

**Modelo de classificação da doença de Parkinson por  
eletroencefalograma**

**Rudson Antônio Negrão Ribeiro de Souza**

Trabalho de Conclusão de Curso  
MBA em Inteligência Artificial e Big Data

**UNIVERSIDADE DE SÃO PAULO**  
**Instituto de Ciências Matemáticas e de Computação**

---

Modelo de classificação da doença de  
Parkinson por eletroencefalograma

*Rudson Antônio Negrão Ribeiro de Souza*

---

Rudson Antônio Negrão Ribeiro de Souza

## Modelo de classificação da doença de Parkinson por eletroencefalograma

Trabalho de conclusão de curso apresentado ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Caetano Mazzoni Ranieri

USP - São Carlos

2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

A636m      Antônio Negrão Ribeiro de Souza, Rudson  
Modelo de classificação da doença de Parkinson por  
eletroencefalograma / Rudson Antônio Negrão Ribeiro  
de Souza; orientador Caetano Ranieri. -- São  
Carlos, 2024.  
35 p.

Trabalho de conclusão de curso (MBA em  
Inteligência Artificial e Big Data) -- Instituto de  
Ciências Matemáticas e de Computação, Universidade  
de São Paulo, 2024.

1. Inteligência Artificial. 2. Doença Parkinson.  
3. Diagnóstico. 4. Eletroencefalograma. I. Ranieri,  
Caetano, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:  
Gláucia Maria Saia Cristianini - CRB - 8/4938  
Juliana de Souza Moraes - CRB - 8/6176

## DEDICATÓRIA

Dedico este trabalho a toda a minha família e amigos por guiarem meus passos e servindo de inspiração.

*A minha mãe, irmãos e namorada  
pela compreensão, carinho e apoio  
incansável.*

## AGRADECIMENTOS

A família e amigos, pelo apoio, suporte e compreensão na realização dessa pesquisa.

Ao Dr. Caetano Mazzoni Ranieri pela orientação e apoio durante a execução desta pesquisa, que muito me ensinou e contribuiu para o meu crescimento científico e pessoal.

Não posso deixar de agradecer aos meus irmãos, pais, amigos e minha namorada, pelo apoio e incentivo constantes e por entenderem minha ausência em muitos momentos, contribuindo assim em mais uma etapa da minha vida.

## RESUMO

Souza, RANR. **Título:** Modelo de classificação da doença de Parkinson por eletroencefalograma. 2024. 35 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

A doença de Parkinson (DP) é uma condição neurodegenerativa complexa que afeta principalmente adultos acima de 60 anos, sendo o segundo distúrbio neurodegenerativo mais comum após a doença de Alzheimer. Com os avanços tecnológicos, as técnicas de neuroimagem e processamento de dados, utilizando aprendizado de máquina (ML) e aprendizado profundo (DL) como técnicas aplicadas para detecção automática a DP, tem se mostrado promissores nos estudos de pesquisa. Este estudo utilizou dados de eletroencefalograma (EEG) disponíveis publicamente para analisar a atividade cerebral de 15 pacientes com doença de Parkinson (DP) e 16 indivíduos de um grupo controle (GC). Os dados foram coletados sem medicação, usando 41 canais, e analisados em estado de repouso. A acurácia foi utilizada como métrica chave de desempenho para o modelo de classificação. A abordagem incluiu a segmentação dos dados, leitura e organização em grupos (GC e DP), aplicação de uma rede neural convolucional (CNN) com camadas como Conv1D, BatchNormalization e LeakyReLU para melhorar o modelo. Como resultado obtivemos uma acurácia de 48,39%, indicando limitações na captura das características distintivas entre casos de Parkinson e controles. Futuras pesquisas devem usar amostras maiores e comparar condições com e sem medicação para aumentar a confiabilidade dos resultados e oferecer insights clínicos mais robustos.

Palavras-chave: Inteligência Artificial; Doença de Parkinson; Diagnóstico; Eletroencefalograma.

## ABSTRACT

SOUZA, RANR. **Title in English:** Parkinson's disease classification model by electroencephalogram. 2024. 35 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Parkinson's disease (PD) is a complex neurodegenerative condition that mainly affects adults over 60 years of age and is the second most common neurodegenerative disorder after Alzheimer's disease. With technological advances, neuroimaging and data processing techniques, using machine learning (ML) and deep learning (DL) as techniques applied for automatic detection of PD, have shown promise in research studies. The study used publicly available electroencephalogram (EEG) data to analyze the brain activity of 15 patients with Parkinson's disease (PD) and 16 individuals from a control group (CG). Data were collected without medication, using 41 channels, and analyzed in a resting state. Accuracy was used as a key performance metric for the classification model. The approach included data segmentation, reading and organization into groups (GC and DP), application of a convolutional neural network (CNN) with layers such as Conv1D, BatchNormalization and LeakyReLU to improve the model. As a result, we obtained an accuracy of 48.39%, indicating limitations in capturing the distinctive characteristics between Parkinson's cases and controls. Future research should use larger samples and compare conditions with and without medication to increase the reliability of results and provide more robust clinical insights.

**Keywords:** Artificial intelligence; Parkinson's disease; Diagnosis; Electroencephalogram.



## **LISTA DE ILUSTRAÇÕES**

Figura 1 – Conexões excitatórias e inibitórias de circuito motor.....	18
---	----

## **LISTA DE TABELAS**

Tabela 1 – Dados demográficos dos participantes no conjunto de dados.....	21
Tabela 2 – Resumo da configuração do modelo.....	22

## LISTA DE ABREVIATURAS E SIGLAS

BG	–	Gânglios da base
CNN	–	Rede neural convolucional
DL	–	Aprendizado profundo
dMSN	–	Neurônios espinhosos médios direto
DP	–	Doença de Parkinson
EEG	–	Eletroencefalograma
GC	–	Grupo controle
Gpe	–	Globo pálido externo
GPi	–	Globo pálido interno
iMSN	–	Neurônios espinhosos médios indireto
ML	–	Aprendizado de máquina
MSN	–	Neurônios espinhosos médios
PUT	–	Putâmen
SNC	–	Substância Negra compacta
SNr	–	Substância Negra reticulada
STN	–	Núcleo Subtalâmico
SVM	–	<i>Support Vector Machine</i>
TH	–	Tálamo

