

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Levi Moreira de Albuquerque

**Uma abordagem embarcada para a detecção de ataques epiléticos
utilizando aprendizado profundo**

Fortaleza – CE

30 de agosto de 2024

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

M835a Moreira de Albuquerque, Levi`.
 Uma abordagem embarcada para a detecção de ataques epiléticos utilizando aprendizado profundo /
Levi` Moreira de Albuquerque. - 2022.
 62 f. : il.

 Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus
Fortaleza, 2022.

 Orientação: Prof. Dr. Elias Teodoro da Silva Jr..

 1. Epilepsia. 2. Redes Neurais Convolucionais. 3. Sistemas Embarcados. I. Título.

CDD 004

Levi Moreira de Albuquerque

Uma abordagem embarcada para a detecção de ataques epilépticos utilizando aprendizado profundo

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Sistemas embarcados

Orientador: Dr. Elias Teodoro da Silva Jr.

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

30 de agosto de 2024



Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)
Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Levi Moreira de Albuquerque

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, sendo aprovada pela Coordenação do Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará e pela banca examinadora:

Orientador: Dr. Elias Teodoro da Silva Jr.

Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

Dr. Antonio Mauro Barbosa de Oliveira

Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

Dr. Edison Pignaton de Freitas

Universidade Federal do Rio Grande do Sul (UFRGS)

Fortaleza – CE
30 de agosto de 2024

Agradecimentos

Em primeiro lugar a Deus por todas as oportunidades que me foram concedidas até aqui. As diversas portas profissionais e acadêmicas que me foram abertas só poderiam ter um fator em comum que as explicassem: Ele.

Em segundo lugar a minha família, representada primeiramente pela minha esposa Jéssica, pelos muitos momentos de apoio e compreensão durante a execução deste projeto. E também aos meus pais, que sempre me ensinaram que a educação é a maior fonte de transformação que se pode ter, me ensinaram a valorizá-la e a buscá-la constantemente.

Gostaria de agradecer ao professor Elias Teodoro que desde a minha passagem pela graduação demonstrou paciência em me acolher na área de sistemas embarcados, naquela época como monitor. Ele também colaborou com o meu trabalho de término de curso e por fim aceitou ser meu orientador e colaborar diretamente para a realização deste trabalho nos últimos três anos.

Por fim gostaria de agradecer a todos que colaboraram direta e indiretamente para que eu pudesse chegar aqui, afinal, cada experiência dentro e fora da academia colaborou para que eu pudesse atingir este objetivo. Aos diversos colegas com quem trabalhei durante a graduação e o mestrado, bem como nas minhas experiências na indústria, que foram fundamentais para o meu amadurecimento como pesquisador.

"Differences of habit and language are nothing at all if our aims are identical and our hearts are open."
(Albus Dumbledore)

Resumo

A epilepsia é uma doença crônica que afeta mais de 40 milhões de pessoas no mundo inteiro. A maioria delas consegue controlar os sintomas que a doença apresenta, contudo 30% desses indivíduos, mesmo se submetendo ao uso de medicamentos, sofre com ataques frequentes. Por isso faz-se necessária a criação de maneiras mais efetivas de detecção da ocorrência dessas crises. Algoritmos de aprendizado de máquina, por exemplo, podem ser aplicados a sinais de Eletroencefalograma (EEG) para realizar tal tarefa. Alguns desses algoritmos, como Máquinas de Vetores de Suporte (SVM), K-vizinhos próximos (KNN) e Transformada Wavelet Discreta (DWT) são aplicados a atributos retirados desses sinais, ou ainda a versões deles no domínio do tempo. As técnicas de *deep learning* também possuem exemplos de utilização nessa tarefa, com alguns resultados atingindo taxas de acerto próximas dos 93%. Entretanto, a maioria dessas soluções considera apenas computadores de propósito geral. Além disso, muitas técnicas desenvolvidas buscam resolver o problema utilizando modelos intra-paciente ou inter-paciente. Este trabalho apresenta uma descrição da utilização de redes neurais convolucionais em sistemas embarcados para a realização dessa tarefa com foco na abordagem inter-paciente. O trabalho demonstra técnicas para a geração e pré-processamento de dados de eletroencefalograma que podem ser utilizados como entrada de uma rede neural convolucional. A arquitetura e desempenho da rede são avaliados e os resultados são compatíveis com o estado da arte. A rede apresentada possui apenas 5 camadas convolucionais, o que representa um tamanho razoável para a utilização em um sistema embarcado vestível.

Palavras-chave: Epilepsia, Redes Neurais Convolucionais, Sistemas Embarcados

Abstract

Epilepsy is a chronic disease that affects over 40 million people around the world. The majority of them are able to control the symptoms with medicine, but 30% of them still suffer from unpredictable attacks. For this reason, there is a real need for creating an effective way of detecting the occurrences of such attacks. Machine Learning algorithms, for instance, can be applied to electroencephalogram signals to help detect seizures. Some algorithms, such as Support Vector Machines (SVM), K-Nearest neighbours (KNN) and Discrete Wavelet Transform (DWT), are applied to attributes extracted to these signals or even time domain data. Deep learning approaches are already used for this task and produce interesting results with an average accuracy of 93%, however, there is not a lot of exploration in applying these algorithms in an embedded systems setting. Moreover, most approaches in the literature concentrate on solving the task using patient-specific models. This work presents a method that uses convolutional neural networks under an embedded systems perspective. Furthermore, a detailed explanation of how to generate and pre-process data from EEG signals is presented with the aim of applying them to a cross-patient task. The proposed architecture performance is evaluated and the results are compatible with state-of-the art approaches. The architecture has 5 convolutional layers which can be reasonably applied to an embedded platform.

Keywords: Epilepsy, Convolutional Neural Networks, Embedded Systems.

Lista de ilustrações

Figura 1 – Ilustração dos tipos de ondas encontradas no EEG	35
Figura 2 – Placa de desenvolvimento utilizada neste trabalho com o processador STM32F746	39
Figura 3 – Amostragem em Épocas com sobreposição de 50%	45
Figura 4 – Sistema internacional 10-20 para montagem de EEG	51

Lista de tabelas

Tabela 1 – Pares de eletrodos usados nesse estudo.	34
Tabela 2 – Arquitetura da rede SeizNet original	37
Tabela 3 – Arquiteturas analisadas neste trabalho	42
Tabela 4 – Arquitetura da rede SeizNet Modificada	43
Tabela 5 – Acurácia média do classificador com dados filtrados - SeizNet Modificada	44
Tabela 6 – Desempenho da amostragem em épocas na SeizNet Modificada usando diferentes taxas de sobreposição	45
Tabela 7 – Resultado dos Modelos Intrapacientes na SeizNet Modificada	47
Tabela 8 – Acurácia média na rede SeizNet Modificada para múltiplos canais . . .	47
Tabela 9 – Uso da RAM pela plataforma STM32F746 para a SeizNet Modificada	48
Tabela 10 – Compressão de Memória Flash em diferentes níveis - SeizNet Modificada	49
Tabela 11 – Desempenho da rede SeizNet Modificada na tarefa inter-paciente . . .	50
Tabela 12 – Desempenho de classificação para cada passo do método na tarefa inter-paciente	50
Tabela 13 – Eventos do protótipo	51
Tabela 14 – Consumo médio e valor de microprocessadores comerciais	52
Tabela 15 – Tempo de execução da rede neural na plataforma embarcada usando a SeizNet Modificada	53

Lista de abreviaturas e siglas

EEG	Electroencephalography
KNN	K-nearest neighbors
DWT	Discrete wavelet transform
CHB-MIT	Children's Hospital Boston - Massachusetts Institute of Technology
BPSVM	Biased p-norm support vector machine
iEEG	Intracranial electroencephalography
MLP	Multilayer perceptron
CNN	Convolutional Neural Network
ADC	Analogue to digital converter
SGD	Stochastic Gradient Descent
MCU	Microcontroller Unit
IoT	Internet of Things

Sumário

1	INTRODUÇÃO	21
1.1	Motivação para a Dissertação	21
1.2	Descrição do problema	23
1.3	Objetivos Geral e Específicos	23
1.4	Produção científica	24
1.5	Estrutura da Dissertação	24
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Epilepsia	25
2.2	Sinais de Eletroencefalograma (EEG)	26
2.3	Aprendizado de Máquina	27
2.4	Redes Neurais	27
2.5	Deep Learning	28
2.6	Dispositivos Vestíveis	29
2.7	Sistemas Embarcados	29
2.8	Sistemas Embarcados com aplicação em Saúde	30
3	TRABALHOS RELACIONADOS	31
4	MATERIAIS E MÉTODOS	33
4.1	Base de Dados	33
4.2	Preparação e Pré-processamento dos Dados	34
4.3	Preparação dos Dados de Entrada	35
4.4	Arquiteturas Avaliadas	36
4.4.1	SeizNet	36
4.4.2	Resnet	36
4.4.3	SqueezeNet	37
4.4.4	SeizNet Modificada	37
4.5	Treinamento	37
4.6	Plataforma embarcada de teste	38
5	RESULTADOS	41
5.1	Seleção de Arquitetura	41
5.2	Efeitos da filtragem	42
5.3	Efeitos da Amostragem em Épocas	44
5.4	Resultados de Modelos Intra-paciente	46

5.5	Efeitos do Número de Canais Usados	46
5.6	Efeitos da Compressão de Pesos	48
5.7	Resultados de Modelos Inter-pacientes	49
5.8	Análise Embarcada	51
6	CONCLUSÃO	55
	REFERÊNCIAS	57

1 Introdução

A epilepsia é uma doença crônica que atinge uma grande parcela da população mundial, com cerca de 40 milhões de pessoas convivendo com a doença no mundo (AL., 2016) e 150 mil somente no Brasil (MINISTÉRIO DA SAÚDE, 2013). Desses indivíduos, uma parcela de 30% não consegue controlar a doença (CHANG; LOWENSTEIN, 2003), sendo acometidos por crises frequentes (ataques) que são características da doença. É relevante dizer ainda que nem todos os tipos de epilepsia se manifestam da mesma forma, os sintomas variam de caso a caso e estão ligados ao local no cérebro onde ocorreu o início da atividade anormal (ELGER; SCHMIDT, 2008). A necessidade de criar ferramentas que possam detectar ou prever esses episódios está diretamente relacionada com a possibilidade de proporcionar uma melhor qualidade de vida às pessoas com essa sintomatologia.

O diagnóstico e acompanhamento da doença se faz através da análise, na maior parte dos casos, de sinais de Eletroencefalograma (EEG). Tais sinais são colhidos de eletrodos implantados na superfície do crânio desses indivíduos, nos quais os padrões de onda apresentados formam o diagnóstico e permitem ao médico avaliador identificar em que parte do cérebro originou-se a crise, podendo assim aplicar as devidas medidas terapêuticas.

A detecção automática de ataques epiléticos vem recebendo bastante atenção nos últimos anos, onde técnicas de aprendizado de máquina foram aplicadas com resultados promissores. Além disso, algoritmos de aprendizado profundo podem ser aplicados onde há uma abundância suficiente de dados. Esses algoritmos, anteriormente limitados a máquinas com alto poder computacional, atualmente podem ser empregados em plataformas com baixo consumo de memória e com pouco poder computacional. Com os avanços na área médica e o crescimento de aplicações de Internet das Coisas (IoT), a aplicação de técnicas de aprendizado profundo utilizado sistemas embarcados é uma realidade em vários aspectos da sociedade atualmente. Ao explorar a aplicação dessa classe de algoritmos em sistemas embarcados abre oportunidades para a criação de dispositivos vestíveis que possam prever/detectar a ocorrência de crises epiléticas.

1.1 Motivação para a Dissertação

A detecção de crises tem uma relevância direta no bem-estar do indivíduo que vive com epilepsia. Considerando a parcela daqueles casos que, mesmo sob tratamento, ainda apresentam casos de convulsões e ataques significativos, uma ferramenta que possa ajudá-los a saber quando a próxima crise irá acontecer pode prepará-los melhor para enfrentá-las, adicionando assim um fator de previsibilidade a uma condição anteriormente

imprevisível.

