



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE TECNOLOGIA**  
**DEPARTAMENTO DE ENGENHARIA ELÉTRICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**  
**MESTRADO ACADÊMICO EM ENGENHARIA**

**BRUNO LUIZ FAUSTINO**

**ESTIMULAÇÃO CEREBRAL PROFUNDA ADAPTATIVA BASEADA EM  
APRENDIZADO POR REFORÇO: UM MODELO COMPUTACIONAL DE  
NEURÔNIOS DE IZHIKEVICH PARA SIMULAÇÃO DO TRATAMENTO DOS  
TREMORES DA DOENÇA DE PARKINSON**

**FORTALEZA**

**2024**

BRUNO LUIZ FAUSTINO

ESTIMULAÇÃO CEREBRAL PROFUNDA ADAPTATIVA BASEADA EM APRENDIZADO  
POR REFORÇO: UM MODELO COMPUTACIONAL DE NEURÔNIOS DE IZHIKEVICH  
PARA SIMULAÇÃO DO TRATAMENTO DOS TREMORES DA DOENÇA DE PARKINSON

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia do Programa de Pós-Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Elétrica. Área de Concentração: Engenharia Elétrica.

Orientador: Prof. Dr. Victor Hugo Costa de Albuquerque.

Coorientador: Prof. Dr. Fabrício Gonzalez Nogueira.

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- F271e Faustino, Bruno Luiz.  
Estimulação cerebral profunda adaptativa baseada em aprendizado por reforço: um modelo computacional de neurônios de Izhikevich para simulação do tratamento dos tremores da doença de Parkinson / Bruno Luiz Faustino. – 2024.  
60 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Fortaleza, 2024.  
Orientação: Prof. Dr. Victor Hugo Costa de Albuquerque.  
Coorientação: Prof. Dr. Fabrício Gonzalez Nogueira.
1. Estimulação Cerebral Profunda. 2. Modelo de Izhikevich. 3. Aprendizado por Reforço. 4. Doença de Parkinson. 5. Tratamento de Tremores. I. Título.
- 

CDD 621.3


BRUNO LUIZ FAUSTINO

ESTIMULAÇÃO CEREBRAL PROFUNDA ADAPTATIVA BASEADA EM APRENDIZADO  
POR REFORÇO: UM MODELO COMPUTACIONAL DE NEURÔNIOS DE IZHIKEVICH  
PARA SIMULAÇÃO DO TRATAMENTO DOS TREMORES DA DOENÇA DE PARKINSON

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia do Programa de Pós-Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Elétrica. Área de Concentração: Engenharia Elétrica.


Aprovada em: 23/02/2024.

BANCA EXAMINADORA

Documento assinado digitalmente  
 VICTOR HUGO COSTA DE ALBUQUERQUE  
Data: 13/05/2024 18:36:33-0300  
Verifique em <https://validar.iti.gov.br>


---

Prof. Dr. Victor Hugo Costa de  
Albuquerque (Orientador)  
Universidade Federal do Ceará (UFC)

Documento assinado digitalmente  
 FABRICIO GONZALEZ NOGUEIRA  
Data: 17/05/2024 11:07:31-0300  
Verifique em <https://validar.iti.gov.br>


---

Prof. Dr. Fabrício Gonzalez Nogueira (Coorientador)  
Universidade Federal do Ceará (UFC)

Documento assinado digitalmente  
 AUZUIR RIPARDO DE ALEXANDRIA  
Data: 21/05/2024 14:39:37-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Auzuir Ripardo de Alexandria  
Instituto Federal do Ceará (IFCE)

Documento assinado digitalmente  
 RENATO WILLIAM RODRIGUES DE SOUZA  
Data: 21/05/2024 14:47:35-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Renato William Rodrigues de Souza  
Instituto Federal do Ceará (IFCE)

Para meus pais, que me mostraram que a educação vai além da formal e me ensinaram sobre respeito, amor e esperança.

## **AGRADECIMENTOS**

Ao Prof. Dr. Victor Hugo Costa de Albuquerque, pela excelente orientação, mostrando-se presente e contribuindo para as ideias aqui implementadas.

Ao Prof. Dr. Fabrício Gonzalez Nogueira, pela coorientação, onde mostrou-se disponível e ajudou na expansão de técnicas para a realização do trabalho.

Aos professores participantes da banca examinadora Dr. Auzuir Ripardo de Alexandria e Dr. Renato William Rodrigues de Souza pelo tempo, pelas valiosas colaborações e sugestões.

Ao Prof. Dr. Guilherme de Alencar Barreto por toda a base passada durante as cadeiras realizadas nesse período, onde pude aprender e, também, reconhecer sua grandeza profissional e humana.

Aos colegas da turma de mestrado, pelas reflexões, críticas e sugestões recebidas, em especial, Eduardo Balbino e Andreza Nascimento.

Ao meu gestor e amigo, Gustavo Souza, pelo apoio no âmbito profissional, o que permitiu a realização dessa dissertação de forma equilibrada às responsabilidades do trabalho.

Às pessoas que são mais próximas de mim, Maria Sueli, José Augusto, Maria Clara, Sônia Maria e Fernanda Sousa, que foram fontes de suporte e contribuíram para que a minha motivação frente a esse desafio fosse inesgotável e me fizeram enxergar a contribuição e o fascínio que é realizar pesquisa científica.

"A qualidade da minha vida aumentou significativamente após o DBS." (Daniel - paciente de Parkinson que realizou procedimento de Estimulação Cerebral Profunda, 2022.)

## RESUMO

A dissertação aprofunda-se na elaboração de um modelo computacional adaptativo para a simulação do tratamento dos tremores associados à doença de Parkinson, utilizando a técnica de Estimulação Cerebral Profunda em combinação com estratégias de aprendizado por reforço e a modelagem de neurônios através do modelo de Izhikevich. O objetivo central é otimizar os resultados terapêuticos potencialmente transformando o paradigma de tratamento para pacientes com Parkinson. Este estudo destaca-se por sua abordagem personalizada ao tratamento, propondo uma metodologia que ajusta dinamicamente os parâmetros de estimulação em resposta às condições variáveis do paciente. Os resultados encontrados são analisados em termos da eficiência da minimização dos sintomas e da energia despendida. Além disso, realiza-se uma comparação com padrões de estimulação cerebral estabelecidos comercialmente, em que o modelo aqui apresentado sugere ser mais eficiente em termos de energia quando acoplado com técnicas de tratamento do sinal, apresentando uma taxa de 80% de energia despendida em relação ao modelo comercial.

**Palavras-chave:** Estimulação Cerebral Profunda; Modelo de Izhikevich; Aprendizado por Reforço; Doença de Parkinson; Tratamento de Tremores.



## ABSTRACT

The dissertation delves into the development of an adaptive computational model for the simulation of treating tremors associated with Parkinson's disease, using Deep Brain Stimulation technique in combination with reinforcement learning strategies and neuron modeling through the Izhikevich model. The primary goal is to optimize therapeutic outcomes potentially transforming the treatment paradigm for Parkinson's patients. This study stands out for its personalized approach to treatment, proposing a methodology that dynamically adjusts stimulation parameters in response to the patient's varying conditions. The findings are analyzed in terms of symptom minimization efficiency and energy expenditure. Furthermore, a comparison is made with commercially established brain stimulation standards, where the model presented here is potentially more energy-efficient when coupled with signal treatment techniques, showing an 80% rate of energy released in comparison to the commercial model.

**Keywords:** Deep Brain Stimulation; Izhikevich Model; Reinforcement Learning; Parkinson's Disease; Tremor Treatment.

## LISTA DE FIGURAS

Figura 1 – Ilustração da aplicação de Deep Brain Stimulation (DBS). . . . .	20
Figura 2 – Representação gráfica de sinal emitido a 130 Hz na DBS. . . . .	20
Figura 3 – Comparação de modelos de disparos neuronais. . . . .	21
Figura 4 – Influência dos coeficientes do modelo nos sinais de potencial e excitação. . .	22
Figura 5 – Padrões de disparos utilizando modelo de Izhikevich. . . . .	22
Figura 6 – Diferença entre padrões de disparos fisiológicos e patológicos em DP. . . . .	23
Figura 7 – Esquemático de Reinforcement Learning. . . . .	25
Figura 8 – Funcionamento da função de clipagem. . . . .	29
Figura 9 – Diagrama de blocos geral. . . . .	33
Figura 10 – Esquemático geral da metodologia. . . . .	34
Figura 11 – Disparo sincronizado de neurônios em sintomas de DP. . . . .	39
Figura 12 – Sincronização de neurônios em sintomas de DP. . . . .	40
Figura 13 – Disparos em condições saudáveis. . . . .	40
Figura 14 – Dessincronização de neurônios em condições saudáveis. . . . .	41
Figura 15 – Média de estados (mV). . . . .	43
Figura 16 – Função de recompensa durante treinamento. . . . .	46
Figura 17 – Atuação do sinal de ação. . . . .	47
Figura 18 – Disparos dessincronizados após ação. . . . .	47
Figura 19 – Atenuação da média dos potenciais. . . . .	48
Figura 20 – Divergência KL durante treinamento. . . . .	48
Figura 21 – Análise gráfica de $k \times S$ . . . . .	49
Figura 22 – Análise gráfica de $k \times$ Energia. . . . .	50
Figura 23 – Aplicação de DBS padrão. . . . .	51

## LISTA DE TABELAS

Tabela 1 – Artigos nos últimos 5 anos com Estimulação Cerebral Profunda adaptativa em DP. . . . .	31
Tabela 2 – Modelos utilizados pela literatura. . . . .	32
Tabela 3 – Configurações da máquina. . . . .	35
Tabela 4 – Configurações de <i>softwares</i> e pacotes essenciais. . . . .	35
Tabela 5 – Parâmetros para modelagem dos neurônios de Izhikevich. . . . .	37
Tabela 6 – Parâmetros de treinamento . . . . .	45
Tabela 7 – Relação de $k$ passos ignorados na ação e métricas de avaliação. . . . .	49
Tabela 8 – Sinal DBS Tradicional. . . . .	50

## LISTA DE ABREVIATURAS E SIGLAS

<i>FLOPS</i>	<i>F</i> loating-point <i>O</i> perations <i>P</i> er <i>S</i> econd
<i>PPO</i>	<i>P</i> roximal <i>P</i> olicy <i>O</i> ptimization
<i>RL</i>	<i>R</i> einforcement <i>L</i> earning
<i>RMS</i>	<i>r</i> oot <i>m</i> ean <i>s</i> quare
DP	doença de Parkinson
GABA	ácido gama-aminobutírico
GPe	Glóbulo Pálido externo
GPi	Glóbulo Pálido interno
SNr	Substância Negra reticulada
STN	Núcleo Subtalâmico
TH	Tálamo

## LISTA DE SÍMBOLOS

$A$	Vantagem do modelo
$I$	Corrente elétrica
$k$	Passos de simulação
$N$	Número de neurônios
$Q$	Valor de ação do modelo
$t$	Tempo
$v$	Potencial de membrana
$V$	Valor de estado do modelo
$\pi$	Política do algoritmo
$\gamma$	Fator de desconto
$\varepsilon$	Limite para clipagem

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>1.1</b>	<b>Justificativa</b>	<b>15</b>
<b>1.2</b>	<b>Objetivos</b>	<b>16</b>
<b>1.3</b>	<b>Organização do Trabalho</b>	<b>17</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>18</b>
<b>2.1</b>	<b>Doença de Parkinson</b>	<b>18</b>
<b>2.2</b>	<b>Estimulação Cerebral Profunda</b>	<b>19</b>
<b>2.2.1</b>	<i>Geração de Sinais e Influência nos Padrões de Potencial de Membrana</i>	<b>19</b>
<b>2.3</b>	<b>Modelo de Izhikevich</b>	<b>21</b>
<b>2.3.1</b>	<i>Padrão de Rajada no GPi na Doença de Parkinson</i>	<b>23</b>
<b>2.3.2</b>	<i>Sincronização Neuronal na Doença de Parkinson</i>	<b>24</b>
<b>2.4</b>	<b>Aprendizado por Reforço</b>	<b>25</b>
<b>2.4.1</b>	<i>Métodos Ator-Crítico em Aprendizagem por Reforço</i>	<b>26</b>
<b>2.4.2</b>	<i>Otimização de Política Proximal</i>	<b>27</b>
<b>2.4.2.1</b>	<i>Função Objetivo Clipada</i>	<b>28</b>
<b>2.4.2.2</b>	<i>Função de Perda</i>	<b>30</b>
<b>2.5</b>	<b>Estado da arte: métodos de estimulação cerebral aplicada no tratamento de DP</b>	<b>30</b>
<b>2.5.1</b>	<i>Estratégias Avançadas de Estimulação Cerebral para DP</i>	<b>31</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>33</b>
<b>3.1</b>	<b>Panorama geral</b>	<b>33</b>
<b>3.2</b>	<b>Configuração Experimental</b>	<b>35</b>
<b>3.3</b>	<b>Criação do ambiente proposto de neurônios de Izhikevich</b>	<b>36</b>
<b>3.3.1</b>	<i>Método step</i>	<b>37</b>
<b>3.3.2</b>	<i>Método reset</i>	<b>37</b>
<b>3.3.3</b>	<i>Métodos render, seed e close</i>	<b>38</b>
<b>3.3.4</b>	<i>Função de Recompensa</i>	<b>38</b>
<b>3.4</b>	<b>Parâmetros e Desempenho do modelo</b>	<b>41</b>
<b>3.5</b>	<b>Análise da atenuação</b>	<b>42</b>

<b>3.6</b>	<b>Estudo de frequência de passos e análise comparativa com sinal DBS padrão . . . . .</b>	<b>43</b>
<b>3.7</b>	<b>Adaptabilidade . . . . .</b>	<b>44</b>
<b>4</b>	<b>RESULTADOS . . . . .</b>	<b>45</b>
<b>4.1</b>	<b>Treinamento do modelo. . . . .</b>	<b>45</b>
<b>4.2</b>	<b>Análise gráfica da dessincronização . . . . .</b>	<b>46</b>
<b>4.3</b>	<b>Visão analítica . . . . .</b>	<b>47</b>
<b>4.4</b>	<b>Estudo da frequência de aplicação do sinal de ação. . . . .</b>	<b>49</b>
<b>4.5</b>	<b>Comparativo com DBS tradicional. . . . .</b>	<b>50</b>
<b>4.6</b>	<b>Discussão dos resultados . . . . .</b>	<b>52</b>
<b>5</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS . . . . .</b>	<b>54</b>
<b>5.1</b>	<b>Trabalhos Futuros . . . . .</b>	<b>54</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>55</b>

## 1 INTRODUÇÃO

A doença de Parkinson (DP), do inglês *Parkinson Disease (PD)*, é um transtorno neurodegenerativo progressivo que afeta predominantemente a população idosa, representando a segunda condição neurodegenerativa mais comum, depois da doença de Alzheimer. Caracteriza-se pela degeneração de neurônios dopaminérgicos na substância negra do cérebro, levando a sintomas motores distintos, como tremores em repouso, rigidez, bradicinesia (lentidão dos movimentos) e instabilidade postural. Além dos sintomas motores, pacientes com DP podem experimentar uma vasta gama de sintomas não motores, incluindo alterações cognitivas, distúrbios do sono, depressão e dificuldades na fala. Segundo a WORLD HEALTH ORGANIZATION (2023b), estima-se que a incidência da doença de Parkinson esteja aumentando, com mais de 10 milhões de pessoas vivendo com a doença em todo o mundo, refletindo tanto o envelhecimento da população quanto a melhoria na capacidade de diagnóstico. Este panorama reforça a necessidade urgente de terapias mais eficazes e personalizadas, como a Estimulação Cerebral Profunda, do inglês *Deep Brain Stimulation (DBS)*, técnica de estímulo de regiões cerebrais por meio de impulsos elétricos que visa melhorar a qualidade de vida dos pacientes afetados por este transtorno debilitante (WORLD HEALTH ORGANIZATION, 2023a).

Diante desse desafio, a intersecção entre a neurociência computacional e a Aprendizagem por Reforço, do inglês *Reinforcement Learning (RL)*, tem aberto novos caminhos para o tratamento de distúrbios neurológicos complexos (WU, 2023; RIBBA, 2023). A capacidade de simular e analisar padrões de atividade neuronal, especialmente em doenças como a doença de Parkinson, é de importância crítica para o avanço das terapias neuromoduladoras. A Estimulação Cerebral Profunda apresenta-se como uma técnica promissora, proporcionando melhorias significativas nos sintomas motores e na qualidade de vida dos pacientes que não respondem ao tratamento farmacológico convencional (KRACK *et al.*, 2019).

No cerne deste avanço tem-se a modelagem dos neurônios de Izhikevich (IZHIKEVICH, 2003), que permite uma representação detalhada dos potenciais de membrana e dos impulsos de corrente no cérebro. Quando combinada com métodos avançados de aprendizagem por reforço, como Otimização de Política Proximal, do inglês *Proximal Policy Optimization (PPO)*, proposta por Schulman *et al.* (2017), surge a possibilidade de se desenvolver algoritmos que não apenas entendem, mas também influenciam a atividade neural. Esses algoritmos abrem a possibilidade de personalizar a DBS para cada paciente ao ajustar os impulsos de corrente em tempo real para otimizar os resultados terapêuticos.



Este trabalho se propõe a explorar a aplicação de um algoritmo de RL baseado em PPO na modulação de correntes em neurônios de Izhikevich (IZHIKEVICH, 2003), com foco no Glóbulo Pálido Interno (GPI), uma região cerebral frequentemente alvo da DBS no tratamento de doença de Parkinson. Através de uma abordagem que combina modelagem computacional detalhada com algoritmos de aprendizagem adaptativa, é investigada a manipulação precisa dos padrões de disparo neuronal que pode levar a uma melhoria nos sintomas de pacientes com doença de Parkinson, potencialmente reduzindo a necessidade de medicamentos e melhorando a qualidade de vida.