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>13</b>
<b>1.1 Objetivo.....</b>	<b>15</b>
<b>2 REFERENCIAL TEÓRICO.....</b>	<b>16</b>
<b>2.1 Características clínicas da doença de Parkinson.....</b>	<b>16</b>
<b>2.2 O circuito motor.....</b>	<b>17</b>
<b>3 METODOLOGIA.....</b>	<b>18</b>
<b>4 RESULTADOS E DISCUSSÃO.....</b>	<b>21</b>
<b>5 CONCLUSÕES.....</b>	<b>24</b>
<b>6 REFERÊNCIAS.....</b>	<b>25</b>
<b>ANEXO A – Código CNN LeavePneGroupOut.....</b>	<b>29</b>

## 1 INTRODUÇÃO

A doença de Parkinson (DP) é uma doença neurodegenerativa complexa e multissistêmica que consiste na destruição da via neuronal do tronco cerebral (OLASODE, 2001). O segundo distúrbio neurodegenerativo mais comum, ficando atrás apenas da doença de Alzheimer, afetando aproximadamente 1% dos adultos com mais de 60 anos (MCGREGOR, 2019). Acredita-se que essa degeneração seja influenciada por uma combinação de fatores genéticos, toxinas ambientais e estresse oxidativo (SHAFIQUE, 2011). Sua condição possui uma série de sintomas, tanto motores quanto não motores, muitos desses sintomas resultam diretamente da neura degeneração, outros são movidos por padrões de atividades em neurônios sobreviventes. Os sistemas motores incluem: tremores, rigidez, movimentos lentos e equilíbrio prejudicado (FAHN, 1989). Os sintomas não motores abrangem distúrbio do sono, distúrbios de controle dos impulsos, apatia, depressão, ansiedade e disfunção cognitiva, impactando significativamente na qualidade de vida (RANA, 2011). Embora os tratamentos atuais possam ajudar a controlar esses sintomas, não há cura para a doença (HEISTERS, 2011).

Por não haver cura, acredita-se que o diagnóstico precoce e o tratamento preventivo podem retardar o surgimento e piora dos sintomas, melhorando significativamente a qualidade de vida dos pacientes (UGRUMOV, 2020). Torna-se necessário encontrar métodos eficazes e precisos para o diagnóstico da DP.

Atualmente o diagnóstico pode ser identificado através de uma série de características clínicas, incluindo sintomas motores e não motores (LAWTON, 2015). Esses sintomas podem ser avaliados por meio de questionários e exames clínicos, sendo a identificação de características atípicas e a resposta aos medicamentos fundamentais no diagnóstico (CAMICIOLI, 2002). Biomarcadores associados à doença, como perfis de expressão, também estão sendo explorados para um diagnóstico precoce (KROKIDIS, 2019). Contudo, esses métodos tradicionais são suscetíveis a subjetividade, levando a possíveis erros de classificação, além de que tais sintomas podem ser leves e despercebidos. Um estudo no Reino Unido identificou uma taxa de 24% de diagnósticos incorretos baseados em observações dos sintomas (PAGAN, 2012).

Com o avanço tecnológico de neuroimagem e processamento de dados, há cada dia mais estudos utilizando técnicas de imagem cerebral não invasiva no auxílio do diagnóstico de doença cerebrais. Tais como tomografia por emissão de pósitrons, ressonância magnética funcional e eletroencefalograma (EEG) (ARBABSHIRANI, 2017). No âmbito da neurociência

clínica, o EEG é um dos mais populares devido as suas vantagens de baixo custo, alta resolução e portabilidade, sendo mundialmente utilizada no diagnóstico de doenças neurológicas e em pesquisas de reabilitação. Não à toa, cada vez mais estudos tem utilizado o EEG para detecção da DP, sendo considerado um grande potencial no diagnóstico precoce de características da DP (WANG, 2020). Pesquisadores identificaram que as anormalidades apresentadas através do EEG para pacientes com DP correspondiam a taxas mais altas do que em adultos mais velhos saudáveis. Em comparativo com grupo controle (GC), os pacientes com DP apresentaram alterações mais lentas em estado de repouso na atividade oscilatória cerebral (SOIKKELI, 1991; NEUFELD, 1994) e no acoplamento fase-amplitude (DE HEMPTINNE, 2013; SWANN, 2015; GONG, 2021), bem como redução  $\beta$  e  $\gamma$  (PEZARD, 2001). Além disso, as anormalidades da DP também se manifestaram em alterações na conectividade funcional do cérebro. Em comparação com o GC, os pacientes DP apresentaram conectividade reduzida na banda  $\alpha$ - $\beta$ , crescimento na conectividade da banda  $\gamma$  (CONTI, 2022) além da perda de conectividade frontotemporal (HASSAN, 2017).

Nos últimos anos, com a evolução e popularidade da inteligência artificial, juntamente com as novas tecnologias de reconhecimento de padrões, as ferramentas baseadas no diagnóstico por processamento de dados trouxeram grande ajuda para o diagnóstico precoce da DP. As quantidades crescentes de estudos foram dedicadas ao acoplamento de vários algoritmos de aprendizado de máquina (ML) ou aprendizado profundo (DL) e sinais de EEG para detectar DP automaticamente (MEI, 2021). Para os métodos ML, Menorca (2017) utilizou áreas aleatórias para classificar os sinais de EEG de 50 pacientes com DP e 41 HC com uma precisão de reconhecimento de 78,0%. Baseando na tecnologia de extração de recursos estatísticos de alta ordem e no classificador de máquina de vetores de suporte (SVM), Yuvaraj (2018) obteve 99,62% de precisão na distinção de sinais de EEG de pacientes com DP e GC. Oliveira (2020) utilizou técnicas aleatórias e seleção de recursos para obter mais de 99% de precisão de classificação para pacientes com DP e GC. Khoshnevis e Sankar (2021) propuseram um método baseado em técnicas estatísticas e obtiveram 87,00% de precisão usando um classificador RUSBoosted para um conjunto com 20 DP e 20 GC. Gunduz (2021) utilizou um classificador *Support Vector Machine* (SVM) e relatou 95,70% de precisão propondo um método eficiente de redução de dimensionalidade. Barua et al. (2021) propuseram a detecção automática de DP em um novo modelo baseado em gráficos utilizando sinais de EEG. Lee et al. (2022) Explorou o segmento de tempo analítico e a faixa de frequência em que o parâmetro Hjorth melhor representa as características do EEG dos pacientes com DP e obteve 89,3% de acurácia.

