UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Alexandre Herrero Matias

Redes Riemannianas Profundas com Abordagem Temporal para Classificação em Sistemas de Interface Cérebro-Computador

Campinas
2025

Alexandre Herrero Matias

Redes Riemannianas Profundas com Abordagem Temporal para Classificação em Sistemas de Interface Cérebro-Computador

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do título de Mestre em Engenharia Elétrica, na Área de Engenharia de Computação.

Orientador: Prof. Dr. Denis Gustavo Fantinato

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO ALEXANDRE HERRERO MATIAS, E ORIENTADA PELO PROF. DR. DENIS GUSTAVO FANTINATO

Campinas
2025

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca da Área de Engenharia e Arquitetura
Vanessa Evelyn Costa - CRB 8/8295

Matias, Alexandre Herrero, 1998-
M427r Redes riemannianas profundas com abordagem temporal para classificação em sistemas de interface cérebro-computador / Alexandre Herrero Matias. – Campinas, SP : [s.n.], 2025.

Orientador: Denis Gustavo Fantinato.
Dissertação (mestrado) – Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Elétrica e de Computação.

1. Redes neurais (Computação). 2. Interfaces cérebro-computador. 3. Eletroencefalografia. 4. Imaginação motora. 5. Geometria riemaniana. I. Fantinato, Denis Gustavo, 1985-. II. Universidade Estadual de Campinas (UNICAMP). Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações complementares

Título em outro idioma: Deep riemannian networks with temporal approach for classification in brain-computer interfaces
Palavras-chave em inglês:
Neural networks
Brain-computer interfaces
Electroencephalography
Motor imagery
Riemannian geometry
Área de concentração: Engenharia de Computação
Titulação: Mestre em Engenharia Elétrica
Banca examinadora:
Denis Gustavo Fantinato [Orientador]
Romis Ribeiro de Faissol Attux
Ricardo Suyama
Data de defesa: 16-07-2025
Programa de Pós-Graduação: Engenharia Elétrica

Objetivos de Desenvolvimento Sustentável (ODS)
Não se aplica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0009-0000-0814-9721
- Currículo Lattes do autor: https://lattes.cnpq.br/0913542934892207

Denis Gustavo Fantinato

Ricardo Suyama

Romis Ribeiro de Faissol Attux

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de PósGraduação da Faculdade de Engenharia Elétrica e de Computação.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil. Código de Financiamento 001. O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processo nº 2023/00640-1

With every simple act of thinking, something permanent, substantial, enters our soul.

Bernhard Riemann

Resumo

No contexto da classificação de sinais de eletroencefalografia (EEG) para tarefas de Imagética Motora (MI), o uso do arcabouço matemático baseado em Geometria Riemanniana (RG) tem demonstrado desempenhos comparáveis aos das redes neurais convolucionais. A aplicação deste arcabouço em dados de EEG normalmente envolve a extração de uma matriz de covariância amostral, que captura informações espaciais (entre os eletrodos). No entanto, é possível também explorar informações temporais, por exemplo, considerando um conjunto de matrizes de covariância com atraso. Nesse sentido, neste trabalho, propomos o uso explícito da informação temporal, principalmente no contexto de Redes Riemannianas Profundas (DRNs). Além disso, apresentamos uma versão modificada da arquitetura SPDNet, capaz de processar um conjunto de matrizes de covariância como entrada. Nossa abordagem amplia a capacidade da SPDNet em aprender padrões espaço-temporais, resultando em uma melhoria no desempenho da classificação dos sinais de EEG.

Palavras-chave: Interface Cérebro-Computador, Eletroencefalografia, Imagética Motora, Geometria Riemanniana.

Abstract

In the context of electroencephalography (EEG) classification for Motor Imagery (MI) tasks, the use of the framework based on Riemannian Geometry (RG) has shown comparable performances to convolutional neural networks. Its application on EEG data usually considers the extraction of a sample covariance matrix, which captures spatial information (between electrodes). However, the temporal information can be used as well, for instance, considering a set of time-delayed covariance matrices. In that sense, in this work, we propose the use of the temporal information, mainly considering Deep Riemannian Networks (DRN). We also present a modified version of SPDNet in order to encompass a set of covariance matrices as input. Our approach enhances SPDNet's ability to learn spatio-temporal patterns, resulting in improved classification performance on EEG data.

Keywords: Brain-Computer Interface, Electroencephalography, Motor Imagery, Riemannian Geometry.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

AEP	Potencial Evocado Auditivo
ANN	Rede Neural Artificial
BCI	Interface Cérebro-Computado
BiMap	Mapeamento Bilinear
CAR	Média de Referência Comum
CNN	Rede Neural Convolucional
CSP	Padrão Espacial Comum
DNN	Rede Neural Profunda
DL	Aprendizado Profundo
DRN	Rede Riemanniana Profunda
EEG	Eletroencefalografia
ECoG	Eletrocorticografia
FC	Totalmente Conexa
FBCSP	Banco de Filtros de Padrão Espacial Comum
fMRI	Imageamento por Ressonância Magnética Funcional
GPU	Unidade de Processamento Gráfico
LogEig	Logaritmo de Autovalor
MCAE	Matriz de Covariância com Atraso Expandida
MCC	Combinação de Máximo Contraste
MEC	Combinação de Mínima Energia
MDM	Miníma Distância a Média
MEG	Magnetoencefalografia
MI	Imagética Motora
MLP	Perceptron Multicamada

NIRS	Espectroscopia de Infravermelho Próximo
PCA	Análise de Componentes Principais
ReEig	Retificação de Autovalor
ReLU	Unidade Linear Retificada
RG	Geometria Riemanniana
SCM	Matriz de Covariância Amostral
SFS	Seleção Sequencial Avançada
SPD	Simétrica Positiva-Definida
SSEP	Potencial Evocado Somato-sensitivo
SVM	Máquina de Vetor de Suporte
TPU	Unidade de Processamento Tensorial
UDA	Adaptação de Domínio não Supervisionada
VEP	Potencial Evocado Visual

Lista de símbolos

β	Letra grega minúscula beta
δ	Letra grega minúscula delta
ϵ	Letra grega minúscula epsilon
ϕ	Letra grega minúscula fi
η	Letra grega minúscula ita
μ	Letra grega minúscula mi
σ	Letra grega minúscula sigma
Σ	Letra grega sigma
$\forall$	Para todo
$\in$	Pertence

Sumário

Parte I

Interfaces Cérebro-Computador

1 Introdução

As Interfaces Cérebro-Computador (BCI, do inglês *Brain-Computer Interfaces*) representam uma área de pesquisa emergente e fascinante, que busca estabelecer uma comunicação direta entre o cérebro humano e os sistemas computacionais (NAM *et al.*, 2018; NICOLAS-ALONSO; GOMEZ-GIL, 2012). Sendo a eletroencefalografia (EEG) uma de suas principais técnicas para captar a atividade elétrica cerebral de forma não invasiva, as BCIs têm o potencial de transformar diversos campos, desde a neurociência até a reabilitação médica (MANE *et al.*, 2020; LEITE, 2016). Com isso, a análise dos sinais de EEG é um desafio complexo devido à sua alta variabilidade e natureza não estacionária.

Nos últimos anos, as técnicas de aprendizado de máquina (ML, do inglês *Machine Learning*), mais especificamente os modelos de aprendizado profundo (DL, do inglês *Deep Learning*), têm revolucionado a análise de dados complexos, incluindo os sinais de EEG (ABIRI *et al.*, 2019; NAM *et al.*, 2018). Contudo, a combinação de múltiplas fontes neurais misturadas, variabilidade temporal, espacial e espectral, além de interações não lineares e influência de ruídos tornam os dado de EEG intrincados e multidimensionais. Em busca de mitigar esta dificuldade, muitos autores desenvolveram abordagens com a função de trabalhar exclusivamente com o processamento de dados de EEG. Exemplos incluem a Shallow ConvNet e Deep ConvNet (SCHIRRMEISTER *et al.*, 2017), a EEGNet (LAWHERN *et al.*, 2018), a EEG-Inception (SANTAMARíA-VáZQUEZ *et al.*, 2020) e a EE(G)-SPDNet (WILSON *et al.*, 2023). Os resultados alcançados com técnicas voltadas exclusivamente ao processamento de sinais de EEG mostraram-se bastante promissores, entretanto, a escassez de dados mostrou-se um fator preponderante na limitação de atuação destes modelos.

Sob outra perspectiva, em competições envolvendo o processamento de sinais de EEG, destacou-se, na última década, abordagens baseadas em Geometria Riemanniana (RG, do inglês *Riemmanian Geometry*) (RODRIGUES *et al.*, 2019). A RG oferece uma abordagem matemática robusta, permitindo a modelagem dos dados em espaços conhecidos como variedades (ou *manifolds*) Riemannianas, que capturam suas propriedades geométricas fundamentais. Isso permite operações em um espaço de dados que, idealmente, possui menor variabilidade, facilitando, portanto, a análise. Nesse âmbito, as Redes Riemannianas Profundas (DRNs, do inglês *Deep Riemannian Networks*) surgem como uma solução inovadora ao combinar os benefícios dos modelos profundos com a Geometria Riemanniana (WILSON *et al.*, 2023).

O uso das DRNs abre, portanto, um caminho promissor para a pesquisa, visando sanar a alta demanda por dados presentes nas abordagens de DL. Contudo, a abordagem usual das DRNs requer o uso direto de matrizes SPD, as quais podem ser obtidas através de matrizes de covariância que oferecem a correlação, uma estatística de

segunda ordem, entre seus canais/eletrodos, sendo essas informações estritamente espaciais dos dados de EEG. Geralmente, as matrizes de covariância são estimadas usando-se uma janela de dados de EEG, assumindo-se uma estacionariedade local dentro da janela, mesmo que o sinal por completo seja não estacionário. Nesse sentido, uma informação complementar à espacial, pode ser aquela referente à estrutura temporal dos dados.

À luz dessa perspectiva, buscamos neste trabalho propor uma abordagem aprimorada para o uso das DRNs em tarefas de classificação de sinais de EEG, com foco em explorar informações temporais que são geralmente negligenciadas pelas abordagens tradicionais. A proposta é integrar essas informações por meio da utilização de um conjunto de matrizes de covariância com atraso, permitindo uma representação mais rica e informativa dos dados.

Como forma de validar a abordagem proposta, serão utilizados dois conjuntos de dados amplamente reconhecidos na área de interfaces cérebro-computador no paradigma de imagética motora (MI, do inglês *Motor Imagery*): o BCI Competition IV Dataset 2A (BRUNNER *et al.*, 2024) e o conjunto Cho2017 (CHO *et al.*, 2017). Esses conjuntos de dados oferecem registros de EEG obtidos durante tarefas de MI, o que permite a avaliação da eficácia do método tanto em cenários padronizados quanto em dados mais realistas. Além disso, ambos são conjuntos de dados abertos, o que viabiliza uma análise comparativa rigorosa, além de garantir reprodutibilidade e relevância prática aos resultados obtidos.

1.1 Objetivos

As abordagens baseadas em RG mostram um grande potencial e a sua aplicação no processamento de sinais de EEG está em constante desenvolvimento, o que implica a necessidade de uma análise mais aprofundada para avaliar o seu desempenho em diferentes cenários e com dados mais complexos. Serão analisadas suas atuais limitações e desafios tendo foco principal em buscar fazer um uso mais extensivo da informação temporal dos dados de EEG.

1.2 Objetivos Específicos

Os objetivos específicos que orientam o desenvolvimento deste trabalho e delimitam as etapas necessárias para alcançar o propósito proposto são os seguintes:

- Investigar as limitações das abordagens baseadas em RG que consideram apenas padrões espaciais nos sinais de EEG.
- Propor a utilização de matrizes de covariância com atraso como forma de incorporar informações temporais à representação dos dados:

1. Propor a utilização de matrizes de covariância expandidas com atraso.
2. Propor a utilização de tensores com atraso.

- Propor uma modificação da arquitetura padrão da SPDNet com o objetivo de otimizar sua utilização das matrizes de covariância com atraso.
- Avaliar, de forma experimental, o impacto da introdução das matrizes de covariância com atraso no desempenho de classificação de sinais EEG.

1.3 Publicações

Trabalhos publicados ou aceitos para publicação em anais de eventos:

1. MATIAS, A H. Deep Riemannian Networks with Temporal Approach for Classification in Brain-Computer Interfaces. In: *XVII Brazilian Conference on Computational Intelligence (CBIC)*, 2025.
2. MATIAS, A H. Motor Imagery Classification with Deep Learning and Riemannian Geometry. In: *10th Brazilian Institute of Neuroscience and Neurotechnology (BRAINN) Congress*, 2024.
3. MATIAS, A H. Análise Comparativa de Métodos Supervisionados para a Classificação dos Atributos Extraídos pela EEGNet. In: *XV DCA/FEEC/University of Campinas (UNICAMP) Workshop (EADCA)*, 2023.

1.4 Organização do Trabalho

Este trabalho está estruturado de maneira a proporcionar uma compreensão progressiva dos conceitos e técnicas envolvidos no desenvolvimento da proposta. O Capítulo 1 introduz o tema das Interfaces Cérebro-Computador (BCI), apresentando os objetivos gerais e específicos do estudo, bem como as publicações relacionadas. No final deste capítulo, descreve-se a organização geral do conteúdo, guiando o leitor sobre a estrutura do texto.

O Capítulo 2 é dedicado à Fundamentação Teórica. Nele são apresentados os conceitos essenciais que sustentam o desenvolvimento da proposta, como os fundamentos das BCIs, os princípios de DL, sua aplicação a sinais de EEG, os fundamentos da RG e, por fim, as chamadas DRNs, que integram os conceitos anteriores de maneira sofisticada.

No Capítulo 3, detalha-se a proposta de pesquisa deste trabalho. Inicialmente, introduz-se a ideia central desta dissertação: o uso de matrizes de covariância com atraso como alternativa para incorporar informações temporais aos métodos tradicionais, enriquecendo a representação dos dados. Em seguida esta ideia é ramificada em duas propostas

de aplicação, matrizes de covariância com atraso expandidas e tensores com atraso, sendo na última onde é apresentada nossa proposta de arquitetura, a SPDNet-Par.

O Capítulo 4 descreve os conjuntos de dados utilizados nos experimentos. Essas bases foram selecionadas por sua relevância e por disponibilizarem sinais reais de MI.

Os resultados experimentais obtidos a partir da aplicação da proposta são apresentados e discutidos no Capítulo 5, comparando o desempenho da abordagem desenvolvida com métodos previamente estabelecidos na literatura.

O Capítulo 6 traz as conclusões, sintetizando as principais contribuições do trabalho, os resultados alcançados e as limitações observadas.

Por fim, o Capítulo 7 aborda as perspectivas futuras, discutindo possíveis melhorias, variações metodológicas e caminhos de pesquisa que podem ser explorados a partir do que foi desenvolvido nesta dissertação. O trabalho é complementado pela seção de Referências que oferece suporte adicional ao conteúdo técnico e teórico apresentado.

2 Fundamentação Teórica

Neste capítulo é apresentada a fundamentação teórica que embasa este trabalho, fornecendo uma visão abrangente sobre os principais conceitos e abordagens relacionados à pesquisa. Inicialmente, serão introduzidos os fundamentos das BCIs, com destaque para o uso de sinais de EEG em conjunto com o paradigma de MI. Em seguida, serão discutidos os desafios associados à análise desses sinais, como sua variabilidade e baixa relação sinal-ruído, e como técnicas de DL têm sido utilizadas para superar essas limitações. Além disso, será abordada a aplicação de RG como um arcabouço matemático robusto para modelagem e análise de matrizes de covariância, que são frequentemente utilizadas no processamento de sinais de EEG. Por fim, será explorada a integração dessas tecnologias, as DRNs, destacando sua relevância para o desenvolvimento de BCIs mais eficientes e precisas.