Com a integração de técnicas computacionais de aprendizado por reforço e conhecimento neurológico, objetiva-se não apenas avançar no tratamento de doença de Parkinson, mas também contribuir para a compreensão mais ampla dos processos neuronais e das possibilidades de sua modulação. Este estudo auxilia no entendimento de uma neurociência aplicada mais interativa e adaptativa, onde a aprendizagem por reforço e a DBS convergem para a melhoria da saúde humana.

## 1.1 Justificativa

A doença de Parkinson, sendo uma das principais desordens neurodegenerativas, afeta significativamente a população global trazendo consigo um significativo ônus tanto para os pacientes quanto para os sistemas de saúde. A complexidade dos sintomas motores e não motores da DP e a progressão variável da doença apresentam desafios únicos para um tratamento eficaz. Dado que a Estimulação Cerebral Profunda tem se mostrado uma terapia promissora para diversos distúrbios, há avanços significativos na personalização do tratamento e na melhoria da qualidade de vida dos pacientes com a utilização de algoritmos de inteligência artificial em processamento de dados e de imagens, como análises de eletroencefalogramas e exames de radiografia (ALIPOUR *et al.*, 2023; OLIVEIRA *et al.*, 2023), contudo, persistem oportunidades para melhorias no processo pós-cirúrgico. A aplicação de métodos de aprendizado de máquina, do inglês *Machine Learning (ML)*, na DBS está sendo investigada como uma forma de melhorar a personalização do tratamento. Esses métodos podem ajudar a otimizar os parâmetros de estimulação e a seleção de alvos dentro do cérebro, proporcionando uma melhoria na qualidade de vida dos pacientes (PERALTA *et al.*, 2021).

Dentro deste contexto, a aplicação de algoritmos de aprendizagem por reforço, especialmente o *Proximal Policy Optimization (PPO)*, destaca-se como um avanço na personalização

do tratamento da DP. A capacidade do PPO de ajustar dinamicamente os parâmetros da estimulação em resposta às mudanças nos padrões de atividade neuronal tem o potencial de superar as limitações inerentes às abordagens tradicionais de DBS. Especificamente, a modelagem de neurônios de Izhikevich fornece uma representação detalhada dos potenciais de membrana e dos impulsos de corrente, permitindo que o algoritmo de RL opere de maneira mais informada e com respostas adaptativas às necessidades fisiológicas do paciente. Assim, este trabalho justifica-se pela necessidade de explorar novas estratégias terapêuticas que abordem as complexidades do tratamento da DP e pelo potencial de contribuição significativa para a neurociência aplicada.

Portanto, a motivação deste trabalho reside na exploração de tecnologias avançadas para abordar as necessidades não atendidas no tratamento da DP, visando uma contribuição significativa à neurociência aplicada e à melhoria da qualidade de vida dos pacientes.

## 1.2 Objetivos

O objetivo principal deste trabalho é desenvolver uma abordagem adaptativa, empregando algoritmos de aprendizado por reforço, para otimizar o *design* de DBS no tratamento de tremores associados à doença de Parkinson. Este tratamento é realizado através da dessincronização dos neurônios no Glóbulo Pálido Interno (GPi), visando a atenuação dos tremores. Os objetivos específicos incluem:

- Implementar um algoritmo adaptável de aprendizado por reforço, especificamente utilizando Otimização de Política Proximal, para modular as correntes emitidas pelo DBS, visando a dessincronização efetiva dos neurônios no GPi e a consequente atenuação dos tremores;
- Conduzir uma análise detalhada da frequência de estimulações aplicadas pelo agente de aprendizado por reforço, buscando um equilíbrio ótimo entre a eficiência do modelo e o consumo energético;
- Comparar o desempenho e eficácia do modelo de sinal de corrente proposto com o modelo padrão de DBS atualmente utilizados em práticas clínicas, destacando as melhorias e inovações.

### 1.3 Organização do Trabalho

Esta dissertação está estruturada da seguinte maneira para facilitar a compreensão do estudo realizado:

- **Capítulo 2:** Este capítulo aborda as fundamentações teóricas essenciais para o desenvolvimento do trabalho, oferecendo um panorama detalhado da doença de Parkinson e dos padrões de comportamento neuronal associados à doença. São discutidos os modelos de Izhikevich, que simulam o comportamento dos neurônios, e os princípios fundamentais da Aprendizagem por Reforço. Essa base conceitual é crucial para entender o contexto e a relevância do projeto de DBS proposto;
- **Capítulo 3:** Descreve a metodologia utilizada no estudo, detalhando a criação do ambiente de simulação, as estratégias de treinamento e validação do modelo, além dos métodos de comparação e análise quantitativa adotados. Este capítulo é essencial para compreender as técnicas e abordagens inovadoras implementadas na pesquisa;
- **Capítulo 4:** Apresenta os resultados alcançados, incluindo os desempenhos de treinamento e a análise dos parâmetros definidos no capítulo anterior. Este capítulo também traz comparações detalhadas com sinais reais e discute as melhorias introduzidas pelo modelo proposto, evidenciando os avanços obtidos na personalização do tratamento da doença de Parkinson através da DBS;
- **Capítulo 5:** Reúne as conclusões derivadas da pesquisa e propõe direções para trabalhos futuros. Este capítulo reflete sobre o impacto do estudo na compreensão e tratamento da doença de Parkinson, destacando a contribuição significativa para a neurociência aplicada e sugerindo caminhos para o aprofundamento da temática.

A organização desta dissertação visa não apenas a apresentação clara e estruturada dos resultados da pesquisa, mas também a demonstração de como o estudo contribui para o avanço no tratamento da doença de Parkinson, enfatizando a importância da inovação tecnológica e da interdisciplinaridade no campo da saúde.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo abordará os aspectos fundamentais para entendimento e realização do trabalho e será dividido em quatro seções. A primeira seção trata da doença de Parkinson (DP) e suas implicações, seguido da apresentação da técnica de atenuação dos sintomas: Estimulação Cerebral Profunda (DBS). Em seguida, é descrito o modelo de Izhikevich, trazido como objeto de estudo neste trabalho e, por fim, o algoritmo proposto de Aprendizagem por Reforço (RL).

### 2.1 Doença de Parkinson

A doença de Parkinson é uma condição neurodegenerativa caracterizada pela degeneração progressiva de neurônios dopaminérgicos na substância negra, parte do mesencéfalo. A perda de dopamina, um neurotransmissor crucial para o controle motor, resulta em sintomas motores característicos como tremor, rigidez e bradicinesia. De acordo com estimativas globais, aproximadamente 8,5 milhões de indivíduos foram afetados pela doença de Parkinson em 2019, destacando sua relevância epidemiológica (WORLD HEALTH ORGANIZATION, 2023b).

A redução nos níveis de dopamina desencadeia alterações críticas nos circuitos do gânglio basal, afetando principalmente o Núcleo Subtalâmico (STN), o Glóbulo Pálido externo (GPe), o Glóbulo Pálido interno (GPi) e o Tálamo (TH). Este desequilíbrio leva à superativação do STN, aumentando a excitação no GPi e na Substância Negra reticulada (SNr), o que resulta em uma inibição excessiva das projeções motoras ao tálamo e tronco cerebral, manifestando-se nos sintomas característicos da DP (OBESO *et al.*, 2000).

Especificamente, a depleção dopaminérgica leva à hipoatividade no GPe e hiperatividade no GPi e no STN. Essas alterações são acompanhadas por padrões anormais de disparos neuronais e oscilações rítmicas síncronas nestas regiões, que estão intimamente relacionadas aos sintomas motores da DP. Pesquisas mostram que o tratamento com levodopa reduz a atividade de baixa frequência e resulta em um novo pico em torno de 70 Hz no espectro de potência do STN e do GPi, alterando a coerência entre esses núcleos (BROWN *et al.*, 2001).

Os tratamentos para a DP variam desde medicamentos que visam aumentar a disponibilidade de dopamina até intervenções avançadas como a Estimulação Cerebral Profunda. A DBS, alvo dos núcleos STN e GPi, modula eficazmente a atividade neuronal anormal, oferecendo melhorias significativas nos sintomas motores e na qualidade de vida dos pacientes (POEWE *et al.*, 2017).

Consequentemente, entender a complexa dinâmica entre a deficiência de dopamina e as mudanças neuronais nos gânglios basais é crucial para o desenvolvimento e aprimoramento de estratégias terapêuticas eficazes contra a doença de Parkinson.

## 2.2 Estimulação Cerebral Profunda

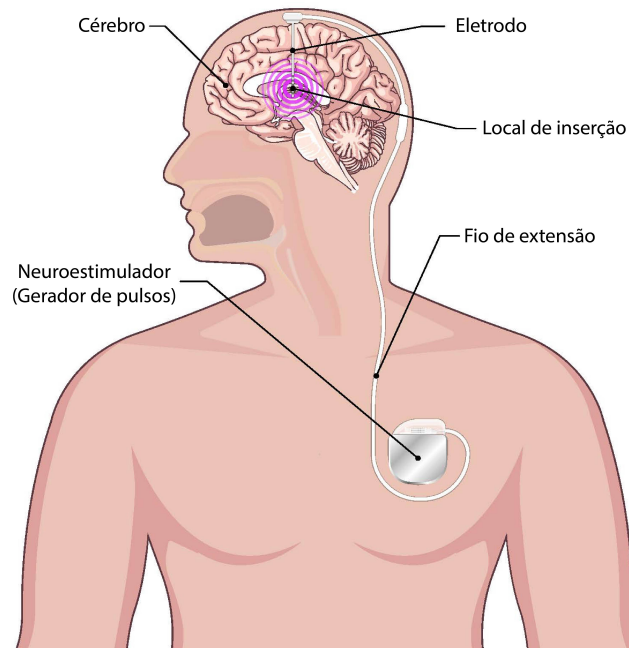
A Estimulação Cerebral Profunda é uma técnica cirúrgica inovadora e menos invasiva utilizada no tratamento de doenças do cérebro, como a DP. A DBS envolve a implantação de eletrodos em áreas específicas do cérebro, com o objetivo de modular a atividade neural. Esta modulação é realizada através de um estimulador externo controlado por bateria, implantado no peito do paciente. A Figura 1 ilustra esquematicamente a aplicação da técnica, destacando o caminho do estímulo elétrico que parte do neuroestimulador, passa pelo fio de extensão, e chega ao eletrodo implantado, geralmente no GPi ou no STN (SIDIROPOULOS *et al.*, 2016; MIRZADEH *et al.*, 2016; BURCHIEL *et al.*, 2013). A seleção do local de implantação dos eletrodos é baseada em uma série de fatores, incluindo as condições específicas do paciente e os sintomas a serem tratados (STARR, 2003).

Comparativamente, a DBS é reconhecida como uma opção terapêutica menos invasiva, especialmente por evitar danos permanentes ao tecido cerebral, diferentemente de outras abordagens cirúrgicas. Exemplos destas incluem a *Palidotomia*, que implica a destruição cirúrgica de uma parte do GPi, e a Terapia por Lesão por Radiofrequência, que cria lesões permanentes em áreas específicas do cérebro (LEVI *et al.*, 2021; SHARMA *et al.*, 2020). Portanto, a DBS se destaca por ser ajustável e reversível, adaptando-se às necessidades individuais de cada paciente, proporcionando uma abordagem personalizada ao tratamento dos sintomas da DP.

### 2.2.1 Geração de Sinais e Influência nos Padrões de Potencial de Membrana

A DBS produz sinais elétricos cujas características, como frequência, largura de pulso e amplitude, podem ser meticulosamente ajustadas. Essas variações são fundamentais para influenciar os padrões de potencial de membrana dos neurônios nas áreas-alvo do cérebro. Notavelmente, estudos demonstram que a DBS modula a interação entre a fase do ritmo beta e a amplitude da atividade de banda larga no córtex motor primário em pacientes com doença de Parkinson. Tal modulação sugere uma potencial melhoria na função cortical (HEMPTINNE *et al.*, 2015). A aplicação de altas frequências de estimulação (superiores a 100 Hz) conforme

Figura 1 – Ilustração da aplicação de Deep Brain Stimulation (DBS).

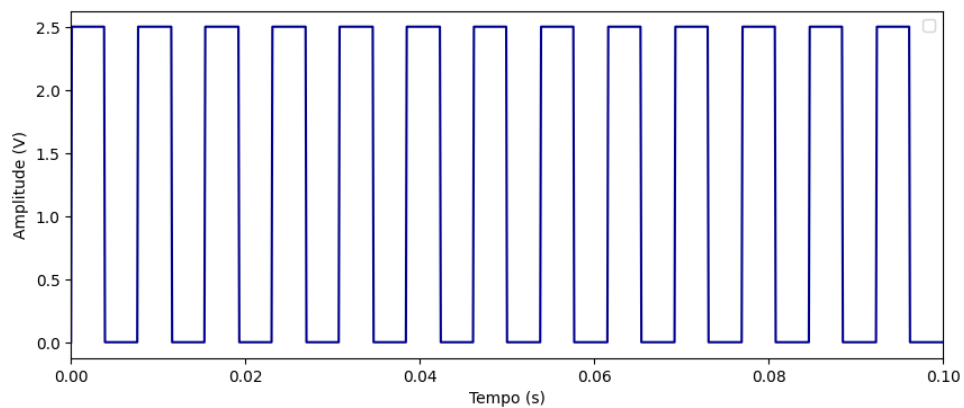


Fonte: Adaptado de PARKINSON's NSW (2024).

ilustrado na Figura 2, que são ajustadas manualmente pelo médico, tem demonstrado eficácia na modulação da atividade neuronal, levando a um maior controle dos sintomas (POPOVYCH *et al.*, 2017).

Portanto, a DBS emerge como uma inovação significativa no tratamento de distúrbios do movimento, incluindo a DP. Essa abordagem terapêutica está em constante evolução, abrangendo a identificação de novos alvos cirúrgicos, aprimoramento das técnicas de implantação e aprofundamento da compreensão sobre os mecanismos neuronais subjacentes à eficácia do tratamento.

Figura 2 – Representação gráfica de sinal emitido a 130 Hz na DBS.

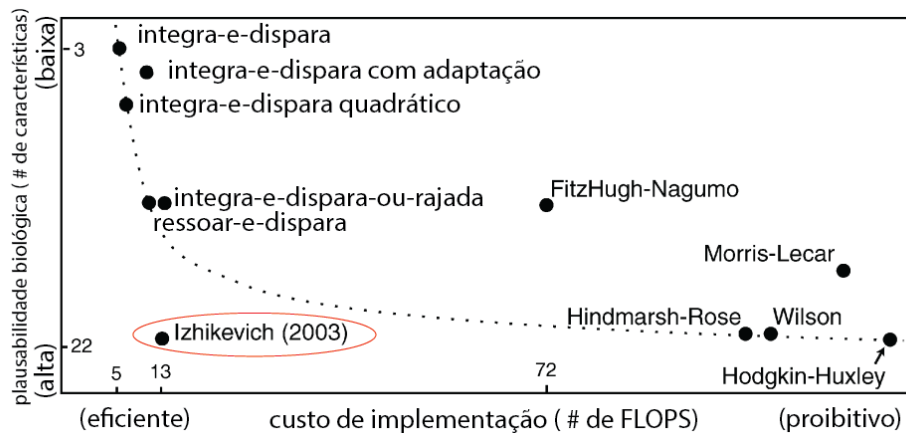


Fonte: Elaboração Própria, 2024.

### 2.3 Modelo de Izhikevich

O modelo de Izhikevich é um modelo computacional de neurônios desenvolvido para replicar a dinâmica dos disparos neuronais de forma eficiente, considerando tanto o custo computacional, medido em operações de ponto flutuante por segundo, do inglês *FLoating-point Operations Per Second (FLOPS)*, quanto a alta plausibilidade biológica, evidenciada pela capacidade de representar características neuronais diversas (IZHIKEVICH, 2003). A Figura 3 oferece uma comparação deste modelo com outros modelos da literatura.

Figura 3 – Comparação de modelos de disparos neuronais.



Fonte: Adaptado de Izhikevich (2004).

Este modelo se destaca na simulação de vários padrões de disparo observados em neurônios corticais, como disparo regular (*regular spiking*), rajada (*bursting*), chocalho (*chattering*), entre outros (IZHIKEVICH, 2004), conforme ilustra a Figura 5. As equações fundamentais do modelo são apresentadas a seguir:

$$v' = 0.04v^2 + 5v + 140 - u + I \quad (2.1)$$

$$u' = a(bv - u) \quad (2.2)$$

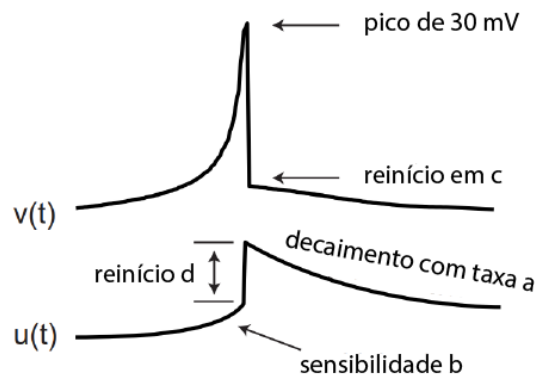
em que  $v'$  representa a variação do potencial de membrana em mV,  $v$  o potencial da membrana no passo atual,  $u$  a variável de recuperação da membrana em mV,  $I$  a corrente equivalente às interações sinápticas dos neurônios em pA, e  $a$  (mV.s<sup>-1</sup>),  $b$  (mV),  $c$  (mV),  $d$  (mV) são parâmetros que determinam o tipo específico de comportamento do neurônio, representado na Figura 4. Ajustando-se estes parâmetros, diferentes padrões de disparo podem ser simulados.