Com o avanço do método de aprendizado de máquina nos últimos anos, cada vez mais estudos exploram a detecção automática da DP baseado em EEG (TANVEER et al., 2022). Gil-Martín et al. (2019) detectaram pacientes com DP usando uma rede neural convolucional (CNN) para analisar os movimentos de desenho e alcançaram 96,5% de precisão. Oh et al. (2020) apresentaram uma arquitetura CNN de 13 camadas para classificar os sinais de EEG em repouso de 20 pacientes com DP e 20 GC e obtiveram uma precisão de 88,25%. Khare et al. (2021) aplicaram o modelo CNN de duas dimensões à transformação Smoothed Pseudo-Wigner Ville Distribution para dois conjuntos de dados de EEG em repouso e obteve precisões de validação de 99,9% e 100%. Loh et al. (2021) também aplicaram um modelo CNN de duas dimensões na transformação Gabor em um conjunto de dados de EEG no estado de repouso para classificar os sujeitos com acurácia de 99,5%. Shaban (2021) usou uma estrutura baseada em rede neural artificial e três canais espaciais de EEG para rastrear indivíduos com DP e GC com 98% de precisão. Mais recentemente, eles introduziram um método de aprendizagem profunda baseado em CNN de 20 camadas para identificar GC, usando *wavelet* do EEG em estado de repouso com 99.9% de acurácia (SHABAN, 2022).

Com base nos estudos levantados, evidenciamos que a ativação cerebral baseada em EEG e a conectividade diferem significativamente nos pacientes DP e GC, além das técnicas ML e DL demonstram potencial para detecção automatizada de DP. Contudo, ainda há questões que precisam ser exploradas. Em primeiro lugar, as características funcionais do cérebro identificadas no estudo do diagnóstico da DP podem ser diversas, carecendo de características válidas e confiáveis. No segundo ponto, a maioria dos estudos anteriores foram baseados em características de um único padrão (ativação ou rede cerebral) para a identificação da DP. Apesar de muitas das estruturas de ML e DL propostas tenham apresentado grande potencial no diagnóstico de DP, a maioria só mostrou bons resultados em um conjunto de pesquisa específico e não foi generalizado para outros conjuntos de dados. Sendo assim, ainda carecemos de modelos eficientes e robustos na detecção automática da DP.

### 1.1. Objetivo

Aplicar dados previamente coletados de EEG utilizando o modelo CNN de uma dimensão para avaliar a acurácia no diagnóstico da doença de Parkinson.

## **2. REFERENCIAL TEÓRICO**

### **2.1. Características clínicas da doença de Parkinson**

A DP é uma doença neurodegenerativa complexa e irreversível. Acomete o sistema nervoso central, e trata-se da segunda doença neurodegenerativa mais comum, apresentando aumento significativo da sua prevalência nas últimas três décadas (CABREIRA; MASSANO, 2019). Ela atinge cerca de 1-3% da população mundial com mais de 60 anos, com predominância no gênero masculino (CHIEN; BARSOTTINI, 2017). No Brasil, estima-se que existam 220.000 indivíduos com DP (BOVOLENTA; FELICIO, 2017). As taxas de prevalência da DP tem projeção, para 2040, de mais de 12 milhões de indivíduos diagnosticados no planeta (DORSEY et al., 2018).

Por acometer predominantemente idosos, e considerando o processo de envelhecimento populacional, é necessário maior atenção e busca de estratégias de prevenção, diagnóstico precoce e tratamento direcionado para esta população potencialmente vulnerável à doença.

Embora as características motoras básicas da DP incluam tremor e rigidez, há muitos outros sintomas motores, dentre eles alterações na maneira de andar, controle nos movimentos dos olhos, fala e até no controle da bexiga. Temos também reduções na amplitude, velocidade e dificuldade no movimento, podendo impactar no controle de muitos grupos musculares. Os pacientes relatam dificuldades em tarefas do dia a dia como abotoar, escrever e utilização de utensílios. (MCGREGOR, 2019)

Atualmente, sabe-se que os indivíduos com DP apresentam degeneração na substância negra, com diminuição da produção de dopamina (SILVA et al., 2021). Ela tem origem multifatorial, combinando fatores genéticos e ambientais, cujo diagnóstico ocorre por meio da identificação dos sintomas clínicos e exclusão de outras doenças neurológicas (BALL et al., 2019).

O diagnóstico baseia-se portanto nos sintomas clínicos e inicia-se com a confirmação da existência de parkinsonismo (designação sindrômica caracterizada por um conjunto de sintomas motores, sendo a bradicinesia a característica fundamental) (MASSANO, BHATIA, 2012). Sabe-se que tremor de repouso e a rigidez não são sintomas obrigatórios para o diagnóstico e, na ausência de bradicinesia, não é possível fazer o diagnóstico clínico de DP. Muitos sintomas da DP resultam diretamente da neuro degeneração, outros são movidos por padrões de atividades aberrantes em neurônios sobreviventes. A disfunção do circuito DP, é uma área de intenso estudo, pois está subjacente a nossa capacidade de tratar muitos sintomas



das doenças de neuro degeneração irreversível (MCGREGOR, 2019). Sendo assim, esse método diagnóstico é subjetivo, levando a possíveis erros de classificação.

Uma série de estudos exploraram o uso de diferentes métodos para detectar a doença de Parkinson. Os autores Braga (2019) e Singh (2022) obtiveram sucesso no uso de algoritmos de aprendizado de máquina, o estudo de Braga concentrou-se na análise da fala e Singh na análise da caligrafia. Andrei (2019) alcançou 100% de precisão na distinção entre indivíduos saudáveis e aqueles com doença de Parkinson usando análise do padrão de marcha. GUO (2010) combinou a programação genética e o algoritmo de maximização de expectativa para criar funções de recurso de aprendizagem para detecção da doença. Esses estudos demonstram coletivamente o aprendizado de máquina e outras técnicas avançadas no diagnóstico preciso da doença de Parkinson, assim como a utilização do EEG. Contudo, ainda falta consenso sobre a caracterização da DP.

Não dispomos ainda de uma forma de diagnosticar com certeza o Parkinson *in vivo*, embora a tomografia e ressonância magnética de crânio e exames moleculares possam auxiliar na detecção da DP (CABREIRA; MASSANO, 2019).

Dessa forma, o diagnóstico continua sendo desafiador, com amplas implicações para o manejo clínico, o que afeta diretamente o custo no sistema de saúde, impossibilitando estimar as despesas futuras com esta doença que cresce exponencialmente entre a população (SILVA et al., 2021).