2.1 Interface Cérebro Computador

Uma BCI é um sistema de comunicação direta entre o cérebro humano e dispositivos externos, como computadores ou próteses (WOLPAW *et al.*, 2002). Seu funcionamento se baseia na tradução de sinais neurais em comandos que podem controlar diversos dispositivos, desde simples movimentos de cursor em uma tela até próteses robóticas complexas (MULLER-PUTZ; PFURTSCHELLER, 2008).

A ideia básica por trás das BCIs é buscar por padrões em dados capturados de atividades elétricas do cérebro para convertê-los em comandos. Para isso, algumas etapas iniciais são necessárias: desta forma, os sistemas BCI são classificados de acordo com estas etapas, como método de aquisição do sinal/registro cerebral e paradigma (LEITE, 2016).

2.1.1 Métodos de Aquisição/Registro de Sinais

Os métodos de aquisição de sinal em sistemas BCI são divididos em duas principais categorias: métodos invasivos e não invasivos (NICOLELIS, 2001).

2.1.1.1 Métodos Invasivos de Registro

Métodos invasivos geralmente envolvem a implantação de eletrodos diretamente no tecido cerebral, oferecem alta resolução espacial e temporal, mas apresentam riscos associados a procedimentos cirúrgicos e possíveis respostas imunológicas. Seus principais exemplos são:

- **Eletrocorticografia (ECoG)**: uma técnica que envolve a colocação de uma matriz de eletrodos diretamente na superfície do cérebro, abaixo da dura-máter, a membrana externa que cobre o cérebro (LEUTHARDT *et al.*, 2004);
- **Microeletrodos Intracorticais**: são pequenos eletrodos inseridos diretamente no tecido cerebral, penetrando no córtex (HOCHBERG *et al.*, 2006).

2.1.1.2 Métodos não Invasivos de Registro

Métodos não invasivos não apresentam a necessidade de implantes e, geralmente, são mais seguros e fáceis de usar, porém oferecem menor resolução espacial. Exemplos incluem:

- **Imageamento por Ressonância Magnética Funcional (fMRI)**: é uma técnica de neuroimagem funcional que utiliza ressonância magnética para observar a atividade cerebral, detectando variações no fluxo sanguíneo associadas ao aumento da atividade em determinadas regiões do cérebro. Apresenta moderada resolução temporal porém resolução espacial muito alta (SLENES *et al.*, 2013);
- **Espectroscopia de Infravermelho Próximo (NIRS)**: uma técnica que monitora as mudanças na concentração de oxi-hemoglobina e deoxi-hemoglobina nas regiões do cérebro, utilizando luz infravermelha. Apresenta resolução temporal moderada e resolução espacial também moderada (COYLE *et al.*, 2007);
- **Magnetoencefalografia (MEG)**: consiste em uma técnica que mede os campos magnéticos gerados pela atividade neuronal. Apresenta alta resolução temporal porém relativamente baixa resolução espacial (MELLINGER *et al.*, 2007);
- **Eletroencefalografia (EEG)**: uma técnica que registra a atividade elétrica do cérebro através de eletrodos colocados no couro cabeludo. Os eletrodos detectam potenciais elétricos gerados pela atividade neuronal, principalmente os potenciais pós-sinápticos de grandes populações de neurônios do córtex cerebral. Apresenta resolução temporal muito alta, na ordem de milissegundos, porém baixa resolução espacial. Além disso podem apresentar mais ruídos e artefatos visto que os sinais elétricos sofrem atenuação e difusão ao atravessar o crânio e os tecidos circundantes (HIRSCH; BRENNER, 2010).

2.1.2 Paradigmas em BCI

Em sistemas BCI, paradigmas referem-se a diferentes métodos ou abordagens utilizados para capturar e interpretar a atividade cerebral, com o objetivo de traduzir os sinais neurais em comandos. Esses paradigmas são essenciais para determinar como os sinais cerebrais são gerados, capturados e processados. Os paradigmas são divididos em

duas principais classes: paradigmas baseados em atenção seletiva e paradigmas baseados em imagética.

2.1.2.1 Paradigmas baseados em atenção seletiva

Paradigmas baseados em atenção seletiva envolvem direcionar a atenção consciente do usuário para estímulos específicos apresentados em um conjunto de opções. Esses estímulos podem ser visuais, auditivos ou táteis. Seus principais exemplos incluem:

- **P300**: baseia-se na resposta P300, um potencial evocado que ocorre aproximadamente 300 milissegundos após a apresentação de um estímulo relevante, neste paradigma os paciente são apresentados a uma série de estímulos onde um estímulo raro e relevante aparece entre estímulos frequentes e irrelevantes (CHAPMAN; BRAGDON, 1964);
- **Potencial Evocado Somato-sensitivo (SSEP do inglês *Somatosensory Evoked Potentials*)**: se baseia na resposta do cérebro a estímulos táteis, como vibrações ou toques, estes estímulos são aplicados em diferentes partes do corpo, e sua resposta neural é monitorada para determinar a intenção do usuário (MULLER-PUTZ *et al.*, 2006);
- **Potencial Evocado Visualmente (VEP, do inglês *Visually Evoked Potentials*)**: são respostas neurais geradas pelo córtex visual do cérebro em resposta a estímulos visuais específicos, como luzes piscando, essas respostas neurais são capturadas e usadas para determinar a intenção do usuário com base no estímulo ao qual ele está prestando atenção (HERRMANN, 2001);
- **Potencial Evocado Auditivo (AEP, do inglês *Auditory Evoked Potentials*)**: utiliza respostas cerebrais a estímulos auditivos, como sons ou tons, e a resposta do cérebro a esses estímulos é usada para inferir a intenção do usuário (NIJBOER *et al.*, 2008).

2.1.2.2 Paradigmas baseados em imagética

Os paradigmas baseados em imagética envolvem a geração consciente de imagens mentais ou a imaginação de ações motoras sem realizar fisicamente essas ações. Essas imagens mentais ativam áreas específicas do cérebro que podem ser capturadas e interpretadas. Exemplos incluem:

- **Imagética Motora (MI)**: os usuários imaginam realizar movimentos específicos sem realmente executar esses movimentos. A imaginação do movimento ativa regiões motoras do cérebro, gerando padrões de atividade neural que podem ser detectados. Usuários mentalmente simulam ações motoras (como mover a mão direita ou

esquerda). Os padrões de atividade elétrica no cérebro, especialmente no córtex motor, são capturados e classificados (PINEDA *et al.*, 2000).

- **Imagética não Motora**: envolve a geração de imagens mentais que não estão relacionadas diretamente com movimentos físicos, mas podem incluir atividades cognitivas, sensoriais ou emocionais. O usuário é instruído a imaginar ou visualizar mentalmente objetos, cenas, cores, sons, sensações táteis, estados emocionais, entre outros (PALANIAPPAN, 2005).

2.1.3 Etapas de um Sistema BCI

Definidos o método de aquisição e o paradigma, busca-se realizar o principal objetivo de um sistema BCI que é a interpretação dos sinais capturados e o acionamento de comandos correspondentes. Este processo é conhecido como classificação e atribui os sinais a categorias ou classes específicas. Por exemplo, em uma BCI para controle de movimento de próteses robóticas, a classificação envolve identificar os padrões de atividade cerebral associados a diferentes tipos de movimentos, como mover a mão direita, a mão esquerda, ou parar. Seu funcionamento pode ser dividido nas etapas mostradas na Fig. 1:

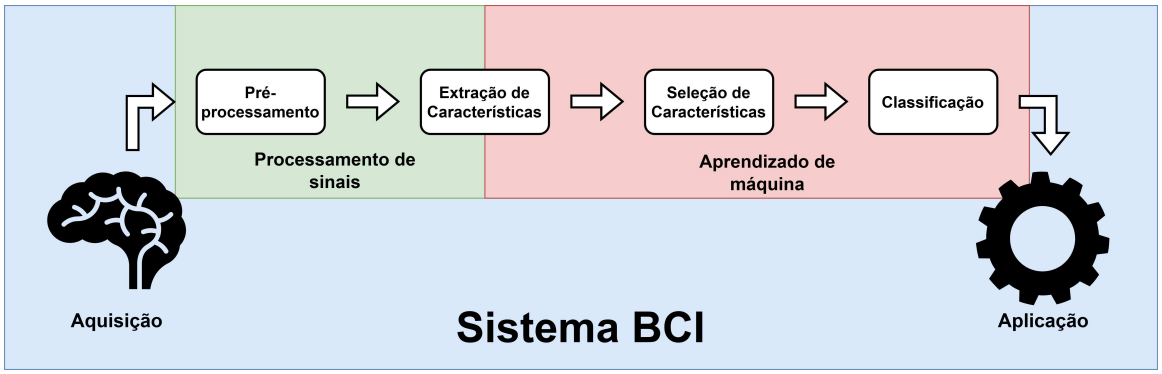

Figura 1 – Etapas de um sistema BCI.

2.1.3.1 Etapa de Pré-processamento dos Sinais

Os sinais cerebrais brutos geralmente passam por uma série de etapas de pré-processamento para remover artefatos, filtrar ruídos e extrair características relevantes dos sinais (MUMTAZ *et al.*, 2021). Embora fontes de ruído sejam complexas e variem de pessoa para pessoa, é possível aplicar técnicas de filtragem que ajudam a eliminar os artefatos e a aumentar a qualidade dos sinais, melhorando sua relação sinal-ruído. Três métodos amplamente utilizados são descritos a seguir:

- **Média de Referência Comum**: a técnica Média de Referência Comum (CAR, do inglês *Common Average Reference*) consiste em redefinir o referencial dos canais para melhorar a relação sinal-ruído. Em um sistema multicanal, o CAR calcula a

média dos potenciais de todos os eletrodos e a utiliza como referencial comum. Em seguida, o potencial de cada canal é então recalculado subtraindo essa média do potencial original do canal. Matematicamente, pode-se definir:

$$V_i' = V_i - \frac{1}{N}\sum_{j=1}^{N} V_j \tag{2.1}$$

onde V_i é o potencial do i-ésimo eletrodo e N é o número total de canais.

- **Combinação de Mínima Energia**: a técnica Combinação de Mínima Energia (MEC, do inglês *Minimun Energy Combination*) visa otimizar a extração de sinais com base em critérios de minimização da energia. A MEC combina os sinais dos diferentes canais em um único sinal com o objetivo de reduzir ruídos e maximizar a concentração de informações relevantes. Isso é feito encontrando uma combinação linear dos sinais que minimiza a energia do sinal resultante. O sinal combinado é dado por:

$$y(t) = \frac{1}{N}\sum_{i=1}^{N} w_i x_i(t) \tag{2.2}$$

onde $x_i(t)$ é o sinal de cada um dos canais e w_i são os pesos atribuídos a cada canal. Os pesos são ajustados, tal que, ao combinar linearmente os canais $x_i(t)$, o sinal resultante $y(t)$ tenha energia mínima, o que ajuda a atenuar ruídos e interferências.

- **Combinação de Máximo Contraste**: a técnica Combinação de Máximo Contraste (MCC, do inglês *Maximum Contrast Combination*) busca maximizar o contraste entre classes ou condições nos sinais, sendo especialmente útil para tarefas de classificação. A MCC combina os sinais dos canais de forma a realçar as diferenças entre dois estados (ou classes) de interesse, como "descanso" versus "movimento". É usado um modelo de otimização para determinar os pesos $\mathbf{w}$ que maximizam a separação entre as médias de sinal μ_1 e μ_2 das classes:

$$J(\mathbf{w}) = \arg\max_{\mathbf{w}} \frac{(\mathbf{w}^\intercal(\mu_1 - \mu_2)^2}{\mathbf{w}^\intercal(\sigma_1 + \sigma_2)\mathbf{w}} \tag{2.3}$$

onde σ_1 e σ_2 são as variâncias dentro de cada classe.

2.1.3.2 Etapa de Extração de Características

São identificadas características distintivas nos sinais cerebrais que possam ser usadas para diferenciar entre diferentes classes de interesse. Conforme ilustra a Fig. 1, isso pode envolver tanto técnicas de processamento de sinal quanto aprendizado de máquina para identificar os padrões relevantes nos dados:

- **Padrão Espacial Comum**: o Padrão Espacial Comum (CSP, do inglês *Common Spatial Pattern*) é uma técnica de análise espacial supervisionada que visa encontrar projeções lineares dos sinais neurais que maximizem a variância para uma classe

enquanto minimizam para outra, aumentando assim a separabilidade entre os estados mentais. Matematicamente, gostaríamos de encontrar a matriz de projeção dos dados que maximiza a razão entre as seguintes variâncias:

$$\mathbf{W} = \arg\max_{\mathbf{W}} \frac{||\mathbf{W}\mathbf{E}_1^2||}{||\mathbf{W}\mathbf{E}_2^2||} \tag{2.4}$$

em que $\mathbf{E}_1$ e $\mathbf{E}_2$ são sinais de EEG pertencentes às classes 1 e 2, e

$$\mathbf{W} = \mathbf{B}^\intercal\mathbf{P} \tag{2.5}$$

sendo $\mathbf{P}$ a matriz de branqueamento (*whitening*), usada para transformar os dados em uma base onde a soma das covariâncias é a identidade, facilitando a separação entre as classes, e $\mathbf{B}$ a matriz de autovetores obtida da decomposição da matriz branqueada de uma das classes, responsável por identificar as componentes que mais distinguem as classes.

- **Banco de Filtros de Padrão Espacial Comum**: o Banco de Filtros de Padrão Espacial Comum (FBCSP, do inglês *Filter Bank Common Spatial Pattern*) é uma extensão direta do CSP que busca capturar informações discriminativas em múltiplas bandas de frequência. O processo começa com a decomposição do sinal neural em diversas bandas usando bancos de filtros, é feito o uso de múltiplos filtros passa-banda (Filtros Chebyshev Tipo II, IIR, com fase zero), com frequências escolhidas manualmente. Para cada sinal filtrado, aplica-se o algoritmo CSP descrito anteriormente, gerando um vetor de características. A concatenação de todos esses vetores forma o vetor final de características do FBCSP.

2.1.3.3 Etapa de Seleção de Características

São determinadas as características mais informativas e as que contribuem pouco para a classificação, desta forma é possível reduzir a dimensionalidade dos dados e melhorar o desempenho do sistema. Essa etapa pode ser realizada por diferentes técnicas, agrupadas principalmente em duas categorias: técnicas de filtro e técnicas de envoltório (do inglês *wrapper*).

- **Técnicas de Filtro**: as técnicas de filtro avaliam cada característica individualmente com base em medidas estatísticas ou métricas de relevância, sem usar um modelo de classificação. No contexto de MI, uma técnica bastante comum é a Informação Mútua (do inglês *Mutual Information*), que mede a dependência entre uma característica e a classe:

$$I(f; y) = \sum_{f \in F} \sum_{y \in Y} p(f, y) \log \left(\frac{p(f, y)}{p(f)p(y)} \right) \tag{2.6}$$

onde f é uma característica extraída, y é a classe e $p(f, y)$ é a distribuição conjunta das variáveis. Quanto maior $I(f; y)$, mais relevante a característica f.