Sistemas que possam detectar ou prever quando uma crise vai acontecer podem melhorar o modo como a sociedade lida atualmente com esses indivíduos. Novos tratamentos baseados nesses sistemas podem agir no momento correto de forma a diminuir os efeitos, ou até mesmo parar um ataque epilético. Ainda existe a possibilidade de sistemas utilizarem a automação para administrar o medicamento antes mesmo que o episódio se inicie. Por fim, uma solução vestível pode ser criada de forma a acompanhar os pacientes diariamente avisando-os da iminência de um ataque.

Dessa forma, uma solução que possa analisar os dados produzidos através de sinais de EEG, utilizando poucos recursos computacionais e visando a sua utilização em um dispositivo embarcado seria interessante. Tal abordagem deve possuir boa taxa de acertos para a aplicação, ao passo que será necessário manter a privacidade dos dados coletados, caso tais dados sejam utilizados para futuras melhorias do algoritmo em uma plataforma de maior poder computacional. Considerando as contribuições científicas da pesquisa, podemos destacar:

- **Desenvolvimento de uma nova aplicação para o aprendizado profundo embarcado:** As técnicas de aprendizado profundo têm ganhado cada vez mais espaço no mundo embarcado, e a cada dia novas aplicações são encontradas e algoritmos anteriormente utilizados apenas em máquinas com grande poder computacional ganharam suas versões embarcadas.
- **Estudo específico com o intuito da criação de uma arquitetura aplicável em dispositivos vestíveis:** Um dos objetivos finais desta pesquisa foi possibilitar a utilização dos algoritmos de aprendizado profundo em uma solução vestível que o indivíduo com epilepsia possa portar a todos os instantes, com o intuito de constantemente monitorar suas atividades cerebrais e avisá-lo no momento em que uma crise pode iniciar. Note que o objetivo não é a criação de tal dispositivo, mas a viabilização do *software* que será executado nele.
- **Comparação entre arquiteturas distintas de CNN:** A pesquisa procurou comparar as abordagens da aplicação de tipos de redes convolucionais diferentes, de forma a descobrir qual o mais adequado à tarefa em questão, além de detectar qual o mais adequado para a aplicação embarcada.
- **Busca de uma nova solução que possa ser aplicada para vários pacientes:** Esta pesquisa buscou ainda encontrar modelos que possam ser utilizadas para diversos pacientes, isto é, após desenvolvido, o modelo poderá ser aplicado para outros pacientes e não somente para aquele sobre o qual o modelo foi desenvolvido, como acontece nos trabalhos mais relevantes acerca do problema de detecção de ataques epiléticos.

1.2 Descrição do problema

Este trabalho busca resolver o seguinte problema:

Deteção de ataques epilépticos utilizando dados de eletroencefalograma.

Entretanto tal problema propõe uma análise mais aprofundada que se alinhe com os objetivos gerais e específico desta pesquisa. Assim, alguns requisitos se fazem necessários para estreitar e isolar o problema:

- **Utilização de dados puros:** Muitas das pesquisas nesta área utilizam dados transformados de eletroencefalograma. Neste trabalho, buscou-se utilizar os dados de EEG no próprio domínio do tempo, sem a necessidade de transformá-los em imagens ou dimensões que não estejam no tempo.
- **Análise intra e inter-pacientes:** O problema pode ser abordado tanto utilizando a técnicas de uso de dados inter-pacientes, ou seja, os dados de um grupo de pacientes podem ser usados em um paciente desconhecido; ou intra-paciente, onde os dados do mesmo paciente fornecem a base de treinamento para dados desconhecidos do mesmo indivíduo. Existem poucos trabalhos que se concentram na abordagem inter-pacientes e, portanto, esta pesquisa fará análises nesse aspecto.
- **Análise embarcada:** A criação de um dispositivo vestível é uma evolução do desenvolvimento da solução para o problema descrito, portanto foi considerada a incorporação da solução produzida em uma plataforma embarcada com a análise de recursos e a proposta de um protótipo.

1.3 Objetivos Geral e Específicos

Como objetivo geral, este trabalho buscou desenvolver um algoritmo de aprendizagem profunda para plataformas embarcadas (preferencialmente vestível), a fim de solucionar o problema de deteção de ataques epilépticos a partir da utilização de sinais de eletroencefalograma. Portanto, uma das contribuições desejadas é a apresentação de um método pelo qual uma rede neural convolucional possa ser embarcada em um dispositivo com baixo consumo de memória e baixo poder computacional.

Os objetivos específicos são:

- Desenvolver uma arquitetura que possibilite a criação de modelos não específicos para um único paciente. Isso significa que a principal arquitetura discutida nesse trabalho deve ter aplicações na tarefa de deteção inter-pacientes.

- Demonstrar como avaliar potenciais arquiteturas para a melhor aplicação no problema de detecção de ataques, levando em consideração uma aplicação embarcada. Considerando o objetivo principal acima, é necessário a demonstração de um método pelo qual outras arquiteturas possam ser avaliadas de forma a viabilizar uma aplicação embarcada.
- Elaboração de um protótipo embarcado que possa acomodar a arquitetura escolhida e levar em consideração a aplicação desta em um futuro sistema vestível.

1.4 Produção científica

Durante este projeto de mestrado, os seguintes trabalhos científicos foram aceitos e publicados, a saber:

- ALBUQUERQUE, L. M. de; SILVA, E. T. da. Using deep-learning to automatically detect epileptic seizure in a wearable device. In: 2020 X Brazilian Symposium on Computing Systems Engineering (SBESC). IEEE, 2020. Disponível em: <<https://doi.org/10.1109/sbesc51047.2020.9277844>>.

1.5 Estrutura da Dissertação

Esta dissertação está organizada em 5 capítulos. O capítulo 1, Introdução, determina os objetivos do trabalho aqui proposto, bem como estabelece o problema a ser solucionado e abre o texto com algumas informações sobre o problema. Já no capítulo 2, Fundamentação Teórica, entramos em um relato detalhado dos termos e conceitos que foram utilizados para embasar este trabalho, sem os quais o seu entendimento seria prejudicado. O capítulo 3 é destinado aos trabalhos que serviram de inspiração e comparação para esta dissertação.

Os capítulos 4, Materiais e Métodos, e 5, Resultados, formam as contribuições principais deste trabalho. No capítulo 4 demonstramos quais foram os dados e métodos utilizados, bem como entramos em maiores detalhes de como os conceitos apresentados no capítulo 2 serão aplicados para a resolução do problema apresentado. É neste capítulo que estabelecemos as arquiteturas que serão avaliadas, além de descrevermos as eventuais mudanças que serão aplicadas às mesmas. No capítulo 5, mostramos quais foram os resultados apresentados pelos métodos desenvolvidos no capítulo 4. Neste capítulo também discutimos os desafios de embarcar a solução apresentada em uma plataforma com baixo consumo de memória e recursos computacionais.

Finalmente o capítulo 6, Conclusão, resume os achados do capítulo 4, bem como estabelece um plano para futuras contribuições e como a pesquisa aqui desenvolvida pode ter continuidade.

2 Fundamentação Teórica

2.1 Epilepsia

A epilepsia é um conjunto de síndromes, e não uma doença única, que possui como uma de suas características mais marcantes a ocorrência de convulsões (CHANG; LOWENSTEIN, 2003), ou crises epiléticas, que afetam os indivíduos em todos os aspectos biopsicossociais (SHOEB; GUTTAG, 2010). É relevante dizer ainda que nem todos os tipos de epilepsia se manifestam da mesma forma, os sintomas variam de caso a caso e estão ligados ao local no cérebro onde ocorreu o início da atividade anormal (ELGER; SCHMIDT, 2008). A maioria dos casos apresenta convulsões, crises de ausência (onde o indivíduo afetado não responde a estímulos externos) e há ainda aqueles que experimentam distúrbios nos sentidos e na memória. As origens da epilepsia são tão diversas quanto seus sintomas (SHOEB; GUTTAG, 2010), estando as mutações genéticas, trauma cerebral, infecções e derrames entre as causas mais conhecidas. Por conta dessa gama de causas e efeitos, a epilepsia ainda representa um desafio de diagnóstico e tratamento até mesmo em países desenvolvidos.

O diagnóstico pode ser feito utilizando resultados de eletroencefalogramas (EEG), onde podem ser detectadas atividades anormais nas ondas cerebrais durante uma crise (JAMESON et al., 2018). Para um melhor diagnóstico, o sinal de EEG coletado deve conter diferentes estágios de consciência (e.g. acordado, sonolento ou dormindo), pois as atividades epiléticas podem ocorrer nesses momentos (STAFSTROM; CARMANT, 2015). Ainda segundo (STAFSTROM; CARMANT, 2015), outros métodos de imagem podem ser utilizados para o diagnóstico de forma complementar ao EEG tais como Ressonância Magnética e tomografia computadorizada.

Uma crise epilética é caracterizada por uma hiperatividade e hipersincronia de uma quantidade de neurônios em uma determinada região do cérebro (ELGER; SCHMIDT, 2008). De uma forma simplificada, uma quantidade de neurônios é ativada em um momento em que não deveria. Onde essa atividade anormal começa determina o tipo de reação que o indivíduo irá apresentar: quando ocorre em uma região responsável pelos movimentos, ele pode apresentar episódios de rigidez muscular e convulsões, enquanto se iniciarem em uma região responsável pela memória, sentimentos de *dejà vu* podem ocorrer. Por fim, vale salientar que o descontrole neuronal originado em uma região pode se espalhar para outras e provocar sintomas diferentes (SHOEB; GUTTAG, 2010).

Os principais tipos de ataques epiléticos, segundo (STAFSTROM; CARMANT, 2015) são:

- Crises de Ausência: Falta de resposta a estímulos externos.
- Tônico-clônicas: Rigidez muscular com eventuais espasmos.
- Mioclônicas: Rápidos espasmos.
- Atônicas: Perda de tônus muscular.

2.2 Sinais de Eletroencefalograma (EEG)

A eletroencefalografia é uma técnica eletrofisiológica que utiliza eletrodos para o monitoramento de sinais elétricos emitidos pelo cérebro. Os diferentes sinais emitidos, chamados canais, são mensurados a partir da diferença de potencial elétrico entre pares distintos de eletrodos (SCHOMER; SILVA, 2010). Normalmente, na utilização para o diagnóstico de patologias, os sinais são gravados durante um intervalo de tempo, e regiões específicas, geralmente aquelas relacionadas à ocorrência de um evento, são analisadas. Existem também aplicações que convertem os sinais para o domínio da frequência analisando os sinais nesse espectro (WHITTINGSTALL; LOGOTHETIS, 2009). Os eletrodos utilizados para a captação dos sinais podem ser classificados em dois tipos: de escalpo e intracranianos.

Na análise de escalpo, uma rede é utilizada e os eletrodos são fixados na superfície da cabeça do paciente. Por estarem fixados fora do crânio, tais eletrodos estão sujeitos a interferências externas, além de não captarem bem a atividade cerebral proveniente dos neurônios mais internos do cérebro (SHOEB; GUTTAG, 2010). Existe ainda a influência atenuante que o próprio crânio oferece, visto que os sinais provenientes dos neurônios são da ordem de milivolts e de fácil interferência.

Por outro lado, na análise intracraniana, o iEEG, os eletrodos são posicionados em pontos específicos no interior ou na superfície do cérebro (BLOUNT et al., 2008). Por ser uma técnica extremamente invasiva, não existem muitos locais onde estes eletrodos possam ser posicionados, o que diminui a área que pode ser monitorada (SHOEB; GUTTAG, 2010). Contudo as interferências externas são quase inexistentes, já que o crânio serve de proteção contra esses sinais.

O eletroencefalograma é o principal método de diagnóstico de epilepsia utilizado atualmente. Entretanto um dos grandes desafios apresentados na análise desses sinais é a sua variabilidade de paciente para paciente (GOTMAN, 1982). Qu e Gotman (1997) especificam três dificuldades na análise automática desses sinais: a grande variabilidade entre indivíduos, o aspecto temporal que caracteriza um evento de epilepsia (isto é, o evento é algo que se desenvolve ao longo de um período de tempo e não simplesmente aparece de forma repentina) e sinais que caracterizam epilepsia em um paciente podem representar uma atividade normal em outro.

2.3 Aprendizado de Máquina

Aprendizado de máquina é uma subárea da Inteligência Artificial que trata de algoritmos capazes de aprender automaticamente a identificar padrões em bases de dados sem receberem comandos explícitos, mas através de poderosas análises estatísticas aplicadas nessas bases (BISHOP, 2006).