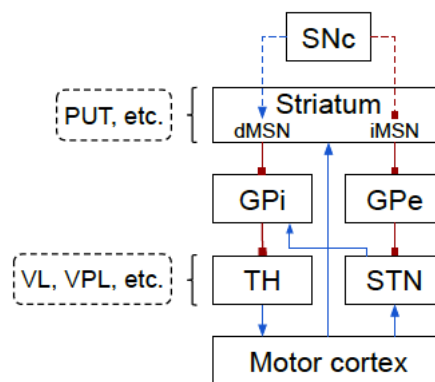
Nos últimos anos, com o desenvolvimento da tecnologia de neuroimagem, mais estudos têm utilizado o EEG para detecção de DP (ARBABSHIRANI et al., 2017). Dentre as vantagens do EEG estão: a portabilidade, baixo custo e aplicabilidade e fato de ser um exame não invasivo, indolor e não submeter o paciente à radiação ionizante, sendo amplamente utilizado no diagnóstico de doenças neurológicas, podendo identificar as características únicas da DP (WANG et al., 2020). Somado a isso, o crescente uso da inteligência artificial e das tecnologias de reconhecimento de padrões, tem sido utilizado como ferramentas para auxiliar o diagnóstico precoce da DP (QIU et al., 2022).

## **2.2. O circuito motor**

O circuito motor do cérebro dos mamíferos é formado pelo córtex motor, o tálamo (TH) e os gânglios da base (BG), este último composto por um subconjunto de estruturas: o striatum, que inclui o putâmen (PUT) e o núcleo caudado e o globo pálido dividido em interna (GPi) e externa (Gpe), o núcleo subtalâmico (STN), e a substância negra, dividida em compacta (SNc)

e reticulada (SNr). O modelo clássico ajuda a explicar as conexões afetadas pela DP conforme Figura 1, destacando as relações entre as projeções de neurônios do SNc para as estruturas BG, principalmente striatum, onde a dopamina é liberada.

Figura. 1: Conexões excitatórias (azul) e inibitórias (vermelhas) de circuito do motor. A DP é causada pela perda de neurônios da substância negra compacta (SNc), que enfraquece as conexões representadas pelas linhas tracejadas. Isso causa mau funcionamento nas vias direta e indireta.



Fonte: RANIERI. 2024. Unveiling Parkinson's Disease Features from a Primate Model with Deep Neural Networks

O caminho começa com a conexão excitatória do córtex para o striatum que projeta seus neurônios de saída, chamados de neurônios espinhosos médios (MSN), para outras estruturas dentro do BG. Na via direta, o MSN direto (dMSN) inibe o GPi, que reduz sua inibição ao TH, que então excita o córtex motor. Na via indireta, o indireto MSN (iMSN) inibe o Gpe, o que reduz sua inibição para o STN, que excita o GPi. Resultando assim na inibição do TH e ausência na saída excitatória para o córtex motor. Em resumo, a via direta excita o córtex, enquanto a via indireta inibe. A DP é caracterizada pela perda progressiva dos neurônios dopaminérgicos, especialmente no SNc, que causa o mau funcionamento de ambas as vias (RANIERI, 2024).

### 3. METODOLOGIA

O conjunto de dados utilizado neste estudo refere-se ao eletroencefalograma (EEG), que analisa a atividade elétrica espontânea do cérebro. Trata-se de informação pública fornecida pela OpenNeuro, obtida pelo laboratório Aron da Universidade da Califórnia em San Diego, com curadoria do laboratório Swann da Universidade de Oregon (ROCKHILL et al., 2020).

A base de dados é composta por 15 pacientes destros com DP, sendo 8 do gênero feminino e 7 do gênero masculino com idade média de 62,6 +- 8,3 anos e 16 do grupo controle (GC), 9 do gênero feminino e 6 do gênero masculino com média de idade 63,5 +- 9,6 anos. Os pacientes foram recrutados na Clínica Scripps em La Jolla, CA nos Estados Unidos e apresentavam DP leve ou moderada, enquanto o GC era composto por voluntários de uma comunidade local.

No conjunto de dados, cada paciente com DP visitou o laboratório para duas sessões, medicação ON e OFF. A ordem das visitas foi contrabalançada entre os pacientes. Para a sessão OFF, o tempo entre a última dose da medicação dopaminérgica e a consulta foi de 12 horas ou mais, e para a sessão ON, os pacientes tomaram a dose matinal habitual antes de virem para a sessão. No entanto, os dados de EEG utilizados neste estudo incluíram apenas os casos sem medicação.

Para as medições de EEG foi utilizado um sistema de 41 canais, coletados nos indivíduos em estado de repouso em uma frequência de amostragem de 512 Hz com pelo menos 3 minutos. A posição dos eletrodos de 41 canais foram Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, Cz, EXG1, EXG2, EXG3, EXG4, EXG5, EXG6, EXG7, EXG8.

Utilizamos a acurácia como métrica para avaliar o desempenho do modelo de classificação. Ela é definida como a proporção de previsões corretas feitas pelo modelo em relação ao número total de previsões realizadas. Ou seja, indica a porcentagem de casos em que o modelo classificou corretamente.

Em um contexto de classificação binária, essa métrica é calculada como: [ Acurácia =  $\frac{(TP + TN)}{(TP + TN + FP + FN)}$  ]

Onde:

- (TP) (True Positives) é o número de exemplos da classe positiva que foram corretamente identificados.
- (TN) (True Negatives) é o número de exemplos da classe negativa que foram corretamente identificados.
- (FP) (False Positives) são os exemplos da classe negativa que foram incorrectamente classificados como positivos.
- (FN) (False Negatives) são os exemplos da classe positiva que foram incorrectamente classificados como negativos.

Utilizamos a ferramenta Colab da Google para manipulação e treinamento dos dados. Importamos as bibliotecas glob (para extração do caminho onde os arquivos de EEG estavam armazenados), numpy (para manipulação de array) e mne (para leitura dos arquivos de extensão. bdf).

Iniciamos com a separação dos arquivos a serem lidos (grupos GC e DP), com cada grupo em um array para leitura. Criamos uma função para a leitura dos dados e lemos os 31 registros de EEG separando-os em duas variáveis, uma para o GC e outra para o DP.

Após a leitura, segmentamos os dados em partes menores de tamanho fixo de 100, considerando os 41 canais coletados e unificamos no final tanto os dados GC quanto os DP.

Para cada arquivo lido, criamos uma lista de array com o tamanho total dos registros e separamos em grupos.