- **Técnicas de Envoltório**: neste caso, utiliza-se um algoritmo de aprendizado (como Máquinas de Vetor-Suporte ou Análise Discriminante Linear) para avaliar diferentes subconjuntos de características, otimizando a combinação que produz melhor desempenho. Técnicas como Seleção Sequencial Avançada (SFS, do inglês *Sequential Forward Selection*) ou Algoritmos Genéticos são utilizadas. Embora eficazes, são computacionalmente custosas, especialmente quando se tem um grande número de características (como em FBCSP, com múltiplas bandas de frequência).

2.1.3.4 Etapa de Classificação

A etapa de classificação em sistemas BCI representa o momento em que as características extraídas e selecionadas dos sinais cerebrais são utilizadas para atribuir uma classe à atividade cerebral registrada como, por exemplo, distinguir se o usuário está imaginando mover a mão direita ou esquerda. Essa etapa é essencial, pois traduz a informação cerebral em comandos compreensíveis por máquinas ou sistemas computacionais.

Diversos classificadores podem ser aplicados nesse contexto. A escolha do modelo de classificação adequado depende de fatores como o número de amostras disponíveis, a dimensionalidade dos dados e a complexidade da separação entre as classes.

Os classificadores utilizados neste trabalhos serão apresentados e discutidos com maior profundidade na seções seguintes, onde serão descritos seus fundamentos teóricos, justificativas para escolha e aplicação prática no contexto de MI com sinais de EEG.

2.1.4 Desafios em Sistemas BCI

Sinais neurais são multidimensionais e não estacionários, apresentando uma alta variabilidade ao longo do tempo. Capturar as informações espaciais e temporais intrínsecas nestes dados é desafiador, e classificadores clássicos têm dificuldade em lidar com essa complexidade. Já técnicas de aprendizado profundo podem aprender automaticamente representações de alto nível dos dados brutos, eliminando a necessidade de extração de características manual, além de serem capazes de capturar representações mais ricas e complexas dos dados, o que melhora significativamente o desempenho dos modelos em tarefas de classificação, desde que providos com um volume de dados suficiente para o treinamento (GOODFELLOW *et al.*, 2016).

2.2 Aprendizado Profundo

O Aprendizado Profundo (DL) é uma subárea do aprendizado de máquina que se destaca pela capacidade de modelar padrões complexos em dados de alta dimensionalidade, utilizando arquiteturas inspiradas no funcionamento das Redes Neurais biológicas.

Esse paradigma tem sido amplamente utilizado no processamento de sinais cerebrais, especialmente no contexto de BCIs, devido à sua habilidade de lidar com a complexidade e a variabilidade inerentes aos dados neurais.

2.2.1 Redes Neurais Artificiais

DL utiliza Redes Neurais Artificiais (ANN, do inglês *Artificial Neural Network*) compostas por múltiplas camadas, que são capazes de aprender representações hierárquicas de seus dados de entrada. Estas redes são compostas por unidades básicas chamadas neurônios artificiais. Um modelo simples e historicamente fundamental de neurônio artificial é conhecido como Perceptron. Para entender o funcionamento destas ANNs, começamos pela definição e funcionamento do Perceptron, que serve como base para a construção de modelos mais complexos.

O Perceptron é um modelo matemático inspirado em um neurônio biológico. Ele recebe um conjunto de entradas, aplica pesos a essas entradas, soma os resultados ponderados, adiciona um viés (*bias*) e passa o resultado por uma função de ativação para gerar uma saída:

$$y = \phi(z), \tag{2.7}$$

onde z é o valor combinado das entradas ponderadas, dado por:

$$z = \sum_{i=1}^{n} w_i x_i + b_i = \mathbf{w}^\intercal \mathbf{x} + \mathbf{b}, \tag{2.8}$$

sendo $\mathbf{x} = [x_1, x_2, ..., x_n]^\intercal$ o vetor de entradas, $\mathbf{w} = [w_1, w_2, ..., w_n]^\intercal$ o vetor de pesos, $\mathbf{b} = [b_1, b_2, ..., b_n]^\intercal$ o vetor de *bias* e n o número de entradas. Por sua vez, $\phi(\cdot)$ é a função de ativação, que define como o Perceptron decide a saída. No Perceptron clássico, a função de ativação é um degrau:

$$\phi(z) = \begin{cases} 1 & \text{se } z \geq 0 \\ 0 & \text{se } z < 0 \end{cases} \tag{2.9}$$

porém, outras funções podem ser utilizadas como a ReLU (*Rectified Linear Unit*), tangente hiperbólica e outras (BISHOP, 2006; GOODFELLOW *et al.*, 2016).

O Perceptron é utilizado para problemas de classificação binária, separando os dados em duas classes de acordo com a função de decisão. Seu treinamento é realizado usando o algoritmo de aprendizado do Perceptron, baseado no gradiente, que ajusta os pesos w_i e os *bias* b_i para minimizar os erros de classificação.

O modelo mais simples de ANN, composta por Perceptrons combinados organizados em camadas, é conhecido como Perceptron Multicamada (MLP, do inglês *Multilayer Perceptron*). Esta rede é composta por três elementos básicos: Camada de Entrada, Camadas Ocultas e Camada de Saída, como pode ser visualizado na Fig. 2. Cada camada pode ser representada da seguinte forma:

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}, \tag{2.10}$$

$$\mathbf{a}^{(l)} = \phi(\mathbf{z}^{(l)}), \tag{2.11}$$

onde $\mathbf{z}^{(l)}$ é o vetor das ativações ponderadas na camada l, $\mathbf{W}^{(l)}$ é a matriz de pesos dos neurônios da camada l, $\mathbf{a}^{(l)}$ é o vetor de ativações da camada l, $\mathbf{b}^{(l)}$ é o vetor de *bias* da camada l e $\phi(\cdot)$ é a função de ativação, sendo as principais ReLU e Sigmoide:

$$\text{ReLU}(x) = \max(0, x) \tag{2.12}$$

$$\sigma(x) = \frac{1}{1 + \mathrm{e}^{-x}} \tag{2.13}$$

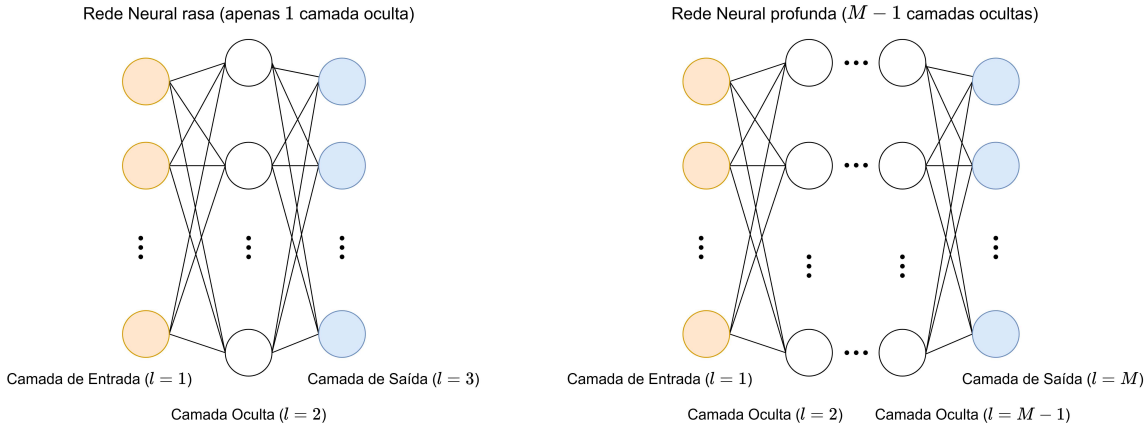

Figura 2 – Representação gráfica de ANNs.

Dois casos que merecem destaque são as saídas das camadas de entrada ($l = 1$) e saída ($l = M$) visto que é onde os dados são processados no início e no final da rede, respectivamente. Na camada de entrada não há transformação do dado, apenas sua leitura:

$$\mathbf{a}^{(1)} = \phi(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) \tag{2.14}$$

por outro lado, a saída da camada de saída é a seguinte:

$$\hat{\mathbf{y}} = \phi(\mathbf{W}^{(M)}\mathbf{a}^{(M-1)} + \mathbf{b}^{(M)}) \tag{2.15}$$

O treinamento das ANNs é realizado por meio de um processo conhecido como retropropagação (*backpropagation*), que utiliza o algoritmo de gradiente descendente. Esse processo ocorre em três etapas principais:

1. Propagação para Frente (*Forward Pass*): Calcula as ativações de cada camada a partir das entradas e dos pesos atuais.

2. Retropropagação (*Backpropagation*): Calcula o gradiente do erro em relação aos pesos e *bias*, usando a regra da cadeia para propagar o erro da camada de saída para as camadas anteriores.

3. Atualização dos Pesos: Ajusta os pesos e *bias* de acordo com a regra de aprendizado adotada (sendo o uso gradientes uma das formas mais comuns):

$$\mathbf{W}' \leftarrow \mathbf{W}' - \eta \frac{\partial L}{\partial \mathbf{W}'}, \tag{2.16}$$

onde $\mathbf{W}' = \begin{bmatrix} \mathbf{W} & \mathbf{b} \end{bmatrix}$ é a matriz expandida para incluir os pesos ($\mathbf{W}$) e *bias* ($\mathbf{b}$), η é taxa de aprendizado (ou *learning rate*) e L é a função custo ou perda que quantifica o erro. Dois exemplos de função custo são o erro quadrático médio e a entropia cruzada. A primeira é geralmente usada para problemas de regressão enquanto que a segunda para problemas de classificação. A função custo para o erro quadrático médio é a seguinte:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{2.17}$$

onde N_c é o número de amostras, $\hat{y}_i$ é o valor previsto pela rede para a i-ésima amostra e y_i é o valor verdadeiro correspondente à i-ésima amostra. Já a função custo para a entropia cruzada é a seguinte:

$$L_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log\left(\hat{y}_i\right) + \left(1 - y_i\right) \log\left(1 - \hat{y}_i\right) \right] \tag{2.18}$$

onde $\hat{y}_i$ é a probabilidade prevista para a classe positiva (saída da função sigmoide) e $y_i \in \{0, 1\}$ é o rótulo verdadeiro (0 ou 1).

Embora as MLPs tenham representado um marco no avanço da modelagem de problemas complexos, suas características estimularam o surgimento de arquiteturas mais especializadas. Em particular, como não possuem mecanismos intrínsecos para explorar padrões específicos nos dados, tais como relações espaciais em imagens ou dependências temporais em séries, muitas vezes é necessário recorrer a técnicas externas de extração ou seleção de atributos, como a análise de componentes principais (PCA, do inglês *Principal Component Analysis*) ou métodos manuais. Esse cenário incentivou o desenvolvimento de novas arquiteturas capazes de integrar tais capacidades de forma mais direta. Além disso, em uma MLP, cada neurônio é totalmente conexo (FC, do inglês *Fully Connected*), ou seja, está conectado a todos os neurônios da camada anterior e essa densidade de conexões faz com que o número de parâmetros cresça rapidamente com o aumento do número de neurônios e camadas e, desta forma, para alcançar um treinamento eficaz as redes mais profundas requerem um grande volume de dados que represente adequadamente seu espaço de busca.

Arquiteturas mais modernas das ANNs, como as Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) e as Redes Neurais Recorrentes (RNNs, do inglês *Recurrent Neural Networks*), oferecem ferramentas poderosas para extrair padrões espaciais e temporais dos dados neurais. As CNNs, por exemplo, são particularmente eficazes para identificar padrões locais em dados estruturados, como aqueles representados em espectrogramas.

2.2.2 Redes Neurais Convolucionais

As CNNs representam uma das mais influentes arquiteturas de DL na atualidade. Elas foram projetadas originalmente para tarefas de visão computacional, como o reconhecimento de imagens, mas seu poder de modelar estruturas espaciais e extrair padrões locais relevantes tem favorecido sua aplicação em diversas áreas, incluindo o processamento de sinais neurais, como os sinais de EEG.

Diferentemente das MLPs, as CNNs aproveitam a estrutura espacial dos dados por meio da operação de convolução. A ideia central é aplicar filtros (do inglês *kernels*) que varrem os dados de entrada e capturam padrões locais.

Seja $\mathbf{X} \in \mathbb{R}^{H \times W}$ uma entrada bidimensional, como uma imagem ou um sinal convertido em representação matricial (por exemplo, tempo x canais no EEG). Um *kernel* convolucional $\mathbf{K} \in \mathbb{R}^{k_h \times k_w}$ de dimensões menores realiza a operação de convolução dada por:

$$(\mathbf{X} * \mathbf{K})_{i,j} = \sum_n \sum_m \mathbf{X}_{i-m,j-n} \mathbf{K}_{m,n} \tag{2.19}$$

O resultado dessa operação é um mapa de ativação, onde os valores mais altos indicam forte correspondência entre o padrão detectado pelo *kernel* e a região da entrada.

As CNNs são compostas tipicamente por três tipos principais de camadas:

- **Camadas Convolucionais**: responsáveis pela detecção de padrões locais. Cada camada convolucional aprende múltiplos *kernels*, permitindo capturar diferentes tipos de informações.

- **Camadas de *Pooling***: as camadas de subamostragem (do inglês *pooling*) reduzem a dimensionalidade dos mapas de ativação, preservando as informações mais relevantes e tornando a rede menos sensível a pequenas variações. A forma mais comum é a subamostragem máxima (*max pooling*), onde se extrai o valor máximo de pequenas regiões:

$$y_{i,j} = \max_{(m,n) \in \mathcal{R}_{i,j}} x_{m,n} \tag{2.20}$$

- **Camadas Totalmente Conexas**: as camadas FC, são utilizadas após a extração de características, são geralmente utilizadas para realizar a classificação a partir das informações extraídas.

Além disso, funções de ativação como ReLU são aplicadas após cada convolução para introduzir não-linearidade, e técnicas como normalização e *dropout* (desativação aleatória de neurônios durante o treino) também são usadas para melhorar a generalização.

2.3 Aprendizado Profundo Aplicado a Sinais de EEG

O DL representa uma evolução das ANNs tradicionais, estendendo a arquitetura das redes para múltiplas camadas ocultas e tornando possível o aprendizado de representações mais abstratas e hierárquicas dos dados. Embora o conceito de MLPs já tenha sido introduzido, DL se destaca por sua capacidade de escalar essas arquiteturas e treinar modelos eficazes em problemas complexos com grande volume de dados.

Essa abordagem ganhou destaque com o avanço computacional (uso de GPUs e TPUs), o crescimento de bases de dados públicas e o desenvolvimento de novas técnicas de treinamento. O resultado é um modelo capaz de lidar com dados não estruturados, como sinais de EEG, imagens, texto e áudio, extraindo padrões que métodos mais rasos dificilmente capturam.

No contexto de BCIs, DL oferece vantagens expressivas ao lidar com dados altamente ruidosos e não estacionários, como os sinais de EEG. Modelos como EEGNet (LAWHERN *et al.*, 2018), a EEG-Inception (SANTAMARíA-VáZQUEZ *et al.*, 2020) e a EE(G)-SPDNet (WILSON *et al.*, 2023), já podem ser considerados modelos profundos, e frequentemente alcançam desempenho superior aos métodos mais rasos ou baseados em extração manual de características

2.3.1 EEGNet

A EEGNet (LAWHERN *et al.*, 2018) é uma CNN amplamente utilizada em BCIs, seu design é otimizado para sinais de EEG, que possuem características temporais e espaciais específicas. A EEGNet utiliza convoluções temporais para aprender os padrões ao longo do tempo em cada canal, seguidas por convoluções espaciais que modelam as interações entre os canais. A combinação dessas operações é essencial para capturar a dinâmica cerebral registrada pelos eletrodos. Além disso, camadas de *pooling* são aplicadas para reduzir a dimensionalidade, preservando as informações mais relevantes e reduzindo o custo computacional. Sua arquitetura é ilustrada na Fig. 3.