Pode-se destacar duas grandes áreas de concentração em aprendizado de máquina: o aprendizado supervisionado e o aprendizado não supervisionado.

No aprendizado supervisionado, os dados utilizados para treinamento possuem classes bem definidas (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012), assim esses algoritmos são capazes de inferir novos exemplos de dados de entrada após terem “visto” um outro conjunto onde ambos os dados e as classes lhes foram apresentados. Tais algoritmos são muito utilizados em tarefas de classificação. Por outro lado, no aprendizado não supervisionado, os dados de treinamento não possuem classes bem definidas, ou simplesmente durante a produção do conjunto de dados as classes não foram estabelecidas, e possuem aplicação em tarefas de clusterização ou de nomeação automática de classes.

2.4 Redes Neurais

Redes Neurais Artificiais são uma classe de algoritmos de aprendizado de máquina que têm por inspiração original o cérebro humano. Os componentes formadores das redes neurais são os neurônios artificiais, e tal qual seus primos biológicos, eles recebem estímulos, neste caso denominados sinais de entrada, e emitem estímulos, sinais de saída, que são conectados como entradas em outros neurônios. Cada neurônio possui ainda uma função de ativação responsável por modelar as não linearidades presentes nos dados de entrada. A definição mais antiga de um neurônio artificial foi escrita por McCulloch e Pitts (1943).

Quando conectados sob a forma de camadas, os neurônios artificiais formam uma rede neural. O Perceptron de multicamadas (MLP) é a implementação de rede neural mais antiga e conhecida, base de muitas outras implementações de redes utilizadas atualmente. Nesta implementação existem tipicamente três ou mais camadas: uma de entrada, uma ou mais camadas ocultas e uma saída (HAYKIN, 1998). Usualmente MLPs são utilizados em tarefas de aprendizado supervisionado, onde o treinamento se dá em duas fases: propagação direta e retropropagação. Na fase de propagação direta, os dados de treinamento são apresentados à camada de entrada que os propaga camada por camada até a saída, produzida de acordo com as necessidades do problema em estudo. Na fase de retropropagação, um erro é calculado de acordo com a saída anteriormente produzida e é utilizado para o ajuste dos pesos sinápticos utilizados em cada camada da rede (HAYKIN, 1998). O processo se repete com o objetivo de minimizar o erro produzido ao fim de cada

fase de propagação direta.

2.5 Deep Learning

Deep learning descreve uma classe de algoritmos de aprendizado de máquina que são baseados em redes neurais e possuem como característica principal o fato de terem várias camadas. Além disso, tais algoritmos são especialistas em aprendizado de representação (LECUN; BENGIO, 1998), isto é, são capazes de aprender padrões e atributos automaticamente a partir de dados puros. Neste caso, usualmente, cada camada se torna responsável por aprender um certo nível de abstração/representação dos dados. Por exemplo, se tomarmos uma imagem e a submetermos a uma rede neural convolucional, dependendo da quantidade de camadas convolucionais, algumas podem se especializar em aprender traços de bordas, que são atributos menos complexos, enquanto outras, geralmente mais internas, podem aprender sobre rostos e objetos, que são atributos mais complexos.

Redes neurais convolucionais (CNNs) são uma subclasse de algoritmos de *deep learning* cuja principal operação/tipo de camada é a convolucional. Uma das principais aplicações desses algoritmos é a classificação de imagens. Uma das características mais promissoras dessas redes é sua baixa susceptibilidade a variações de ângulo, posicionamento e distorção. Segundo LeCun e Bengio (1998), três propriedades arquiteturais das CNNs colaboram para essa característica: campos receptivos locais, o compartilhamento de pesos e, finalmente, a subamostragem espacial ou temporal.

Uma camada convolucional é composta por um conjunto de mapas de atributos. Tais mapas são o resultado de uma operação de convolução (produto de matrizes) entre os dados de entrada e a matriz de *kernels* de cada camada (LECUN; BENGIO, 1998). Tais *kernels* são o objeto de aprendizado das redes convolucionais e podem ser vistos como extratores de características que aprenderam automaticamente, durante a etapa de treinamento, quais atributos da entrada são importantes para a tarefa apresentada. A subamostragem espacial/temporal é executada pelas camadas de *pooling*, responsáveis por reduzir a resolução dos mapas de características produzidos pela camada convolucional (SCHERER; MÜLLER; BEHNKE, 2010). Ao final de uma rede convolucional geralmente encontram-se algumas camadas completamente conectadas, uma versão de um MLP, que são responsáveis por classificar o mapa de características produzido pela última camada convolucional (ou de *pooling*) da rede.

2.6 Dispositivos Vestíveis

A definição de Raad (2021) para dispositivos vestíveis é a que mais se aproxima da definição usada neste trabalho: um dispositivo que pode ser incorporado em acessórios ou vestimentas que podem ser utilizados *confortavelmente* por seres humanos. Uma importante característica do dispositivo vestível é a sua habilidade de se tornar imperceptível ao usuário, não adicionando complexidades à utilização do acessório ao qual está incorporado.

Além desta característica principal e daquelas relacionadas a sistemas embarcados, os dispositivos vestíveis possuem propriedades particulares a sua aplicação e interação com o usuário (RAAD, 2021):

- **Interface de comunicação:** Geralmente em dispositivos vestíveis uma interface de comunicação direta é necessária (um *app* ou dispositivo companheiro).
- **Fatores físicos:** Durabilidade, conforto, estética e ergonomia são características especialmente importantes em dispositivos vestíveis.
- **Coleta de dados:** Em dispositivos vestíveis, a coleta e armazenamento de dados faz parte do próprio dispositivo que processa os dados. Em alguns casos, os dados podem ser transmitidos e processados em outro dispositivo (e.g. servidor na nuvem).

2.7 Sistemas Embarcados

De acordo com Wolf (2012), um Sistema Embarcado pode ser definido, superficialmente, como qualquer computador programável que não é de propósito geral. A definição indica que o sistema embarcado deve possuir um propósito ou uma utilidade específica. Ainda acompanhando a definição de Wolf (2012), as características principais desses sistemas demonstram que embora, por definição, tais sistemas sejam indicados a único propósito, eles devem ser capazes de executar algoritmos de diferentes complexidades.

Um fator chave para o desenvolvimento de aplicações e sistemas embarcados é a eficiência (MARWEDEL, 2022). Isso significa que, para embarcar uma aplicação de forma bem sucedida, é necessário avaliar certos fatores que irão influenciar esta eficiência:

- **Energia:** O consumo de energia necessário para a execução da aplicação, levando em consideração possíveis restrições de tamanho e uso de bateria.
- **Tamanho da aplicação e sua eficiência:** Se a aplicação foi escrita de forma eficiente a consumir somente o necessário de recursos de memória e CPU.
- **Peso:** O peso irá implicar na utilização de um possível dispositivo com a aplicação embarcada. Fatores como memória adicional, baterias e unidades de entrada e saída de dados devem ser levados em consideração.
- **Custo:** O custo individual de cada componente necessário para a execução da

aplicação embarcada.

Sistemas embarcados se tornaram cada vez mais especializados na atualidade e podem ser encontrados em diversas áreas: sistemas de controle de transporte, telecomunicações, eletrodomésticos, aplicações médicas, residências inteligentes, etc.

2.8 Sistemas Embarcados com aplicação em Saúde

Uma aplicação bastante conhecida de sistemas embarcados (e sistemas vestíveis) é na área da saúde. Na verdade, a computação em geral vem sendo utilizada como forma de melhorar o diagnóstico, acompanhamento e tratamento em diversos setores da medicina. Exemplos clássicos como a tomografia computadorizada, bombas de infusão automáticas e monitores cardíacos são claras definições de sistemas computacionais com propósito específicos. Entretanto, a medicina não está mais se limitando a grandes máquinas que geralmente não podem sair de dentro do hospital para fazer o seu trabalho.

Como discutido por (KELLY et al., 2020), sistemas embarcados sob a forma de Internet das Coisas já vem sendo amplamente explorado em aplicações no meio médico. A aplicação desses sistemas se dá não somente na coleta dos dados de saúde, mas sim em todas as etapas do diagnóstico médico. O trabalho ainda cita em detalhes onde se dá a presença desses dispositivos em uma arquitetura IoT: camada de percepção, onde os dados são coletados; camada de comunicação, onde os dados são transferidos; e camada de aplicação onde esses dados são utilizados para o diagnóstico.

Como explicado por (DUNN; RUNGE; SNYDER, 2018) a utilização de dispositivos vestíveis no acompanhamento do cuidado médico proporciona uma personalização ao cuidado e tratamento, o que promove um bem estar geral a estes indivíduos. Produtos como medidores glicêmicos, temporizadores cardíacos e *smartwatches* já são realidade em muitos mercados e a tendência da utilização desses dispositivos embarcados só aumenta a medida que eles se tornam mais especializados e com melhores resultados.

Ainda no trabalho desenvolvido por (DUNN; RUNGE; SNYDER, 2018), percebe-se como a introdução de algoritmos de aprendizado de máquina podem potencializar o uso desses dispositivos não só por usuários finais, mas para a mediação no meio clínico de um hospital. Portanto é de grande relevância qualquer tipo de pesquisa que se proponha a expandir o campo da utilização de dispositivos embarcados no meio da saúde.

3 Trabalhos Relacionados

A detecção automática de crises de epilepsia a partir da análise de sinais de EEG é uma tarefa que vem sendo explorada por diversos trabalhos nos últimos anos, com avanços relevantes e resultados promissores. Um dos trabalhos mais interessantes envolvendo técnicas de aprendizado de máquina e análise de dados de eletroencefalograma com o objetivo de detecção de crises foi desenvolvido por Shoeb e Gutttag (2010). Nesse estudo, foram utilizados os dados produzidos pelo Hospital Infantil de Boston contendo 844 horas de gravações com 163 episódios de crises epiléticas. Os autores utilizaram uma variação de máquina de vetores de suporte (SVM) com atributos extraídos dos sinais de EEG, no domínio do tempo e da frequência. Como métricas de performance, foram utilizadas o tempo entre a detecção e o início dos sintomas de uma crise, sensibilidade e alarmes falsos por hora. Como resultado, 91% dos eventos de ataque foram detectados com 10 segundos de latência, além disso, a sensibilidade encontrada chegou a 96%. Por sua relevância para a tarefa, esses achados ratificam a necessidade de automatizar a detecção de crises nesses sinais, bem como destacam de forma sistemática as dificuldades que outros pesquisadores podem encontrar ao explorar essa tarefa.

Com o acúmulo de dados produzidos atualmente por nossa sociedade, algoritmos envolvendo técnicas de *deep learning* têm se tornado cada vez mais comuns, já que tais técnicas são mais utilizadas em ambientes com abundância de dados para treinamento. O trabalho desenvolvido por Zhou et al. (2018), por exemplo, utilizou redes neurais convolucionais para detecção de ataques epiléticos em sinais de EEG. Na pesquisa supracitada, os autores utilizaram a análise desses sinais tanto no domínio do tempo quanto na frequência em três tipos de experimentos diferentes: duas tarefas de classificação binária, com detecção de períodos antes e durante os ataques e outros durante e após esses; e uma tarefa com três classes para detecção de períodos antes, durante e depois dos ataques. Duas bases de dados foram utilizadas, a CHB-MIT com resultados entre 49,0% e 95,6% de acurácia e os dados do hospital de Freiburg, com acurácia variando entre 85,1% e 96,7%. Este trabalho mostra uma clara aplicação de redes convolucionais na tarefa de detecção de crises. Os resultados são demonstrados em uma base de dados conhecida do problema, o que auxilia na comparação com pesquisas futuras, apesar de os autores não detalharem a arquitetura utilizada.