Também foi necessário a identificação dos dados, no caso se aquele dado era de um GC ou DP e para isso percorremos os dados já segmentados, fizemos a transpose da matriz para adequar aos parâmetros de entrada do modelo CNN e para os 16 primeiros identificamos como GC e os demais DP.

Para treinar o modelo, utilizamos o conjunto de bibliotecas da Sklearn e Tensorflow Keras. Do Sklearn, temos a função LeaveOneGroupOut que separa os grupos de treinamento e testes. Essa função faz a separação dos dados conforme o grupo estabelecido para que não seja utilizado no treinamento parte de dado de um EEG e parte de outro.

No caso do Tensorflow, utilizamos Conv1D por se tratar de um modelo de uma dimensão. O BatchNormalization para normalização e a função LeakyReLU para melhorar o refinamento do modelo, pois ela permite um pequeno gradiente para valores negativos de entrada ao invés de mapeá-los para zero evitando o problema de “neurônio morto”.

Precisamos implementar a função AveragePooling1D também do Tensorflow usada em redes neurais para implementar o pooling das sequências temporais. O conceito de pooling é utilizado nas redes neurais para reduzir a dimensionalidade dos mapas e introduzir invariância a pequenas translações nos dados de entrada, mantendo as características mais relevantes. Para este estudo utilizamos o parâmetro pool\_size de valor 2, este parâmetro é o tamanho da janela de pooling, o que significa que a média é calculada sobre cada segmento de 2 valores. Utilizamos também o parâmetro strides de valor 2, que corresponde ao número de passos que a janela de pooling move em cada iteração.

Foi utilizado também o método GlobalAveragePooling1D para reduzir os mapas de características unidimensionais a valores de características únicas por mapa através da média global do conjunto de dados.

Por fim, foi adicionado o método Dense, que é uma camada totalmente conectada (ou camada densa). Em uma camada como essa, cada neurônio está conectado a todos os neurônios da camada anterior. Utilizamos a função de ativação sigmoid, que transforma a saída linear da camada dense em um valor entre 0 e 1 e é ideal para a camada de saída em tarefas de classificação binária, já que a saída pode ser interpretada como a probabilidade de pertencer a classe positiva.

## 4 Resultados e Discussão

Na amostra estudada, os 15 pacientes DP e os 16 participantes de GC foram pareadas demograficamente por idade e gênero. A média de idade do GC foi  $63,5 \pm 9,6$  anos e DP  $63,2 \pm 6,3$  anos. Para caracterizar as amostras, também foram utilizados: *Scores from the North American Adult Reading Test*, *Mini Mental State Exam* e o *United Parkinson's Disease Rating Scale*, conforme ilustrado na tabela abaixo.

Tabela 1 - Dados demográficos dos participantes no conjunto de dados

UC San Diego		
	DP	GC
Número	15	16
Gênero (feminino/masculino)	8f/7m	9f/7m
Idade (anos)	$63,2 \pm 8,2$	$63,5 \pm 9,6$
NAART	$46,1 \pm 6,3$	$49,1 \pm 7,1$
MMSE	$28,9 \pm 1,0$	$29,2 \pm 1,1$
UPDRS	$39,2 \pm 9,7$ (OFF)	

Legenda: NAART = Scores from the North American Adult Reading Test; MMSE 0 Mini Mental State Exam; UPDRS = United Parkinson's Disease Rating Scale (motor); DP = Doença de Parkinson; GC = Grupo Controle.

Com base na configuração da sessão do modelo CNN de uma dimensão e considerando a quantidade de 5000 segmentos, obtivemos os resultados da tabela 2.

Tabela 2 – Resumo da configuração do modelo

Layer (type)	Output Shape	Param
conv1d (Conv1D)	(None, 4998, 5)	620
batch_normalization (BatchNormalization)	(None, 4998, 5)	20
leaky_re_lu (LeakyReLU)	(None, 4998, 5)	0
max_pooling1d (MaxPooling1D)	(None, 2499, 5)	0
conv1d_1 (Conv1D)	(None, 2497, 5)	80
leaky_re_lu_1 (LeakyReLU)	(None, 2497, 5)	0
max_pooling1d_1 (MaxPooling1D)	(None, 1248, 5)	0
dropout (Dropout)	(None, 1248, 5)	0
conv1d_2 (Conv1D)	(None, 1246, 5)	80
leaky_re_lu_2 (LeakyReLU)	(None, 1246, 5)	0
average_pooling1d (AveragePooling1D)	(None, 623, 5)	0
dropout_1 (Dropout)	(None, 623, 5)	0
conv1d_3 (Conv1D)	(None, 621, 5)	80
leaky_re_lu_3 (LeakyReLU)	(None, 621, 5)	0
average_pooling1d_1 (AveragePooling1D)	(None, 310, 5)	0
conv1d_4 (Conv1D)	(None, 308, 5)	80
leaky_re_lu_4 (LeakyReLU)	(None, 308, 5)	0
global_average_pooling1d (GlobalAveragePooling1D)	(None, 5)	0
dense (Dense)	(None, 1)	6
<b>Total params:</b> 966 (3.77 KB)		
<b>Trainable params:</b> 956 (3.73 KB)		
<b>Non-trainable params:</b> 10 (40.00 B)		

A aplicação de dados de EEG previamente coletados utilizando um modelo CNN unidimensional com o objetivo de avaliar a acurácia no diagnóstico da doença de Parkinson resultou em uma acurácia de 48,39% (Tabela 3). Este resultado destaca algumas considerações importantes:

1. **Desempenho do Modelo:** A acurácia obtida indica um desempenho abaixo do esperado para um diagnóstico confiável de Parkinson, sugerindo que o modelo pode não estar capturando adequadamente as características distintivas dos sinais de EEG entre os pacientes com DP e o grupo controle.
2. **Possíveis Causas:** Esta baixa acurácia pode ser atribuída a vários fatores, incluindo a complexidade insuficiente do modelo, possível falta de representatividade ou qualidade do conjunto de dados, e a necessidade de ajuste fino de hiperparâmetros ou de melhorias na preparação e pré-processamento dos dados.
3. **Direções Futuras:** Para melhorar o diagnóstico automatizado, seria recomendável explorar arquiteturas de rede mais complexas ou híbridas, incorporar técnicas avançadas de pré-processamento de sinais de EEG e realizar otimizações de hiperparâmetros mais robustas. Além disso, considerar o uso de dados adicionais ou características complementares pode ajudar a criar um modelo mais robusto.
4. **Implicações Clínicas:** Apesar dos desafios, o uso de EEG na combinação com redes neurais continua a ser uma abordagem promissora. Investimentos adicionais em pesquisa e desenvolvimento podem potencialmente levar a melhorias significativas na acurácia e utilidade clínica, contribuindo para um diagnóstico mais precoce e preciso da doença de Parkinson.
5. **Problemas de generalização:** O modelo pode estar memorizando o conjunto de treinamento, mas não generalizando bem para dados não vistos, um clássico sinal de overfitting.