Quando um disparo é gerado, ou seja, quando  $v$  atinge 30 mV,  $v$  é resetado para  $c$  e  $u$

é incrementado por  $d$  para que o ciclo seja reiniciado:

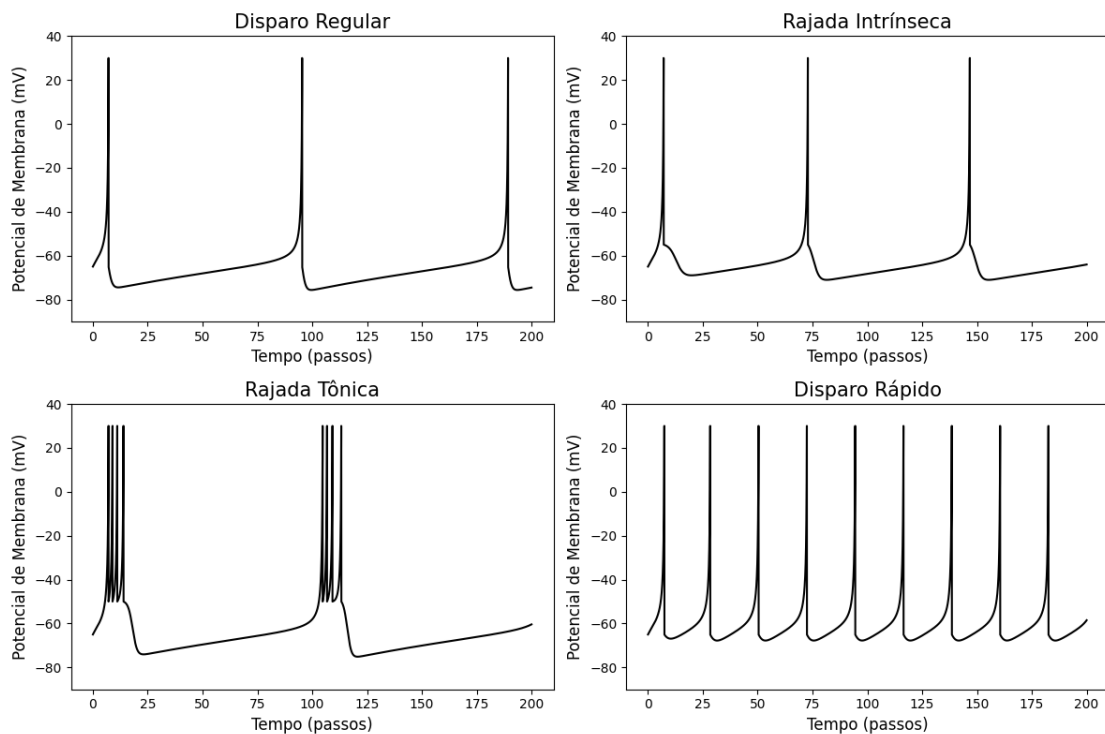
$$\text{se } v \geq 30 \text{ mV, então } \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \quad (2.3)$$

Figura 4 – Influência dos coeficientes do modelo nos sinais de potencial e excitação.



Fonte: Adaptado de Izhikevich (2024).

Figura 5 – Padrões de disparos utilizando modelo de Izhikevich.



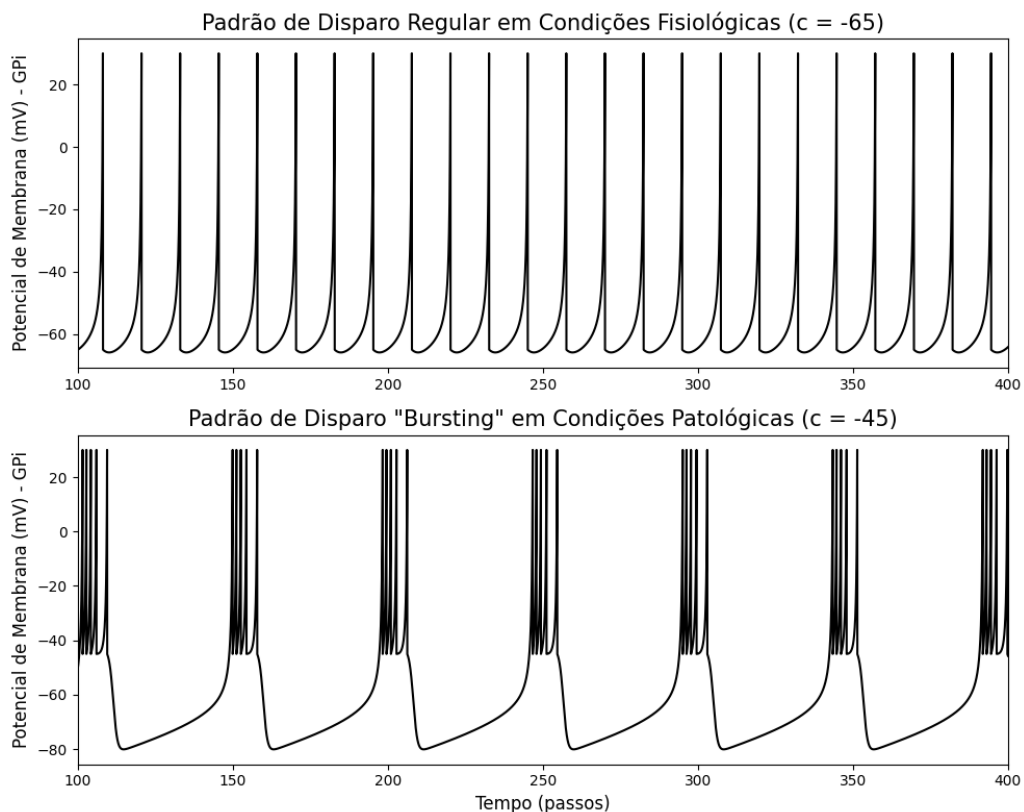
Fonte: Elaboração Própria, 2024.



### 2.3.1 Padrão de Rajada no GPi na Doença de Parkinson

Na doença de Parkinson, a diminuição dos níveis de dopamina contribui significativamente para a hiperatividade neuronal, frequentemente manifestada em forma de rajadas. Alterações nos canais de íons, particularmente nos canais de cálcio e sódio, desempenham um papel crucial na geração desses padrões de disparo (HALLWORTH *et al.*, 2003; KADALA *et al.*, 2015). Estudos comparativos entre condições fisiológicas e patológicas (LIU *et al.*, 2015) utilizaram o modelo de Izhikevich para demonstrar diferenças nos padrões de disparo, especialmente no GPi, padrões esses que também foram observados em células de pacientes humanos (TANG *et al.*, 2005; ALAM *et al.*, 2016). Tais sinais são captados por técnicas e instrumentos de medição, como o Eletroencefalograma (EEG), que emprega eletrodos posicionados no couro cabeludo, e a Magnetoencefalografia (MEG), que detecta os campos magnéticos gerados pela atividade elétrica neuronal.

Figura 6 – Diferença entre padrões de disparos fisiológicos e patológicos em DP.



Fonte: Adaptado de Liu *et al.* (2015).

### 2.3.2 Sincronização Neuronal na Doença de Parkinson

A inibição sináptica mediada pelo ácido gama-aminobutírico (GABA) é fundamental no sistema nervoso central, regulando a excitabilidade neuronal e mantendo o equilíbrio entre as atividades excitatórias e inibitórias no cérebro. Esse processo é orquestrado pelo neurotransmissor inibitório GABA, que ao interagir com seus receptores nos neurônios, induz a hiperpolarização da membrana através da entrada de íons cloreto (BELELLI *et al.*, 2009). Tal hiperpolarização reduz a probabilidade de geração de potenciais de ação, desempenhando um papel crucial na modulação da atividade neural.

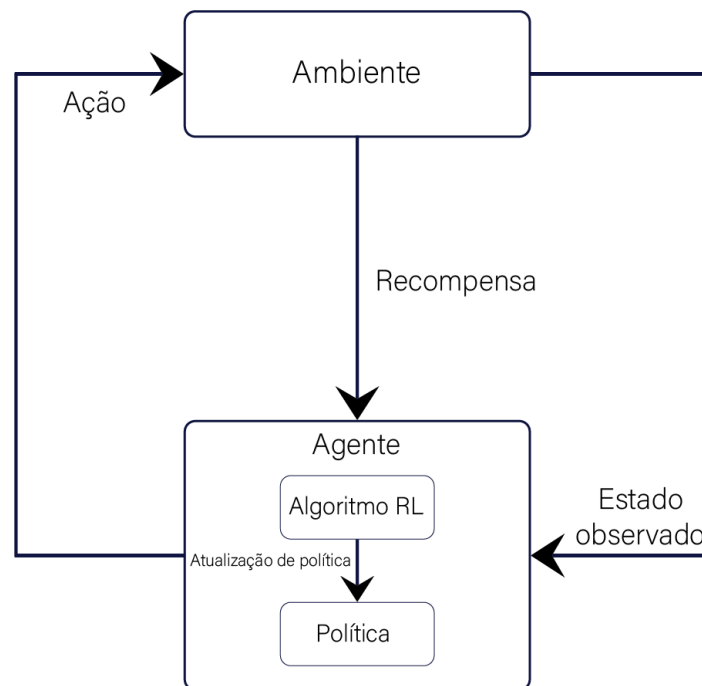
Os principais receptores GABAérgicos,  $GABA_A$  e  $GABA_B$ , fornecem respostas inibitórias rápidas e lentas, respectivamente, sendo essenciais para o funcionamento cerebral adequado, desde o processamento de informações até a regulação do humor e a prevenção de atividades anormais, como convulsões.

Especificamente na doença de Parkinson, alterações na neurotransmissão GABAérgica, consequentes à depleção dopaminérgica, indicam modificações substanciais na dinâmica do circuito e na transmissão sináptica associadas à DP (RUBI; FRITSCHY, 2020). Além disso, a observação de sincronização oscilatória entre neurônios no GPi e no GPe em pacientes exibindo tremor de membros destaca a existência de padrões sincronizados característicos da DP (LEVY *et al.*, 2002). Pesquisas complementares revelam que o STN e o GPe constituem um loop de feedback responsável por disparos oscilatórios sincronizados, fundamentais para compreender as dinâmicas neuronais dos gânglios basais em condições normais e patológicas (PLENZ; KITAL, 1999).

## 2.4 Aprendizado por Reforço

O Aprendizado por Reforço, do inglês *Reinforcement Learning (RL)*, é um método de aprendizagem de máquina onde um agente aprende a tomar decisões através de tentativa e erro, recebendo recompensas ou penalidades para suas ações. Historicamente, a RL destacou-se por sua capacidade de controlar agentes em ambientes de alta complexidade (MNIH *et al.*, 2015).

Figura 7 – Esquemático de Reinforcement Learning.



Fonte: Adaptado de Sutton e Barto (2018).

A partir da Figura 7, é possível conceituar os principais componentes de um projeto de aprendizado por reforço (SUTTON; BARTO, 2018):

- **Ambiente:** o contexto ou o espaço onde o agente opera. Inclui todos os fatores externos que influenciam e são influenciados pelas ações do agente.
- **Recompensa:** um sinal numérico que o agente recebe do ambiente após cada ação. Indica o sucesso ou falha da ação em relação ao objetivo a ser alcançado. Pode ser representado por uma função matemática ou um valor numérico absoluto a depender de cada estado.
- **Estado:** uma representação do ambiente em um determinado momento. O estado pode incluir a posição do agente, objetos ao redor, ou qualquer outra informação relevante para a tomada de decisão.

- **Agente:** a entidade que aprende e toma decisões no ambiente. O agente seleciona ações com base na política atual com o objetivo de maximizar a soma das recompensas futuras.
- **Algoritmo RL:** um método ou conjunto de regras que o agente utiliza para aprender a melhor política. O algoritmo atualiza a política com base na experiência adquirida (recompensas recebidas e transições de estado observadas).
- **Política:** uma estratégia que o agente usa para determinar a próxima ação com base no estado atual do ambiente. Pode ser uma função simples ou um modelo complexo, como uma rede neural.
- **Ação:** uma decisão ou movimento específico que o agente executa no ambiente. A escolha da ação é influenciada pela política do agente.

Em suma, através das ações impostas ao ambiente, busca-se otimizar a política para que a recompensa seja a melhor possível.

#### ***2.4.1 Métodos Ator-Crítico em Aprendizagem por Reforço***

Os métodos Ator-Crítico são uma abordagem fundamental na aprendizagem por reforço, combinando as vantagens dos métodos baseados em valor e em política. Estes métodos são compostos por duas partes: o ator que é responsável por escolher as ações, e o crítico que avalia essas ações. O ator atualiza a política de ações, enquanto o crítico estima o valor dessas ações, fornecendo um feedback ao ator sobre a qualidade das decisões tomadas. Tais componentes são conceituados da seguinte maneira:

- **Ator:** representa a política de ações do agente, onde o objetivo é aprender a escolher ações que maximizem a recompensa total. O ator é geralmente implementado por um modelo de rede neural que mapeia estados para ações.
- **Crítico:** avalia as ações escolhidas pelo ator por meio de uma função de valor, como a função de valor de estado ou a função de valor de ação. O crítico ajuda a guiar o ator para melhores políticas.
- **Atualizações de Política:** a política é atualizada com base tanto nas previsões do crítico quanto nos resultados das ações tomadas pelo ator, combinando as abordagens baseadas em modelo e sem modelo.

### 2.4.2 Otimização de Política Proximal

A Otimização de Política Proximal, do inglês *Proximal Policy Optimization (PPO)*, é uma abordagem de aprendizagem por reforço que foi desenvolvida para superar algumas das dificuldades encontradas em métodos anteriores, como a Otimização de Política por Gradiente de Confiança, do inglês *Trusted Region Policy Optimization (TRPO)*. A PPO busca simplificar o processo de otimização enquanto mantém a robustez e eficiência. Introduzida por Schulman *et al.* (2017), a PPO se tornou popular devido à sua facilidade de implementação, facilidade de otimização e desempenho estável em uma variedade de tarefas de controle, incluindo aplicações complexas como robótica e jogos. Notavelmente, a PPO permite a realização de múltiplas atualizações de política com o mesmo conjunto de dados, otimizando a eficiência de amostragem do algoritmo (HAMALAINEN *et al.*, 2020; YU *et al.*, 2021).

O funcionamento da PPO é ilustrado no Algoritmo 1, onde, a cada iteração, múltiplos atores coletam dados ao executar a política vigente no ambiente por um número predeterminado de passos. A partir dos dados obtidos, o algoritmo calcula estimativas de vantagem para avaliar o desempenho relativo das ações comparadas a uma política base. Com essas vantagens, a PPO efetua várias passadas de otimização (ou épocas) para ajustar os parâmetros da política. Este ajuste visa maximizar uma função objetivo composta, que engloba termos para a política (com *clipping*), o valor estimado e um bônus de entropia para promover a exploração. Após a otimização, a política é atualizada com novos parâmetros, prosseguindo iterativamente o aperfeiçoamento da política do agente em busca da maximização das recompensas acumuladas (SCHULMAN *et al.*, 2017).

---

#### Algoritmo 1: PPO

---

```

for  $iteration = 1, 2, \dots$  do
  for  $ator = 1, 2, \dots, N$  do
    Executa política  $\pi_{\theta_{antiga}}$  no ambiente para  $T$  passos de tempo
    Calcula vantagem estimada  $\hat{A}_1, \dots, \hat{A}_T$ 
  end
  Otimiza substituto  $L$  com respeito a  $\theta$ , com  $K$  épocas e minilotes de tamanho
   $M \leq NT$ 
   $\theta_{antiga} \leftarrow \theta$ 
end

```

---

### 2.4.2.1 Função Objetivo Clipada

A PPO emprega uma estratégia de clipagem na função objetivo para evitar a necessidade de resolver problemas de otimização complexos a cada atualização da política. Essa abordagem é projetada para aprimorar a política de decisão do agente de maneira eficiente e estável. A formulação da função objetivo clipada do PPO é dada por:

$$L_t^{CLIP}(\theta) = \max_{\theta} \mathbb{E}_{s,a \sim \pi_{\theta_{\text{antiga}}}} \left[ \min \left( r_t(\theta) A^{\pi_{\theta_{\text{antiga}}}}(s,a), \text{clip}(r_t(\theta), 1 - \varepsilon, 1 + \varepsilon) A^{\pi_{\theta_{\text{antiga}}}}(s,a) \right) \right], \quad (2.4)$$

onde  $\varepsilon$  é um hiperparâmetro que delimita o intervalo de clipagem.

A fim de facilitar o entendimento da equação, tem-se que a finalidade do algoritmo é maximizar a função objetivo em relação aos parâmetros da política, denotados por  $\theta$ . O algoritmo ajusta  $\theta$  para obter o máximo retorno possível. Dessa forma, tem-se que:

- $\mathbb{E}_{s,a \sim \pi_{\theta_{\text{antiga}}}}$  é a expectativa matemática de recompensas sob a política antiga  $\pi_{\theta_{\text{antiga}}}$ , em que  $s$  representa o estado e  $a$  representa a ação. O agente avalia as ações baseando-se na experiência passada.
- $r_t(\theta) = \frac{\pi_{\theta}(a|s)}{\pi_{\theta_{\text{antiga}}}(a|s)}$  é a razão de probabilidade, comparando a nova política  $\pi_{\theta}$  com a política antiga  $\pi_{\theta_{\text{antiga}}}$  para uma ação  $a$  em um estado  $s$ .
- $A^{\pi_{\theta_{\text{antiga}}}}(s,a)$  é a função de vantagem, indicando a qualidade de uma ação  $a$  em um estado  $s$  em relação à média sob a política antiga. É um indicador crítico de quais ações devem ser escolhidas e é definida por:

$$A^{\pi}(s,a) = Q^{\pi}(s,a) - V^{\pi}(s), \quad (2.5)$$

em que  $V(s)$  é a função de valor de estado, estimada pelo retorno total esperado (ou seja, as recompensas cumulativas descontadas) que um agente pode esperar começando do estado  $s$  e seguindo uma política  $\pi$ , definida como:

$$V^{\pi}(s) = \mathbb{E}_{\pi} \left[ \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s \right], \quad (2.6)$$

em que:

- $\gamma$  é o fator de desconto, com  $0 \leq \gamma < 1$ , que modula a importância das recompensas futuras. Um  $\gamma$  mais próximo de 1 faz com que o agente valorize as recompensas futuras quase tanto quanto as imediatas.