Na linha de pesquisa em *deep learning*, o trabalho apresentado por Avcu, Zhang e Chan (2019) também se destaca pela utilização de um exemplar de rede neural convolucional, entretanto, nesse trabalho os autores propõem uma arquitetura bem definida utilizando quatro camadas convolucionais. Eles aplicam sua rede em uma base de dados privada e comparam seus resultados com uma técnica que utiliza Máquina de Vetores

de Suporte com viés p-norm (BPSVM) para extração de características dos sinais de EEG. Neste trabalho existe também uma comparação entre duas quantidades distintas de canais (sinais) a serem utilizadas durante o treinamento, 2 ou 18, o que ajuda a visualizar a influência da quantidade de canais na tarefa de detecção. Os resultados descobertos são promissores, com sensibilidade perto dos 96%. O destaque deste trabalho está na arquitetura bem definida e métodos de treinamento explícitos, o que proporciona uma melhor reprodutibilidade dos experimentos e futuras adaptações.

O trabalho desenvolvido por (GODOY,) apresenta a aplicação de arquiteturas de redes de *deep learning* no problema de predição de crises epiléticas. A utilização de redes convolucionais demonstra a proximidade das pesquisas. Após analisar 5 arquiteturas candidatas os autores apresentaram resultados na faixa de 98% para a detecção utilizada a CNN. Embora os resultados sejam individualizados não está claro se uma abordagem inter-pacientes foi utilizada no momento de treinamento dos modelos escolhidos.

Explorando abordagens que utilizam sistemas embarcados na tarefa de detectar crises em sinais de EEG, o trabalho de Sopic, Aminifar e Atienza (2018) propõe uma solução intitulada e-Glass que utiliza uma quantidade reduzida de eletrodos escondidos em óculos que podem ser utilizados por portadores de epilepsia. O dispositivo utiliza um ARM Cortex-M3 para fazer a detecção de novos episódios de crise através da leitura de quatro eletrodos posicionados na armação dos óculos. O algoritmo de extração de características utilizado foi de transformada wavelet discreta (DWT), com atributos não lineares que depois foram submetidos a uma floresta aleatória (Random Forest). Os resultados obtidos foram muito promissores, com sensibilidade chegando a 93,0% na base de dados do CHB-MIT. Embora os autores utilizem outros algoritmos, diferentes daqueles que serão explorados nesta pesquisa, eles trouxeram resultados interessantes e destacaram uma plataforma embarcada que pode ser candidata para a exploração.

Quando analisamos implementações de redes neurais convolucionais em sistemas embarcados, o trabalho de Wang et al. (2019) traz implementações das redes mais conhecidas em dispositivos ARM com análises de tempo de computação e gasto energético. Entretanto, por avaliar aplicações de imagem, a versão do processador utilizado foi a família ARM Cortex A, com um bom poderio computacional. Outros trabalhos, como o desenvolvido por Venieris, Kouris e Bouganis (2018), utilizam FPGA, o que, para a aplicação desta pesquisa, seria muito além do poder computacional necessário. Esta avaliação bibliográfica sugere que existem poucos esforços em se implementar tais redes convolucionais em plataformas com um baixo volume de recursos.

4 Materiais e Métodos

4.1 Base de Dados

A base de dados utilizada neste trabalho foi a *CHB-MIT Scalp Database* (GOLDBERGER et al., 2000). A versão original deste *dataset* contém dados de eletroencefalograma de 23 pacientes com idade de 18 anos ou menos. Os sinais foram coletados utilizando uma frequência de 256 Hz e uma resolução de 16 bits. Tanto o arranjo dos eletrodos quanto a nomenclatura obedecem o sistema internacional de arranjo de eletroencefalograma 10-20 (ROJAS et al., 2018). Ao todo foram coletadas 844 horas de sinais contendo 163 episódios de crise epiléptica. O paciente número 1 está representado duas vezes na base de dados, tanto como paciente 1 quanto como paciente 21. Para os experimentos desenvolvidos neste estudo, considerou-se os dados deste paciente individualmente, como sendo um outro paciente, totalizando 24 subconjuntos de dados (1 para cada paciente).

O conjunto eletrônico de dados contém uma pasta para cada paciente e cada pasta contém de 9 a 46 arquivos no formato *.edf*, formato padrão para a gravação de dados de EEG. Na maioria dos casos, cada arquivo contém uma hora de gravação, embora isso não seja o padrão para todos os arquivos. Além disso, os dados disponibilizados para a comunidade não contêm informações diretas dos pacientes originais. Como estas informações são protegidas, foram ofuscadas nos arquivos disponíveis. Vale ainda ressaltar que a dependência temporal dos dados não foi ofuscada, sendo esta de extrema importância para a identificação de um episódio epiléptico e não contendo informações protegidas dos pacientes.

Um outro aspecto importante de ser destacado é que o *dataset* original apresenta diversos problemas que dificultam a sua utilização direta em uma tarefa de aprendizado de máquina:

- Faixas de 10 segundos onde não há qualquer tipo de sinal (*gaps*).
- Sinais extras que foram adicionados ao EEG e que foram coletados juntamente com os sinais (tais como sinais do nervo vago).
- Sinais de marcação, adicionados para facilitar a visualização dos dados.

Os sinais coletados são formados pela diferenciação de eletrodos vizinhos. O nome do sinal se dá pela nomenclatura dos eletrodos usados, assim poderemos utilizar nesta dissertação as nomenclaturas sinal, par de eletrodos e canal para designar um dos sinais do conjunto de dados utilizado.

A maioria dos arquivos contém 23 sinais, contudo, como veremos na seção seguinte,

nem todos poderão ser utilizados na tarefa de classificação, pois se encaixam nas categorias acima ou simplesmente não foram coletados em todos os pacientes. Ao final do processo de pré-processamento, 18 sinais serão utilizados.

4.2 Preparação e Pré-processamento dos Dados

Para preparar os dados para a utilização na classificação, algumas medidas precisaram ser tomadas para remover os artefatos indesejados do conjunto de dados original.

- Alguns pares de eletrodos não estão presentes em todos os pacientes, por exemplo, o par T7-FT9 está presente no paciente 12, mas não está presente no paciente 19.
- Alguns sinais não são de EEG, mas sim sinais que foram adicionados durante a coleta do sinal original (e.g. sinais do nervo vago).
- Alguns sinais foram adicionados para facilitar a visualização dos dados quando plotados em conjunto (e.g. canais 5 a 13 do paciente 12).
- Faixas com nenhum sinal coletado (*gaps*).

Para padronizar o conjunto de dados, os sinais marcadores foram removidos e foram utilizados apenas canais disponíveis em todos os pacientes. Por fim, os sinais que não são de EEG não foram utilizados no estudo e as faixas de sinais vazios também foram desconsideradas. Ao final da preparação, 18 canais restaram para serem utilizados nos passos seguintes deste estudo. Na Tabela 1 estão listados os pares utilizados.

Tabela 1 – Pares de eletrodos usados nesse estudo.

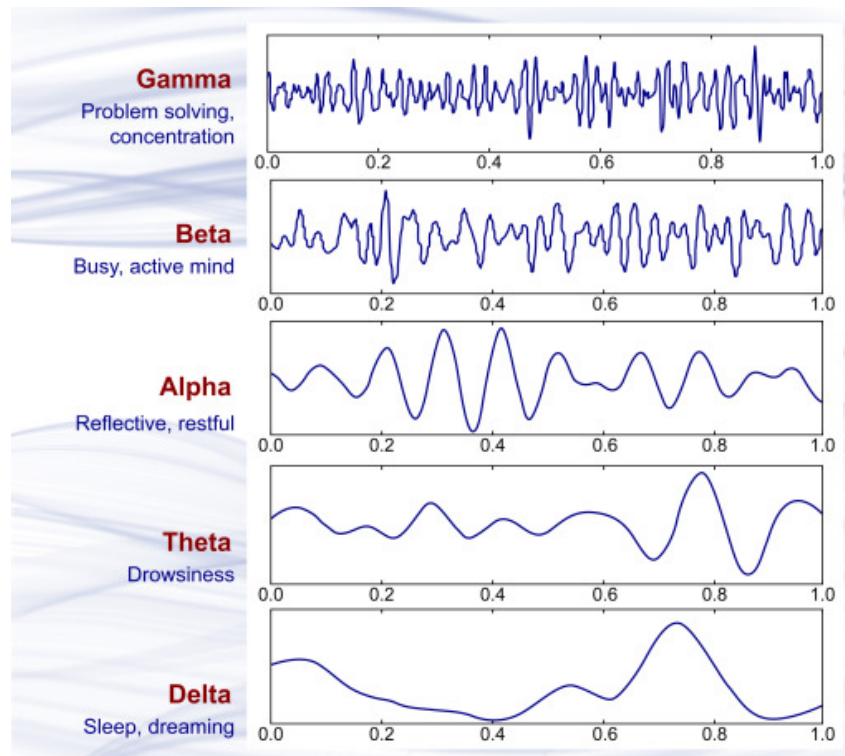
FP1-F7	F7-T7	T7-P7
P7-O1	FP1-F3	F3-C3
C3-P3	P3-O1	FP2-F4
F4-C4	C4-P4	P4-O2
FP2-F8	F8-T8	T8-P8
P8-O2	FZ-CZ	CZ-PZ

Fonte: Autor.

Cada eletrodo é posicionado em quatro lóbulos do crânio: temporal, frontal, parietal e occipital. De acordo com Kumar e Bhuvaneswari (2012), o sinal de EEG coletado tem componentes de frequências diferentes (e.g. alfa, teta e gama) e esses componentes estão relacionados a diferentes atividades e anormalidades encontradas no cérebro humano. A Figura 1 demonstra exemplos de ondas desses componentes encontradas comumente nos sinais de EEG ao longo do tempo.

Finalmente, para normalizar os dados provenientes dos 18 canais escolhidos, todos os sinais foram reescalados para ter variância unitária e média zero. Isso ajudou a eliminar discrepâncias relacionadas às diferenças de tensão produzidas por cada eletrodo

Figura 1 – Ilustração dos tipos de ondas encontradas no EEG



Fonte: Abhang, Gawali e Mehrotra (2016).

do arranjo. Além disso, os dados foram tratados com um filtro passa faixa para remover frequências indesejadas providos pela linha de energia e outras interferências. A faixa de frequência usada foi a de 4 Hz a 40 Hz, pois esta é a mais adequada para descrever artefatos de ataques epilépticos (TSIPOURAS, 2019). É importante ressaltar que a filtragem e o tratamento dos sinais precisam ser feitos em relação a todo o sinal, ou seja, devem ser realizados antes do passo de separação de dados (ver próxima seção).

4.3 Preparação dos Dados de Entrada

Uma rede neural convolucional normalmente trabalha com dados de entrada de tamanho fixo, portanto o *dataset* descrito acima precisa ser apresentado à CNN utilizando pedaços fixos. O tamanho escolhido será o equivalente a janelas de 5 segundos, ou 1280 amostras. Tendo cada amostra 18 canais, o tamanho máximo de cada amostra de entrada será de 1280 por 18.

Para dividir os sinais de EEG em partições de tamanho fixo, foi utilizada a técnica descrita por Avcu, Zhang e Chan (2019) e Hossain et al. (2019), onde uma janela de tamanho fixo é usada para percorrer o sinal original e capturas são coletados após certas distâncias. O processo é formalmente conhecido como Amostragem baseada em épocas ou *Epoch-based sampling* (ZIYABARI et al., 2017). Essa abordagem de separação de dados pode ser desenvolvida utilizando janelas com ou sem sobreposição. A utilização

com sobreposição pode ser considerada uma abordagem de aumento de dados (*data augmentation*). Neste estudo utilizamos janelas de 5 segundos com diferentes níveis de sobreposição (i.e., 0%, 20%, 50%, e 80%), e analisamos com isso o impacto da sobreposição nos resultados de taxas de acerto dos modelos desenvolvidos.

4.4 Arquiteturas Avaliadas

Um dos objetivos desta investigação é identificar uma arquitetura que tenha uma boa taxa de acerto aliada a uma baixa utilização de memória. Neste trabalho, a taxa de acerto será considerada como a média de cada modelo individual para cada paciente no conjunto de dados. O uso de memória é definido como a quantidade de memória necessária para conter os pesos do modelo necessários para fazer a inferência de novos sinais de EEG.