Essas ações podem permitir que modelos futuros alcancem uma performance que atenda melhor às necessidades clínicas e diagnósticas, contribuindo para uma melhor gestão da doença de Parkinson.

Futuras pesquisas devem considerar amostras maiores e comparações envolvendo condições com e sem medicação, para melhorar a volumetria de dados e a confiabilidade dos resultados, oferecendo insights mais robustos para a caracterização clínica da doença.

## 5 CONCLUSÃO

A aplicação de EEG com 41 canais em formato de dados em bdf, divididos em dois grupos (GC e DP) treinados em um modelo CNN unidimensional para diagnosticar a doença de Parkinson revelou uma acurácia de 48,39%, indicando limitações na captura das características distintivas entre casos de Parkinson e controles. As potenciais causas dessa baixa performance incluem a complexidade insuficiente do modelo, representatividade limitada do conjunto de dados e possíveis ajustes necessários na pré-processamento e nos hiperparâmetros. Além disso, a utilização de uma quantidade menor de canais nos treinamentos podem trazer melhorias na performance tendo em vista a utilização de atributos mais específicos para detecção.

Outros modelos como as redes neurais recorrentes, que são utilizadas em aplicações como reconhecimento de fala, tradução e processamento de linguagem natural podem ser muito promissoras na detecção do padrão da DP.



## REFERÊNCIAS

- ANDREI, A.-G.; TAUTAN, A.-M.; IONESCU, B. Parkinson's Disease Detection from Gait Patterns. **2019 E-Health and Bioengineering Conference (EHB)**, nov. 2019.
- ARBABSHIRANI, M. R. et al. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. **NeuroImage**, v. 145, p. 137–165, jan. 2017.
- BALL, N. et al. Parkinson's Disease and the Environment. **Frontiers in Neurology**, v. 10, n. 218, 19 mar. 2019.
- BARUA, P. D. et al. Novel automated PD detection system using aspirin pattern with EEG signals. **Computers in Biology and Medicine**, v. 137, p. 104841, out. 2021.
- BRAGA, D. et al. Automatic detection of Parkinson's disease based on acoustic analysis of speech. **Engineering Applications of Artificial Intelligence**, v. 77, p. 148–158, jan. 2019.
- CABREIRA, V.; MASSANO, J. Doença de Parkinson: Revisão Clínica e Atualização. **Acta Médica Portuguesa**, v. 32, n. 10, p. 661, 1 out. 2019.
- CAMICIOLI, R. Identification of parkinsonism and Parkinson's disease. **Drugs of Today**, v. 38, n. 10, p. 677, 2002.
- CHATURVEDI, M. et al. Quantitative EEG (QEEG) Measures Differentiate Parkinson's Disease (PD) Patients from Healthy Controls (HC). **Frontiers in Aging Neuroscience**, v. 9, 23 jan. 2017.
- CHIEN, H. F. et al. Rehabilitation in patients with cerebellar ataxias. **Arquivos de Neuro-Psiquiatria**, v. 80, n. 3, p. 306–315, 25 fev. 2022.
- CONTI, M. et al. Brain Functional Connectivity in de novo Parkinson's Disease Patients Based on Clinical EEG. **Frontiers in Neurology**, v. 13, 15 mar. 2022.
- DE HEMPTINNE, C. et al. Exaggerated phase-amplitude coupling in the primary motor cortex in Parkinson disease. **Proceedings of the National Academy of Sciences**, v. 110, n. 12, p. 4780–4785, 7 mar. 2013.
- DORSEY, E. R. et al. The Emerging Evidence of the Parkinson Pandemic. **Journal of Parkinson's Disease**, v. 8, n. s1, p. S3–S8, 18 dez. 2018.
- FAHN, S. The history of parkinsonism. **Movement Disorders**, v. 4, n. S1, p. S2–S10, 1989.
- GIL-MARTÍN, M.; MONTERO, J. M.; SAN-SEGUNDO, R. Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks. **Electronics**, v. 8, n. 8, p. 907, 17 ago. 2019.

GONG, R. et al. Spatiotemporal features of  $\beta$ - $\gamma$  phase-amplitude coupling in Parkinson's disease derived from scalp EEG. **Brain: A Journal of Neurology**, v. 144, n. 2, p. 487–503, 3 mar. 2021.

GUNDUZ, H. An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification. **Biomedical Signal Processing and Control**, v. 66, p. 102452, abr. 2021.

HASSAN, M. et al. Functional connectivity disruptions correlate with cognitive phenotypes in Parkinson's disease. **NeuroImage: Clinical**, v. 14, p. 591–601, 2017.

HEISTERS, D. Parkinson's: symptoms, treatments and research. **British Journal of Nursing**, v. 20, n. 9, p. 548–554, 9 maio 2011.

KHARE, S. K.; BAJAJ, V.; ACHARYA, U. R. **PDCNNNet: An Automatic Framework for the Detection of Parkinson's Disease Using EEG Signals** | **IEEE Journals & Magazine | IEEE Xplore**. Disponível em: <<https://ieeexplore.ieee.org/document/9430513>>.

KHOSHNEVIS, S. A.; SANKAR, R. Classification of the stages of Parkinson's disease using novel higher-order statistical features of EEG signals. **Neural Computing and Applications**, 24 nov. 2020.

KROKIDIS, M. G. Identification of biomarkers associated with Parkinson's disease by gene expression profiling studies and bioinformatics analysis. **AIMS Neuroscience**, v. 6, n. 4, p. 333–345, 26 dez. 2019.

LAWTON, M. et al. Parkinson's Disease Subtypes in the Oxford Parkinson Disease Centre (OPDC) Discovery Cohort. **Journal of Parkinson's Disease**, v. 5, n. 2, p. 269–279, 1 jun. 2015.

LEE, S.-B. et al. Predicting Parkinson's disease using gradient boosting decision tree models with electroencephalography signals. **Parkinsonism & Related Disorders**, v. 95, p. 77–85, fev. 2022.

LOH, H. W. et al. GaborPDNet: Gabor Transformation and Deep Neural Network for Parkinson's Disease Detection Using EEG Signals. **Electronics**, v. 10, n. 14, p. 1740, 20 jul. 2021.