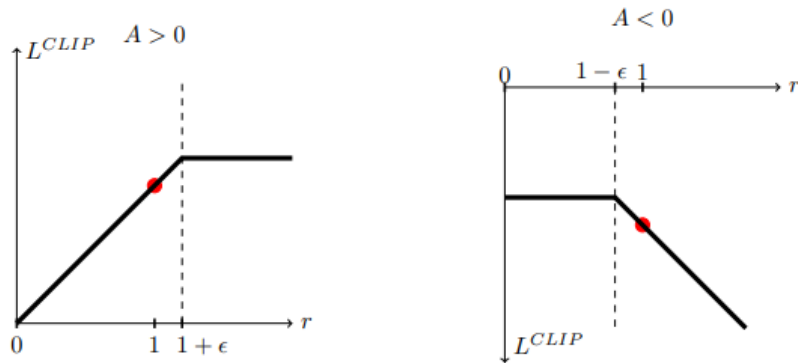
- $R_{t+k+1}$  é a recompensa recebida no passo de tempo  $t + k + 1$ .
- $S_t$  é o estado no passo de tempo  $t$ .
- O somatório  $\sum_{k=0}^{\infty}$  calcula a soma total das recompensas futuras, ponderadas pelo fator de desconto  $\gamma$ .

Similarmente,  $Q(s, a)$  é a função valor de ação, que representa o retorno ocorrido dada a ação imposta. Assim, definida por:

$$Q^{\pi}(s, a) = \mathbb{E}_{\pi} \left[ \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s, A_t = a \right]. \quad (2.7)$$

- **clip**  $\left( \frac{\pi_{\theta}(a|s)}{\pi_{\theta_{\text{antiga}}}(a|s)}, 1 - \epsilon, 1 + \epsilon \right)$  limita a razão de probabilidade dentro de um intervalo especificado por  $\epsilon$ , prevenindo ajustes excessivos na política que possam comprometer a estabilidade do aprendizado. Como exposto na Figura 8, quando a vantagem (A) é positiva, ou seja, quando a ação é considerada boa e sua probabilidade aumenta, a função a clipa, mantendo-a numa faixa segura para que não atualize excessivamente e piore. Do contrário, caso a vantagem seja negativa, ou seja, com uma ação indesejada e menos provável, a função impede que sua probabilidade seja reduzida demasiadamente.

Figura 8 – Funcionamento da função de clipagem.



Fonte: Schulman *et al.* (2017)

- $\min(\dots)$  é a operação de minimização que garante que o menor valor entre a vantagem multiplicada pela razão de probabilidade não clipada e a vantagem multiplicada pela razão clipada seja usado. Isso efetivamente limita a atualização da política, equilibrando a exploração de novas estratégias e a estabilidade do aprendizado.

Em resumo, a função objetivo do PPO é projetada para ajustar a política do agente de maneira que maximize a recompensa total esperada, mantendo a estabilidade do processo de aprendizado. O uso do *clipping* na razão de probabilidade é uma característica distintiva do PPO, que o ajuda a evitar grandes desvios na política que poderiam prejudicar o desempenho.

### 2.4.2.2 Função de Perda

O PPO utiliza uma função de perda especializada que busca maximizar a eficácia do agente enquanto mantém as mudanças na política relativamente pequenas para evitar instabilidades no aprendizado. Assim, a fim de equilibrar eficiência do treinamento e a estabilidade, permite que o agente explore o espaço de ação e aprenda políticas robustas sem grandes oscilações. A função é definida por:

$$L_t^{CLIP+VF+S}(\theta) = \hat{\mathbb{E}}_t \left[ L_t^{CLIP}(\theta) - c_1 L_t^{VF}(\theta) + c_2 S[\pi_\theta](s_t) \right], \quad (2.8)$$

onde  $c_1$  e  $c_2$  são coeficientes,  $S$  denota um bônus de entropia, e  $L_t^{VF}$  é uma perda de erro quadrático médio para a função de valor dada por, com definição em 2.9:

$$L_t^{VF} = \left( V_\theta(s_t) - V_t^{alvo} \right)^2. \quad (2.9)$$

O primeiro termo,  $L_t^{CLIP}(\theta)$ , representa a função objetivo 2.4. O segundo termo,  $-c_1 L_t^{VF}(\theta)$ , é a perda de valor que tenta minimizar o erro na estimativa do valor de estado. O terceiro termo,  $c_2 S[\pi_\theta](s_t)$ , adiciona um bônus de entropia para incentivar a exploração adequada pelo agente.

## 2.5 Estado da arte: métodos de estimulação cerebral aplicada no tratamento de DP

As abordagens de Estimulação Cerebral Profunda (DBS) no tratamento da Doença de Parkinson (DP) abrangem diversas fases: pré-operatória, operatória e pós-operatória. O foco deste trabalho reside na adaptação dos estímulos em caráter pós-operatório.

No contexto atual, a inovação em estímulos adaptativos emerge como um desafio significativo, com várias abordagens sendo exploradas na literatura recente. Para compreender o estado da arte, realizou-se um levantamento bibliográfico nos últimos 5 anos, destacando as principais abordagens identificadas, conforme sumarizado na Tabela 1.

A metodologia de busca incorporou análises em duas plataformas com três *queries* distintas:

- No PubMed, com as consultas: “((deep brain stimulation) AND (deep learning) AND (parkinson))”; e “(parkinson AND machine learning AND (adaptive deep brain stimulation OR closed-loop deep brain stimulation))”;



- No Google Scholar, com a consulta: “(‘deep brain stimulation’ and ‘deep learning’ and ‘parkinson’)”;

Essas consultas visaram identificar trabalhos que abordassem o uso de técnicas avançadas de aprendizado de máquina e aprendizado profundo na otimização da DBS para pacientes com DP, especialmente no contexto pós-operatório.

Além dessas buscas principais, realizou-se consultas pontuais por artigos envolvendo estratégias de controle.

### 2.5.1 Estratégias Avançadas de Estimulação Cerebral para DP

A evolução das estratégias de Estimulação Cerebral Profunda (DBS) em resposta à Doença de Parkinson (DP) tem sido influenciada significativamente pelo avanço das tecnologias de aprendizado de máquina, aprendizado por reforço e sistemas de controle. Estas tecnologias abrangem uma variedade de métodos computacionais que permitem a personalização e otimização do tratamento de DBS. A Tabela 1 apresenta uma compilação dos métodos mais recentes e suas métricas principais, refletindo a diversidade e a inovação no campo.

Tabela 1 – Artigos nos últimos 5 anos com Estimulação Cerebral Profunda adaptativa em DP.

Autores	Método	Métrica principal
Su <i>et al.</i> (2021)	Controlador P	Frequência de oscilação
Bahadori-Jahromi <i>et al.</i> (2023)	Controlador P	Frequência de oscilação
Liu <i>et al.</i> (2021)	Controlador PI	Taxa de disparos de potencial e energia expandida
Yu <i>et al.</i> (2022)	<i>Direct Delayed Feedback</i>	Frequência de oscilação
Stojsavljevic <i>et al.</i> (2023)	Estímulo On-Off	Taxa de disparos de potencial
He <i>et al.</i> (2021)	Estímulo On-Off	Frequência de oscilação
Mohammed <i>et al.</i> (2020)	Fuzzy e Método Ator-Crítico	Acurácia e Latência
Su <i>et al.</i> (2023)	Controlador CAR-fuzzy	Frequência de oscilação
Kerman <i>et al.</i> (2022)	<i>Spiking Neural Networks (SNNs)</i>	Latência e energia expandida
Liu <i>et al.</i> (2020)	Rede Neural + Controlador PID	Energia expandida
Boutet <i>et al.</i> (2021)	ML com processamento de imagens	Teste de imagens
Shah <i>et al.</i> (2024)	ML com processamento de imagens	Observação de eletroencefalografia
Hirschmann <i>et al.</i> (2022)	ML com processamento de imagens	Observação de magnetoencefalografia
Merk <i>et al.</i> (2022)	ML com dados de estímulo cerebral	Frequência de oscilação
Kleinholdermann <i>et al.</i> (2023)	ML (Floresta Aleatória)	Validado clinicamente
Gao <i>et al.</i> (2020)	Aprendizado por Reforço	Validado clinicamente
Lu <i>et al.</i> (2020)	Aprendizado por Reforço	Energia expandida

Fonte: Elaboração Própria, 2024.

A revisão aborda técnicas de aprendizado por reforço e outras metodologias de *machine learning* e de controle devido às relevâncias e aplicabilidades no ajuste fino dos parâmetros de DBS. Estes métodos oferecem benefícios como a capacidade de adaptar-se em

tempo real às condições do paciente, o que é vital para o tratamento de uma doença tão complexa e variável como a DP.

A variabilidade de metodologia dos artigos chama atenção mesmo para aqueles em que os métodos são similares, visto que a abordagem varia entre modelos neuronais e bases de dados reais. Comparativamente à proposta deste trabalho, os seguintes artigos representados na Tabela 2 usaram em sua maioria modelos próprios e complexificados entre mais de uma região do cérebro.

Tabela 2 – Modelos utilizados pela literatura.

Autores	Método	Modelo
Su <i>et al.</i> (2021)	Controlador P	Córtex-Basal Gânglio-Tálamo
Bahadori-Jahromi <i>et al.</i> (2023)	Controlador P	Cortico-BG-Talâmico
Liu <i>et al.</i> (2021)	Controlador PI	Modelo Gânglio-Basal próprio
Yu <i>et al.</i> (2022)	<i>Direct Delayed Feedback</i>	Modelo Gânglio-Basal próprio
Stojsavljevic <i>et al.</i> (2023)	Estímulo On-Off	Modelo Gânglio-Basal próprio
Mohammed <i>et al.</i> (2020)	Fuzzy e Método Ator-Crítico	Modelo Gânglio-Basal próprio
Su <i>et al.</i> (2023)	Controlador CAR-fuzzy	Córtex-Basal Gânglio-Tálamo
Liu <i>et al.</i> (2020)	Rede Neural + Controlador PID	Modelo Gânglio-Basal próprio
Lu <i>et al.</i> (2020)	Aprendizado por Reforço	Hodgkin-Huxley

Fonte: Elaboração Própria, 2024.

Dessa forma, a metodologia aplicada neste trabalho e explicada na próxima seção visa elaborar um modelo de agente de RL particular para o GPi e não explora a relação entre esse estímulo e outras regiões do cérebro, tendo em vista o foco em entender e projetar, primeiramente, um agente de aprendizado por reforço e sua capacidade de adaptar as correntes de estímulo numa região específica.

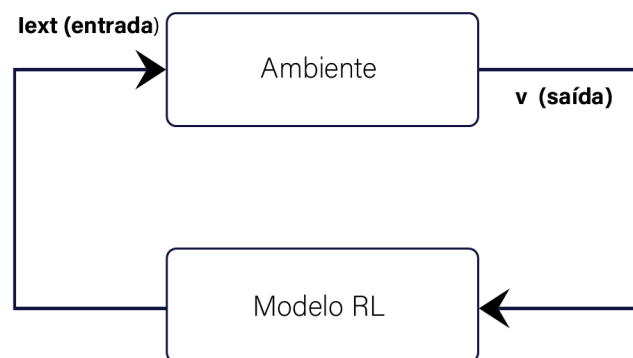
### 3 METODOLOGIA

Neste capítulo, são detalhadas as metodologias computacionais e o arcabouço teórico utilizados para o desenvolvimento e a validação do algoritmo de Aprendizado por Reforço aplicado ao tratamento da doença de Parkinson através da DBS. Primeiramente, é abordado o panorama do trabalho, configurações experimentais e linha de projeto adotada. Em seguida, o ambiente criado a partir do modelo de Izhikevich e seus métodos. Logo após, são descritas a aplicação da Otimização de Política Proximal e a arquitetura escolhida para o agente de treinamento. Por fim, são descritos os parâmetros utilizados na avaliação de desempenho do modelo.

#### 3.1 Panorama geral

A Figura 9 representa o panorama geral do trabalho. Uma corrente externa é aplicada ao ambiente moldado por neurônios de Izhikevich e a saída ou estados do ambiente, em potenciais de membrana, é aplicada ao modelo treinado de RL, o qual emite uma nova corrente adaptada ao novo estado. O bloco “Modelo RL” é detalhado na Figura 10.

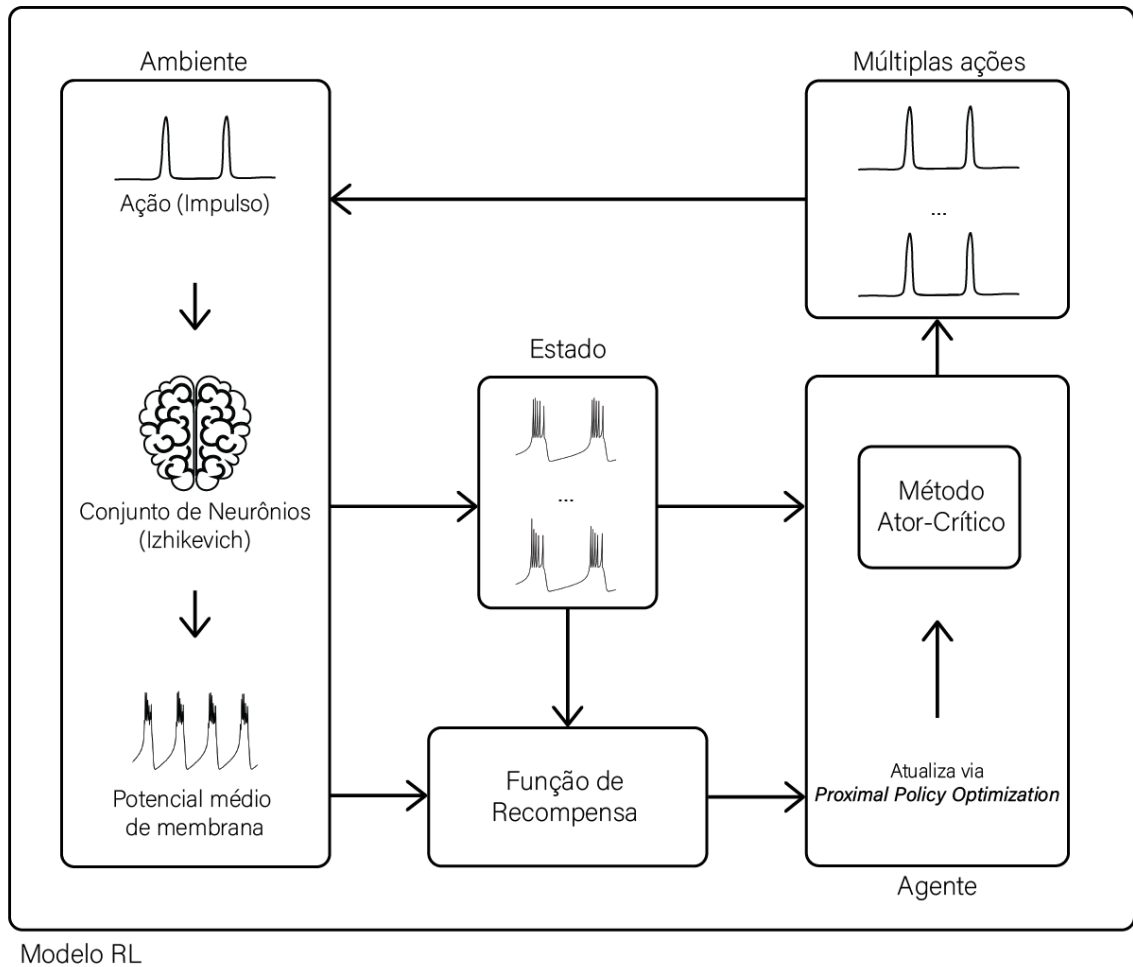
Figura 9 – Diagrama de blocos geral.



Fonte: Elaboração Própria, 2024.

O projeto teve como objetivo a criação de um modelo de aprendizado por reforço que otimizou a aplicação de correntes elétricas em neurônios do modelo de Izhikevich, visando o controle dos sintomas de tremor da doença de Parkinson. A metodologia aplicada é adaptada a partir da ideia proposta por Krylov *et al.* (2021), onde houve a utilização de aprendizado por reforço para atenuação de oscilações nos modelos de Hindmarsh-Rose e Bonhoeffer-van der Pol

Figura 10 – Esquemático geral da metodologia.



Fonte: Elaboração Própria, 2024.

(HINDMARSH *et al.*, 1984; BONHOEFFER, 1948).

Na Figura 10 é mostrado como o trabalho propôs abordar esse problema através de uma visão de blocos. O Ambiente é o bloco responsável por receber os impulsos elétricos (ações) e aplicar sobre o conjunto de neurônios de Izhikevich. Com essas aplicações, pôde-se observar os potenciais de membrana dos neurônios, no bloco Estado, e realizar a medição do potencial médio. A partir dessas informações, calculou-se uma recompensa baseada na variância dos estados e nas ações aplicadas. No bloco Agente, o *PPO* atua em conjunto com o método Ator-Crítico para atualizar a política de ações a serem aplicadas novamente no Ambiente e, assim, fechar a malha.

Como linha de base do projeto, implementou-se os passos:

1. Importação das bibliotecas necessárias;
2. Criação do ambiente utilizando o modelo de Izhikevich;

3. Testagem do ambiente para garantir que está funcionando corretamente;
4. Criação e treinamento do modelo utilizando PPO;
5. Ajuste fino dos parâmetros usando busca em grade;
6. Armazenamento do modelo treinado;
7. Avaliação e teste do modelo para verificar o desempenho.

### 3.2 Configuração Experimental

Este trabalho foi redigido em numa máquina com as configurações mostradas na Tabela 3.

Tabela 3 – Configurações da máquina.

Processador	Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz 2.50 GHz
Memória RAM	8 GB
Placa de vídeo dedicada	NVIDIA® GeForce GTX 1650 4GB
Sistema Operacional	Windows® 11 Pro 64 bits Versão 22h2

Fonte: Elaboração Própria, 2024.

Além disso, houve necessidade de instalação dos *softwares* mostrados na Tabela 4, com destaque para as bibliotecas e pacotes essenciais (RAFFIN *et al.*, 2021).

Tabela 4 – Configurações de *softwares* e pacotes essenciais.

Python	Versão 3.9.0 64 bits
Pytorch	Stable 2.2.0 CUDA 11.8
Stable Baselines3	Versão 2.2.1

Fonte: Elaboração Própria, 2024.