As arquiteturas analisadas de acordo com os critérios, taxa de acerto e baixo consumo de memória, foram:

- SeizNet
- Resnet
- SqueezeNet
- SeizNet Modificada

4.4.1 SeizNet

Será utilizada a versão original desta rede apresentada por Avcu, Zhang e Chan (2019). A Tabela 2 abaixo mostra a arquitetura desta rede: quatro camadas convolucionais e duas camadas completamente conectadas. Este trabalho considera a taxa de acerto desta rede tanto no eixo intra-paciente quanto no eixo inter-pacientes. A versão original da SeizNet utilizou um *dataset* privado para explorar o eixo intra-paciente e o trabalho desenvolvido em Albuquerque e Silva (2020) aplicou esta rede ao CHB-MIT considerando o mesmo eixo.

4.4.2 Resnet

Por ser uma rede conhecida na comunidade, a Resnet (HE et al., 2015) será utilizada como *benchmark* para a avaliação da taxa de acerto dos modelos produzidos. Adaptações deverão ser realizadas para que esta rede, comumente utilizada em tarefas de reconhecimento de imagens, possa aceitar um sinal unidimensional como entrada. Neste trabalho foi utilizada a versão de 34 camadas, a menor disponível, para a execução dos experimentos.

Tabela 2 – Arquitetura da rede SeizNet original

Camada	Kernel/Neurônios/Parâmetros
Conv 1	8x1x10
MaxPool1	1x2
Dropout	0,2
Conv 2	16x1x10
MaxPool 2	1x2
Dropout	0,2
Conv 3	32x1x10
MaxPool 3	1x2
Dropout	0,2
Conv 4	64x1x10
Maxpool 4	1x2
Dropout	0,2
Flatten	3456
Dense 1	1x50
Dropout	0,5
Dense 2	1

Fonte: Autor.

4.4.3 SqueezeNet

Esta rede desenvolvida por Iandola et al. (2016) obteve bons resultados na tarefa de reconhecimento de imagens, além de ocupar pouco espaço de memória. A rede foi então adaptada para receber sinais de 1 dimensão, pois mostrou-se uma boa candidata para ser embarcada.

4.4.4 SeizNet Modificada

Esta é a evolução da SeizNet proposta neste trabalho que apresenta resultados satisfatórios para a tarefa tanto no eixo intra-paciente quanto no eixo inter-pacientes.

4.5 Treinamento

Neste trabalho foi utilizado o Keras (CHOLLET et al., 2015) para modelagem, treinamento e teste dos modelos de aprendizado de máquina. A máquina básica de treinamento foi um *desktop* com Intel Core i5 e 8GB de memória RAM sem GPU. A máquina de treinamento para os modelos mais complexos foi uma instância da Amazon com uma GPU NVIDIA T4, 8 CPUs virtuais e 32 GB de RAM. O algoritmo de otimização utilizado pelas arquiteturas SeizNet e SeizNet Modificada foi o Adam (KINGMA; BA, 2014), com uma taxa de aprendizado de 0,00041. Já para a Resnet, utilizou-se o algoritmo SGD com uma taxa de aprendizado de 0,01. Finalmente, para a Squeezenet, os parâmetros

foram SGD para o algoritmo e 0,04 para a taxa de aprendizado. Cada modelo foi treinado por 100 épocas, o processo foi repetido 10 vezes e a média dos resultados são os valores demonstrados na seção seguinte.

Quando apropriado, o treinamento seguiu o procedimento intitulado *leave-one-patient-out* indicado por Muhammad et al. (2018). Como o objetivo principal dos experimentos era desenvolver um modelo generalizado que possa ser utilizado por diferentes pacientes, seria benéfico então utilizar os dados de $N - 1$ pacientes para treinamento e usar os dados do N -ésimo paciente como teste.

Um outro aspecto importante do processo de treinamento é o balanceamento dos dados. No conjunto de dados CHB-MIT, o número de amostras negativas é muito superior ao de amostras positivas. Portanto utilizamos a técnica de subamostragem da classe majoritária, usada por Muhammad et al. (2018), onde o número de amostras positivas e negativas para treinamento é a mesma. Para tanto, as amostras negativas são coletadas aleatoriamente do conjunto de dados disponível. Contudo, durante a fase de teste, mostraremos uma análise comparativa dos casos com e sem subamostragem.

4.6 Plataforma embarcada de teste

Para a realização dos testes em plataforma embarcada foi utilizada uma placa de desenvolvimento da *ST Electronics* contendo um micro-processador STM32F746. A placa, uma STM32 Nucleo, possibilita a gravação rápida do programa de teste e a avaliação do porte de uma rede neural convolucional para a estrutura embarcada, emitindo resultados de tempo de computação e utilização de memória. As especificações da placa podem ser encontradas abaixo:

- 144 pinos de conexão.
- ARM Cortex-M7 com 216 MHz de *clock*.
- 1 MB de Flash e 320 kbytes de SRAM.
- Conexão USB e Ethernet.

Outras plataformas embarcadas foram avaliadas, a título de informação, e comparadas com a apresentada acima. Nesta comparação, avaliamos o custo e consumo médio da plataforma, bem como seu poder computacional e memória disponível para considerá-las como alternativas à Nucleo com STM32F746. As plataformas analisadas foram: ARM M7 - F745, ARM A53 - RBP, ARM M4 - F407 e FPGA Spartan7. A Figura 2 abaixo ilustra a plataforma STM32F746.

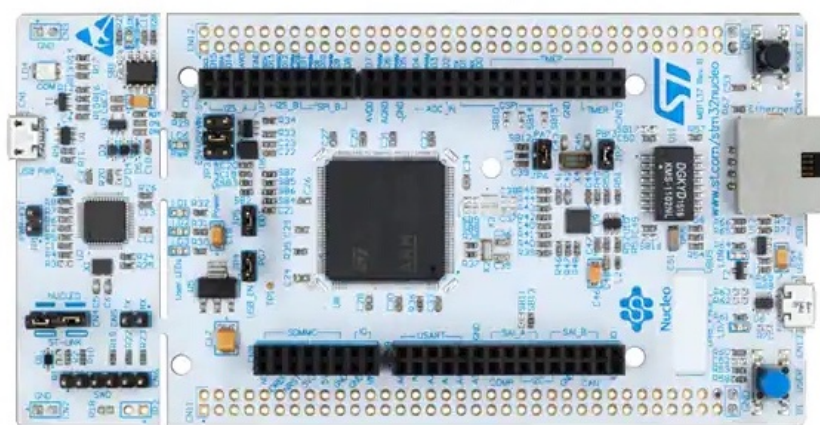


Figura 2 – Placa de desenvolvimento utilizada neste trabalho com o processador STM32F746

Fonte: Autor.

5 Resultados

Este capítulo mostra os resultados dos experimentos executados com a base de dados CHB-MIT e as redes Resnet, SeizNet, SqueezeNet e SeizNet Modificada. No total, oito seções compõem este capítulo. A seção 4.1 trata da avaliação e seleção de uma arquitetura candidata para a realização do restante dos experimentos. Nela comparamos algumas arquiteturas conhecidas, além de explicar eventuais modificações feitas nessas arquiteturas para a melhoria dos resultados. Na seção 4.2, discutimos os efeitos da filtragem no sinal de entrada e os seus benefícios para o resultado geral da rede proposta. A seção 4.3 destaca o efeito da técnica de amostragem com janela deslizante em épocas com sobreposição, técnica relevante para a tarefa de detecção de ataques epiléticos em EEG.

As seções 4.4 e 4.7 falam dos resultados finais obtidos para a classificação utilizando a melhor rede nos cenários intra e inter-pacientes. Nas seções 4.5 e 4.6, tratamos de duas características importantes ao considerar o objetivo de embarcar uma rede neural dessa categoria: a quantidade de canais e a sua influência no resultado da classificação e a compressão dos pesos quando embarcados. Finalmente a seção 4.8 faz uma análise sobre a implementação de um protótipo embarcado para a detecção de crises epiléticas.

5.1 Seleção de Arquitetura

As arquiteturas base foram avaliadas de forma a selecionar a mais adequada para a tarefa de detecção de ataques epiléticos. Aqui definimos como mais adequada a arquitetura que possui a melhor taxa de acerto média e o menor consumo de memória.

Três arquiteturas foram avaliadas: Resnet, SeizNet e SqueezeNet. Uma versão modificada da SeizNet (subseção 4.1.1) também foi considerada.

O consumo de memória foi determinado pela estimativa da soma de todos os pesos necessários para a execução de um passo de inferência caso a rede seja embarcada em um microcontrolador. A taxa de acerto média é dada pela média de todas as acurácias de todos os pacientes quando considerada a abordagem inter-pacientes. O resultado da análise pode ser encontrado na Tabela 3.

No resultado apresentado, o gasto de memória apresentado pela Resnet pode ser considerado proibitivo para a maioria dos sistemas embarcados que foram considerados neste estudo. Por outro lado, a SqueezeNet possui um baixo consumo de memória se comparado à Resnet. Finalmente a SeizNet foi a rede que apresentou menor utilização de recursos de memória. Seu desempenho não foi o melhor, devido ao seu tamanho reduzido, mas com as necessárias modificações foi possível melhorar a sua acurácia média para

Tabela 3 – Arquiteturas analisadas neste trabalho

Rede	Memória	Desempenho Médio
Resnet	271,0 MB	89,2%
SeizNet	697,0 kB	83,3%
SqueezeNet	8,8 MB	89,2%
SeizNet Modificada	702,5 kB	87,9%

Fonte: Autor.

se aproximar àquelas das redes maiores que possuem os melhores resultados. A versão modificada da SeizNet obteve melhor resultado se comparado com a versão original e com pouca adição ao consumo de memória, o que a caracteriza como uma boa candidata para estudos mais aprofundados, algo que será abordado nas próximas seções.

A evolução da SeizNet, aqui nomeada SeizNet Modificada, foi criada a partir de um estudo que se propôs a aumentar o desempenho geral da rede, representada pela acurácia média, sem a realização de modificações drásticas em sua estrutura original. Esta é uma versão proposta neste trabalho, que se compara a versão original da SeizNet e recebeu um tamanho de entrada diferente e uma camada convolucional a mais:

- Os dados de entrada serão apresentados à rede no tamanho de 1280 por N, onde N é o número de canais escolhidos do sinal de EEG original. O número de amostras aumentou de 1000 (AVCU; ZHANG; CHAN, 2019) para 1280, pois utilizou-se amostras de 5 segundos com a frequência original de 256 Hz. A utilização do sinal completo sem o processo de subamostragem originalmente proposto pela SeizNet é justificado, pois reduz a probabilidade de erros de *aliasing* serem introduzidos no processo, além de utilizar todos os dados que a base oferece.
- Uma camada convolucional foi adicionada para melhorar a capacidade da rede ao considerar a tarefa inter-pacientes. O número de mapas de características por conta da camada adicional será o dobro da versão original da rede, o que possibilita que a rede contenha mais informações sobre mais tipos de sinais de EEG. Como os sinais de eletroencefalograma são muito específicos de cada paciente, um número maior de mapas de características ajudou a incluir informações sobre mais pacientes.

A nova arquitetura possui um gasto total de memória de 702.520 kB. Além disso, a acurácia média para a tarefa inter-pacientes subiu para 87.9%. A Tabela 4 mostra o resultado final das modificações da arquitetura SeizNet:

5.2 Efeitos da filtragem

Uma importante técnica para melhorar a taxa de acerto de qualquer modelo é a filtragem. O sinal de EEG original continha diferentes tipos de artefatos indesejados que

Tabela 4 – Arquitetura da rede SeizNet Modificada

Camada	Kernel/Neurônios/Parâmetros
Conv 1	8x1x10
MaxPool1	1x2
Dropout	0,2
Conv 2	16x1x10
MaxPool 2	1x2
Dropout	0,2
Conv 3	32x1x10
MaxPool 3	1x2
Dropout	0,2
Conv 4	64x1x10
Maxpool 4	1x2
Dropout	0,2
Conv 5	128x1x10
MaxPool 5	1x2
Dropout	0,2
Flatten	3456
Dense 1	1x50
Dropout	0,5
Dense 2	1

Fonte: Autor.

adicionavam complexidades que a rede não conseguia eliminar durante o treinamento. Tais artefatos tendem a se concentrar em altas frequências, ao passo que a atividade cerebral e epiléptica tende a ocorrer em baixas frequências (TSIPOURAS, 2019).