MASSANO, J.; BHATIA, K. P. Clinical Approach to Parkinson's Disease: Features, Diagnosis, and Principles of Management. **Cold Spring Harbor Perspectives in Medicine**, v. 2, n. 6, p. a008870–a008870, 21 fev. 2012.

MCGREGOR, M. M.; NELSON, A. B. Circuit Mechanisms of Parkinson's Disease. **Neuron**, v. 101, n. 6, p. 1042–1056, 20 mar. 2019.

MEI, J.; DESROSIERS, C.; FRASNELLI, J. Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. **Frontiers in Aging Neuroscience**, v. 13, 6 maio 2021.

NEUFELD, M. Y. et al. EEG Frequency Analysis in Demented and Nondemented Parkinsonian Patients. **Dementia and Geriatric Cognitive Disorders**, v. 5, n. 1, p. 23–28, 1994.

OH, S. L. et al. A deep learning approach for Parkinson's disease diagnosis from EEG signals. **Neural Computing and Applications**, 30 ago. 2018.

OLASODE, B. J. Idiopathic Parkinson's disease: a review of clinical types and pathology. **Nigerian Journal of Medicine: Journal of the National Association of Resident Doctors of Nigeria**, v. 10, n. 3, p. 116–120, 2001.

PAGAN, F. L. Improving outcomes through early diagnosis of Parkinson's disease. **The American Journal of Managed Care**, v. 18, n. 7 Suppl, p. S176-182, 1 set. 2012.

PEZARD, L.; JECH, R.; RŮŽIČKA, E. Investigation of non-linear properties of multichannel EEG in the early stages of Parkinson's disease. **Clinical Neurophysiology**, v. 112, n. 1, p. 38–45, 1 jan. 2001.

QIU, L.; LI, J.; PAN, J. Parkinson's disease detection based on multi-pattern analysis and multi-scale convolutional neural networks. **Frontiers in Neuroscience**, v. 16, 27 jul. 2022.

RANA, A. Q. et al. Parkinson's disease: a review of non-motor symptoms. **Expert Review of Neurotherapeutics**, v. 15, n. 5, p. 549–562, 4 maio 2015.

REES, R. N. et al. An early diagnosis is not the same as a timely diagnosis of Parkinson's disease. **F1000Research**, v. 7, p. 1106, 18 jul. 2018.

SHABAN, M.; AMARA, A. W. Resting-state electroencephalography based deep-learning for the detection of Parkinson's disease. **PLOS ONE**, v. 17, n. 2, p. e0263159, 24 fev. 2022.

SHAFIQUE, H. Causes of Parkinson's disease: Literature Review. **Journal of Parkinsonism and Restless Legs Syndrome**, p. 5, dez. 2011.

SWANN, N. C. et al. Elevated synchrony in Parkinson disease detected with electroencephalography. **Annals of Neurology**, v. 78, n. 5, p. 742–750, 2 set. 2015.

SILVA, A. B. G. et al. Doença de Parkinson: revisão de literatura / Parkinson's Disease: literature review. **Brazilian Journal of Development**, v. 7, n. 5, p. 47677–47698, 7 jun. 2021.

SINGH, S. et al. Detection of Parkinson's Disease using Machine Learning Algorithm. **International Journal of Computer Applications**, v. 184, n. 6, p. 24–29, 2022.

SOIKKELI, R. et al. Slowing of EEG in Parkinson's disease. **Electroencephalography and Clinical Neurophysiology**, v. 79, n. 3, p. 159–165, 1991.

TANVEER, M. et al. Parkinson's disease diagnosis using neural networks: Survey and comprehensive evaluation. **Information Processing & Management**, v. 59, n. 3, p. 102909, maio 2022.

UGRUMOV, M. Development of early diagnosis of Parkinson's disease: Illusion or reality? **CNS Neuroscience & Therapeutics**, v. 26, n. 10, p. 997–1009, 29 jun. 2020.

WANG, Q. et al. Characterization of EEG Data Revealing Relationships With Cognitive and Motor Symptoms in Parkinson's Disease: A Systematic Review. **Frontiers in Aging Neuroscience**, v. 12, 10 nov. 2020.

YUVARAJ, R.; RAJENDRA ACHARYA, U.; HAGIWARA, Y. A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals. **Neural Computing and Applications**, v. 30, n. 4, p. 1225–1235, 9 dez. 2016.

## ANEXO A – Código CNN LeavePneGroupOut

```
!pip install mne
```

```
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->pooch>=1.5->mne) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->pooch>=1.5->mne) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->pooch>=1.5->mne) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->pooch>=1.5->mne) (2024.8.30)
Downloading mne-1.8.0-py3-none-any.whl (7.4 MB)
```

7.4/7.4 MB 70.8 MB/s eta 0:00:00

```
Installing collected packages: mne
Successfully installed mne-1.8.0
```

```
from glob import glob
import numpy as np
import mne
import itertools
```

```
all_file_path = glob('dataverse_files/*.bdf')
print(len(all_file_path))
```

31

```
healthy_file_path = [i for i in all_file_path if 'h' in i.split('/')[1]]
patient_file_path = [i for i in all_file_path if 'p' in i.split('/')[1]]
print(len(healthy_file_path), len(patient_file_path))
```

16 15

```
def read_data(file_path):
    raw = mne.io.read_raw_bdf(file_path, preload=True)
    epochs = raw.get_data()
    return epochs #trials,channel,length
```

```
control_epochs_array=[read_data(subject) for subject in healthy_file_path]
patients_epochs_array=[read_data(subject) for subject in patient_file_path]
```

```
Extracting EDF parameters from /content/dataverse_files/sub-hc31_ses-hc_task-rest_eeg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 95743 = 0.000 ... 186.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-hc18_ses-hc_task-rest_eeg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 95231 = 0.000 ... 185.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-hc4_ses-hc_task-rest_eeg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 92671 = 0.000 ... 180.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-hc10_ses-hc_task-rest_eeg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 102911 = 0.000 ... 200.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-hc21_ses-hc_task-rest_eeg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 96767 = 0.000 ... 188.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-hc20_ses-hc_task-rest_eeg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 94207 = 0.000 ... 183.998 secs...
```

[illegible]

```

BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 96767 = 0.000 ... 188.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-pd5_ses-off_task-rest_ccg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 100863 = 0.000 ... 196.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-pd28_ses-off_task-rest_ccg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 105471 = 0.000 ... 205.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-pd9_ses-off_task-rest_ccg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 97791 = 0.000 ... 190.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-pd12_ses-off_task-rest_ccg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 105983 = 0.000 ... 206.998 secs...
Extracting EDF parameters from /content/dataverse_files/sub-pd23_ses-off_task-rest_ccg.bdf...
BDF file detected
Setting channel info structure...
Creating raw.info structure...
Reading 0 ... 104447 = 0.000 ... 203.998 secs...