O tempo de simulação do ambiente e de treinamento são diretamente influenciados pelas configurações da máquina. Dessa forma, buscou-se otimizar os recursos e utilizou-se o *Jupyter notebook* pela facilidade de importação de bibliotecas e registro automático de manutenção do código.

### 3.3 Criação do ambiente proposto de neurônios de Izhikevich

---

#### Algoritmo 2: Ambiente de Izhikevich para RL

---

**Dados:** Número de neurônios, tamanho do episódio, desvio padrão do ruído, parâmetro

beta

**Resultado:** Ambiente de simulação de Izhikevich para uso em RL

/\* Definição da classe do ambiente de Izhikevich \*/

**Classe *IzhikevichEnv*:**

/\* Inicialização do ambiente \*/

**Inicialização *\_\_init\_\_(num\_neurons, ep\_length, noise\_std, beta):***

Defina os parâmetros da simulação

Inicialize os parâmetros do modelo de Izhikevich

Defina os espaços de ação e observação

**Método *seed(seed):***

Defina a semente do gerador de números aleatórios

**Método *step(action):***

Aplique as perturbações como correntes externas

Atualize a dinâmica dos neurônios usando método de Euler

Implemente o mecanismo de disparo

Calcule a recompensa

Atualize o passo atual e verifique se o episódio terminou

Organize as observações para retornar

**Método *reset():***

Redefina os estados dos neurônios

**Método *render(mode, close):***

/\* Renderização do ambiente, se necessário \*/

**Função *\_calculate\_reward(action):***

Calcule a recompensa com base na ação e no estado

---

Para criação e inicialização do ambiente, a classe de neurônios obedeceu ao parâmetros estabelecidos na Tabela 5, com base nos valores fornecidos em (LIU *et al.*, 2015) e adaptados para essa aplicação.

Para a inicialização do modelo, o espaço de ações (corrente externa em pA) foi adotado como variável *float* entre -0,5 e 0,5. O espaço de observações (potencial de membrana

Tabela 5 – Parâmetros para modelagem dos neurônios de Izhikevich.

Parâmetro	Conceito	Valor
N	Número de neurônios	1000
a	Taxa de decaimento de variável de recuperação $u$ .	0,01
b	Sensibilidade de variável de recuperação $u$ .	0,585
c	Valor de <i>reset</i> do potencial $v$ .	-45
d	Valor de <i>reset</i> da variável de recuperação $u$ após disparo.	4
I	Corrente (pA) composta pela soma de interações sinápticas.	15
Desvio padrão	Perturbação adicionada seguindo a distribuição normal.	0,005

Fonte: Adaptado de (LIU *et al.*, 2015).

em mV) foi definido como variável *float* entre -100 e 40. Além disso, o tamanho do episódio foi de 5000 *steps*, escolhido como um valor suficiente para o reconhecimento e exploração da política.

### 3.3.1 Método step

Durante esse trecho, a ação  $I_{ext}$  representa a corrente externa a ser aplicada na soma do potencial de membrana, como exposto na Equação 3.1:

$$v' = \underbrace{0,04v^2 + 5v + 140 - u + I}_{\text{equação padrão}} + \underbrace{I_{ext}}_{\text{ação}}. \quad (3.1)$$

Assim, a corrente total aplicada ao neurônio é a soma entre as interações sinápticas  $I$  e a corrente externa  $I_{ext}$  advinda do DBS, esta aplicada pelo agente do modelo. Como o máximo adotado nesse caso para a aplicação foi 0,5 pA, o que o algoritmo de aprendizado por reforço teve como missão foi definir a intensidade e a frequência da aplicação. Esta ação é ajustada iterativamente com o objetivo de restaurar o padrão de disparo regular dos neurônios afetados pelos disparos em rajada associados à doença de Parkinson.

A atualização da dinâmica dos neurônios foi realizada através do método de Euler, com base na Equação 3.2:

$$y_{n+1} = y_n + f(t_n, y_n) \cdot \Delta t, \quad (3.2)$$

em que  $\Delta t$  é o tamanho do passo, definido em 0,01 s, e  $n$  indica o índice do passo atual.

### 3.3.2 Método reset

Uma vez que o episódio é concluído, os valores de potencial de membrana e corrente sináptica são atualizados para os valores iniciais. Processo exposto a seguir tendo os valores da Tabela 5 como parâmetros.

$$\text{se } v \geq 30 \text{ mV, então } \begin{cases} v \leftarrow -45 \text{ mV} \\ u \leftarrow u + 4 \end{cases}$$

### 3.3.3 Métodos *render*, *seed* e *close*

Os demais métodos compreendem conceitos de implementação do algoritmo de RL (SUTTON; BARTO, 2018). O método *render* é responsável pela visualização do estado atual do ambiente. Nesse caso, optou-se por uma visualização posterior ao treinamento. O método *seed* estabelece uma semente para o gerador de números aleatórios do ambiente. Isso é crucial para garantir a reprodutibilidade dos experimentos, pois a simulação pode envolver processos estocásticos, como a introdução de ruído nos parâmetros dos neurônios. Ao definir uma semente fixa, assegura-se que cada execução da simulação produza os mesmos resultados sob as mesmas condições iniciais. Por fim, o método *close* é utilizado para finalizar a simulação e liberar recursos, especialmente para fechar janelas gráficas. Embora não haja necessidade nesse código, para futuras expansões, é uma boa prática manter esse método para compatibilidade com padrões de ambientes de simulação.

Esses métodos são parte integrante do ciclo de vida do ambiente de simulação, permitindo controle preciso sobre a execução, visualização e finalização das simulações realizadas.

### 3.3.4 Função de Recompensa

A função de recompensa é projetada para promover a sincronização neuronal próxima ao padrão natural, penalizando desvios causados por padrões anormais de disparo. A recompensa é calculada com base na diferença entre os padrões de disparo observados e os padrões desejados. Com efeito, o objetivo é recompensar a dessincronização entre os neurônios, visto que é um padrão encontrado em condições de DP (LEVY *et al.*, 2002; PLENZ; KITAL, 1999).

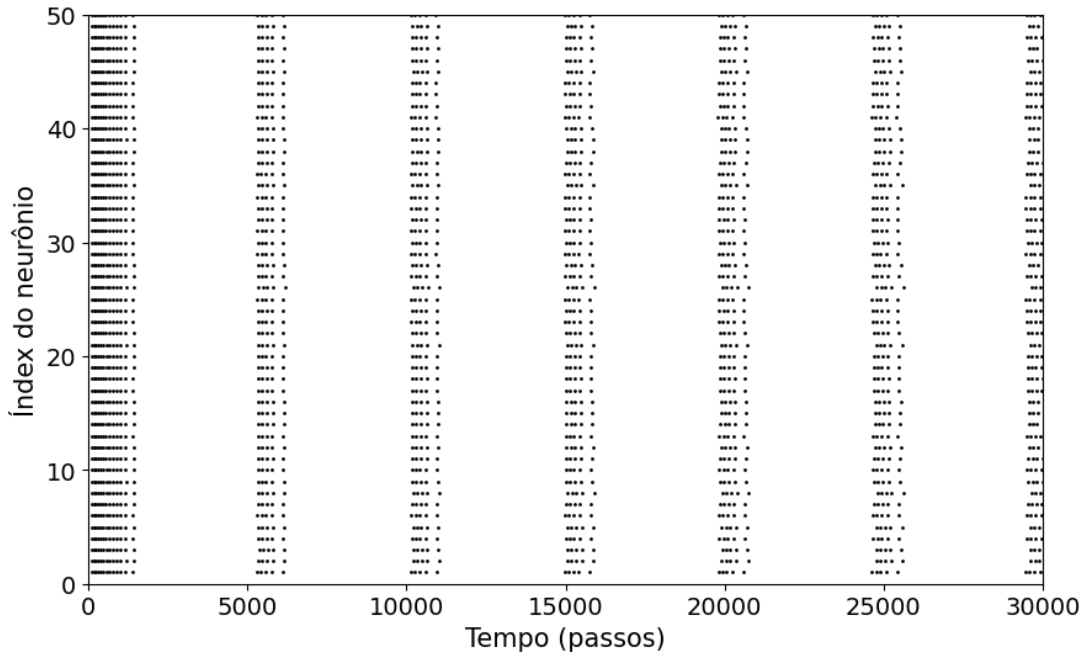
A fim de ilustrar esse comportamento, a Figura 11 demonstra a sincronização para 50 neurônios do GPi em estado parkinsoniano, a partir dos valores da Tabela 5.

Com ênfase a facilitar o entendimento da sincronização, expandido para 100 neurônios e focado nos intervalos de 200 ms a 650 ms, o comportamento dos potenciais de membrana dos neurônios em sincronia é observado na Figura 12.

Por fim, para fins de comparação, a Figura 13 demonstra o comportamento de 50



Figura 11 – Disparo sincronizado de neurônios em sintomas de DP.



Fonte: Elaboração Própria, 2024.

neurônios em condições fisiológicas (LIU *et al.*, 2015), com  $c = -65$  no modelo de Izhikevich.

Similarmente ao exposto na Figura 12, para condições fisiológicas, a Figura 14 demonstra os disparos regulares encontrados para condições saudáveis. Nota-se que os picos de tensão são também mais espaçados.

Tendo em vista que a intenção da recompensa é promover o comportamento dessincronizado entre os neurônios, foi adotada a Equação 3.3, adaptada de Krylov *et al.* (2021):

$$R[t] = \text{Var}(v(t)) + \beta \cdot |A(t)|, \quad (3.3)$$

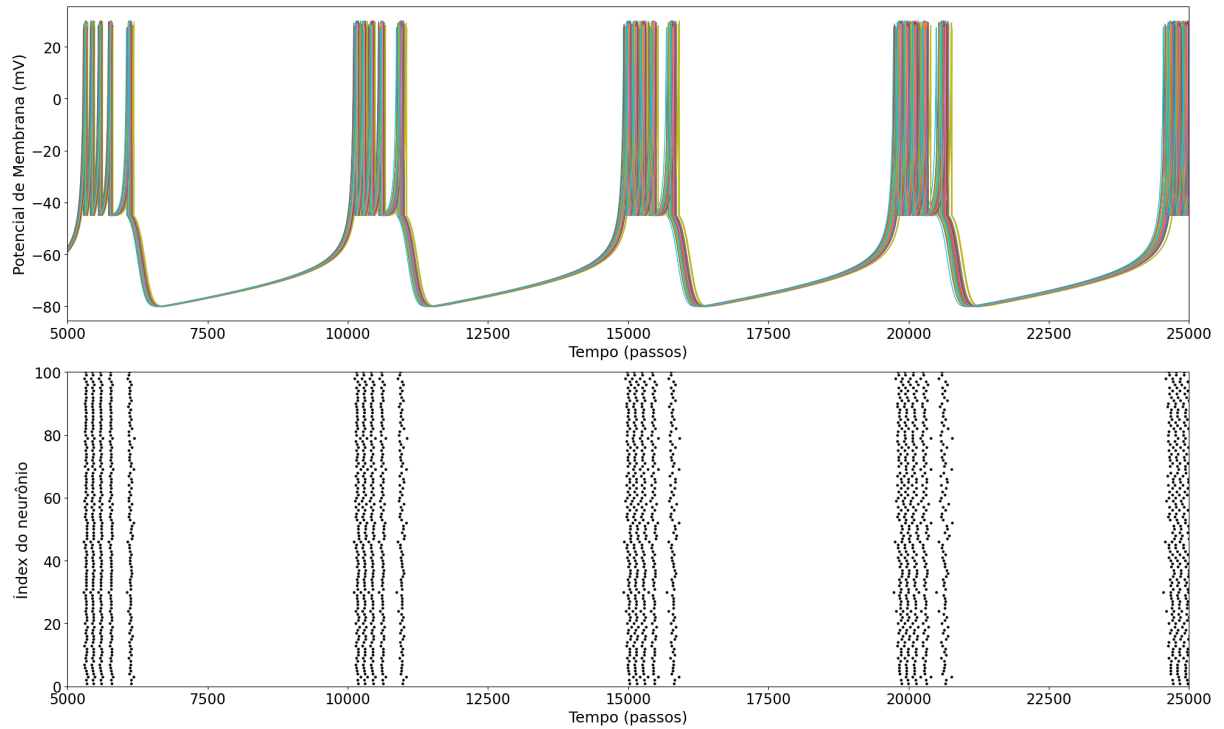
onde o primeiro valor representa a variância entre os potenciais no instante  $t$ , definida por:

$$\text{Var}(v(t)) = \frac{\sum (v_i - \bar{v})^2}{N}, \quad (3.4)$$

em que  $v_i$  é o potencial no neurônio  $i$ ,  $\bar{v}$  é a média de potencial no instante  $t$  e  $N$  é o número de neurônios. Uma vez que a variância é desejada para uma maior dessincronização, esse fator é recompensado na política.

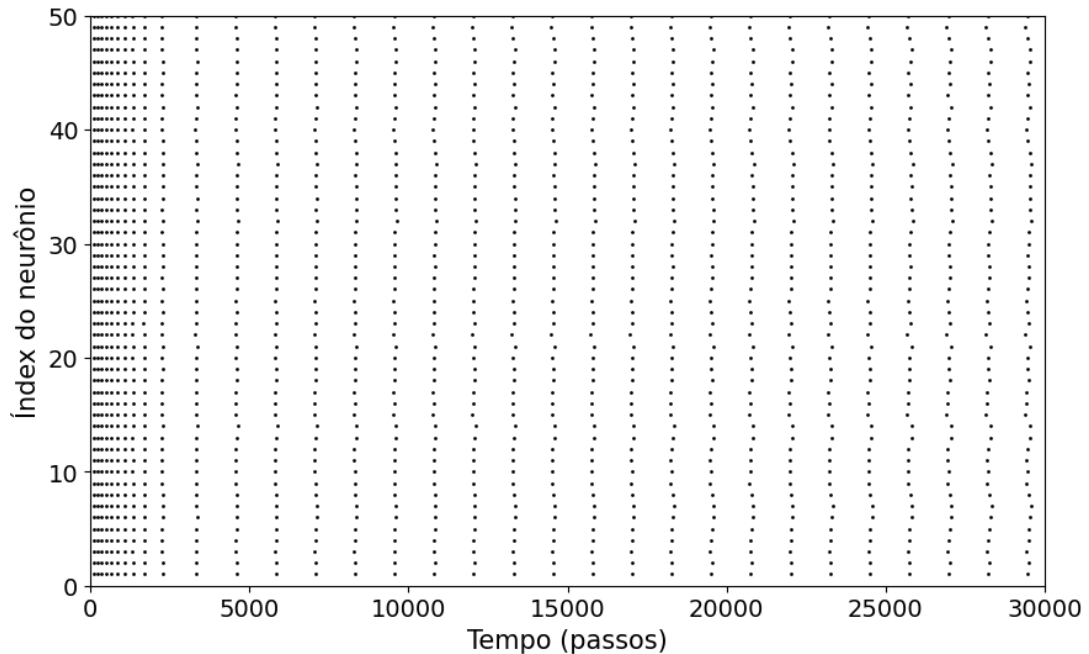
O segundo componente recompensa valores de ação  $A$  relacionados com os valores observados nos potenciais onde a recompensa é maior. O coeficiente  $\beta$  permite a adição de *bias* a fim de padronizar as ordens de grandeza entre as duas componentes da função.

Figura 12 – Sincronização de neurônios em sintomas de DP.



Fonte: Elaboração Própria, 2024.

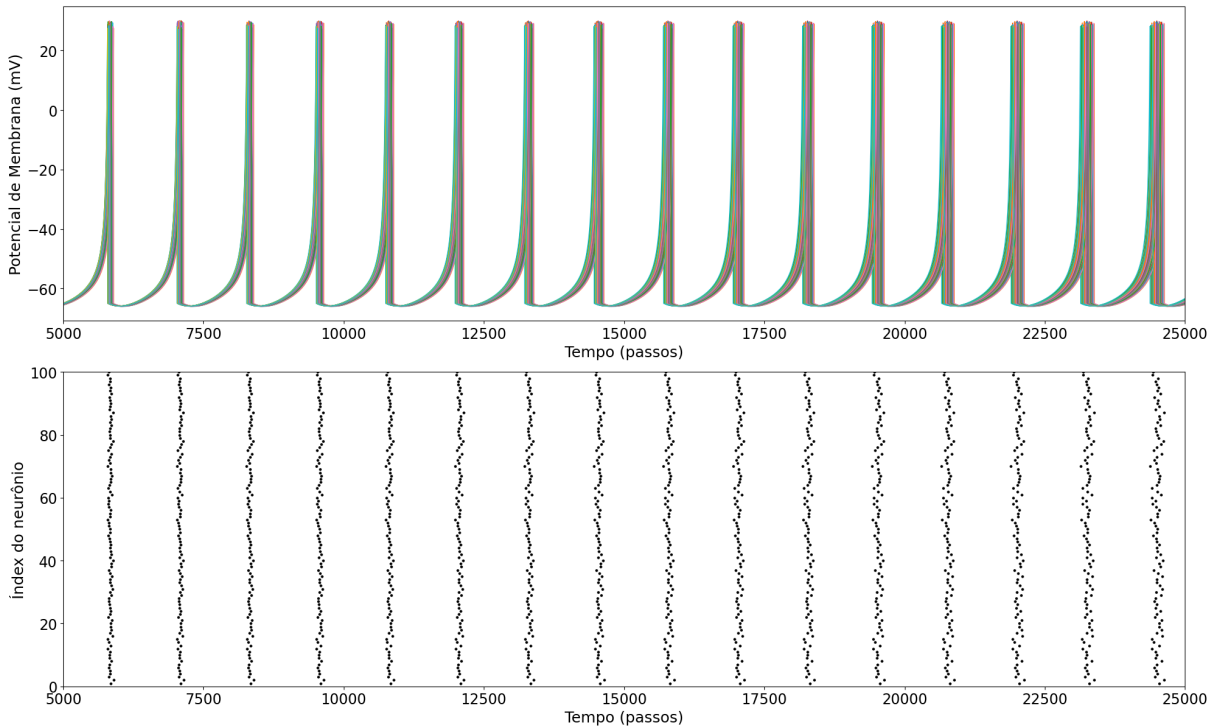
Figura 13 – Disparos em condições saudáveis.