Além de filtrar o sinal original na faixa de interesse para ataques epilépticos (4 Hz - 40 Hz), vários experimentos envolvendo a filtragem do sinal nas outras faixas de interesse foram executados. Tais faixas estão diretamente relacionadas às frequências especialmente afetadas no EEG durante os ataques epilépticos:

- **Teta:** 4 Hz – 7 Hz
- **Alfa:** 7 Hz – 12 Hz
- **Baixo Beta:** 12 Hz – 19 Hz
- **Alto Beta:** 19 Hz – 30 Hz
- **Gama:** 30 Hz – 40 Hz

Além de desenvolver modelos com dados nas faixas principais, as faixas secundárias e combinações de faixas foram testadas a fim de identificar as áreas que melhor expressariam um ataque epiléptico. As combinações utilizadas foram:

- **Alfa + Beta:** 7 Hz – 30 Hz
- **Betas:** 12 Hz – 30 Hz
- **Teta + Alfa + Betas:** 4 Hz – 30 Hz
- **Alfa + Beta + Gama:** 7 Hz – 40 Hz
- **Teta + Alfa + Baixo Beta:** 4 Hz – 19 Hz

O objetivo principal do experimento foi identificar as faixas que melhor caracterizam um ataque epiléptico e assim melhorar a taxa de acerto dos modelos desenvolvidos. Foram treinados e testados 240 modelos (10 faixas de frequência para cada um dos 24 pacientes), além dos 24 modelos nos quais não foi aplicada nenhuma filtragem. A Tabela 5 mostra a acurácia média de cada combinação de faixa. O efeito da filtragem é bastante relevante, com um aumento de quase 7% no resultado apresentado.

Tabela 5 – Acurácia média do classificador com dados filtrados - SeizNet Modificada

Sem Filtro	87,9%
4-7	93,8%
7-12	93,6%
12-19	94,2%
19-30	93,3%
30-40	94,2%
7-30	94,4%
12-30	93,8%
4-30	94,4%
7-40	95,2%
4-19	94,1%

Fonte: Autor.

Este incremento está atribuído à atenuação de sinais que não contribuem para a identificação da atividade epiléptica, tais como sinais da linha de distribuição de energia, frequências altas captadas pelo EEG e faixas que não afetaram os indivíduos utilizados nesse estudo. A melhor faixa encontrada é a que está entre as faixas Alfa e Gama, de 7 Hz a 40 Hz. Vale ressaltar que a natureza variável dos sinais de EEG pode fazer com que esta faixa seja diferente para um grupo distinto de indivíduos. Entretanto a natureza das atividades epilépticas sugere que as faixas aqui analisadas podem ser utilizadas também em um grupo distinto de indivíduos para encontrar resultados similares.

5.3 Efeitos da Amostragem em Épocas

A geração de dados utilizando janela deslizante, também conhecida como Amostragem em Épocas, é utilizada pelos autores Avcu, Zhang e Chan (2019), Ziyabari (2017) e Muhammad et al. (2018) com diferentes níveis de sobreposição. Essa técnica também pode

ser considerada como uma técnica de aumento da base de dados, visto que, dependendo do tamanho da sobreposição, novas amostras podem ser geradas. A Figura 3 mostra como as amostras podem ser criadas utilizando a janela deslizante com sobreposição de 50%. Neste estudo, buscou-se identificar o efeito da sobreposição na taxa de acerto final do classificador; para tanto, três níveis de sobreposição foram analisados: 20%, 50% e 80%.

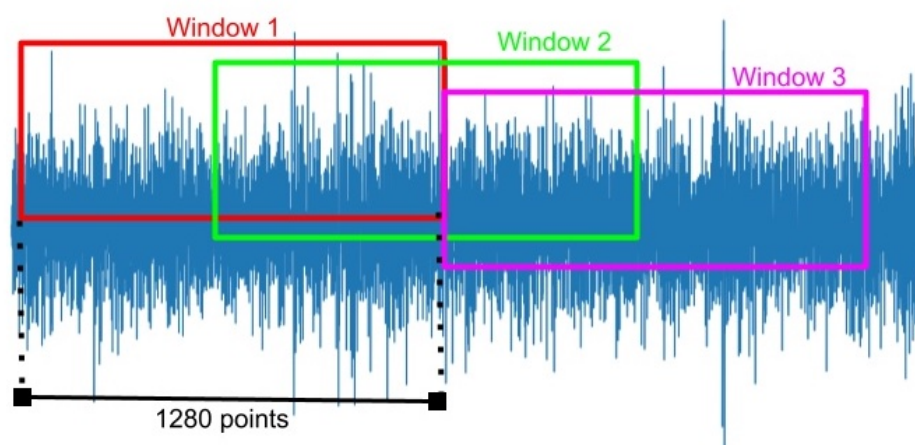


Figura 3 – Amostragem em Épocas com sobreposição de 50%
Fonte: Autor.

Para analisar cada configuração de sobreposição, 72 modelos foram desenvolvidos, um para cada paciente no conjunto de dados. A Tabela 6 mostra os resultados da acurácia média para cada tipo de sobreposição, bem como o valor original dos modelos sem sobreposição.

Ao analisar os resultados apresentados na Tabela 6, pode-se verificar que todos os tipos de sobreposição sugeridos são capazes de melhorar o resultado final da classificação. Esta melhoria advém da quantidade maior de dados produzidos, portanto apresentando mais dados às redes desenvolvidas. Contudo a melhoria é limitada; é perceptível que não houve melhoria quando aumentou-se a sobreposição de 50% para 80%, indicando um limite para a aplicação desta técnica nesta tarefa. Esse resultado alinha-se com o fato de que técnicas de *data augmentation* podem ser limitadas, especialmente se, como no caso desse conjunto de dados, a quantidade de amostras a serem aumentadas também for pequena.

Tabela 6 – Desempenho da amostragem em épocas na SeizNet Modificada usando diferentes taxas de sobreposição

Sem sobreposição	20%	50%	80%
87,9%	88,0%	89,0%	89,0%

Fonte: Autor.

5.4 Resultados de Modelos Intra-paciente

A abordagem intra-pacientes é utilizada por um dos primeiros trabalhos na área de detecção de ataques epiléticos (SHOEB; GUTTAG, 2010) e, portanto, representa um importante fator a ser analisado no desenvolvimento de uma rede neural para este problema. A SeizNet original foi testada em um conjunto de dados privado e o resultado do artigo de (2020) está contido neste trabalho e demonstra a sua utilização no conjunto de dados CHB-MIT. Este trabalho também realizou experimentos para verificar o resultado quando é utilizada a versão modificada da SeizNet.

Para o desenvolvimento dos modelos intra-paciente, parte dos dados de um paciente são reservados para teste e validação, enquanto o restante é inserido na base de dados de treinamento juntamente com os dados dos outros $N-1$ pacientes. Assim a rede neural desenvolvida conhece os padrões de onda daquele paciente e poderá identificar novos padrões apresentados por ele. Para a realização do experimento, 24 modelos foram criados seguindo a separação de dados descrita anteriormente. A Tabela 7 mostra a acurácia média de cada modelo desenvolvido.

Os trabalhos desenvolvidos por Avcu, Zhang e Chan (2019) e Sopic, Aminifar e Atienza (2018), também considerando a aplicação intra-paciente do problema, apresentaram resultados similares ou piores aos aqui apresentados. A taxa de acerto média para o trabalho de Avcu, Zhang e Chan (2019) ficou em 96% e para o de Sopic, Aminifar e Atienza (2018) em 93%, enquanto a taxa média deste trabalho na aplicação intra-paciente chegou a 98%. Além disso, a acurácia média aqui apresentada também supera o valor sugerido no trabalho de Albuquerque e Silva (2020) com a implementação original da SeizNet.

5.5 Efeitos do Número de Canais Usados

A influência do número de canais, a princípio, parece clara: a maior quantidade de canais representa mais dados para a rede e, portanto, maior taxa de acerto. Contudo faz-se necessária uma análise mais aprofundada, já que a maior quantidade de canais também representa uma maior quantidade de memória consumida. Além disso, a complexidade de um possível dispositivo vestível também está relacionada à quantidade de canais que serão utilizados para coletar dados do usuário.

Os dados de todos os pacientes foram combinados em um único conjunto de dados, e este foi separado em conjuntos de treinamento, teste e validação. A partir desses dados, três modelos foram desenvolvidos, com variadas quantidade de canais: 2, 13 e 18. Essas quantidades representam o mínimo, um ponto médio e o máximo de canais disponíveis. A Tabela 8 mostra a acurácia média exibida por cada um dos modelos, bem como o desvio padrão (STDEV), já que utilizou-se uma técnica de validação cruzada com 10 rodadas. A

Tabela 7 – Resultado dos Modelos Intrapacientes na SeizNet Modificada

Paciente	Resultado	Desvio Padrão
1	99,543%	1,31
2	99,434%	1,22
3	98,865%	1,18
4	99,356%	1,16
5	99,667%	0,88
6	98,735%	1,28
7	99,604%	1,55
8	98,874%	1,36
9	96,389%	1,35
10	99,526%	1,29
11	99,435%	1,04
12	98,237%	1,62
13	98,316%	1,60
14	97,572%	1,33
15	98,114%	1,44
16	96,764%	1,51
17	95,598%	1,64
18	91,468%	1,53
19	97,855 %	1,55
20	96,962%	1,64
21	99,691%	1,74
22	96,744%	1,44
23	96,753%	1,78
24	99,397%	1,56
Média	98,037%	1,42

Fonte: Autor.

acurácia cresceu em mais de 10 pontos percentuais quando o número de canais aumentou de 2 para 13. Este aumento não foi visualizado na outra faixa, o que indica que existe um limite ao próximo de 13 canais onde a maior quantidade de dados não tem um impacto proporcional ao resultado final do classificador.

Tabela 8 – Acurácia média na rede SeizNet Modificada para múltiplos canais

	Acurácia Média	STDEV
2 Canais	81,471%	1,92
13 Canais	97,753%	1,25
18 Canais	98,037%	1,42

Fonte: Autor.

A partir deste resultado, pode-se concluir que a utilização de 13 canais pode ser suficiente para obter um classificador com bons resultados. Entretanto, de forma a melhor ilustrar a utilização de todos os dados disponíveis, todos os canais foram usados nos outros experimentos executados.

5.6 Efeitos da Compressão de Pesos

O consumo de memória foi avaliado diretamente em um microcontrolador candidato, já que o objetivo final seria o desenvolvimento de um dispositivo vestível que provavelmente utilizará baterias. A ferramenta *Cube MX* (STMICROELECTRONICS, 2022) foi utilizada para desenvolver o algoritmo em C++ e transferi-lo para a plataforma embarcada. Neste experimento, a quantidade de memória RAM usada pelos modelos foi a mesma, independente do número de canais utilizados. Isso é esperado, visto que o formato da rede (e portanto os pesos) será o mesmo para todas as quantidades de canais. A diferença está na quantidade de RAM utilizada para a entrada de dados na rede, que varia de acordo com o número de canais. A Tabela 9 mostra o resumo da quantidade de RAM usada para cada quantidade de canais analisada. A própria ferramenta *CubeMX* informa a quantidade de RAM que será usada na aplicação. É importante notar que os dados apresentados nesta seção são diretamente relacionados ao formato da rede utilizada, SeizNet Modificada.

Tabela 9 – Uso da RAM pela plataforma STM32F746 para a SeizNet Modificada

	Modelo	Entrada
2 canais	59,56Kb	10,00Kb
13 canais	59,56Kb	65,00Kb
18 canais	59,56Kb	90,00Kb

Fonte: Autor.

Ao embarcar o algoritmo e a rede no microcontrolador, a análise da compressão dos pesos, promovida pela própria plataforma, também foi conduzida de forma a analisar o impacto desta ferramenta no resultado final da classificação. A compressão só poderia ser realizada nos dados salvos na memória *flash*, já que estes são considerados dados estáticos e poderiam ser comprimidos durante o processo de transferência para a plataforma. O algoritmo utilizado foi o de compartilhamento de pesos proposto por Nowlan e Hinton (1992). Dois níveis de compressão foram utilizados (compressão de quatro e oito vezes), além do nível sem compressão onde os pesos são transferidos utilizando otimizações padrão (sem algoritmo de otimização de pesos).