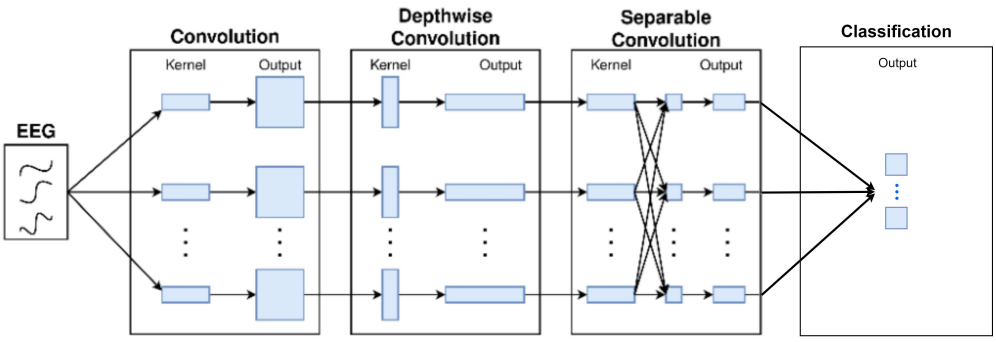

Figura 3 – Arquitetura da EEGNet.

Adaptado de (LAWHERN *et al.*, 2018)

A primeira camada convolucional aplica filtros ao longo do eixo temporal, permitindo extrair padrões dinâmicos no tempo para cada canal de EEG. A operação pode ser representada como:

$$\mathbf{Z}_{f,t}^{(1)} = \sum_{\tau=1}^{F} \mathbf{X}_{c,t+\tau} \mathbf{W}_{\tau,f}^{(1)} + \mathbf{b}_f^{(1)} \tag{2.21}$$

onde $\mathbf{X}_{c,t}$ representa o sinal EEG no canal c e tempo t, $\mathbf{W}_{\tau,f}^{(1)}$ é o *kernel* convolucional para o filtro f, F é o tamanho do filtro temporal e $\mathbf{b}_f^{(1)}$ é o *bias* associado ao filtro f. Essa camada opera independentemente para cada canal.

Após a convolução temporal, a convolução espacial é aplicada para capturar as interações entre canais. Essa camada pode ser formulada como:

$$\mathbf{Z}_{f,t}^{(2)} = \sum_{c=1}^{C} \mathbf{Z}_{c,t}^{(1)} \mathbf{W}_{c,f}^{(2)} + \mathbf{b}_f^{(2)} \tag{2.22}$$

aqui, C é o numero de canais, $\mathbf{W}_{c,f}^{(2)}$ representa os pesos que conectam os canais c ao filtro f e $\mathbf{b}_f^{(2)}$ é o *bias* associado.

Para reduzir a dimensionalidade e melhorar a estabilidade do treinamento, uma camada de *pooling* é aplicada, como *average pooling* (subamostragem média) ou a *max pooling*:

$$\mathbf{Z}_{f,t}^{(3)} = \text{Pooling}(\mathbf{Z}_{f,t}^{(2)}, \text{stride}) \tag{2.23}$$

isso é seguido por técnicas como normalização de lotes (*batch normalization*) e *dropout* para evitar o sobreajuste (*overfitting*).

Além disso, a EEGNet utiliza convoluções em profundidade (*depthwise*) separáveis para reduzir a complexidade computacional. A convolução em profundidade aplica um filtro em cada canal separadamente, e a convolução pontual (*pointwise*) combina as saídas:

$$\mathbf{Z}_{f,t}^{\text{depthwise}} = \sum_{\tau=1}^{F} \mathbf{X}_{c,t+\tau} \mathbf{W}_{\tau,c}^{\text{depthwise}} \tag{2.24}$$

$$\mathbf{Z}_{f,t}^{\text{pointwise}} = \sum_{c=1}^{C} \mathbf{Z}_{c,t}^{\text{depthwise}} \mathbf{W}_{c,f}^{\text{pointwise}} \tag{2.25}$$

As saídas convolucionais e reduzidas pelo *pooling* são achatadas e passadas por camadas FC. A última camada geralmente utiliza uma função de ativação *softmax* para realizar a classificação, dada por:

$$\hat{y}_k = \frac{\exp(a_k)}{\sum_{j=1}^{K} \exp(a_j)} \tag{2.26}$$

onde a_k é a entrada para a classe k, e K é o número total de classes.

2.4 Geometria Riemanniana

Além do uso de DL, outra poderosa ferramenta para o processamento de sinais cerebrais é a aplicação de Geometria Riemanniana (RODRIGUES, 2019). RG é um ramo da matemática que estuda espaços diferenciáveis, onde conceitos como métricas e curvaturas são definidos (RODRIGUES *et al.*, 2019). No contexto das BCIs, a RG tem sido explorada para melhorar a compreensão e análise dos padrões neurais (CONGEDO *et al.*, 2017).

A RG possibilita a utilização de conjuntos de matrizes SPD, as quais sinais neurais podem ser mapeados, onde as propriedades estatísticas dos sinais serão completamente descritas, desde que se assuma que o sinal é estacionário e suas estatísticas seguem a suposição de Gaussianidade (MOAKHER, 2005). Matematicamente, para uma matriz $\mathbf{P} \in \mathbb{R}^{N \times N}$ ser SPD, ela deve estar contida no espaço de matrizes SPD, definido da seguinte forma:

$$\text{SPD}(N) = S(N) \cap P(N) \tag{2.27}$$

onde $S(N) = \left\{\mathbf{P} \in \mathbb{R}^{N \times N}, \mathbf{P} = \mathbf{P}^{\intercal}\right\}$ é o espaço das matrizes simétricas, enquanto $P(N) = \left\{\mathbf{P} \in \mathbb{R}^{N \times N}, \mathbf{u}^{\intercal}\mathbf{P}\mathbf{u} > 0, \forall \mathbf{u} \in \mathbb{R}^{N}\right\}$ é o espaço das matrizes definidas positivas.

O espaço das matrizes SPD, dotado da métrica Riemanniana, é uma variedade (ou *manifold*) diferenciável Riemanniana $\mathcal{M}$. Os conceitos de distância Riemanniana e espaço tangente desempenham um papel importante na aplicação de variedades Riemannianas. Denotadas por duas matrizes SPD $\mathbf{A}$, $\mathbf{B} \in \text{SPD}(N)$, a distância Riemanniana é definida como:

$$\delta_R(\mathbf{A}, \mathbf{B}) = \left\|\log\left(\mathbf{A}^{-\frac{1}{2}}\mathbf{B}\mathbf{A}^{-\frac{1}{2}}\right)\right\|_F \tag{2.28}$$

onde, $\|\cdot\|_F$ é a norma de Frobenius de uma matriz:

$$\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|m_{ij}|^2} \tag{2.29}$$

A distância Riemanniana é o menor comprimento da curva que conecta dois pontos na variedade Riemanniana. Ela satisfaz as três propriedades fundamentais do espaço métrico: positividade, simetria e a desigualdade triangular.

O espaço tangente de uma variedade Riemanniana é um espaço linear que frequentemente é utilizado para projetar dados da variedade Riemanniana para um espaço Euclidiano. O espaço tangente $\mathcal{T}(N)$ em um ponto $\mathbf{P} \in \text{SPD}(N)$ é definido como:

$$\mathcal{T}(N) = \left\{\mathbf{s}_i = \text{upper}(\mathbf{P}^{-\frac{1}{2}}\text{Log}_{\mathbf{P}}(\mathbf{P}_i)\mathbf{P}^{-\frac{1}{2}}) \in \mathbb{R}^{N(N+1)/2}\right\} \tag{2.30}$$

onde $\mathbf{P}$ é um ponto tangente, i é o índice do ponto $\mathbf{P}_i \in \mathcal{M}$, $\text{upper}(\cdot)$ é o operador que mantém a parte triangular superior da matriz e a vetoriza, $\text{Log}_{\mathbf{P}}(\mathbf{P}_i) = \mathbf{P}^{\frac{1}{2}}\log\left(\mathbf{P}^{-\frac{1}{2}}\mathbf{P}_i\mathbf{P}^{-\frac{1}{2}}\right)\mathbf{P}^{\frac{1}{2}}$ é o operador de mapeamento logarítmico.

Uma vez obtido o espaço tangente, podemos aproximar a distância Riemanniana entre duas matrizes SPD utilizando a sua distância euclidiana no espaço tangente, como ilustrado na Fig. 4:

$$\delta_R(\mathbf{P}, \mathbf{P}_i) \approx \|\mathbf{s} - \mathbf{s}_i\|_F \tag{2.31}$$

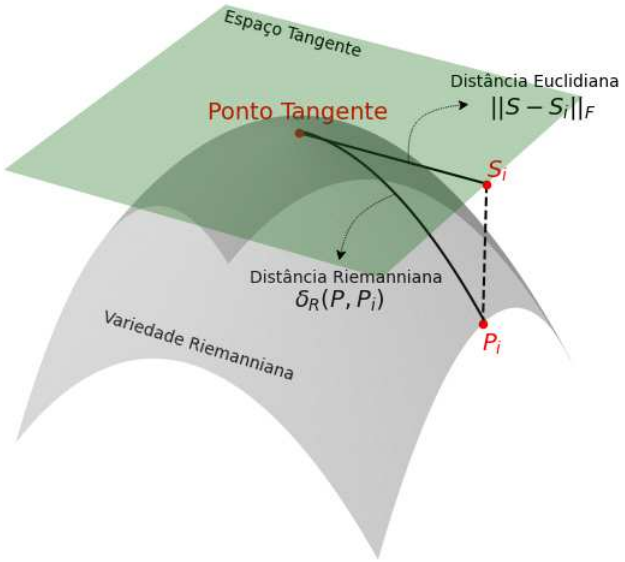

Figura 4 – Visualização da variedade Riemanniana.

Além disso, uma vez que os dados estejam projetados no espaço tangente, também é possível fazer o uso de classificadores tradicionais nestes dados, recurso amplamente explorado pelas Redes Riemannianas Profundas, cujo funcionamento será discutido mais adiante.

2.4.1 Algoritmo MDM

O algoritmo K-NN (do inglês *K-Nearest Neighbors*) (BISHOP, 2006) é um método clássico muito conhecido de classificação não supervisionado baseado na proximidade entre amostras no espaço de características. De forma análoga, no contexto Riemanniano, o método Mínima Distância à Média (MDM, do inglês *Minimum Distance to Mean*) (BARACHANT *et al.*, 2012) também utiliza distâncias para classificar matrizes SPD. Dada uma matriz SPD $\mathbf{P} \in \text{SPD}(N)$ não rotulada, sua classe será a mesma das matrizes já rotuladas cujo centroide apresenta a menor distância Riemanniana, conforme ilustra a Fig. 5:

O calculo do centroide é feito utilizando a média geométrica Riemanniana das posições das matrizes de uma mesma classe:

$$\mathfrak{G}_r(\mathbf{P}_{r1}, \ldots, \mathbf{P}_{rm}) = \arg\min_{\mathbf{P}_r} \sum_{k=1}^{m} \delta_R^2(\mathbf{P}_r, \mathbf{P}_{rk}) \tag{2.32}$$

onde r é a classe das matrizes que desejamos calcular o centroide. Com isso, é possível descrever a classificação de uma matriz SPD $\mathbf{P} \in \text{SPD}(N)$ não rotulada pelo algoritmo

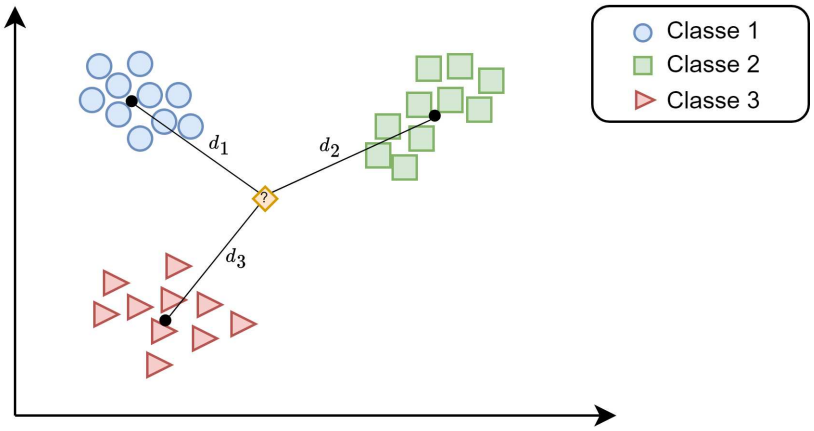

Figura 5 – Funcionamento do MDM.

MDM da seguinte forma:

$$c = \arg\min_{l} \delta_R(\mathbf{P}, \mathfrak{G}_l) \tag{2.33}$$

onde c é a classe com a qual a matriz $\mathbf{P}$ será rotulada.

2.5 Redes Riemannianas Profundas

Uma vez que ambas as abordagens, DL e RG mostraram-se promissoras no contexto de classificação, o próximo passo mais natural foi combiná-las, dando origem às Redes Riemannianas Profundas (DRN, do inglês *Deep Riemannian Network*) (HUANG; GOOL, 2016; WILSON *et al.*, 2023; XIE *et al.*, 2022).

Destacam-se, entre elas, as arquiteturas SPDNet (HUANG; GOOL, 2016), EE(G)-SPDNet (WILSON *et al.*, 2023) e TSMNet (KOBLER *et al.*, 2022). As DRNs alcançam altas taxas de acurácia em diversos conjuntos de dados de EEG, proporcionando uma nova perspectiva para o desenvolvimento de sistemas BCI.

2.5.1 SPDNet

A SPDNet (HUANG; GOOL, 2016), sendo uma DRN, também foi projetada especificamente para processar dados representados por matrizes SPD. Diferentemente do algoritmo MDM, ao invés de mapear os dados para espaços euclidianos, a SPDNet utiliza operações compatíveis com a RG, como mapeamentos logarítmicos e exponenciais, permitindo a extração de representações discriminativas a partir das matrizes SPD sem comprometer sua natureza geométrica. Suas componentes são descritas a seguir:

- **Entrada de Dados**: A SPDNet aceita como entrada matrizes SPD, como as já descritas matrizes de covariância ou matrizes de similaridade, que também são comuns em tarefas de reconhecimento de padrões e processamento de imagens médicas.

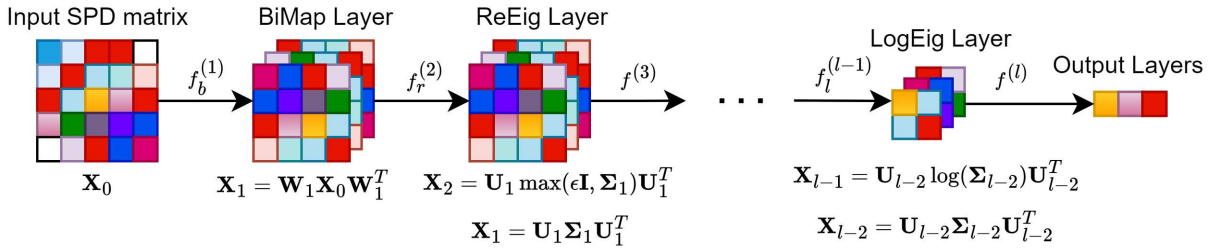

Figura 6 – Arquitetura da SPDNet.
Adaptado de (HUANG; GOOL, 2016).