Fonte: Elaboração Própria, 2024.

É esperado que a recompensa inicie de um número de alta escala, dado que é diretamente proporcional ao número de neurônios do treino. Assim, as recompensas individuais são somadas e representam a recompensa do conjunto.

Figura 14 – Dessincronização de neurônios em condições saudáveis.



Fonte: Elaboração Própria, 2024.

### 3.4 Parâmetros e Desempenho do modelo

Os hiperparâmetros do algoritmo de RL, incluindo taxa de aprendizado, desconto de recompensa futura e parâmetros específicos da política PPO, são ajustados durante a fase de validação para alcançar a convergência ótima e o equilíbrio entre exploração e exploração no espaço de ações.

A metodologia proposta é validada em um ambiente computacional controlado, permitindo a replicação e a verificação dos resultados obtidos. O desempenho do modelo de RL é avaliado com base na sua capacidade de regular os padrões de disparo dos neurônios de maneira a mitigar os sintomas da doença de Parkinson.

A seguir, são detalhados os aspectos levados em consideração:

- A divergência de Kullback-Leibler (KL), uma medida estatística utilizada pelo PPO para quantificar a diferença entre a política antiga e a nova. A divergência KL é usada para regular a magnitude das atualizações da política, garantindo que as mudanças sejam pequenas e estáveis, evitando oscilações extremas no aprendizado (??). Uma divergência alta indica uma alta exploração.
- O número de lotes, referente ao número de subconjuntos de dados em que o conjunto

de treinamento é dividido. No contexto do PPO, um tamanho de lote adequado ajuda a balancear a eficiência do treinamento e a qualidade das atualizações da política. Um tamanho de lote maior pode aumentar a estabilidade do treinamento, mas também exige mais recursos computacionais (WU, 2022).

- Gamma ( $\gamma$ ), fator de desconto utilizado no cálculo do retorno descontado. Ele influencia o quanto as recompensas futuras são valorizadas em relação às recompensas imediatas. Um valor de  $\gamma$  próximo de 1 faz com que o agente valorize mais as recompensas futuras, enquanto um valor menor dá mais ênfase às recompensas imediatas (ZOU, 2016).
- Passos de treinamento, do inglês *Time steps*, referente ao número de passos de tempo que o agente interage com o ambiente durante o treinamento. Um número maior de timesteps permite que o agente explore mais o ambiente e aprenda com uma gama mais ampla de experiências, mas também aumenta o tempo de treinamento (SONG, 2021).
- A taxa de aprendizado, determinante da magnitude das atualizações dos parâmetros da política a cada passo do treinamento. Uma taxa de aprendizado mais alta pode acelerar o treinamento, mas também pode causar instabilidade e impedir a convergência. Por outro lado, uma taxa mais baixa pode resultar em convergência mais estável, mas torna o processo de treinamento mais lento (JAQUES *et al.*, 2019).

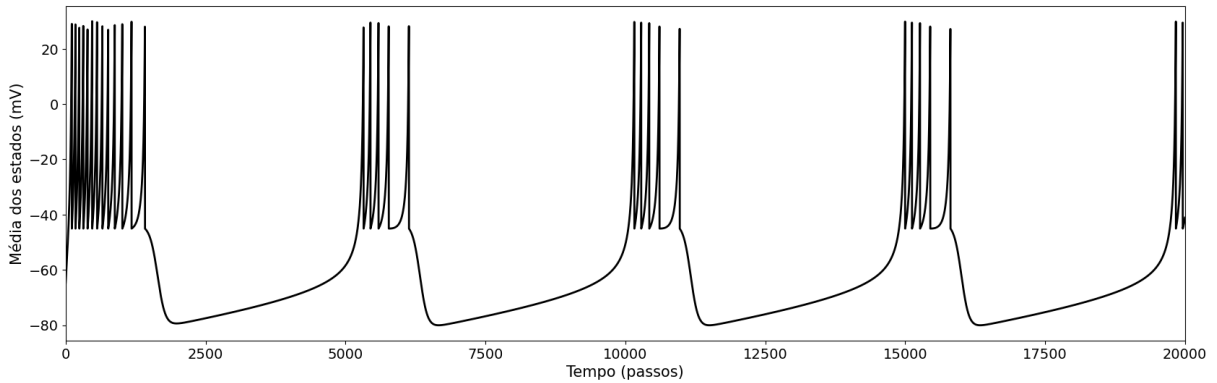
### 3.5 Análise da atenuação

A implementação do modelo avalia se o algoritmo efetivamente age no processo de dessincronização e, comparativamente a outros métodos, age de maneira mais eficiente. Para analisar essa ação, foi utilizada a média entre os potenciais das membranas ao longo do tempo para observar-se a atenuação, além do índice de gasto energético e a relação de desvios padrões das médias dos estados.

A média dos potenciais dos neurônios é uma forma de observar em conjunto como os neurônios se comportam em relação à sua sincronia, como mostrado na Figura 15.

O índice de gasto energético é utilizado para avaliar o custo energético do estimulador utilizado nos sistemas de Estimulação Cerebral Profunda. No contexto da DBS em circuito fechado, o gasto energético é proporcional ao quadrado do valor eficaz, do inglês *root mean square (RMS)*, da corrente, calculado por:

Figura 15 – Média de estados (mV).



Fonte: Elaboração Própria, 2024.

$$E_{\text{RMS}} = N \sqrt{\frac{1}{t_{\text{max}} - t_1} \int_{t_1}^{t_{\text{max}}} I_{\text{ext}}^2(t) dt} \quad (3.5)$$

em que  $N$  é o número de neurônios estimulados,  $t_{\text{max}}$  é o tempo máximo de estimulação e  $t_1$  é o início da estimulação. Uma vez que há comparações entre o estímulo uniforme e o proposto neste trabalho, a energia é uma métrica importante para indentificar custo-benefício entre os métodos, além de ser proporcional ao valor efetivo da corrente de estimulação, sendo válida para a análise (GORZELIC *et al.*, 2013).

Outra forma de analisar a efetividade do sistema é através da relação entre desvios padrões das médias dos estados antes e após a ação da corrente de controle. A medida é proposta em Krylov *et al.* (2021) e é definida por:

$$S = \frac{\text{std}[X_{\text{antes}}]}{\text{std}[X_{\text{depois}}]}. \quad (3.6)$$

em que  $X_{\text{antes}}$  é a média dos estados antes da ação do agente e  $X_{\text{depois}}$  é a média dos estados após a ação. O operador *std* representa o cálculo de desvio padrão sobre as médias.

### 3.6 Estudo de frequência de passos e análise comparativa com sinal DBS padrão

Após análise e validação do modelo, foi estendida uma modificação na aplicação a fim de minimizar o Índice de Gasto Energético, representado na Equação 3.5 ao passo que a eficiência, representada pela equação 3.6, não fosse impactada demasiadamente. O objetivo foi entender se a aplicação da ação do agente em todos os passos de tempo afeta a efetividade da solução.

Por fim, a fim de comparar a eficiência do modelo frente um sinal DBS padrão, foram realizadas comparações de aplicações frente ao sinal DBS-RL, proposto neste trabalho.

### **3.7 Adaptabilidade**

Na definição dos neurônios de Izhikevich, foi adicionado um ruído para os parâmetros  $a$ ,  $b$ ,  $c$  e  $d$  dos neurônios do conjunto, seguindo uma distribuição normal. O ruído foi definido durante a execução da classe e atualizado a cada iteração. Dessa forma, o modelo foi treinado e validado levando em consideração a variabilidade dos neurônios de um conjunto.

A partir dessa metodologia definida, buscou-se implementar o *framework* proposto em Krylov *et al.* (2021) para o modelo de Izhikevich em conjunto a métodos de avaliação encontrados na literatura, como a análise de energia, contribuindo para os resultados encontrados na próxima seção.

## 4 RESULTADOS

Neste capítulo, são apresentados os resultados do aprendizado do algoritmo com relação ao controle de impulsos elétricos para um conjunto de 1000 neurônios localizados no Glóbulo Pálido interno (GPi), local escolhido para simular a inserção do eletrodo de Estimulação Cerebral Profunda. Além disso, são expostos os resultados de simulação para o controle realizado e a redução do potencial médio dos neurônios. Em seguida, é realizada a análise dos parâmetros escolhidos considerando a frequência da estimulação e a energia expendida. Na última seção, são discutidos os resultados.

### 4.1 Treinamento do modelo.

O processo de treinamento do modelo foi meticulosamente elaborado, iniciando com uma análise exploratória de dados para identificar tendências, padrões e anomalias que poderiam influenciar o comportamento do algoritmo. Em seguida, implementou-se um *grid search* sistemático para experimentar uma gama de valores candidatos e determinar a configuração ótima de parâmetros, conforme ilustrado na Tabela 6.

Tabela 6 – Parâmetros de treinamento

Parâmetro	Instância	Valor
Tamanho de episódio	Ambiente	5000 passos
$\beta$	Função de Recompensa	0,5
$\gamma$	Agente	0,99
Passos de Treinamento	Agente	100000 passos
Tamanho de lote	Agente	1024 passos
Taxa de aprendizagem	Agente	Função: [0,01 $\rightarrow$ 0,001]

Fonte: Elaboração Própria, 2024.

Cada parâmetro foi selecionado com base na sua contribuição para a eficiência e eficácia do algoritmo.

O tamanho do episódio foi escolhido tendo em vista que 5000 passos foram suficientes para que as equações de Izhikevich promovessem os potenciais de ação e que a ação imposta fosse observada, equilibrando, também, um tamanho de episódio eficiente computacionalmente.

O parâmetro  $\beta$  foi escolhido através do método *grid search* e, também, foi levado em consideração a padronização de ordens de grandeza entre os componentes da função de recompensa 3.3: variância dos estados e ações implementadas. Utilizou-se o método *evaluate – policy* da biblioteca *Stable-baselines3* para avaliar o desvio padrão e a recompensa em 10 épocas.

Dessa forma, o valor final escolhido que melhor equilibrou os resultados foi 0,5.

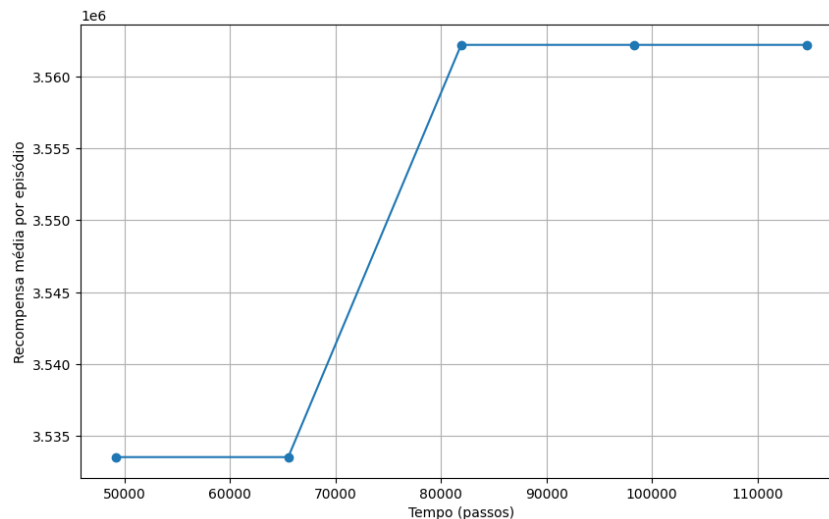
O parâmetro  $\gamma$  foi escolhido dado um critério de projetista, escolhido próximo a 1 para que as recompensas futuras fossem compensadas tanto quanto as imediatas, a fim de promover um aprendizado mais rápido.

Os passos de treinamento foram definidos em 100000 para equilibrar a taxa de aprendizagem e a média de recompensa por episódio treinado. Na Figura 16 é demonstrada uma inclinação positiva durante o treinamento e atingimento do platô ao final, demonstrando um tamanho de passos de treinamento propício para a estabilidade do aprendizado.

O tamanho de lote foi escolhido de forma a acentuar a velocidade de iteração do algoritmo. Usualmente, é utilizado um tamanho equivalente a 64 passos, mas a fim de promover a rapidez do treinamento, foi observado que 1024 passos mostrou-se um valor propício sem que interferisse na função de recompensa de forma acentuada.

A taxa de aprendizagem foi escolhida como uma função linear que começa em um valor mais alto para explorar o espaço de ação amplamente e, gradualmente, diminui para permitir uma convergência refinada. Este método assegura um equilíbrio entre exploração e exploração, crucial para a aprendizagem em ambientes dinâmicos.

Figura 16 – Função de recompensa durante treinamento.



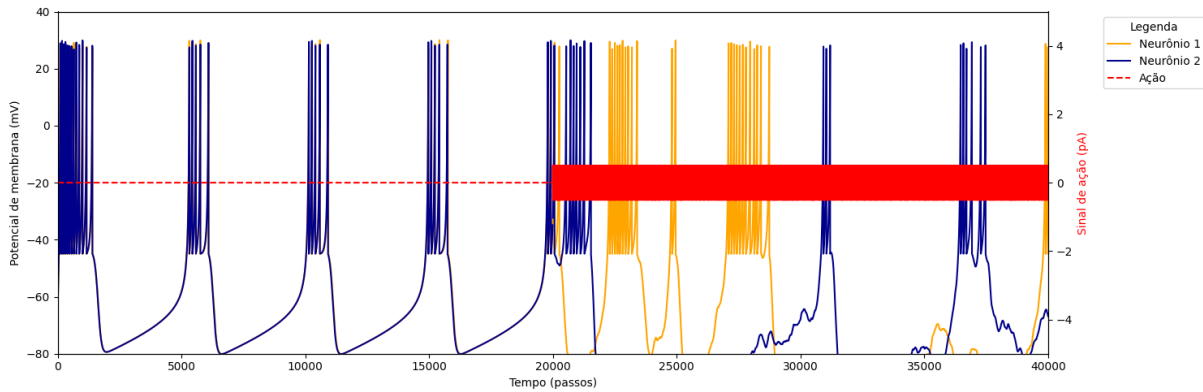
Fonte: Elaboração Própria, 2024.

## 4.2 Análise gráfica da dessincronização

A análise visual da atuação do modelo na dessincronização é representada na Figura 17, demonstrada em apenas 2 neurônios para facilitar a visualização.



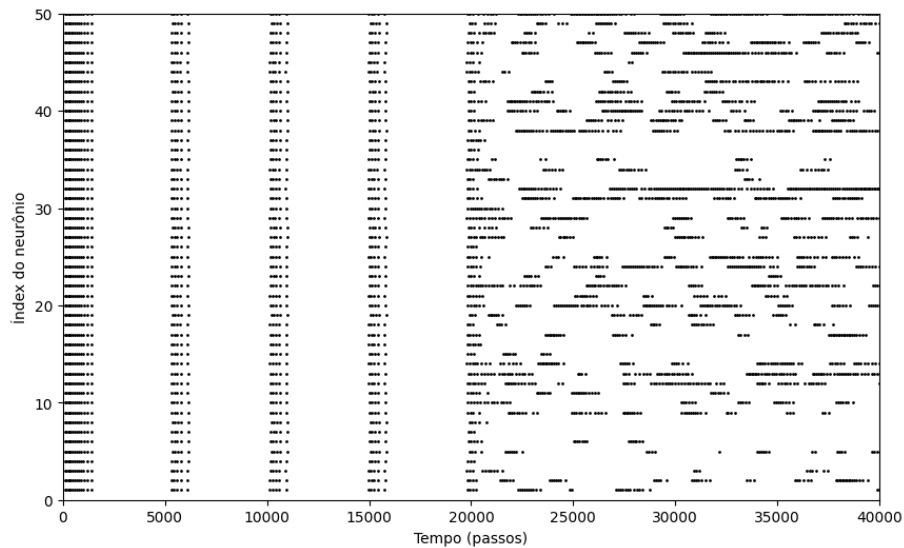
Figura 17 – Atuação do sinal de ação.



Fonte: Elaboração Própria, 2024.

Ainda, a fim de visualizar a dessincronização do conjunto, a Figura 18 representa para 50 neurônios a atuação do sinal de ação no passo 20000.

Figura 18 – Disparos dessincronizados após ação.



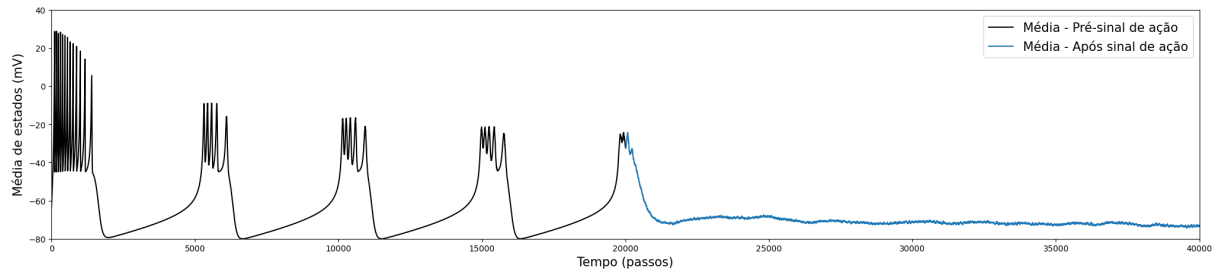
Fonte: Elaboração Própria, 2024.

Em linhas gerais, a dessincronização também pode ser visualizada através da média dos potenciais do conjunto de neurônios, ilustrado para 1000 neurônios na Figura 19. Semelhante à média, a variância é onde a função de recompensa - Equação 3.3, atua para que essa atenuação ocorra.

### 4.3 Visão analítica

Fazendo uso da Equação 3.6, foi calculada a relação de desvios padrões das médias dos estados antes e após a ação da corrente aplicada:

Figura 19 – Atenuação da média dos potenciais.



Fonte: Elaboração Própria, 2024.