Esta avaliação impacta diretamente a utilização de energia pela plataforma, já que a quantidade de memória *flash* usada é um fator que influencia no consumo de energia pelo SoC (*System on Chip*). A Tabela 10 mostra o consumo de memória *flash* considerando os diferentes níveis de compressão oferecidos pela plataforma. Esta é a memória utilizada apenas pela rede e não inclui a memória utilizada pelo código necessário para a sua execução. É importante notar que o resultado da classificação permaneceu praticamente o mesmo com o uso de diferentes níveis de compressão, apenas os modelos que utilizaram 2 canais apresentaram maior sensibilidade ao processo com pequenas mudanças na acurácia média calculada (menos de 2%).

Tabela 10 – Compressão de Memória Flash em diferentes níveis - SeizNet Modificada

	1x	4x	8x
2 canais	604,25Kb	474,00Kb	451,19Kb
13 canais	607,69Kb	477,44Kb	454,63Kb
18 canais	609,45Kb	479,20Kb	456,39Kb

Fonte: Autor.

O bom resultado de classificação, aliado a alta compressão disponível, permite a implementação de uma rede neural convolucional utilizando menos de 500 kB de *flash* e 70 kB de RAM (para 2 canais), o que se encaixa nas especificações de uma grande gama de microcontroladores comercialmente disponíveis.

Os trabalhos anteriores não realizam a análise do consumo de memória das suas soluções. Por exemplo, em Avcu, Zhang e Chan (2019), a rede foi executada em uma estação de trabalho utilizando linguagens de alto nível, enquanto a abordagem embarcada aqui é otimizada para rodar em plataformas com poucos recursos de memória.

5.7 Resultados de Modelos Inter-pacientes

Finalmente avaliamos a taxa de acerto da classificação utilizando a abordagem inter-pacientes. Foi desenvolvido um modelo para cada paciente utilizando a técnica *leave-one-patient-out*, onde o modelo é treinado com os dados de $N - 1$ pacientes e todos os dados do último paciente são utilizados para testar o modelo.

A Tabela 11 mostra o resultado para cada paciente quando utilizamos todos os dados disponíveis para teste ou quando utilizamos subamostragem da classe majoritária (como foi explicado na seção 3). A utilização de todos os dados disponíveis tem a vantagem de conferir ao experimento as mesmas condições de um evento real.

Ao comparar os resultados apresentados com aqueles demonstrados por Hossain et al. (2019), a diferença é de menos de 3%, o que põe o método aqui apresentado no nível de estado da arte com a vantagem de ser um método melhor explicado e facilmente reproduzível. Além disso, como demonstrado na seção anterior, a SeizNet Modificada pode também ser aplicada à técnicas intra-paciente com resultados semelhantes à versão original da SeizNet (AVCU; ZHANG; CHAN, 2019).

Por fim, na Tabela 12 mostramos a taxa de acerto média encontrada para cada um dos passos do método apresentado aqui, considerando o objetivo final de classificar amostras inter-pacientes.

Tabela 11 – Desempenho da rede SeizNet Modificada na tarefa inter-paciente

ID do Paciente	Com Subamostragem	D. Padrão	Sem Subamostragem	D. Padrão
1	99,2%	1,77	97,8%	1,99
2	98,4%	1,61	94,9%	1,36
3	100,0%	1,74	96,3%	2,18
4	94,1%	1,35	96,6%	1,91
5	99,2%	2,07	97,0%	2,06
6	85,9%	3,03	97,8%	3,12
7	100,0%	1,97	95,6%	1,68
8	92,2%	2,34	96,8%	2,13
9	100,0%	2,70	95,5%	2,38
10	93,8%	1,54	61,4%	1,86
11	98,9%	1,90	97,2%	1,99
12	76,6%	1,46	96,1%	2,22
13	78,1%	2,06	98,3%	2,08
14	89,8%	2,46	96,5%	4,51
15	91,9%	1,76	95,4%	2,04
16	96,9%	1,28	96,8%	2,50
17	99,2%	1,46	96,2%	1,38
18	100,0%	1,98	98,8%	2,04
19	100,0%	1,97	99,1%	1,95
20	92,2%	2,12	97,8%	2,49
21	100,0%	3,22	99,2%	2,57
22	100,0%	2,00	97,9%	1,88
23	99,2%	2,31	97,4%	2,59
24	92,2%	2,03	96,9%	2,63
Média	94,9%	2,01	95,6%	2,23

Fonte: Autor.

Tabela 12 – Desempenho de classificação para cada passo do método na tarefa inter-paciente

Passo	Performance
SeizNet	83,3%
SeizNet Modificada	87,9%
SeizNet Modificada com Filtro e Subamostragem	94,9%
SeizNet Modificada com Filtro sem Subamostragem	95,6%

Fonte: Autor.

5.8 Análise Embarcada

Para a análise embarcada, uma proposta de protótipo foi desenvolvida. A tabela 13 mostra o fluxo de eventos e onde eles devem acontecer.

Tabela 13 – Eventos do protótipo

Local	Evento
Dispositivo na cabeça	Coleta de Dados
Dispositivo na cabeça/ MCU	Pré-processamento
MCU	Preparação dos dados
MCU	Inferência

Fonte: Autor.

O fluxo de dados do protótipo se dá em duas etapas, em um dispositivo anexo na cabeça do usuário ou no microcontrolador; a imagem destaca onde cada passo do processo acontece. Para a coleta de dados, um dispositivo precisa ser utilizado pelo usuário, um tipo de chapéu ou fita pode ser necessário para conter o aparato. O tamanho dele está diretamente relacionado à quantidade de eletrodos que serão utilizados para coletar os dados. Além disso, o passo de tratamento inicial dos dados (normalização e filtragem) pode ser realizado nesse dispositivo. Os dados captados pelo dispositivo auxiliar serão preparados (divididos em épocas de acordo com o método do capítulo 3) e, por fim, serão inseridos na rede para inferência que emitirá um resultado.

Um outro objeto que pode ser usado são óculos comuns, como foi demonstrado por Sopic, Aminifar e Atienza (2018). Quatro eletrodos foram utilizados no trabalho mencionado. Isso reduzirá a capacidade de classificação da rede, mas deixará o dispositivo mais confortável para a utilização do usuário final. Seguindo o sistema 10-20 de montagem de eletrodos em EEG, os eletrodos T3 até T6, F7 e F8 poderão ser utilizados para capturar as informações necessárias em óculos comuns. Se um chapéu for utilizado, as áreas F, C e P também poderão ser consideradas. A Figura 4 mostra o sistema 10-20, os nomes dos eletrodos e suas localizações.

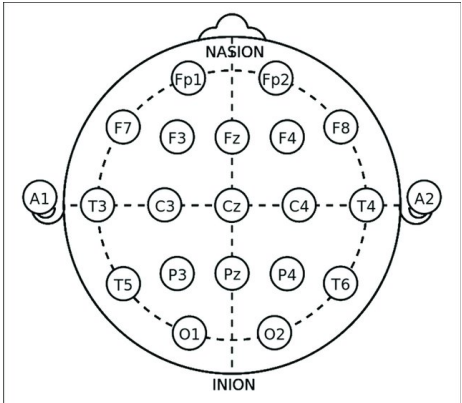


Figura 4 – Sistema internacional 10-20 para montagem de EEG
Fonte: Rojas et al. (2018).

O microcontrolador STM32F746ZG tem três conversores AD (ADC) de 13 bits com 24 canais cada. A frequência de amostragem é de 2,4 MPS, o que é mais que suficiente para amostrar os 256 Hz necessários das ondas emitidas pelo usuário. Portanto os sinais dos eletrodos podem ser conectados ao ADC e amostrados na frequência desejada. A voltagem produzida por eles é entre $10\mu\text{V}$ e $100\mu\text{V}$, então será necessário amplificar os sinais antes de serem submetidos ao ADC.

O ciclo de recarga da bateria depende do período de amostragem da aplicação. Para um simples diagnóstico, por exemplo, um período de 24 horas pode se utilizado de forma a coletar e validar os dados de EEG. O paciente deverá utilizar o dispositivo e os dados antes e durante os ataques podem ser coletados para avaliações mais profundas pelo médico. É estimado que para uma leitura a cada minuto o dispositivo pode operar por 60 horas com uma bateria de 300mAh(dimensões 4x20x40mm e 8g), que pode ser utilizado no dispositivo vestível e não influenciará exageradamente no peso geral do dispositivo.

Tabela 14 – Consumo médio e valor de microprocessadores comerciais

	Preço (USD)	Consumo (mW)
ARM M7 - F745	20,2	465,0
ARM A53 - RBP	35,0	800,0
ARM M4 - F407	19,9	175,0
FPGA Spartan7	99,0	-

Fonte: Autor.

Uma comparação foi feita entre o microcontrolador usado nos experimentos e algumas plataformas comerciais disponíveis usadas na literatura para tarefas envolvendo *deep learning*. Os aspectos de valor e consumo de energia foram avaliados. Ao considerar um dispositivo vestível, a menor quantidade de consumo aliado ao melhor preço são fatores chave a serem considerados. A Tabela 14 mostra os valores coletados. A plataforma usada nos experimentos é uma das mais baratas disponíveis com poder de computação suficiente para embarcar uma rede convolucional completa.

Finalmente, a Tabela 15 demonstra o tempo médio para executar uma inferência quando se tem todos os dados disponíveis para tanto na plataforma embarcada. Assim, analisando o ciclo de recarregamento de uma bateria para protótipo, sabe-se que esta dependerá da quantidade de amostras que serão recolhidas do usuário e do tempo de execução da aplicação. Em um estimativa para uma leitura a cada minuto (tempo suficiente para executar uma inferência na rede embarcada) o dispositivo seria capaz de operar por 60 horas usando uma bateria de 300mAh (tamanho 4x20x40mm e 8 gramas), o que poderia ser incorporado ao dispositivo vestível sem prejudicar o seu peso total.

Tabela 15 – Tempo de execução da rede neural na plataforma embarcada usando a SeizNet Modificada

2 Canais	13 Canais	18 Canais
27,2ms	35,8ms	37,4ms

Fonte: Autor.

6 Conclusão

O objetivo principal deste trabalho foi demonstrar as análises necessárias para o desenvolvimento de um método de detecção de ataques epiléticos usando técnicas de *deep learning*. Três arquiteturas foram analisadas: SeizNet, Resnet e SqueezeNet. As análises consideravam o desempenho geral e a quantidade de memória utilizada. Uma das arquiteturas, SeizNet, foi escolhida como *baseline* de uma série de melhorias que ajudariam a melhorar os parâmetros anteriormente mencionados. A arquitetura original foi desenvolvida para a abordagem intra-paciente e não apresentou resultados satisfatórios para a abordagem inter-paciente. Foi apresentada uma versão modificada da SeizNet que apresentou um melhor desempenho geral quando foi aplicada ao problema inter-pacientes.

Os resultados mostraram que a SeizNet Modificada consegue atingir resultados comparáveis ao estado da arte com pouca utilização de memória, especialmente em condições reais, com um conjunto de dados não balanceado. Ao comparar os resultados para a tarefa intra-pacientes com a versão original da SeizNet apresentada por (AVCU; ZHANG; CHAN, 2019) podemos ver que a SeizNet modificada apresentou leve melhora nos resultados, além de pode ser aplicada na tarefa inter-paciente.

A CNN conseguiu generalizar para pacientes não conhecidos (aplicação inter-paciente) o que abre uma boa oportunidade para a criação de dispositivos vestíveis que podem utilizar dados de vários pacientes. Ao compará-la com os resultados apresentados por (HOSSAIN et al., 2019), podemos verificar que a SeizNet modificada apresentou resultados semelhantes, com poucos pontos percentuais de diferença e com a vantagem de ter o método facilmente reproduzível.

A utilização da amostragem em épocas com sobreposição de 50% melhorou o desempenho geral dos modelos criados. Os resultados também destacaram a importância da utilização de técnicas de filtragem nos dados de entrada para melhorar a acurácia média dos modelos. A combinação de filtros e sobreposição produziu modelos com resultados comparáveis ao estado da arte na resolução do problema inter-pacientes.

Para aprofundar as descobertas apresentadas neste trabalho, sugerimos a realização de experimentos e pesquisas nas seguintes áreas:

- **Criação do protótipo embarcado:** A criação de um dispositivo final que possa ler os sinais de EEG, pré-processá-los e por fim executar uma inferência usando os modelos treinados no PC.
- **Criação de uma abordagem com redes adaptativas:** O desempenho da rede selecionada não foi ideal no caso de alguns pacientes, portanto é necessário explorar abordagens que possam beneficiar diferentes pacientes ao mesmo tempo.

