipos diferentes, no caso o ciano e o verde. A análise exposta nesse trabalho servirá de base para o desenvolvimento da matriz de confusão, o que permitirá determinar fatores como sensibilidade e sensibilidade do aparelho de bancada. Além disso, válida todos os dados obtidos para o banco de dados.

5. REFERÊNCIAS

¹ LAKOWICZ, Joseph R. **Principles of Fluorescence Spectroscopy**. New York: Kluwer Academic/Plenum Publishers, 1999.

² CHA, Deokjea et al. **Point-of-care diagnostic (POCD) method for detecting Bursaphelenchus xylophilus in pinewood using recombinase polymerase amplification (RPA) with the portable optical isothermal device (POID)**. Plos one, v. 15, n. 1, p. e0227476, 2020.

³ BELGRADER, Phillip et al. **A battery-powered notebook thermal cycler for rapid multiplex real-time PCR analysis**. Analytical chemistry, v. 73, n. 2, p. 286-289, 2001.

⁴ SCOTT, David W. **Multivariate density estimation: theory, practice, and visualization**. John Wiley & Sons, 2015.

⁵ SHEIKHPOUR, Razieh; SARRAM, Mehdi Agha; SHEIKHPOUR, Robab. **Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer**. Applied Soft Computing, v. 40, p. 113-131, 2016.

⁶ TUKEY, John W. et al. **Exploratory data analysis**. 1977.

⁷ PYTHON SOFTWARE FOUNDATION. **Python**, 2001. Site da linguagem de programação Python. Disponível em: <<https://www.python.org/>>. Acesso em: fevereiro de 2021.

⁸ NUMFOCUS. **Pandas**, 2008. Site sobre a biblioteca Pandas utilizada para importar e manipular dados. Disponível em: <<https://pandas.pydata.org/>>. Acesso em: fevereiro de 2021.

⁹ MICHAEL WASKOM. **Seaborn: statistical data visualization**, 2012. Site sobre a biblioteca Seaborn, utilizada para tratar os dados. Disponível em: <<https://seaborn.pydata.org/index.html>>. Acesso em: fevereiro de 2021.

¹⁰ MATPLOTLIB DEVELOPMENT TEAM. **Matplotlib**, 2002. Site sobre a biblioteca Matplotlib utilizada para plotar dados. Disponível em: <<https://matplotlib.org/>>. Acesso em: fevereiro de 2021.