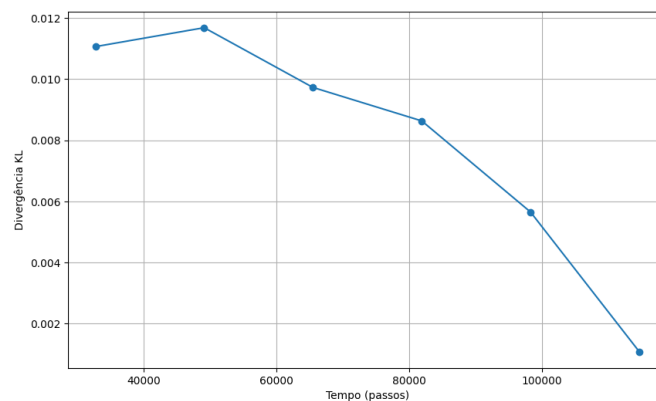
$$S = \frac{std[X_{antes}]}{std[X_{depois}]} = \frac{19,75}{5,95} = 3,32.$$

A energia expendida pela corrente do agente também foi calculada através da Equação 3.5:

$$E_{RMS} = 0,68 \text{ J.}$$

Observou-se uma diminuição significativa na divergência KL, conforme ilustrado na Figura 20, indicativo de uma redução na exploração do agente à medida que o treinamento avança. Inicialmente, com uma taxa de aprendizagem elevada, o agente demonstra uma tendência a explorar amplamente. Contudo, com a diminuição linear desta taxa, a divergência KL também decresce, indicando uma abordagem mais conservadora na relação entre as políticas nova e antiga, refletindo o ajuste fino entre exploração e exploração no aprendizado por reforço.

Figura 20 – Divergência KL durante treinamento.



Fonte: Elaboração Própria, 2024.

#### 4.4 Estudo da frequência de aplicação do sinal de ação.

Na busca por equilibrar a eficácia do tratamento e a conservação de energia, o modelo adota uma estratégia de sinal de pulso com passagem definida pelo intervalo  $k$ . Esta abordagem permite que o estímulo seja aplicado somente após  $k$  passos, com o objetivo de analisar a relação custo-benefício entre a dessincronização neuronal eficaz e o consumo energético.

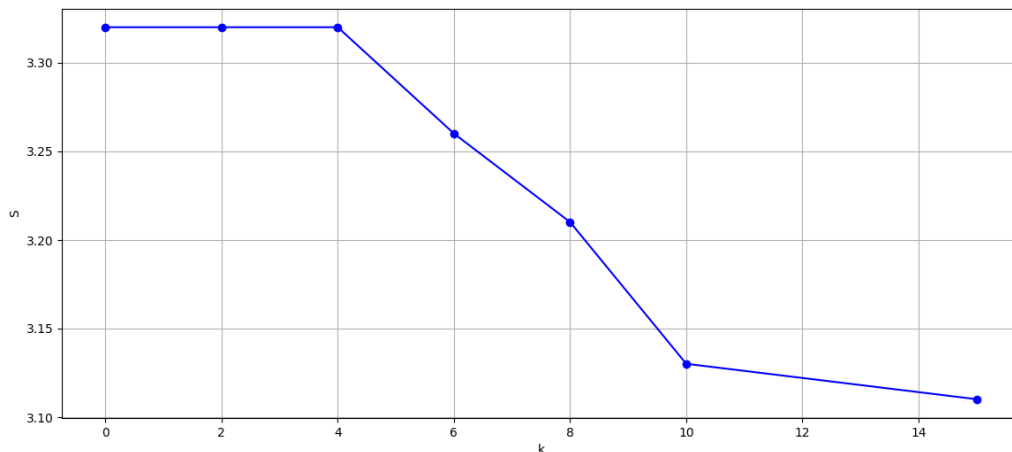
Na Tabela 7 são apresentados os valores de  $k$  utilizados e as consequentes relações entre os desvios padrões das médias dos estados neuronais ( $S$ ) e a energia  $RMS$ , fornecendo um comparativo quantitativo da economia energética versus eficiência do estímulo.

Tabela 7 – Relação de  $k$  passos ignorados na ação e métricas de avaliação.

$k$	$S$	$E_{RMS}$
0	3,32	0,68
2	3,32	0,34
4	3,32	0,16
6	3,26	0,11
8	3,21	0,09
10	3,13	0,06
15	3,11	0,04
100	2,31	0,006

Fonte: Elaboração Própria, 2024.

Figura 21 – Análise gráfica de  $k$  x  $S$ .

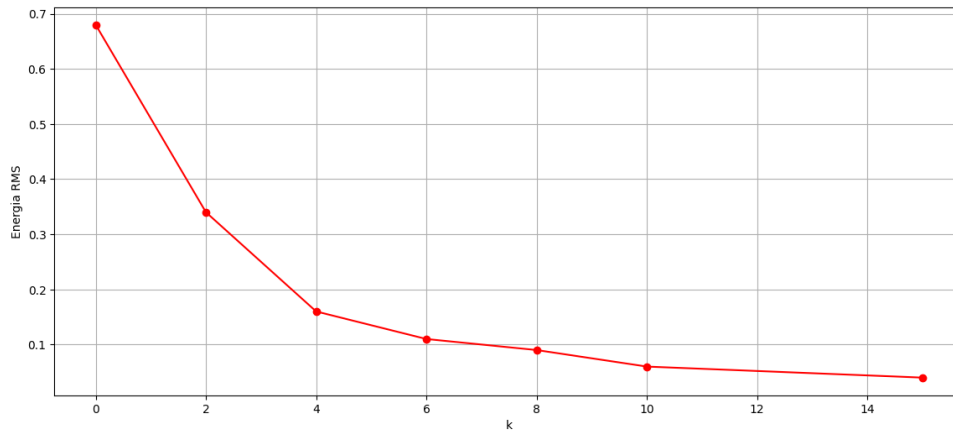


Fonte: Elaboração Própria, 2024.

As Figuras 21 e 22 mostram:

- O valor de  $S$ , representado na Equação 3.6, permanece constante em 3,32 para  $k = 0, 2, 4$  e depois diminui ligeiramente à medida que  $k$  aumenta.

Figura 22 – Análise gráfica de k x Energia.



Fonte: Elaboração Própria, 2024.

- A energia, representada pela Equação 3.5, diminui com o aumento de k, sugerindo uma relação inversa entre essas variáveis. Levando em consideração que o tempo de aplicação do sinal apresentado na Figura 17 e para todas as simulações dos sinais na Tabela 7 foi equivalente a 20000 passos, o cálculo da energia para N=1000 foi calculado por:

$$E_{RMS} = 1000 \sqrt{\frac{1}{20000} \int_0^{20000} I_{ext}^2(t) dt.}$$

- Para fins de visualização, k=100 foi omitido das figuras 21 e 22.

#### 4.5 Comparativo com DBS tradicional.

Para avaliar o impacto do modelo proposto, uma análise comparativa foi realizada contra o sinal DBS padrão, cujos parâmetros são delineados na Tabela 8, refletindo as configurações convencionais utilizadas na prática clínica (POPOVYCH *et al.*, 2017).

Tabela 8 – Sinal DBS Tradicional.

Parâmetro	Valor
Amplitude	0,5 pA
Frequência	130 Hz

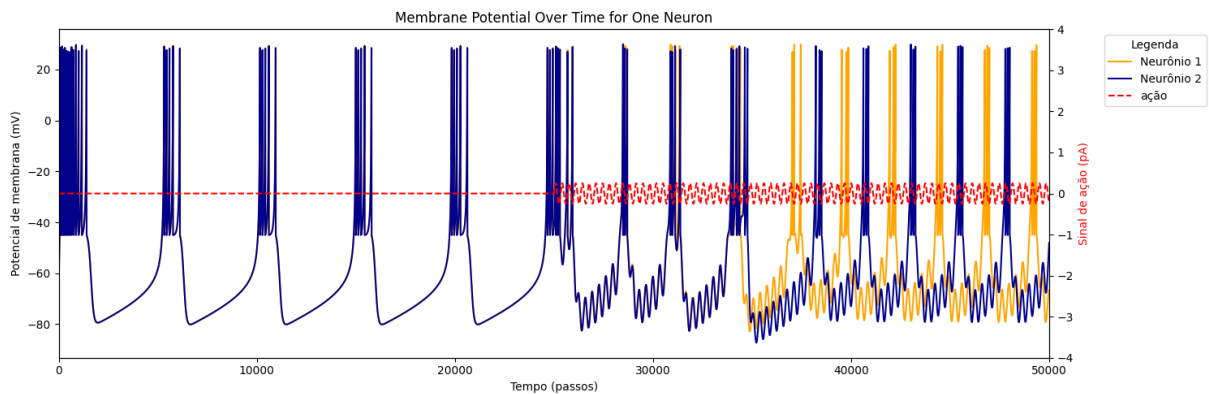
**Fonte:** Elaboração Própria, 2024.

Conforme ilustrado na Figura 23, o sinal DBS padrão é eficaz na dessincronização neuronal. Entretanto, a análise da energia consumida, utilizando a Equação 3.5, revelou um gasto

energético de 0,05 J para o sinal padrão. Em contraste, o modelo proposto visa uma abordagem mais eficiente, tanto em termos de eficácia quanto de conservação energética.

$$E_{RMS} = 1000 \sqrt{\frac{1}{20000} \int_0^{20000} I_{DBS_{normal}}^2(t) dt} = 0,05 \text{ J.}$$

Figura 23 – Aplicação de DBS padrão.



Fonte: Elaboração Própria, 2024.

A eficiência do sinal padrão pode ser quantificada pela taxa  $S$ , que compara a variação no desvio padrão dos estados neuronais antes e após a aplicação do DBS. Para o sinal padrão, a taxa  $S$  calculada foi de 2,04, indicando uma alteração significativa na atividade neuronal. A discussão subsequente avaliará como esses achados se comparam com os resultados do modelo proposto, enfatizando o potencial para melhorias na prática clínica da DBS.

$$S_{DBS_{normal}} = \frac{std[X_{antes}]}{std[X_{depois}]} = \frac{19,65}{9,6} = 2,04.$$

#### 4.6 Discussão dos resultados

Após avaliação do modelo, os resultados encontrados apresentaram-se satisfatórios com os objetivos traçados. O modelo, por ser adaptativo, exige processamento de máquina para treinamento acima da média, visto que de acordo com os parâmetros escolhidos, como passo de simulação, tamanho de lotes, tamanho de episódio e passos de treinamento.

O treinamento do modelo de aprendizagem por reforço mostrou-se satisfatório frente aos resultados encontrados, principalmente ao analisar a curva de recompensa ao longo do treinamento, exposta na Figura 16, que, como esperado, atingiu valores elevados durante o treinamento dado que o número de neurônios é diretamente proporcional à sua ordem de grandeza. Para um conjunto de 1000 neurônios, a variância entre os estados foi somada para cada passo de treinamento, bem como a soma das ações aplicadas. A função atingiu um platô ao final do treinamento, o que demonstra uma estabilidade no aprendizado do agente.

Em termos analíticos, a relação de desvios padrões das médias (S) foi equivalente a 3,32. Comparada ao padrão comercial analisado através do sinal senoidal proposto, a capacidade de dessincronização demonstrou ser superior. No entanto, a taxa de energia expendida foi superior ao DBS padrão, como observado na Tabela 7, em que:

$$\frac{E_{k=0}}{E_{DBS_{normal}}} = \frac{0,68}{0,05} = 13,6.$$

Dessa forma, apenas analisando o modelo proposto sem um tratamento posterior do sinal aplicado, a energia gasta é cerca de 14 vezes maior. Por isso, fez-se necessário o estudo da frequência de aplicação do sinal de ação.

A investigação subsequente sobre a frequência de aplicação do estímulo identificou um ponto ótimo em 15 passos ignorados, equilibrando eficiência energética e eficácia de dessincronização. Este achado sugere que, apesar do consumo energético inicialmente mais alto, ajustes na estratégia de estímulo podem oferecer uma alternativa viável e até superior ao DBS padrão em termos de economia energética e eficácia terapêutica, como demonstrado em:

$$\frac{E_{k=15}}{E_{DBS_{normal}}} = \frac{0,04}{0,05} = 0,8,$$

em que a energia da corrente proposta despense 80% da energia do DBS padrão.

Contudo, a comparação direta com o modelo proposto por Krylov *et al.* (2021) foi limitada pela diferença nos critérios de avaliação e na dinâmica neuronal dos modelos, destacando a necessidade de futuras pesquisas para explorar essas diferenças em detalhes.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

Foi realizada a implementação de um algoritmo de aprendizagem por reforço para o tratamento de estimulação cerebral profunda utilizando Otimização de Política Proximal a fim de auxiliar no tratamento de sintomas da doença de Parkinson, em especial tremores, causados pela sincronização de atividade neuronal em regiões cerebrais como o Glóbulo Pálido interno (GPi). Conclui-se que:

- Foi implementado um algoritmo adaptável de aprendizado por reforço utilizando Otimização de Política Proximal para modular as correntes emitidas pelo DBS, atingindo a dessincronização dos neurônios no GPi e a atenuação dos tremores;
- Foi conduzida uma análise detalhada da frequência de estimulações aplicadas pelo agente de aprendizado por reforço, em que houve um equilíbrio ótimo encontrado entre eficiência e consumo energético;
- Foi comparado o desempenho e a eficácia do modelo de sinal corrente proposto pelo agente RL e o DBS tradicional, em que a metodologia adotada sugere maior vantagem na utilização do modelo proposto.

Como limitações, ressalta-se o alto processamento requerido para as simulações, visto que o treinamento do modelo de aprendizagem por reforço demandou diversas iterações e épocas de treinamento com diferentes parâmetros a fim de atingir-se um resultado satisfatório.

### 5.1 Trabalhos Futuros

Como temas de trabalhos futuros, é sugerida a implementação da metodologia aplicada para outros modelos de neurônios, como Hodgkin-Huxley, com alta plausibilidade biológica. Além disso, sugere-se uma adaptação da linha de projeto para que o algoritmo seja treinado considerando dois ambientes de simulação com equações de mais de um modelo, garantindo o caráter híbrido. Por fim, aponta-se o estudo das interações e influência do modelo treinado em outras regiões cerebrais conectadas com o Glóbulo Pálido interno, como o Tálamo e o Glóbulo Pálido externo.



## REFERÊNCIAS

- ALAM, M.; SANGHERA, M.; SCHWABE, K.; LÜTJENS, G.; JIN, X.; SONG, J.; WRANGEL, C.; STEWART, R. M.; JANKOVIC, J.; GROSSMAN, R.; DARBIN, O.; KRAUSS, J. Globus pallidus internus neuronal activity: a comparative study of linear and non-linear features in patients with dystonia or parkinson's disease. **Journal of Neural Transmission**, 2016. Acesso em: 25 jan. 2024. Disponível em: <https://link.springer.com/article/10.1007/s00702-015-1484-3>.
- ALIPOUR, M.; VERDOM, B. H.; AKHLAGHDOUST, M. *et al.* Artificial intelligence in deep brain stimulation: A brief review. **Interventional Pain Medicine and Neuromodulation**, IEEE, 2023. Acesso em: 05 dez. 2023. Disponível em: <https://brieflands.com/articles/ipmn-134133>.
- BAHADORI-JAHROMI, F.; SALEHI, S.; ASL, M. M.; VALIZADEH, A. Efficient suppression of parkinsonian beta oscillations in a closed-loop model of deep brain stimulation with amplitude modulation. **Frontiers in Human Neuroscience**, v. 16, p. 1–19, 01 2023. Acesso em: 17 fev. 2024. Disponível em: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2022.1013155/full>.
- BELELLI, D.; HARRISON, N.; MAGUIRE, J.; MACDONALD, R.; WALKER, M.; COPE, D. Extrasynaptic gabaa receptors: Form, pharmacology, and function. **The Journal of Neuroscience**, v. 29, p. 12757–12763, 2009. Acesso em: 14 dez. 2023. Disponível em: <https://www.jneurosci.org/content/29/41/12757>.
- BONHOEFFER, K. Activation of passive iron as a model for the excitation of nerve. *J gen physiol*. **The Journal of general physiology**, v. 32, p. 69–91, 10 1948. Acesso em: 25 jan. 2024. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/18885679/>.
- BOUTET, A.; MADHAVAN, R.; ELIAS, G. *et al.* Predicting optimal deep brain stimulation parameters for parkinson's disease using functional mri and machine learning. **Nature Communications**, v. 12, p. 3043, 2021. Acesso em: 19 fev. 2024. Disponível em: <https://www.nature.com/articles/s41467-021-23311-9>.
- BROWN, P.; OLIVIERO, A.; MAZZONE, P.; INSOLA, A.; TONALI, P.; LAZZARO, V. D. Dopamine dependency of oscillations between subthalamic nucleus and pallidum in parkinson's disease. **The Journal of Neuroscience**, v. 21, n. 3, p. 1033–1038, 2001. Acesso em: 18 dez. 2023. Disponível em: <https://www.jneurosci.org/content/21/3/1033.long>.
- BURCHIEL, K.; MCCARTNEY, S.; LEE, A.; RASLAN, A. Accuracy of deep brain stimulation electrode placement using intraoperative computed tomography without microelectrode recording. **Journal of neurosurgery**, Journal of Neurosurgery Publishing Group, v. 119 2, p. 301–306, 2013. Acesso em: 10 jan. 2024. Disponível em: <https://thejns.org/view/journals/j-neurosurg/119/2/article-p301.xml>.
- GAO, Q.; NAUMANN, M.; JOVANOVIĆ, I.; LESI, V.; KAMARAVELU, K.; GRILL, W. M.; PAJIC, M. Model-based design of closed loop deep brain stimulation controller using reinforcement learning. In: **2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS)**. Sydney, NSW, Australia: [S. n.], 2020. p. 108–118.
- GORZELIC, P.; SCHIFF, S. J.; SINHA, A. Model-based rational feedback controller design for closed-loop deep brain stimulation of parkinson's disease. **Journal of Neural Engineering**, 2013. Acesso em: 15 dez. 2023. Disponível em: <https://dx.doi.org/10.1088/1741-2560/10/2/026016>.