Por exemplo, uma abordagem com redes treinadas de forma diferente executando em paralelo pode ser suficiente para compensar a queda no desempenho.

- **Outros tipos de pré-processamento:** O pré-processamento utilizando a filtragem e geração de dados com sobreposição foi essencial para a melhoria dos resultados, mas outros tipos de pré-processamento poderiam ser utilizados, como: verificação da importância de cada canal utilizado e utilização de bandas de frequência mais adequadas para certos tipos de ataque.
- **Utilização de outros conjuntos de dados:** O conjunto de dados CHB-MIT é um dos mais conhecidos na aplicação de ataques epiléticos. Contudo existem outros conjuntos de dados e uma exploração da utilização deles juntos pode ajudar a esclarecer outras melhorias necessárias na rede desenvolvida. Por exemplo, o conjunto de dados da *Temple University* (OBEID; PICONE, 2016) possui uma grande quantidade de dados de EEG (mais de 30 mil gravações) com dados de diversos tipos, incluindo ataques epiléticos.

Referências

ABHANG, P. A.; GAWALI, B. W.; MEHROTRA, S. C. Technological basics of EEG recording and operation of apparatus. In: *Introduction to EEG- and Speech-Based Emotion Recognition*. Elsevier, 2016. p. 19–50. Disponível em: <<https://doi.org/10.1016/b978-0-12-804490-2.00002-6>>. Citado na página 35.

AL., T. V. et. *Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015*. 2016. Disponível em: <[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(16\)31678-6](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(16)31678-6)>. Citado na página 21.

ALBUQUERQUE, L. M. de; SILVA, E. T. da. Using deep-learning to automatically detect epileptic seizure in a wearable device. In: *2020 X Brazilian Symposium on Computing Systems Engineering (SBESC)*. IEEE, 2020. Disponível em: <<https://doi.org/10.1109/sbesc51047.2020.9277844>>. Citado 2 vezes nas páginas 36 e 46.

AVCU, M. T.; ZHANG, Z.; CHAN, D. W. S. Seizure detection using least eeg channels by deep convolutional neural network. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019. Disponível em: <<https://doi.org/10.1109/icassp.2019.8683229>>. Citado 8 vezes nas páginas 31, 35, 36, 42, 44, 46, 49 e 55.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. 1. ed. New York, NY: Springer, 2006. (Information Science and Statistics). Citado na página 27.

BLOUNT, J. P. et al. Advances in intracranial monitoring. *Neurosurgical Focus*, Journal of Neurosurgery Publishing Group (JNSPG), v. 25, n. 3, p. E18, set. 2008. Disponível em: <<https://doi.org/10.3171/foc/2008/25/9/e18>>. Citado na página 26.

CHANG, B.; LOWENSTEIN, D. Epilepsy. *New England Journal of Medicine*, v. 349, p. 1257–1266, 2003. Citado 2 vezes nas páginas 21 e 25.

CHOLLET, F. et al. *Keras*. 2015. <<https://keras.io>>. Citado na página 37.

DUNN, J.; RUNGE, R.; SNYDER, M. Wearables and the medical revolution. *Personalized Medicine*, Future Medicine Ltd, v. 15, n. 5, p. 429–448, sep 2018. Disponível em: <<https://doi.org/10.2217/2Fpme-2018-0044>>. Citado na página 30.

ELGER, C.; SCHMIDT, D. Modern management of epilepsy: A practical approach. *Epilepsy Behavior*, v. 4, p. 501–539, 2008. Citado 2 vezes nas páginas 21 e 25.

GODOY, R. V. de. *Predição de crises epiléticas utilizando técnicas de Deep Learning*. Tese (Doutorado). Disponível em: <<https://doi.org/10.11606/d.18.2021.tde-31012022-114038>>. Citado na página 32.

GOLDBERGER, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, Ovid Technologies (Wolters Kluwer Health), v. 101, n. 23, jun. 2000. Disponível em: <<https://doi.org/10.1161/01.cir.101.23.e215>>. Citado na página 33.

- GOTMAN, J. Automatic recognition of epileptic seizures in the EEG. *Electroencephalography and Clinical Neurophysiology*, Elsevier BV, v. 54, n. 5, p. 530–540, nov. 1982. Disponível em: <[https://doi.org/10.1016/0013-4694\(82\)90038-4](https://doi.org/10.1016/0013-4694(82)90038-4)>. Citado na página 26.
- HAYKIN, S. O. *Neural networks*. 2. ed. Upper Saddle River, NJ: Pearson, 1998. Citado na página 27.
- HE, K. et al. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. Disponível em: <<http://arxiv.org/abs/1512.03385>>. Citado na página 36.
- HOSSAIN, M. S. et al. Applying deep learning for epilepsy seizure detection and brain mapping visualization. *ACM Transactions on Multimedia Computing, Communications, and Applications*, Association for Computing Machinery (ACM), v. 15, n. 1s, p. 1–17, fev. 2019. Disponível em: <<https://doi.org/10.1145/3241056>>. Citado 3 vezes nas páginas 35, 49 e 55.
- IANDOLA, F. N. et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. *CoRR*, abs/1602.07360, 2016. Disponível em: <<http://arxiv.org/abs/1602.07360>>. Citado na página 37.
- JAMESON, J. L. et al. *Harrison's principles of internal medicine, twentieth edition (vol.1 & vol.2)*. 20. ed. Columbus, OH: McGraw-Hill Education, 2018. Citado na página 25.
- KELLY, J. T. et al. The internet of things: Impact and implications for health care delivery. *Journal of Medical Internet Research*, JMIR Publications Inc., v. 22, n. 11, p. e20135, nov. 2020. Disponível em: <<https://doi.org/10.2196/20135>>. Citado na página 30.
- KINGMA, D. P.; BA, J. *Adam: A Method for Stochastic Optimization*. 2014. Citado na página 37.
- KUMAR, J. S.; BHUVANESWARI, P. Analysis of electroencephalography (EEG) signals and its categorization—a study. *Procedia Engineering*, Elsevier BV, v. 38, p. 2525–2536, 2012. Disponível em: <<https://doi.org/10.1016/j.proeng.2012.06.298>>. Citado na página 34.
- LECUN, Y.; BENGIO, Y. Convolutional networks for images, speech, and time series. In: _____. *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1998. p. 255–258. ISBN 0262511029. Citado na página 28.
- MARWEDEL, P. *Embedded system design*. 4. ed. Cham, Switzerland: Springer Nature, 2022. (Embedded Systems). Citado na página 29.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, Springer Science and Business Media LLC, v. 5, n. 4, p. 115–133, dez. 1943. Disponível em: <<https://doi.org/10.1007/bf02478259>>. Citado na página 27.
- MINISTÉRIO DA SAÚDE. *Protocolos Clínicos e Diretrizes Terapêuticas. v. 1. Brasília: Ministério da Saúde*. 2013. Citado na página 21.
- MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of machine learning*. London, England: MIT Press, 2012. (Adaptive Computation and Machine Learning series). Citado na página 27.

- MUHAMMAD, G. et al. Automatic seizure detection in a mobile multimedia framework. *IEEE Access*, Institute of Electrical and Electronics Engineers (IEEE), v. 6, p. 45372–45383, 2018. Disponível em: <<https://doi.org/10.1109/access.2018.2859267>>. Citado 2 vezes nas páginas 38 e 44.
- NOWLAN, S. J.; HINTON, G. E. Simplifying neural networks by soft weight-sharing. *Neural Computation*, v. 4, n. 4, p. 473–493, 1992. Citado na página 48.
- OBEID, I.; PICONE, J. The temple university hospital EEG data corpus. *Frontiers in Neuroscience*, Frontiers Media SA, v. 10, maio 2016. Disponível em: <<https://doi.org/10.3389/fnins.2016.00196>>. Citado na página 56.
- QU, H.; GOTMAN, J. A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device. *IEEE Transactions on Biomedical Engineering*, Institute of Electrical and Electronics Engineers (IEEE), v. 44, n. 2, p. 115–122, 1997. Disponível em: <<https://doi.org/10.1109/10.552241>>. Citado na página 26.
- RAAD, H. *Fundamentals of IoT and wearable technology design*. Hoboken, NJ: Wiley-Blackwell, 2021. Citado na página 29.
- ROJAS, G. M. et al. Study of resting-state functional connectivity networks using EEG electrodes position as seed. *Frontiers in Neuroscience*, Frontiers Media SA, v. 12, abr. 2018. Disponível em: <<https://doi.org/10.3389/fnins.2018.00235>>. Citado 2 vezes nas páginas 33 e 51.
- SCHERER, D.; MÜLLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. In: *Artificial Neural Networks – ICANN 2010*. Springer Berlin Heidelberg, 2010. p. 92–101. Disponível em: <https://doi.org/10.1007/978-3-642-15825-4_10>. Citado na página 28.
- SCHOMER, D. L.; SILVA, F. Lopes da (Ed.). *Niedermeyer's electroencephalography*. 6. ed. Philadelphia, PA: Lippincott Williams and Wilkins, 2010. Citado na página 26.
- SHOEB, A.; GUTTAG, J. Application of machine learning to epileptic seizure detection. In: . [S.l.: s.n.], 2010. p. 975–982. Citado 4 vezes nas páginas 25, 26, 31 e 46.
- SOPIC, D.; AMINIFAR, A.; ATIENZA, D. e-glass: A wearable system for real-time detection of epileptic seizures. In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018. Disponível em: <<https://doi.org/10.1109/iscas.2018.8351728>>. Citado 3 vezes nas páginas 32, 46 e 51.
- STAFSTROM, C. E.; CARMANT, L. Seizures and epilepsy: An overview for neuroscientists. *Cold Spring Harbor Perspectives in Medicine*, Cold Spring Harbor Laboratory, v. 5, n. 6, p. a022426–a022426, jun. 2015. Disponível em: <<https://doi.org/10.1101/cshperspect.a022426>>. Citado na página 25.
- STMICROELECTRONICS. *STMCubeMX*. 2022. <<https://www.st.com/en/development-tools/stm32cubemx.html>>. Citado na página 48.
- TSIPOURAS, M. G. Spectral information of EEG signals with respect to epilepsy classification. *EURASIP Journal on Advances in Signal Processing*, Springer Science and Business Media LLC, v. 2019, n. 1, fev. 2019. Disponível em: <<https://doi.org/10.1186/s13634-019-0606-8>>. Citado 2 vezes nas páginas 35 e 43.

- VENIERIS, S. I.; KOURIS, A.; BOUGANIS, C.-S. Toolflows for mapping convolutional neural networks on FPGAs. *ACM Computing Surveys*, Association for Computing Machinery (ACM), v. 51, n. 3, p. 1–39, jun. 2018. Disponível em: <<https://doi.org/10.1145/3186332>>. Citado na página 32.
- WANG, S. et al. High-throughput CNN inference on embedded ARM big.LITTLE multi-core processors. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Institute of Electrical and Electronics Engineers (IEEE), p. 1–1, 2019. Disponível em: <<https://doi.org/10.1109/tcad.2019.2944584>>. Citado na página 32.
- WHITTINGSTALL, K.; LOGOTHETIS, N. K. Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron*, Elsevier BV, v. 64, n. 2, p. 281–289, out. 2009. Disponível em: <<https://doi.org/10.1016/j.neuron.2009.08.016>>. Citado na página 26.
- WOLF, M. *Computers as components*. 3. ed. Oxford, England: Morgan Kaufmann, 2012. (The Morgan Kaufmann Series in Computer Architecture and Design). Citado na página 29.
- ZHOU, M. et al. Epileptic seizure detection based on EEG signals and CNN. *Frontiers in Neuroinformatics*, Frontiers Media SA, v. 12, dez. 2018. Disponível em: <<https://doi.org/10.3389/fninf.2018.00095>>. Citado na página 31.
- ZIYABARI, S. et al. Objective evaluation metrics for automatic classification of EEG events. *CoRR*, abs/1712.10107, 2017. Disponível em: <<http://arxiv.org/abs/1712.10107>>. Citado 2 vezes nas páginas 35 e 44.