```

```

data_list=control_epochs_array+patients_epochs_array
print(data_list[0].shape)
(41, 95744)

```

```

def segment_time_series(data, m):
    # Ensure the shape is correct
    if data.shape[0] != 41:
        raise ValueError("Input data must have 41 channels.")

    # Number of timesteps
    n = data.shape[1]

    # Calculate the number of segments
    num_segments = n // m

    segments = []

    # Iterate over the number of segments and slice the array
    for i in range(num_segments):
        # Calculate the start and end indices for the slice
        start_idx = i * m
        end_idx = start_idx + m
        segment = data[:, start_idx:end_idx] # Slice the array
        segments.append(segment)

    # Convert the list of segments to a numpy array of shape (num_segments, 41, m)
    segments_array = np.array(segments)

    return segments_array

```

```

m = 5000
teste_array = segment_time_series(data_list[0], m)
teste_full_array = teste_array.transpose(0, 2, 1)
groups_list = [0] * teste_full_array.shape[0]
label_list = [0] * teste_full_array.shape[0]
#print(teste_full_array.shape)

for num in range(1, len(data_list)):
    semented_data = segment_time_series(data_list[num], m)
    rearranged_segment = semented_data.transpose(0, 2, 1)
    teste_full_array = np.concatenate((teste_full_array, rearranged_segment), axis=0)
    teste_full_array
    #print(teste_full_array.shape)
    group_list = [num] * rearranged_segment.shape[0]
    groups_list += group_list

if num > 16:
    label_list += [1] * rearranged_segment.shape[0]
else:
    label_list += [0] * rearranged_segment.shape[0]

print(len(groups_list))
print(len(label_list))
31043
31043

from sklearn.model_selection import LeaveOneGroupOut
from sklearn.preprocessing import StandardScaler
from tensorflow.keras.layers import Conv1D, BatchNormalization, LeakyReLU, MaxPool1D,
GlobalAveragePooling1D, Dense, Dropout, AveragePooling1D
from tensorflow.keras.models import Sequential
from tensorflow.keras.backend import clear_session

logo = LeaveOneGroupOut()
logo.get_n_splits(teste_full_array, label_list, groups_list)
logo.get_n_splits(groups=groups_list) # 'groups' is always required

def cnnmodel():
    clear_session()
    model = Sequential()
    model.add(Conv1D(filters=5, kernel_size=3, strides=1, input_shape=(5000, 41))) #1
    model.add(BatchNormalization())
    model.add(LeakyReLU())

```



```

model.add(MaxPool1D(pool_size=2, strides=2)) #2
model.add(Conv1D(filters=5,kernel_size=3,strides=1)) #3
model.add(LeakyReLU())
model.add(MaxPool1D(pool_size=2, strides=2)) #4
model.add(Dropout(0.5))
model.add(Conv1D(filters=5,kernel_size=3,strides=1)) #5
model.add(LeakyReLU())
model.add(AveragePooling1D(pool_size=2,strides=2))#6
model.add(Dropout(0.5))
model.add(Conv1D(filters=5,kernel_size=3,strides=1))#7
model.add(LeakyReLU())
model.add(AveragePooling1D(pool_size=2,strides=2))#8
model.add(Conv1D(filters=5,kernel_size=3,strides=1))#9
model.add(LeakyReLU())
model.add(GlobalAveragePooling1D())#10
model.add(Dense(1,activation='sigmoid'))#11

model.compile('adam',loss='binary_crossentropy',metrics=['accuracy'])

return model

model=cnnmodel()
model.summary()

```

Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 4998, 5)	620
batch_normalization (BatchNormalization)	(None, 4998, 5)	20
leaky_re_lu (LeakyReLU)	(None, 4998, 5)	0
max_pooling1d (MaxPooling1D)	(None, 2499, 5)	0
conv1d_1 (Conv1D)	(None, 2497, 5)	80
leaky_re_lu_1 (LeakyReLU)	(None, 2497, 5)	0
max_pooling1d_1 (MaxPooling1D)	(None, 1248, 5)	0
dropout (Dropout)	(None, 1248, 5)	0
conv1d_2 (Conv1D)	(None, 1246, 5)	80
leaky_re_lu_2 (LeakyReLU)	(None, 1246, 5)	0
average_pooling1d (AveragePooling1D)	(None, 623, 5)	0
dropout_1 (Dropout)	(None, 623, 5)	0
conv1d_3 (Conv1D)	(None, 621, 5)	80
leaky_re_lu_3 (LeakyReLU)	(None, 621, 5)	0
average_pooling1d_1 (AveragePooling1D)	(None, 310, 5)	0
conv1d_4 (Conv1D)	(None, 308, 5)	80
leaky_re_lu_4 (LeakyReLU)	(None, 308, 5)	0
global_average_pooling1d	(None, 5)	0

(GlobalAveragePooling1D)			
dense (Dense)	(None, 1)	6	

**Total params:** 966 (3.77 KB)  
**Trainable params:** 956 (3.73 KB)  
**Non-trainable params:** 10 (40.00 B)

```

accuracy = []
for i, (train_index, test_index) in enumerate(logo.split(teste_full_array, label_list,
groups_list)):
    train_index = np.array(train_index, dtype=int)
    test_index = np.array(test_index, dtype=int)
    label_list = np.array(label_list)
    #groups_list = np.array(groups_list)

    train_features, train_labels = teste_full_array[train_index], label_list[train_index]
    val_features, val_labels = teste_full_array[test_index], label_list[test_index]

    # Standardize features
    scaler = StandardScaler()
    train_features = scaler.fit_transform(train_features.reshape(-1, train_features.shape[-
1])).reshape(train_features.shape)
    val_features = scaler.transform(val_features.reshape(-1, val_features.shape[-
1])).reshape(val_features.shape)
    model = cnnmodel()

    model.fit(train_features,train_labels,epochs=200,batch_size=100,validation_data=(val_featur
es,val_labels))
    accuracy.append(model.evaluate(val_features,val_labels)[1])
# Print the accuracies of each fold
print("Accuracies:", accuracy)

```