- **Camadas Ocultas**: A SPDNet geralmente inclui camadas projetadas especificamente para operar com matrizes SPD. Essas camadas convolucionais preservam a estrutura da matriz original e são capazes de extrair características relevantes dos dados. Três camadas importantes que trabalham com a variedade SPD são as camadas mapeamento bilinear (BiMap), retificação de autovalor (ReEig) e logaritmo de autovalor (LogEig).

  1. **Camadas BiMap** (representadas no segundo bloco da Fig. 6): Convertem matrizes SPD em novas matrizes, que idealmente contêm mais informações. Matematicamente, é definida da seguinte forma:

  $$\mathbf{X}_k = \mathbf{W}_k\mathbf{X}_{k-1}\mathbf{W}_k^\intercal \tag{2.34}$$

  onde $\mathbf{X}_{k-1}$ é a matriz SPD de entrada na k-ésima camada, $\mathbf{W}_k$ é a matriz de transformação e $\mathbf{X}_k$ é a matriz resultante.

  2. **Camadas ReEig** (representadas no terceiro bloco da Fig.6): Introduzem uma não linearidade (semelhante à ReLU) ao modificar os autovalores da matriz SPD:

  $$\mathbf{X}_k = \mathbf{U}_{k-1} \max(\epsilon\mathbf{I}, \mathbf{\Sigma}_{k-1})\mathbf{U}_{k-1}^\intercal \tag{2.35}$$

  onde $\mathbf{U}_{k-1}$ e $\mathbf{\Sigma}_{k-1}$ são obtidas por decomposição de autovalores ($\mathbf{X}_{k-1} = \mathbf{U}_{k-1}\mathbf{\Sigma}_{k-1}\mathbf{U}_{k-1}^\intercal$) e ϵ é um limiar escolhido. A utilização da camada ReEig impede que as matrizes se tornem não-positivas.

  3. **Camadas LogEig** (representadas no penúltimo bloco da Fig. 6): Transformam a matriz de entrada para um espaço onde operações lineares podem ser aplicadas de maneira mais eficaz, enquanto ainda respeita a estrutura geométrica intrínseca do espaço SPD:

  $$\mathbf{X}_k = \mathbf{U}_{k-1} \log(\mathbf{\Sigma}_{k-1})\mathbf{U}_{k-1}^\intercal \tag{2.36}$$

onde $\mathbf{U}_{k-1}$ e $\mathbf{\Sigma}_{k-1}$ são obtidas da mesma forma por decomposição de autovalores, e a base do logaritmo é e. O mapeamento para o espaço tangente via logaritmo, realizado pela Eq. 2.36 permite que a matriz SPD resultante seja manipulada em um espaço onde operações lineares são válidas e mais fáceis de realizar. Tais propriedades são importantes para a classificação nas camadas finais da rede.

- **Camadas de *Pooling***: Assim como as CNNs, a SPDNet pode incluir camadas de *pooling* para reduzir a dimensionalidade dos dados e extrair características importantes de forma mais eficiente.
- **Camadas de Normalização e Ativação**: A rede pode conter camadas de normalização para garantir que as matrizes de saída mantenham suas propriedades de positividade-definida e simetria. Além disso, as camadas de ativação são aplicadas para introduzir não-linearidades na rede.
- **Camadas Totalmente Conexas**: No final da rede, podem existir camadas FC para realizar a classificação ou regressão com base nas características extraídas das matrizes de entrada.

2.5.2 EE(G)-SPDNet

A EE(G)-SPDNet (*End-to-End* EEG SPDNet) (WILSON *et al.*, 2023) é uma DRN capaz de ser treinada de ponta a ponta (*end-to-end*), equipada com um banco de filtros adaptado aos conjuntos de dados, oferecendo, portanto, maior potencial de aplicação em diferentes paradigmas.

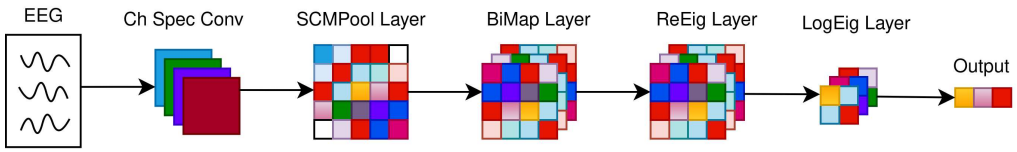

Figura 7 – Arquitetura da EE(G)-SPDNet com filtragem espacífica.

A solução encontrada para a incorporação do banco de filtros foi a inserção de camadas convolucionais na entrada de uma SPDNet. Dessa forma, a operação de convolução unidimensional (ao longo do eixo temporal) atua diretamente sobre os sinais de EEG, funcionando como um filtro passa-faixa:

1. **Camada Convolucional**: A camada convolucional inicial segue uma estrutura similar à da arquitetura EEGNet, atuando como filtros temporais. No entanto, existem duas abordagens possíveis:

- **Filtragem convencional (*channel independent*)**: um único filtro é aplicado a todos os canais — cada filtro gera uma saída com o mesmo comprimento do sinal de EEG de entrada (convolução do tipo *same*).
- **Filtragem específica (*channel specific*)**: um ou mais filtros são aplicados por canal — essa abordagem possui mais parâmetros treináveis, mas oferece maior flexibilidade na seleção de frequências específicas para cada canal. Nesse caso, as saídas de cada convolução são empilhadas, como mostrado na Fig. 7.

2. **Camada SCM *Pooling***: Em seguida, a camada SCM (*Sample Covariance Matrix*) *Pooling* extrai as estimativas das matrizes de covariância.

   - **Filtragem específica**: se essa opção foi utilizada, a estimativa da matriz de covariância é direta.
   - **Filtragem convencional**: nessa opção, cada filtro, após a convolução, gera um sinal com a mesma dimensão da entrada, sendo possível extrair uma estimativa da matriz de covariância para cada filtro.

     Sejam $\mathbf{S}_n$ e $\mathbf{S}_m$ duas matrizes SPD de dimensões n e m, respectivamente:

     $$\text{Conc}(\mathbf{S}_n, \mathbf{S}_m) = \begin{bmatrix} \mathbf{S}_n & \mathbf{0}_{n\times m} \\ \mathbf{0}_{m\times n} & \mathbf{S}_m \end{bmatrix} \tag{2.37}$$

     Esse procedimento permite que matrizes SPD de diferentes dimensões sejam combinadas, gerando uma nova matriz $\text{Conc}(\mathbf{S}_n, \mathbf{S}_m)$ — uma matriz diagonal em blocos — de maior dimensão, que também é SPD.

     Para este trabalho, a estratégia de filtragem adaptada foi a filtragem específica.

2.5.3 TSMNet

A TSMNet (KOBLER *et al.*, 2022) é uma arquitetura Riemanniana projetada para enfrentar desafios de adaptação de domínio não supervisionada (UDA, do inglês *Unsupervised Domain Adaptation*) em sinais de EEG, especialmente em cenários de transferência inter-sessão e inter-sujeito. Utiliza técnicas de adaptação de domínio e normalização de momentum em lote (*batch momentum normalization*). Sua arquitetura pode ser resumida da seguinte forma:

Onde **TC** e **SC** são respectivamente convoluções temporal e espacial, **BiMap**, **ReEig** e **LogEig** são as mesmas camadas SPD discutidas na SPDNet, **SCMPool** é apresentada na EE(G)-SPDNet e **SPDDSMBN** (*Domain-Specific Momentum Batch Normalization*) é uma nova camada SPD que transforma entradas específicas de domínio no espaço SPD em saídas invariantes ao domínio.

O bloco SPDDSMBN é uma extensão da *batch normalization* aplicada diretamente ao espaço de matrizes SPD, visando UDA em sinais de EEG. É composto por

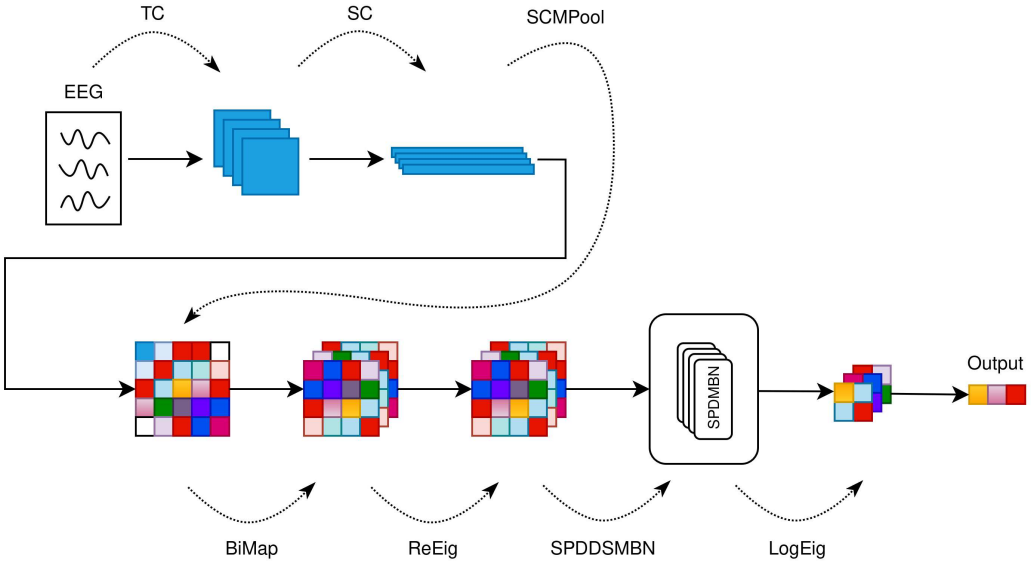

Figura 8 – Arquitetura da TSMNet.

dois subcomponentes principais: o SPDMBN (*Momentum Batch Normalization* específico para SPD) e o mecanismo DSBN (*Domain-Specific Batch Normalization*), trabalhando em conjunto.

- **SPDMBN**: O objetivo do SPDMBN é normalizar entradas SPD mantendo a estrutura geométrica do espaço:

  1. Calculo da média de Fréchet $\mathbf{G}_i$ e a variância ν_i dos dados em cada domínio i, atualizadas por momentum:

  $$\mathbf{G}_i^{(t)} = \alpha \mathbf{G}_i^{(t-1)} + (1-\alpha)\bar{\mathbf{Z}}_i^{(t)} \tag{2.38}$$

  $$\nu_i^{(t)} = \alpha \nu_i^{(t-1)} + (1-\alpha)\mathrm{Var}_F(\mathbf{Z}_i^{(t)}) \tag{2.39}$$

  onde $\bar{\mathbf{Z}}_i^{(t)}$ é o baricentro de Fréchet (tendendo à média geométrica), α é o fator de momentum, e Var_F é a variância no espaço SPD.

  2. Cada matriz $\mathbf{Z}_j$ do domínio i é centralizada e reescalada conforme:

  $$\tilde{\mathbf{Z}}_j = \exp\left(\gamma \log\left((\mathbf{G}_i)^{-\frac{1}{2}} \mathbf{Z}_j (\mathbf{G}_i)^{-\frac{1}{2}}\right)\right) \tag{2.40}$$

  onde log e exp são aplicados no domínio SPD e γ é um parâmetro de escala aprendido.

- **DSBN**: O mecanismo consiste em aplicar o SPDMBN de forma separada para cada domínio i, gerando estatísticas próprias $\mathbf{G}_i$ e ν_i. Essa estratégia permite capturar características estatísticas específicas de cada domínio durante o treinamento.

- **SPDDSMBN**: A camada SPDDSMBN é definida como:

$$\mathrm{SPDDSMBN}(\mathbf{Z}_j, i) = \mathrm{SPDMBN}_i(\mathbf{Z}_j, \mathbf{G}_i, \nu_i, \gamma, \alpha)\ \forall \mathbf{Z}_j \in B_k, i \in I_{B_k} \tag{2.41}$$

onde onde B_k é um lote contendo amostras de um mesmo domínio, e I_{B_k} indica quais domínios estão representados no lote.

2.6 Estimação de Matrizes de Covariância

Para trabalhar com RG, nossos dados precisam estar contidos no espaço das matrizes SPD. No entanto, como os dados de EEG são séries temporais, é necessário mapeá-los de alguma forma. A maneira mais simples de fazer isso é estimando a matriz de auto-covariância do sinal de EEG.

Para um vetor aleatório $\mathbf{X} \in \mathbb{R}^n$ com esperança matemática $\mu = \mathbb{E}[\mathbf{X}]$, a matriz de covariância verdadeira é dada por:

$$\mathbf{\Sigma} = \mathbb{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^\intercal] \tag{2.42}$$

Como não conhecemos a distribuição verdadeira, estimamos $\hat{\mathbf{\Sigma}}$ a partir de um conjunto finito de amostras. Dado um sinal de EEG de entrada $\mathbf{X} \in \mathbb{R}^{N \times L}$, sua matriz de auto-covariância amostral pode ser estimada por:

$$\hat{\mathbf{\Sigma}} = \left(\frac{1}{L-1}\right) \mathbf{X}\mathbf{X}^\intercal \tag{2.43}$$

onde N é o número de canais e L o número de amostras por canal.

Entretanto, antes de aplicar RG, é necessário garantir que essa matriz de covariância seja uma matriz SPD. Isso é válido se, e somente se, $\mathbf{X}$ possuir uma distribuição não degenerada (ou seja, todas as variáveis aleatórias em $\mathbf{X}$ precisam ser linearmente independentes), o que podemos assumir, uma vez que nossos dados de entrada são sinais neurais.

No contexto da análise de EEG, a matriz de covariância é estimada com base nos sinais dos eletrodos. Ao explorar a estrutura geométrica da variedade (*manifold*) de matrizes SPD, é possível capturar padrões e relações mais significativas nos dados de EEG, o que pode aprimorar o desempenho nas tarefas de classificação e extração de características.

Parte II

Atributos Temporais em Sinais de EEG

3 Informação Temporal em Sinais de EEG

As DRNs emergiram como uma abordagem inovadora e promissora, combinando os poderosos recursos de DL com as propriedades fundamentais da RG. Essa fusão as permite uma capacidade superior de classificação e generalização em tarefas de processamento de sinais de EEG.

Apesar de seu grande potencial, as DRNs se concentram predominantemente na identificação de padrões espaciais presentes nos dados, o que é uma característica fundamental de seu funcionamento. No entanto, essa abordagem deixa de explorar a riqueza das informações temporais que podem ser cruciais para uma compreensão mais completa dos sinais. Em tarefas como a análise de EEG, as variações temporais ao longo do tempo podem conter informações essenciais sobre a dinâmica neural, especialmente em tarefas de classificação de movimentos ou processos cognitivos.

Com base nesse entendimento, nosso objetivo é aprimorar os resultados das DRNs de forma a permitir a extração não apenas dos padrões espaciais, mas também dos padrões temporais presentes nos dados de entrada. Ao integrar essas duas dimensões de informação, espera-se que os modelos sejam capazes de capturar uma gama mais ampla de características dos dados, resultando em um desempenho significativamente melhorado. Essa modificação visa explorar o potencial completo dos dados, aproveitando tanto a variação espacial quanto a evolução temporal dos sinais para melhorar a precisão da classificação e a capacidade de generalização das redes.