HALLWORTH, N.; WILSON, C. J.; BEVAN, M. Apamin-sensitive small conductance calcium-activated potassium channels, through their selective coupling to voltage-gated calcium channels, are critical determinants of the precision, pace, and pattern of action potential generation in rat subthalamic nucleus neurons in vitro. **The Journal of Neuroscience**, v. 23, p. 7525–7542, 2003. Acesso em: 23 jan. 2024. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/12930791/>.

HAMALAINEN, P.; BABADI, A.; MA, X.; LEHTINEN, J. Ppo-cma: Proximal policy optimization with covariance matrix adaptation. In: **2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)**. Espoo, Finland: [S. n.], 2020. p. 1–6. Disponível em: <https://ieeexplore.ieee.org/document/9231618>.

HE, S.; BAIG, F.; MOSTOFI, A.; POGOSYAN, A.; DEBARROS, J.; GREEN, A. L.; AZIZ, T. Z.; PEREIRA, E.; BROWN, P.; TAN, H. Closed-loop deep brain stimulation for essential tremor based on thalamic local field potentials. **Movement Disorders**, v. 36, n. 4, p. 863–873, 2021. Acesso em: 15 fev. 2024. Disponível em: <https://movementdisorders.onlinelibrary.wiley.com/doi/abs/10.1002/mds.28513>.

HEMPTINNE, C. de; SWANN, N. C.; OSTREM, J. L.; RYAPOLOVA-WEBB, E. S.; LUCIANO, M. S.; GALIFIANAKIS, N. B.; STARR, P. A. Therapeutic deep brain stimulation reduces cortical phase-amplitude coupling in parkinson’s disease. **Nature neuroscience**, Nature Publishing Group, v. 18, p. 779–786, 2015. Acesso em: 12 dez. 2023. Disponível em: <https://www.nature.com/articles/nn.3997>.

HINDMARSH, J. L.; ROSE, R. M.; HUXLEY, A. F. A model of neuronal bursting using three coupled first order differential equations. **Proceedings of the Royal Society of London. Series B. Biological Sciences**, v. 221, n. 1222, p. 87–102, 1984. Acesso em: 20 dez. 2023. Disponível em: <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.1984.0024>.

HIRSCHMANN, J.; STEINA, A.; VESPER, J.; FLORIN, E.; SCHNITZLER, A. Neuronal oscillations predict deep brain stimulation outcome in parkinson’s disease. **Brain Stimulation**, v. 15, n. 3, p. 792–802, 2022. ISSN 1935-861X. Acesso em: 18 fev. 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1935861X22000870>.

IZHIKEVICH, E. M. Simple model of spiking neurons. **IEEE Transactions on Neural Networks, IEEE**, v. 14, n. 6, p. 1569–1572, 2003. Acesso em: 01 dez. 2023. Disponível em: <https://ieeexplore.ieee.org/document/1257420>.

IZHIKEVICH, E. M. Which model to use for cortical spiking neurons? **IEEE Transactions on Neural Networks, IEEE**, v. 15, n. 5, p. 1063–1070, 2004. Acesso em: 01 dez. 2023. Disponível em: <https://ieeexplore.ieee.org/document/1333071>.

IZHIKEVICH, E. M. **Which Model to Use for Cortical Spiking Neurons?** 2024. Electronic version available at <https://izhikevich.org/publications/spikes.htm>. Acesso em: 25 jan. 2024.

JAQUES, N.; GHANDEHARIOUN, A.; SHEN, J.; FERGUSON, C.; LAPEDRIZA, A.; JONES, N.; GU, S.; PICARD, R. **Way Off-Policy Batch Deep Reinforcement Learning of Implicit Human Preferences in Dialog**. 2019. Acesso em: 06 jan. 2024. Disponível em: <https://arxiv.org/abs/1906.09345>.

KADALA, A.; VERDIER, D.; MORQUETTE, P.; KOLTA, A. Ion homeostasis in rhythmogenesis: The interplay between neurons and astroglia. **Physiology**, v. 30 5, p. 371–88, 2015. Acesso em: 23 jan. 2024. Disponível em:

<https://consensus.app/papers/homeostasis-rhythmogenesis-interplay-between-neurons-kadala/d4908209d1cb54afabfa6542481494ff/>.

KERMAN, Z.; YU, C.; AN, H. Beta oscillation detector design for closed-loop deep brain stimulation of parkinson's disease with memristive spiking neural networks. In: **2022 23rd International Symposium on Quality Electronic Design (ISQED)**. Santa Clara, CA, USA: [S. n.], 2022. p. 1–6.

KLEINHOLDERMANN, U.; BACARA, B.; TIMMERMANN, L.; PEDROSA, D. J. Prediction of movement ratings and deep brain stimulation parameters in idiopathic parkinson's disease. **Neuromodulation: Technology at the Neural Interface**, v. 26, n. 2, p. 356–363, 2023. ISSN 1094-7159. Acesso em: 12 dez. 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1094715922012892>.

KRACK, P.; VOLKMANN, J.; TINKHAUSER, G.; DEUSCHL, G. Deep brain stimulation in movement disorders: From experimental surgery to evidence-based therapy. **Movement Disorders**, Wiley Online Library, v. 34, p. 1795–1810, 2019. Acesso em: 12 dez. 2023. Disponível em: <https://movementdisorders.onlinelibrary.wiley.com/doi/10.1002/mds.27860>.

KRYLOV, D.; COMBES, R. T. D.; LAROCHE, R.; ROSENBLUM, M.; DYLOV, D. V. Reinforcement learning framework for deep brain stimulation study. In: **Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence**. Yokohama, Yokohama, Japan: [S. n.], 2021. (IJCAI'20). ISBN 9780999241165.

LEVI, V.; ZORZI, G.; MESSINA, G.; ROMITO, L.; TRAMACERE, I.; DONES, I.; NARDOCCI, N.; FRANZINI, A. Deep brain stimulation versus pallidotomy for status dystonicus: a single-center case series. **Journal of Neurosurgery JNS**, American Association of Neurological Surgeons, v. 134, n. 1, p. 197 – 207, 2021. Acesso em: 05 dez. 2023. Disponível em: <https://thejns.org/view/journals/j-neurosurg/134/1/article-p197.xml>.

LEVY, R.; HUTCHISON, W.; LOZANO, A.; DOSTROVSKY, J. Synchronized neuronal discharge in the basal ganglia of parkinsonian patients is limited to oscillatory activity. **The Journal of Neuroscience**, 2002. Acesso em: 15 jan. 2024. Disponível em: <https://www.jneurosci.org/content/22/7/2855.long>.

LIU, C.; WANG, J.; YU, H.; DENG, B.; WEI, X.; LI, H.; LOPARO, K. A.; FIETKIEWICZ, C. Dynamical analysis of parkinsonian state emulated by hybrid izhikevich neuron models. **Communications in Nonlinear Science and Numerical Simulation**, Elsevier, v. 28, p. 10–26, 2015. Acesso em: 15 dez. 2023. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1007570415001021>.

LIU, C.; ZHAO, G.; MENG, Z.; ZHOU, C.; ZHU, X.; ZHANG, W.; WANG, J.; LI, H.; WU, H.; FIETKIEWICZ, C.; LOPARO, K. A. Closing the loop of dbs using the beta oscillations in cortex. **Cognitive neurodynamics**, v. 15, n. 6, p. 1157—1167, December 2021. ISSN 1871-4080. Acesso em: 17 fev. 2024. Disponível em: <https://europepmc.org/articles/PMC8572178>.

LIU, C.; ZHAO, G.; WANG, J.; WU, H.; LI, H.; FIETKIEWICZ, C.; LOPARO, K. A. Neural network-based closed-loop deep brain stimulation for modulation of pathological oscillation in parkinson's disease. **IEEE Access**, v. 8, p. 161067–161079, 2020. Acesso em: 16 fev. 2024. Disponível em: <https://ieeexplore.ieee.org/document/9180308>.

LU, M.; WEI, X.; CHE, Y.; WANG, J.; LOPARO, K. A. Application of reinforcement learning to deep brain stimulation in a computational model of parkinson's disease. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 28, n. 1, p. 339–349, 2020. Acesso em: 15 fev. 2024. Disponível em: <https://ieeexplore.ieee.org/document/8895773>.

MERK, T.; PETERSON, V.; KÖHLER, R.; HAUFE, S.; RICHARDSON, R. M.; NEUMANN, W.-J. Machine learning based brain signal decoding for intelligent adaptive deep brain stimulation. **Experimental Neurology**, 2022. Acesso em: 15 fev. 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0014488622000188>.

MIRZADEH, Z.; CHAPPLE, K.; LAMBERT, M.; EVIDENTE, V.; MAHANT, P.; OSPINA, M. C.; SAMANTA, J.; MOGUEL-COBOS, G.; SALINS, N.; LIEBERMAN, A.; TRÖSTER, A.; DHALL, R.; PONCE, F. Parkinson's disease outcomes after intraoperative ct-guided "asleep" deep brain stimulation in the globus pallidus internus. **Journal of neurosurgery**, Journal of Neurosurgery Publishing Group, v. 124 4, p. 902–907, 2016. Acesso em: 10 jan. 2024. Disponível em: <https://thejns.org/view/journals/j-neurosurg/124/4/article-p902.xml>.

MNIH, V.; KAVUKCUOGLU, K.; SILVER, D.; OUTROS. Human-level control through deep reinforcement learning. **Nature**, v. 518, p. 529–533, 2015. Acesso em: 17 dez. 2023. Disponível em: <https://www.nature.com/articles/nature14236>.

MOHAMMED, A.; BAYFORD, R.; DEMOSTHENOUS, A. A framework for adapting deep brain stimulation using parkinsonian state estimates. **Frontiers in Neuroscience**, v. 14, 2020. ISSN 1662-453X. Acesso em: 18 fev. 2024. Disponível em: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2020.00499>.

OBESO, J. A.; RODRIGUEZ-OROZ, M. C.; RODRIGUEZ, M.; MACIAS, R.; ALVAREZ, L.; GURIDI, J.; VITEK, J.; DELONG, M. R. Pathophysiologic basis of surgery for parkinson's disease. **Neurology**, v. 55, n. 12 Suppl 6, p. S7–S12, 2000. Acesso em: 18 dez. 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/11188978/>.

OLIVEIRA, A.; COELHO, L.; CARVALHO, E. *et al.* Machine learning for adaptive deep brain stimulation in parkinson's disease: closing the loop. **Journal of Neurology**, v. 270, p. 5313–5326, 2023. Acesso em: 15 fev. 2024. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10576725/>.

PARKINSON'S NSW. **3D Software to Improve Success of DBS Surgery**. 2024. <https://www.parkinsonsnsw.org.au/3d-software-to-improve-success-of-dbs-surgery/>. Acesso em: 25 jan. 2024.

PERALTA, M.; JANNIN, P.; BAXTER, J. S. H. Machine learning in deep brain stimulation: a systematic review. **Artificial Intelligence in Medicine**, v. 117, p. 102083, 2021. Acesso em: 18 dez. 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0933365721001913>.

PLENZ, D.; KITAL, S. T. A basal ganglia pacemaker formed by the subthalamic nucleus and external globus pallidus. **Nature**, 1999. Acesso em: 14 dez. 2023. Disponível em: <https://www.nature.com/articles/23281>.

POEWE, W.; SEPPI, K.; TANNER, C. M.; HALLIDAY, G. M.; BRUNDIN, P.; VOLKMANN, J.; SCHRAG, A.-E.; LANG, A. E. Parkinson's disease. **Nature Reviews Disease Primers**, 2017. Acesso em: 15 dez. 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28332488/>.

- POPOVYCH, O.; LYSYANSKY, B.; ROSENBLUM, M.; PIKOVSKY, A.; TASS, P. Pulsatile desynchronizing delayed feedback for closed-loop deep brain stimulation. **PLoS ONE**, Public Library of Science, v. 12, 2017. Acesso em: 21 jan. 2024. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28273176/>.
- RAFFIN, A.; HILL, A.; GLEAVE, A.; KANERVISTO, A.; ERNESTUS, M.; DORMANN, N. Stable-baselines3: Reliable reinforcement learning implementations. **Journal of Machine Learning Research**, v. 22, n. 268, p. 1–8, 2021. Acesso em: 29 nov. 2023. Disponível em: <http://jmlr.org/papers/v22/20-1364.html>.
- RIBBA, B. Reinforcement learning as an innovative model-based approach: Examples from precision dosing, digital health and computational psychiatry. **Frontiers in Pharmacology**, 2023. Acesso em: 13 dez. 2023. Disponível em: <https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2022.1094281/full>.
- RUBI, L.; FRITSCHY, J. Increased gabaergic transmission in neuropeptide  $\gamma$ -expressing neurons in the dopamine-depleted murine striatum. **Journal of Neurophysiology**, 2020. Acesso em: 15 jan. 2024. Disponível em: <https://journals.physiology.org/doi/full/10.1152/jn.00059.2020>.
- SCHULMAN, J.; WOLSKI, F.; DHARIWAL, P.; RADFORD, A.; KLIMOV, O. **Proximal Policy Optimization Algorithms**. 2017. Acesso em 12 dez. 2023. Disponível em: <https://arxiv.org/abs/1707.06347>.
- SHAH, S. A. A.; BAIS, A.; ZHANG, L. Optimization of patient specific stimulus for deep brain stimulation using spatially distributed neural sources. **IEEE Transactions on Artificial Intelligence**, v. 5, n. 2, p. 786–800, 2024. Acesso em: 18 fev. 2024. Disponível em: <https://ieeexplore.ieee.org/document/10143254>.
- SHARMA, V. D.; PATEL, M.; MIOCINOVIC, S. Surgical treatment of parkinson's disease: Devices and lesion approaches. **Neurotherapeutics**, v. 17, n. 4, p. 1525–1538, 2020. ISSN 1878-7479. Acesso em: 13 dez. 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1878747923012576>.
- SIDIROPOULOS, C.; RAMMO, R. A.; MERKER, B.; MAHAJAN, A.; LEWITT, P.; KAMINSKI, P.; WOMBLE, M.; ZEC, A.; TAYLOR, D.; WALL, J.; SCHWALB, J. Intraoperative mri for deep brain stimulation lead placement in parkinson's disease: 1 year motor and neuropsychological outcomes. **Journal of Neurology**, v. 263, p. 1226–1231, 2016. Acesso em: 25 jan. 2024. Disponível em: <https://link.springer.com/article/10.1007/s00415-016-8125-0>.
- SONG, Y. Information-loss-bounded policy optimization. In: \_\_\_\_\_. **Reinforcement Learning Algorithms: Analysis and Applications**. Cham: Springer International Publishing, 2021. p. 85–95.
- STARR, P. Placement of deep brain stimulators into the subthalamic nucleus or globus pallidus internus: Technical approach. **Stereotactic and Functional Neurosurgery**, Karger Publishers, Basel, Switzerland, v. 79, p. 118–145, 2003.
- STOJSAVLJEVIC, T.; GUO, Y.; MACALUSO, D. Adaptive stimulations in a biophysical network model of parkinson's disease. **International Journal of Molecular Sciences**, v. 24, n. 6, 2023. ISSN 1422-0067. Acesso em: 17 fev. 2024. Disponível em: <https://www.mdpi.com/1422-0067/24/6/5555>.

SU, F.; CHEN, M.; ZU, L.; LI, S.; LI, H. Model-based closed-loop suppression of parkinsonian beta band oscillations through origin analysis. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 29, p. 450–457, 2021. Acesso em: 17 fev. 2024. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9344835>.

SU, F.; WANG, H.; ZU, L.; CHEN, Y. Closed-loop modulation of model parkinsonian beta oscillations based on car-fuzzy control algorithm. **Cognitive Neurodynamics**, v. 17, p. 1185–1199, 2023. Acesso em: 18 fev. 2024. Disponível em: <https://link.springer.com/article/10.1007/s11571-022-09820-3>.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning**: An introduction. Cambridge, MA, USA: MIT Press, 2018.

TANG, J. K. H.; MAHANT, N.; HUTCHISON, W.; MORO, E.; LOZANO, A.; LANG, A.; DOSTROVSKY, J. Alterations in globus pallidus internus firing patterns are associated with different movement disorders. **Journal of Neurophysiology**, 2005. Acesso em: 16 dez. 2023. Disponível em: [https://link.springer.com/chapter/10.1007/0-387-28066-9\\_37](https://link.springer.com/chapter/10.1007/0-387-28066-9_37).

WORLD HEALTH ORGANIZATION. **Brain Health**. 2023. [https://www.who.int/health-topics/brain-health#tab=tab\\_1](https://www.who.int/health-topics/brain-health#tab=tab_1). Acesso em: 2 fev. 2024.

WORLD HEALTH ORGANIZATION. **Parkinson Disease**. 2023. <https://www.who.int/news-room/fact-sheets/detail/parkinson-disease>. Acesso em: 2 fev. 2024.

WU, H. Reinforcement learning inspired by psychology and neuroscience. **Journal of Education, Humanities and Social Sciences**, 2023. Acesso em: 2 fev. 2024. Disponível em: <https://drpress.org/ojs/index.php/EHSS/article/view/4673>.

WU, S. Motion control of unmanned surface vehicle based on improved reinforcement learning proximal policy optimization algorithm. In: NIKHATH, K. (Ed.). **2nd International Conference on Information Technology and Intelligent Control (CITIC 2022)**. Kunming, China: SPIE, 2022. v. 12346, p. 1234615. Disponível em: <https://doi.org/10.1117/12.2653441>.

YU, C.; VELU, A.; VINITSKY, E.; GAO, J.; WANG, Y.; BAYEN, A.; WU, Y. **The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games**. 2021. Acesso em: 16 jan. 2024. Disponível em: <https://arxiv.org/abs/2103.01955>.

YU, H.; MENG, Z.; LI, H.; LIU, C.; WANG, J. Intensity-varied closed-loop noise stimulation for oscillation suppression in the parkinsonian state. **IEEE Transactions on Cybernetics**, v. 52, n. 9, p. 9861–9870, 2022. Acesso em: 17 fev. 2024. Disponível em: <https://ieeexplore.ieee.org/document/9513987>.

ZOU, Y. Reinforcement learning-based real-time energy management for a hybrid tracked vehicle. **Applied Energy**, v. 171, p. 372–382, 2016. Acesso em: 22 dez. 2023. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0306261916304081>.