3.1 Matrizes de Covariância com Atraso

No contexto da RG, uma possibilidade para a incorporação de informações temporais aos dados de entrada é mapear o sinal de EEG para um conjunto composto por suas matrizes de covariância amostrais defasadas no tempo (atrasadas). Basicamente, ao considerar um conjunto de passos no tempo, podemos calcular um conjunto de "Matrizes de Covariância com Atraso". Seja $\mathbf{X}_d \in \mathbb{R}^{N \times L}$ com uma quantidade d de atraso:

$$\mathbf{X}_d = \begin{bmatrix} x_{e_1}(t-d) & x_{e_1}(t-d+1) & \cdots & x_{e_1}(t-d+L) \\ x_{e_2}(t-d) & x_{e_2}(t-d+1) & \cdots & x_{e_2}(t-d+L) \\ \vdots & \vdots & \ddots & \vdots \\ x_{e_N}(t-d) & x_{e_N}(t-d+1) & \cdots & x_{e_N}(t-d+L) \end{bmatrix}_{N \times L} \tag{3.1}$$

onde e_n representa o n-ésimo canal, L é o número de amostras por canal, t o instante da amostra e $x_{e_n}(t)$ a amostra do canal n no tempo t. Então, adicionamos um atraso d aos dados de entrada.

Agora, uma matriz de covariância com atraso (para o atraso d) pode ser estimada como:

$$\hat{\mathbf{M}}_d = \left(\frac{1}{L-1}\right) \mathbf{X}_0 \mathbf{X}_d{}^\intercal \tag{3.2}$$

sendo $\hat{\mathbf{M}}_d \in \mathbb{R}^{N \times N}$. Considerando que nenhum dado fora da janela de EEG pode ser utilizado, apenas d amostras não são usadas para a estimativa de $\hat{\mathbf{M}}_d$.

Além disso, como $\mathbf{X}_0$ e $\mathbf{X}_d$ não são idênticas, para garantir que $\hat{\mathbf{M}}_d$ ainda seja uma matriz SPD, utiliza-se uma relação que assegura a simetria em relação à diagonal principal:

$$\mathbf{M}_d = \frac{\hat{\mathbf{M}}_d + \hat{\mathbf{M}}_d^\intercal}{2} \tag{3.3}$$

A seguir, para demonstrar que $\mathbf{M}_d$ é definida positiva, aproximando $\mathbf{M}_d \approx \hat{\mathbf{\Sigma}}$ (equação 2.43), é possível mostrar que, para qualquer vetor não nulo $\mathbf{z} \in \mathbb{R}^n$, $\mathbf{z}^\intercal \mathbf{M}_d \mathbf{z} > 0$:

$$\mathbf{z}^\intercal \mathbf{M}_d \mathbf{z} = \mathbf{z}^\intercal (\mathbb{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^\intercal]) \mathbf{z} \tag{3.4}$$

utilizando a linearidade da esperança matemática:

$$\mathbf{z}^\intercal \mathbf{M}_d \mathbf{z} = \mathbb{E}[\mathbf{z}^\intercal (\mathbf{X} - \mu)(\mathbf{X} - \mu)^\intercal \mathbf{z}] \tag{3.5}$$

como $\mathbf{z}^\intercal(\mathbf{X} - \mu)$ é um escalar, denotamos como $\mathbf{Y} = \mathbf{z}^\intercal(\mathbf{X} - \mu)$, assim obtemos:

$$\mathbf{z}^\intercal \mathbf{M}_d \mathbf{z} = \mathbb{E}[\mathbf{Y}^2] \tag{3.6}$$

Como a esperança de uma quantidade elevada ao quadrado é sempre não-negativa, concluímos que $\mathbf{z}^\intercal \mathbf{M}_d \mathbf{z} \geq 0$. Por fim, para que $\mathbf{z}^\intercal \mathbf{M}_d \mathbf{z} > 0$, é necessário que $\mathbf{M}_d$ seja de posto completo, o que novamente podemos assumir, dado que nossos dados de entrada são sinais neurais.

3.1.1 Matrizes de Covariância com Atraso Expandidas

Após a geração das matrizes de covariância com atraso, passamos à fase de treinamento com RG. No entanto, alguns métodos, como a SPDNet, recebem como entrada uma única matriz SPD por vez. Desta forma se faz necessário utilizar uma combinação das matrizes de covariância com atraso em uma única entrada. Para isso, propomos o uso de Matrizes de Covariância com Atraso Expandidas (MCAE).

As MCAE são matrizes em blocos compostas por p matrizes de covariância com atraso organizadas em uma configuração análoga a uma matriz de covariância:

$$\mathbf{A}_p = \begin{bmatrix} \mathbf{M}_0 & \mathbf{M}_1 & \cdots & \mathbf{M}_p \\ \mathbf{M}_1{}^\intercal & \mathbf{M}_0 & \cdots & \mathbf{M}_{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_p{}^\intercal & \mathbf{M}_{p-1}{}^\intercal & \cdots & \mathbf{M}_0 \end{bmatrix}_{(p+1)N \times (p+1)N} \tag{3.7}$$

Com o uso das MCAE, busca-se capturar possíveis correlações entre as diferentes quantidades de atraso representadas por cada matriz de sua composição.

3.1.2 Tensores com Atraso

A utilização de MCAEs, apesar de eficaz em diversas aplicações, apresenta limitações notáveis no que diz respeito à escalabilidade e ao tempo de processamento. Tais limitações tornam-se ainda mais evidentes em cenários que envolvem o uso de maiores quantidades de matrizes de covariância com atraso. Por exemplo, para uma entrada com 22 canais, a inclusão de apenas 5 matrizes resulta em uma matriz expandida com dimensões de 132×132, o que implica um aumento significativo na carga e tempo computacional. Mesmo utilizando um hardware moderadamente potente, o treinamento para cada usuário demandaria mais de 1h30, evidenciando o alto custo computacional envolvido.

Uma maneira mais intuitiva de agrupar múltiplas matrizes de covariância com atraso em uma única entrada seria organizá-las em um único tensor:

$$\mathbf{T}_p = (\mathbf{M}_0, \mathbf{M}_1, \ldots, \mathbf{M}_p) \tag{3.8}$$

onde $\mathbf{T}_p \in \mathbb{R}^{N \times N \times p}$. Neste formato, as informações de atraso são representadas como uma terceira dimensão do tensor, permitindo que o modelo capture as dependências temporais entre os diferentes atrasos.

No entanto, para que isso seja possível, o modelo precisa ser capaz de processar essa estrutura de dados tensorial, o que não é suportado diretamente pela arquitetura de diversos modelos. Com isso em mente, propomos uma modificação na SPDNet capaz de receber tensores como entrada, preservando a geometria do espaço de matrizes SPD ao longo dos diferentes atrasos temporais.

3.1.2.1 SPDNet-Par

A fim de viabilizar o processamento das matrizes de covariância com atraso na forma de um tensor, realizamos modificações na arquitetura padrão da SPDNet com o objetivo de paralelizar o processamento de cada matriz contida no tensor, ao mesmo tempo em que evitamos um crescimento excessivo no número de parâmetros ajustáveis. Essa abordagem, ilustrada na Fig. 9, busca otimizar a capacidade da rede de lidar com a dinâmica temporal, sem comprometer sua eficiência ou viabilidade prática.

A arquitetura padrão da SPDNet não foi projetada para processar estruturas de entrada na forma tensorial, uma vez que opera diretamente sobre matrizes SPD individuais. Para superar essa restrição, reestruturamos a rede de modo que cada uma das matrizes contidas no tensor é submetida às mesmas operações realizadas nos blocos SPD da rede, preservando assim as propriedades geométricas associadas ao espaço de variedades de matrizes SPD.

Após o processamento independente de cada matriz, os resultados intermediários são agregados — através de operações de concatenação, então, passam pelas camadas subsequentes da rede, que realizam o processamento unificado. Esse fluxo permite que

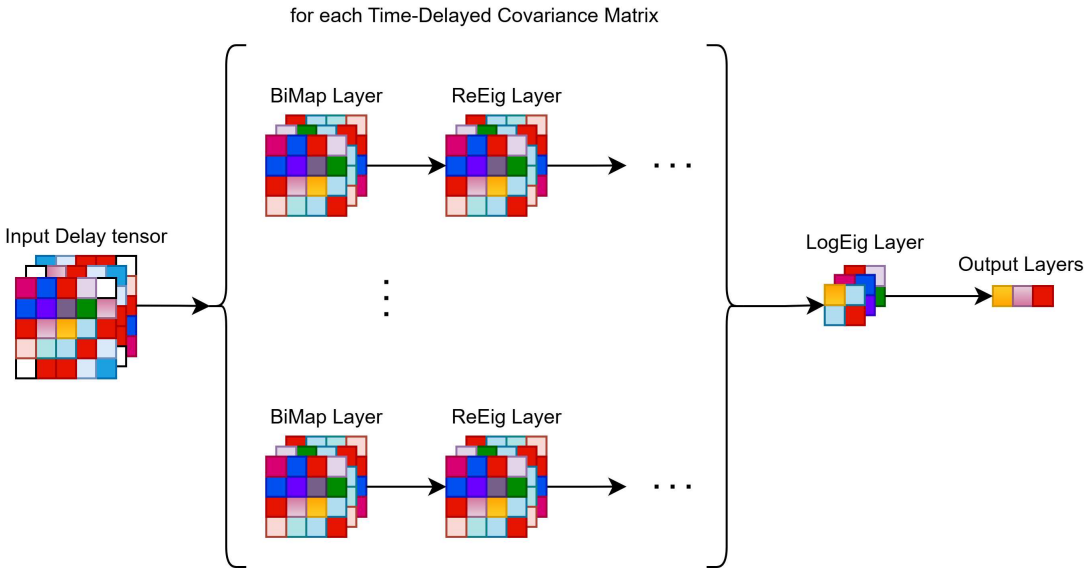

Figura 9 – Arquitetura da SPDNet-Par.

a rede combine de maneira eficiente as informações espaciais e temporais presentes nos sinais de EEG, capturadas através das matrizes de covariância com diferentes atrasos.

Dessa forma, a SPDNet-Par mantém a robustez das operações no espaço Riemanniano, ao mesmo tempo em que incorpora uma modelagem explícita da dinâmica temporal dos sinais.

4 Conjuntos de Dados

Para a validação e análise de desempenho das abordagens propostas, foram empregados dois conjuntos de dados amplamente reconhecidos na literatura e de acesso público: BCI Competition IV Dataset 2A (BRUNNER *et al.*, 2024) e Cho2017 (CHO *et al.*, 2017). Ambos se caracterizam por fornecer sinais de EEG de alta qualidade, associados a protocolos experimentais rigorosos, configurando-se como conjuntos de referência (*benchmarks*) robustos para pesquisas em BCI utilizando o paradigma de MI.

4.0.1 BCI Competition IV Dataset 2A

O conjunto de dados BCI Competition IV Dataset 2A (BRUNNER *et al.*, 2024) compreende registros de EEG de 9 participantes, cada um executando quatro tarefas distintas de MI: movimento da mão esquerda, mão direita, ambos os pés e língua. As aquisições foram realizadas utilizando 22 canais de EEG, com taxa de amostragem de 250 Hz. Durante a coleta dos sinais, foram aplicados um filtro passa-faixa de 0,5 a 100 Hz, visando a retenção das componentes relevantes do sinal, e um filtro notch centrado em 50 Hz, com o objetivo de mitigar interferências provenientes da rede elétrica. Cada sujeito participou de duas sessões distintas, realizadas em dias diferentes. Cada sessão foi organizada em 6 blocos (denominados *runs*), sendo que cada bloco contém 48 capturas (12 por classe), totalizando 288 capturas por sessão e 576 capturas por participante.

4.0.2 Cho2017

O conjunto Cho2017 (CHO *et al.*, 2017), se diferencia principalmente pela sua escala, contendo registros de EEG de 52 sujeitos, o que proporciona uma diversidade substancial e maior variabilidade inter-sujeito, refletindo desafios mais próximos de cenários do mundo real. Durante os experimentos, os participantes foram instruídos a realizar duas tarefas de MI, correspondentes à imaginação dos movimentos da mão esquerda e da mão direita. A coleta dos sinais foi realizada por meio de 64 canais de EEG, com taxa de amostragem de 512 Hz. Cada sujeito participou de um número variável de capturas por classe, totalizando entre 100 a 120 capturas por classe, dependendo da sessão de cada participante. Assim, o número total de capturas por sujeito varia entre 200 e 240 capturas.

Conjunto	**Indivíduos**	**Canais**	**Taxa de Amostragem (Hz)**	**Classes**	**Capturas por indivíduo**
BCI Competition IV Dataset 2A	9	22	250	4	576
Cho2017	52	64	512	2	200-240

Tabela 1 – Detalhes dos conjuntos de dados utilizados

4.1 Pré-processamento dos dados

Inicialmente, para garantir a qualidade dos dados e a efetividade dos testes, foi realizada uma filtragem de frequências no intervalo de 4 Hz a 38 Hz. Essa faixa foi escolhida com base nas características do sinal de EEG, buscando remover frequências indesejadas e ruídos que poderiam interferir na análise dos dados. A filtragem foi fundamental para isolar as bandas de interesse, como a banda de frequência alfa, que está frequentemente associada a processos cognitivos relevantes em experimentos de MI.

Além disso, uma modificação de escala foi aplicada nos dados. Originalmente, os dados estavam em volts (V), mas devido à baixa amplitude dos sinais registrados, uma conversão para microvolts (μV) foi realizada utilizando um fator de escala de 10^6. Essa transformação tem como objetivo melhorar a visualização e análise dos sinais, uma vez que a unidade de μV proporciona uma resolução mais adequada para sinais de EEG, que são geralmente de pequena amplitude. Essa mudança torna os dados mais consistentes com as unidades comuns em estudos de EEG e facilita o processamento subsequente.

Após o pré-processamento, foi realizada uma etapa de janelamento nos dados, no qual o intervalo de tempo dos sinais foi ajustado para que fosse considerada janela de captura de IM em sua totalidade (de $3s$ a $6s$ para BCI Competition IV Dataset 2A e de $0s$ a $3s$ para Cho2017). O janelamento tem como objetivo focar nas janelas de interesse, descartando dados fora desse intervalo e, assim, otimizando o processamento e a análise dos sinais. Essa técnica é importante para garantir que apenas as partes relevantes do sinal sejam analisadas, melhorando a eficácia da modelagem subsequente. Este processo está ilustrado na Fig. 10.

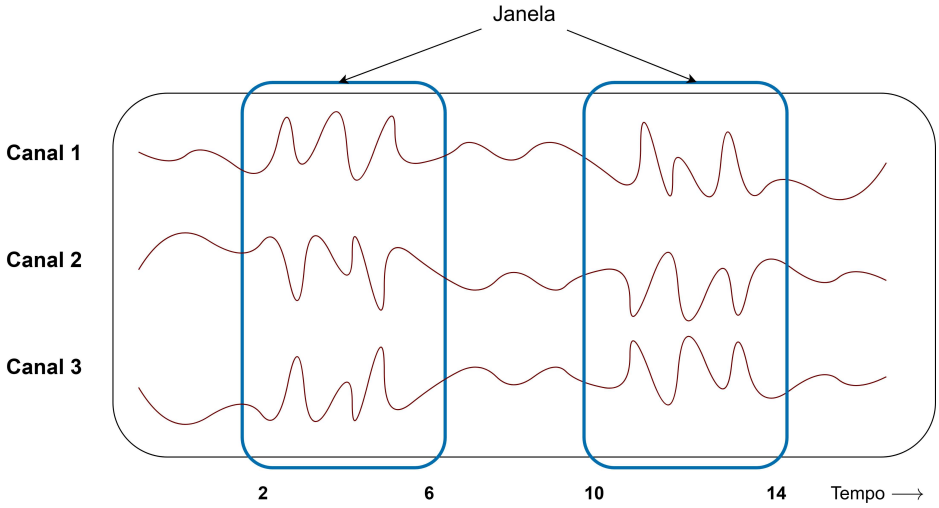

Figura 10 – Janelamentos de um sinal de EEG.

5 Resultados

A fim de avaliar o impacto da introdução das matrizes de covariância com atraso, conduzimos experimentos utilizando cinco modelos de classificação amplamente reconhecidos na classificação de EEG baseados em MI (além do nosso próprio modelos proposto). Esses modelos abrangem diferentes paradigmas de processamento, desde classificadores clássicas baseadas em RG até arquiteturas de DL capazes de explorar características espaciais e temporais dos sinais de EEG.

Cada modelo foi selecionado por representar uma linha distinta de desenvolvimento na área, permitindo uma análise comparativa robusta sobre como a informação temporal influencia o desempenho em diferentes estratégias de modelagem. Além disso, para garantir a reprodutibilidade dos resultados, todos os hiperparâmetros utilizados durante os experimentos estão detalhados na Tabela 2.

Hiperparâmetros	SPDNet	SPDNet-Par	EEGNet	EE(G)-SPDNet	TSMNet
Taxa de Aprendizado (*Learning Rate*)	0.01	0.001	0.01	0.007	0.009
Tamanho de Batelada (*Batch Size*)	64	32	32	32	16
Èpocas (*Epochs*)	500	500	75	100	1000

Tabela 2 – Hiperparâmetros

Para avaliar a eficácia dos métodos propostos, comparamos a acurácia de classificação dos cinco modelos utilizados. Utilizamos o cenário sem atraso ($p = 0$) como linha de base para nossas comparações. Ao analisar com os dois conjuntos de dados apresentados anteriormente, é possível observar que a adição da informação temporal com atrasos $p = 1$ e $p = 10$ superou consistentemente os resultados dos modelos de base, como apresentado na Tabela 3.

Dataset	MDM	SPDNet	SPDNet-Par	EEGNet	EE(G)-SPDNet	TSMNet
BNCI2014001 $p = 0$	70.54 ± 16.18	71.94 ± 15.88	68.50 ± 16.20	$\mathbf{68.98 \pm 13.79}$	71.47 ± 16.97	72.37 ± 16.73
BNCI2014001 $p = 1$	$\mathbf{72.90 \pm 14.55}$	$\mathbf{73.98 \pm 13.95}$	$\mathbf{70.40 \pm 15.50}$	64.43 ± 05.47	$\mathbf{72.58 \pm 16.71}$	70.96 ± 14.24
BNCI2014001 $p = 10$	71.12 ± 15.90	73.30 ± 14.45	64.52 ± 13.33	66.04 ± 11.36	72.14 ± 16.82	$\mathbf{74.20 \pm 15.83}$
Cho2017 $p = 0$	53.38 ± 12.74	61.01 ± 12.26	58.78 ± 11.69	$\mathbf{52.48 \pm 07.04}$	60.94 ± 11.86	62.21 ± 12.38
Cho2017 $p = 1$	53.38 ± 12.86	61.62 ± 12.58	$\mathbf{59.09 \pm 12.13}$	51.28 ± 05.82	61.29 ± 11.85	62.15 ± 12.50
Cho2017 $p = 10$	$\mathbf{54.42 \pm 13.05}$	$\mathbf{62.28 \pm 12.78}$	56.52 ± 08.15	51.57 ± 05.47	$\mathbf{61.59 \pm 12.46}$	$\mathbf{63.79 \pm 12.75}$

Tabela 3 – Resultados

Para que fosse possível incorporar informações com atraso temporal nos dados de entrada da arquitetura EEGNet, adotamos uma estratégia de concatenação das representações temporais ao longo da dimensão dos canais. Nessa abordagem, para cada atraso pré-definido, é gerada uma cópia do sinal de EEG original, deslocada no tempo, e suas respectivas características são empilhadas com os canais originais. Como resultado, a rede recebe uma entrada expandida, onde cada canal original é complementado com suas versões defasadas no tempo. Contudo, a EEGNet ainda foi o único modelo cujo melhor desempenho foi obtido sem o uso de atraso ($p = 0$). Esse resultado era esperado, uma vez que a EEGNet é o único modelo do experimento que não é baseado em RG. Diferente dos modelos baseados em RG, que se beneficiam de representações baseadas em matrizes de covariância enriquecidas com informações temporais, a EEGNet aprende diretamente os padrões espaço-temporais por meio de suas operações convolucionais.

Os resultados também nos mostram que o parâmetro de atraso ideal varia de acordo com as características de cada conjunto. No caso do BCI Competition IV Dataset 2A, o melhor desempenho foi obtido com um atraso de 1 ($p = 1$), indicando que este conjunto de dados se beneficia principalmente da captura de dependências temporais de curto prazo. Por outro lado, no conjunto Cho2017, a melhor configuração foi aquela com atraso de 10 ($p = 10$), sugerindo que este conjunto contém dinâmicas temporais mais relevantes a longo prazo.

Um ponto relevante a ser destacado é o desempenho da nossa solução proposta, a SPDNet-Par, apesar de ter apresentado resultados consistentes e competitivos em ambos os conjuntos, ela não conseguiu superar a arquitetura original da SPDNet. Esse resultado sugere que, embora o processamento paralelo das informações temporais tenha potencial, nossa estratégia ainda não captura as características discriminativas de forma tão eficiente quanto as transformações hierárquicas aplicadas na arquitetura padrão da SPDNet.

Além da análise geral, também realizamos uma avaliação detalhada por indivíduo para os modelos MDM, SPDNet e SPDNet-Par, uma vez que esses representam os principais enfoques desta pesquisa. Essa análise individual visa identificar variações de desempenho entre os sujeitos e compreender melhor como cada modelo responde às diferenças interindividuais nas tarefas de MI com EEG.

Os gráficos de barras conjuntos ilustrados nas Fig. 11 e Fig. 12 foram construídos para facilitar a visualização comparativa do desempenho por indivíduo dos modelos. Nestes, cada indivíduo da base de dados é representado por um agrupamento de três barras, correspondentes às acurácias médias obtidas com os atrasos $p = 0$, $p = 1$ e $p = 10$, onde a barra do atraso com a maior acurácia é representada com a cor mais opaca (ambas as barras em caso de empate). Essa visualização permite uma análise direta do impacto da informação temporal adicional no desempenho por sujeito. Um gráfico foi gerado para cada conjunto de dados, destacando as variações individuais de acurácia e facilitando a identificação de padrões específicos de resposta dos sujeitos frente às diferentes estratégias e modelos avaliados.

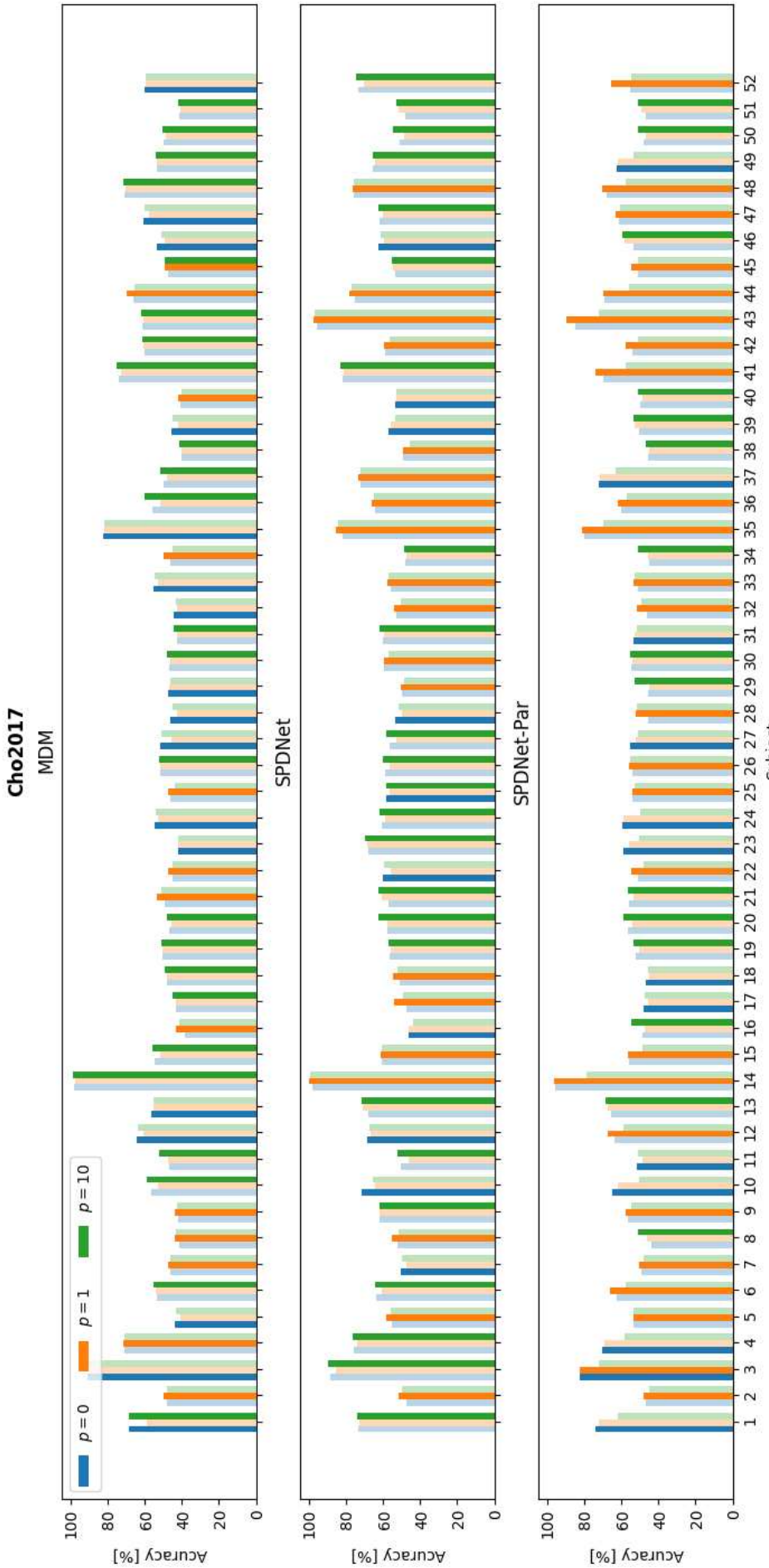

Figura 11 – Resultados por indivíduo para o conjunto Cho2017.

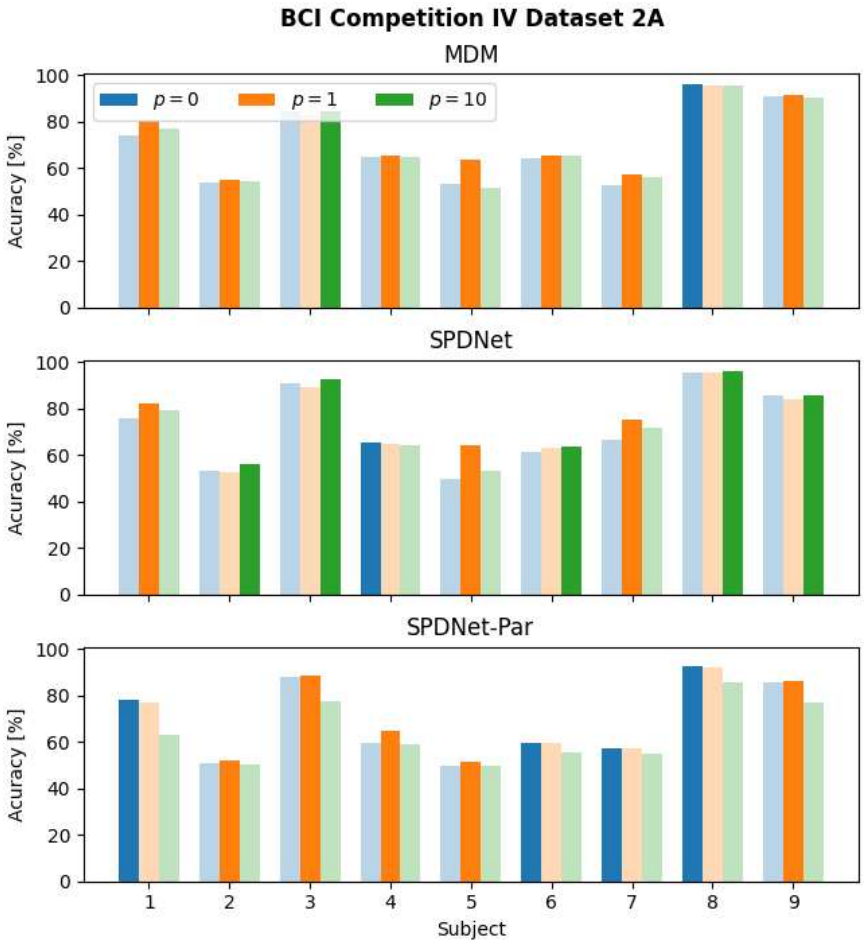

Figura 12 – Resultados por indivíduo para o BCI Competition IV Dataset 2A.

A análise dos gráficos reforça como o uso da informação temporal, para grande maioria dos indivíduos, melhora a classificação. Além disso fica mais claro como cada conjunto de dados se beneficia mais de informações de curto ou longo prazo ($p = 10$ para o Cho2017 e $p = 1$ para BCI Competition IV Dataset 2A). Também fica mais claro como a nossa proposta da SPDNet-Par captura padrões diferentes comparada com a arquitetura padrão da SPDNet, mesmo que apresentando um desempenho inferior.

Além dos gráficos de acurácia média por indivíduo, também foram elaborados os gráficos de barras apresentados nas Fig. 13 e Fig. 14 que contabilizam, para cada modelo, o número de indivíduos cuja maior acurácia foi obtida em cada uma das condições de atraso. Para isso, os sujeitos foram agrupados conforme o valor de p que proporcionou seu melhor desempenho individual, permitindo avaliar de forma agregada qual configuração temporal foi mais vantajosa em cada modelo. Essa análise contribui para uma compreensão mais ampla do comportamento dos modelos em relação à variação temporal do sinal.

Esses gráficos corroboram que os modelos mais robustos, nesse caso a SPDNet, se beneficiam das informações temporais quaisquer que sejam elas, da mesma forma que também demonstra como nossa proposta não foi capaz de capturar tão bem as informações de longo prazo como a arquitetura padrão da SPDNet.

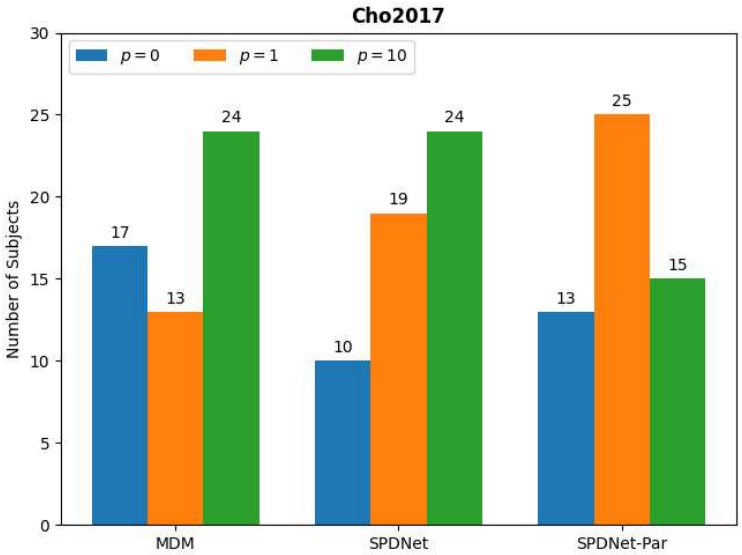

Figura 13 – Total de indivíduos por atraso para o conjunto Cho2017.

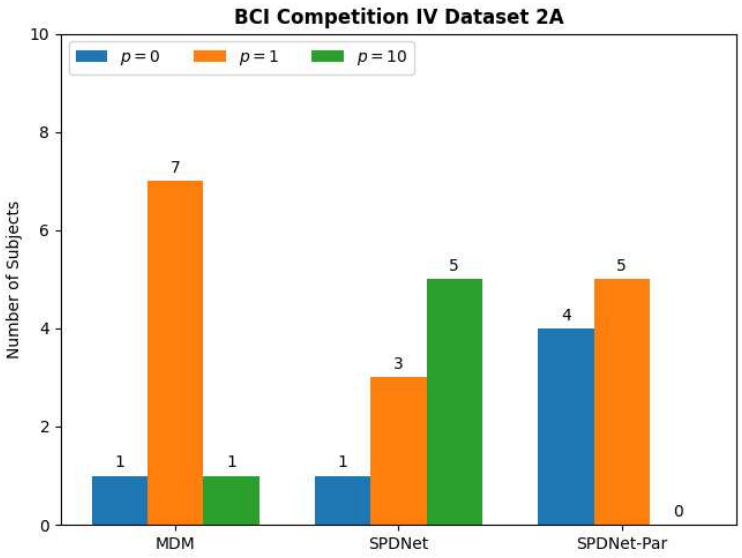

Figura 14 – Total de indivíduos por atraso para o BCI Competition IV Dataset 2A.

De forma geral, observamos uma melhora média de aproximadamente 2 pontos percentuais na acurácia dos sujeitos para as soluções baseadas em RG. Isso demonstra que a captura de dependências temporais por meio das matrizes de covariância com atraso melhora significativamente a capacidade dos modelos em aprender características discriminativas, confirmando, assim, a importância da modelagem espaço-temporal em tarefas de classificação de EEG.

6 Conclusões

Neste projeto, investigamos a interseção entre BCIs, DL e RG, destacando como a integração dessas tecnologias pode transformar significativamente o desempenho e a robustez das BCIs. Essa abordagem multidisciplinar é essencial para enfrentar os desafios inerentes à análise de sinais cerebrais, promovendo avanços na interpretação de dados complexos e no desenvolvimento de sistemas mais eficientes e adaptativos.

O uso do EEG como ferramenta não invasiva para captar a atividade cerebral é amplamente reconhecido como um método eficaz e acessível. No entanto, os sinais de EEG apresentam desafios significativos devido à sua alta variabilidade entre sujeitos, ruído inerente e natureza não estacionária. Essas características tornam difícil capturar padrões consistentes e demandam soluções computacionais avançadas que possam lidar com a complexidade intrínseca dos dados. Nesse cenário, as técnicas de DL surgem como uma abordagem poderosa, oferecendo a capacidade de generalizar e identificar padrões ocultos em dados complexos e de alta dimensionalidade. Contudo, essas técnicas necessitam de adaptações específicas para explorar a estrutura única dos sinais de EEG, que possuem propriedades temporais e espaciais interdependentes.

A RG, por sua vez, oferece um arcabouço matemático sólido que permite modelar os sinais de EEG em espaços de covariância. Ao representar os dados como matrizes SPD, a RG preserva propriedades geométricas essenciais, garantindo que a análise respeite as características inerentes da estrutura dos dados. Essa abordagem não apenas melhora a interpretação dos sinais, mas também oferece um meio mais eficaz de lidar com a variabilidade dos mesmos, promovendo maior estabilidade nos modelos de aprendizado.

Nesse contexto, a introdução das DRNs representa um marco significativo ao unir o poder da DL com os princípios da RG. Essa integração permite que os modelos extraiam características mais representativas e informativas dos sinais de EEG, resultando em uma análise mais robusta e precisa. Entretanto, apesar de sua eficácia na modelagem de padrões espaciais, as DRNs tradicionalmente desconsideram a dimensão temporal, limitando-se ao uso de uma única matriz de covariância por amostra.

Para contornar essa limitação, foram propostas estratégias que exploram a informação temporal de forma mais abrangente. As matrizes de covariância com atraso foram integradas por meio de suas versões expandidas, as MCAEs, estruturadas em blocos simétricos que preservam as relações entre diferentes janelas temporais. Também foi adotada uma representação em formato tensorial, reunindo múltiplas matrizes de covariância com atraso em uma estrutura tridimensional. Para processar tais tensores, foi proposta uma modificação da arquitetura SPDNet, denominada SPDNet-Par, que processa cada matriz de forma paralela e agrega suas saídas antes das camadas finais. Essa abordagem mostrou-se tecnicamente viável e amplia o potencial das DRNs ao permitir a modelagem

conjunta de informações espaciais e temporais.

Os resultados experimentais indicam que a introdução de informação temporal melhora o desempenho de modelos de classificação baseados em RG. Verificou-se ganho consistente de acurácia média com o uso de atrasos apropriados, cujos valores ótimos variaram de acordo com as características dos conjuntos de dados utilizados. A arquitetura SPDNet-Par, apesar de não superar a versão original da SPDNet, apresentou desempenho competitivo, evidenciando o potencial do processamento paralelo de estruturas temporais.

De forma geral, as soluções propostas contribuem para o desenvolvimento de BCIs mais robustas e eficazes, com potencial aplicação em contextos práticos e em pesquisas futuras focadas na personalização e adaptação de modelos.

7 Perspectivas Futuras

Este trabalho reforça o potencial dos métodos baseados em RG aplicados à análise de sinais de EEG, especialmente quando combinados com estratégias que incorporam informações temporais. Embora os resultados obtidos sejam promissores, diversas direções podem ser exploradas em trabalhos futuros para aprimorar ainda mais o desempenho e a aplicabilidade dos modelos desenvolvidos.

Um dos principais avanços deste estudo foi a proposta de mecanismos para inserir informações temporais nos modelos RG, como as matrizes de covariância com atraso, suas formas expandidas, as MCAEs, e os tensores com atraso. No entanto, a modelagem da dinâmica temporal ainda pode ser aprofundada. Futuras pesquisas podem investigar arquiteturas que combinem explicitamente o espaço Riemanniano com modelos sequenciais, como Redes Recorrentes ou Redes Transformers adaptados ao espaço de matrizes SPD, criando modelos que considerem tanto dependências espaciais quanto temporais de forma mais robusta.

Outra limitação observada foi que, apesar da proposta SPDNet-Par apresentar melhorias em alguns cenários, ela não conseguiu superar consistentemente sua versão convencional nos experimentos avaliados. Isso sugere que há espaço para otimizações na forma como as informações temporais são incorporadas. Investigações futuras podem explorar diferentes estratégias de fusão temporal, como operações mais sofisticadas de atenção ou mecanismos de aprendizado hierárquico no espaço Riemanniano.

Além disso, este trabalho demonstrou que a escolha do parâmetro de atraso ideal varia significativamente entre os conjuntos de dados. No BCI Competition IV Dataset 2A, atrasos curtos (atraso = 1) proporcionaram melhor desempenho, enquanto no conjunto Cho2017, atrasos longos (atraso = 10) foram mais adequados. Este fenômeno abre espaço para pesquisas focadas em métodos automáticos de seleção ou otimização dos parâmetros de atraso, possivelmente adaptativos em tempo real, levando em consideração as características estatísticas específicas de cada sujeito ou tarefa.

Em resumo, embora este trabalho avance na integração de informações temporais em modelos baseados em RG para EEG, há um amplo espaço para pesquisa futura, tanto na melhoria das arquiteturas quanto na aplicação em cenários práticos mais desafiadores.

Referências

ABIRI, R.; BORHANI, S.; SELLERS, E. W.; JIANG, Y.; ZHAO, X. A comprehensive review of EEG-based brain-computer interface paradigms. *Journal of neural engineering*, v. 16, n. 1, p. 011001, 2019.

BARACHANT, A.; BONNET, S.; CONGEDO, M.; JUTTEN, C. Multiclass Brain–Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering*, v. 59, n. 4, p. 920–928, 2012.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. [S.l.]: Springer, 2006.

BRUNNER, C.; LEEB, R.; MüLLER-PUTZ, G. *BCI Competition 2008–Graz data set A*. [S.l.]: IEEE Dataport, 2024.

CHAPMAN, R.; BRAGDON, H. Evoked Responses to Numerical and Non-Numerical Visual Stimuli while Problem Solving. *Nature*, v. 4950, n. 203, p. 1155–1157, 1964.

CHO, H.; AHN, M.; AHN, S.; KWON, M.; JUN, S. C. *EEG datasets for motor imagery brain-computer interface*. [S.l.]: Gigascience, 2017.

CONGEDO, M.; BARACHANT, A.; BHATIA, R. Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, v. 4, p. 1–20, 2017.

COYLE, S. M.; WARD, T. E.; MARKHAM, C. M. Brain-computer interface using a simplified functional near-infrared spectroscopy system. *Journal of neural engineering*, v. 4, n. 3, p. 219–226, 2007.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.

HERRMANN, C. Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, v. 137, n. 3-4, p. 346–353, 2001.

HIRSCH, L. J.; BRENNER, R. P. *Atlas of EGG in Critical Care*. [S.l.]: John Wiley Sons, Ltd., 2010.

HOCHBERG, L. R. *et al.* Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, v. 442, n. 7099, p. 164–171, 2006.

HUANG, Z.; GOOL, L. V. *A Riemannian Network for SPD Matrix Learning*. 2016.

KOBLER, R. J.; HIRAYAMA, J.-i.; ZHAO, Q.; KAWANABE, M. SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2022. (NIPS '22).

LAWHERN, V. J.; SOLON, A. J.; WAYTOWICH, N. R.; GORDON, S. M.; HUNG, C. P.; LANCE, B. J. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, v. 15, n. 5, p. 056013, 2018.

LEITE, S. N. d. C. *Contribuições ao Desenvolvimento de Interfaces Cérebro-Computador Baseadas em Potenciais Evocados Visualmente em Regime Estacionário*. Tese (PhD thesis) — University of Campinas, 2016.

LEUTHARDT, E. C.; SCHALK, G.; WOLPAW, J. R.; OJEMANN, J. G.; MORAN, D. W. A brain-computer interface using electrocorticographic signals in humans. *Journal of neural engineering*, v. 1, n. 2, p. 63–71, 2004.

MANE, R.; CHOUHAN, T.; GUAN, C. BCI for stroke rehabilitation: motor and beyond. *Journal of neural engineering*, v. 17, n. 4, p. 041001, 2020.

MELLINGER, J. *et al.* An MEG-Based Brain–Computer Interface (BCI). *NeuroImage*, v. 36, n. 3, p. 581–593, 2007.

MOAKHER, M. A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices. *SIAM J. Matrix Analysis Applications*, v. 26, p. 735–747, 01 2005.

MULLER-PUTZ, G.; SCHERER, R.; NEUPER, C.; PFURTSCHELLER, G. Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, v. 14, n. 1, p. 30–37, 2006.

MULLER-PUTZ, G. R.; PFURTSCHELLER, G. Control of an Electrical Prosthesis With an SSVEP-Based BCI. *IEEE Transactions on Biomedical Engineering*, v. 55, n. 1, p. 361–364, 2008.

MUMTAZ, W.; RASHEED, S.; IRFAN, A. Review of challenges associated with the EEG artifact removal methods. *Biomedical Signal Processing and Control*, v. 68, p. 102741, 2021.

NAM, C.; NIJHOLT, A.; LOTTE, F. *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. [S.l.]: CRC Press, 2018.

NICOLAS-ALONSO, L. F.; GOMEZ-GIL, J. Brain computer interfaces, a review. Sensors. *Sensors*, v. 12, n. 2, p. 1211–1279, 2012.

NICOLELIS, M. A. L. Actions from thoughts. *Nature*, v. 409, n. 6818, p. 403–407, 2001.

NIJBOER, F.; FURDEA, A.; GUNST, I.; MELLINGER, J.; MCFARLAND, D. J.; BIRBAUMER, N.; KüBLER, A. An auditory brain-computer interface (BCI). *Journal of Neuroscience Methods*, v. 167, n. 1, p. 43–50, 2008.

PALANIAPPAN, R. Identifying Individuality Using Mental Task Based Brain Computer Interface. In: *2005 3rd International Conference on Intelligent Sensing and Information Processing*. [S.l.: s.n.], 2005. p. 238–242.

PINEDA, J. A.; ALLISON, B. Z.; VANKOV, A. The effects of self-movement, observation, and imagination on mu rhythms and readiness potentials (RP's): toward a brain-computer interface (BCI). *IEEE Transactions on Rehabilitation Engineering*, v. 8, n. 2, p. 219–222, 2000.

RODRIGUES, P. L. C. *Exploring invariances of multivariate time series via Riemannian geometry : validation on EEG data*. Tese (Doutorado) — Université Grenoble Alpes, 2019. Disponível em: <https://theses.hal.science/tel-02905408>.

RODRIGUES, P. L. C.; JUTTEN, C.; CONGEDO, M. Riemannian Procrustes Analysis: Transfer Learning for Brain–Computer Interfaces. *IEEE Transactions on Biomedical Engineering*, v. 66, n. 8, p. 2390–2401, 2019.

SANTAMARíA-VáZQUEZ, E.; MARTíNEZ-CAGIGAL, V.; VAQUERIZO-VILLAR, F.; HORNERO, R. EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-Based Brain-Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, v. 28, n. 12, p. 2773–2782, 2020.

SCHIRRMEISTER, R. T.; SPRINGENBERG, J. T.; FIEDERER, L. D. J.; GLASSTETTER, M.; EGGENSPERGER, K.; TANGERMANN, M.; HUTTER, F.; BURGARD, W.; BALL, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, v. 38, n. 11, p. 5391–5420, 2017.

SLENES, G. F.; BELTRAMINI, G. C.; LIMA, F. O.; LI, L. M.; CASTELLANO, G. The use of fMRI for the evaluation of the effect of training in motor imagery BCI users. In: *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. [S.l.: s.n.], 2013. p. 686–690.

WILSON, D.; SCHIRRMEISTER, R. T.; GEMEIN, L. A. W.; BALL, T. *Deep Riemannian Networks for EEG Decoding*. 2023.

WOLPAW, J. R.; BIRBAUMER, N.; MCFARLAND, D. J.; PFURTSCHELLER, G.; VAUGHAN, T. M. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, v. 113, n. 6, p. 767–791, 2002.

XIE, X.; XIAOKUN, Z.; TANG, R.; HOU, Y.; QI, F. Multiple graph fusion based on Riemannian geometry for motor imagery classification. *Applied Intelligence*, v. 52, 2022.