

**UNIVERSIDADE ESTADUAL PAULISTA**  
**"JÚLIO DE MESQUITA FILHO"**

Faculdade de Ciências  
Câmpus Bauru

PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**Auxílio ao diagnóstico de Acidente Vascular  
Cerebral utilizando Redes de Boltzmann e  
agregação de informação com a transformada de  
Fourier**

Mateus Roder  
Prof. Dr. João Paulo Papa (Orientador)

Bauru  
07/06/2023

Mateus Roder

**Auxílio ao diagnóstico de Acidente Vascular Cerebral  
utilizando Redes de Boltzmann e agregação de informação  
com a transformada de Fourier**

Tese de doutorado, para o curso de Pós-Graduação em Ciência da Computação da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências, Câmpus Bauru.

Banca Examinadora

**Prof. Dr. João Paulo Papa**  
Orientador

**Prof. Dr. Kelton Augusto Pontara da Costa**

**Prof. Dr. Aparecido Nilceu Marana**

**Prof. Dr. Alexandre Luis Magalhães Levada**

**Prof. Dr. Ricardo Cerri**

Bauru, 07 de Junho de 2023.

R688a Roder, Mateus  
Auxílio ao diagnóstico de Acidente Vascular Cerebral  
utilizando Redes de Boltzmann e agregação de informação com  
a transformada de Fourier / Mateus Roder. -- Bauru, 2023  
75 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),  
Faculdade de Ciências, Bauru  
Orientador: João Paulo Papa

1. Inteligência artificial. 2. Redes neurais (Computação). 3.  
Inteligência artificial Aplicação médica. 4. Acidente vascular  
cerebral. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da  
Faculdade de Ciências, Bauru. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

**ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE MATEUS RODER, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO, DA FACULDADE DE CIÊNCIAS - CÂMPUS DE BAURU.**

Aos 07 dias do mês de junho do ano de 2023, às 14:00 horas, por meio de Videoconferência, realizou-se a defesa de TESE DE DOUTORADO de MATEUS RODER, intitulada **Auxílio ao diagnóstico de Acidente Vascular Cerebral utilizando Redes de Boltzmann e agregação de informação com a transformada de Fourier**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. JOAO PAULO PAPA (Orientador - Participação Presencial) do Departamento de Computação/FC/UNESP/Bauru (SP), Prof. Dr. KELTON AUGUSTO PONTARA DA COSTA (Participação Presencial) do Departamento de Computação/FC/UNESP/Bauru (SP), Prof. Dr. APARECIDO NILCEU MARANA (Participação Presencial) do Departamento de Computação/FC/UNESP/Bauru (SP), Prof. Dr. ALEXANDRE LUIS MAGALHÃES LEVADA (Participação Virtual) do Departamento de Computação/UFSCar/São Carlos (SP), Prof. Dr. RICARDO CERRI (Participação Virtual) do Departamento de Computação/UFSCar/ São Carlos (SP). Após a exposição pelo doutorando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final: APROVADO. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo Presidente da Comissão Examinadora.



Prof. Dr. JOAO PAULO PAPA

# Agradecimentos

Agradeço inicialmente a Deus, que me guia e possibilita o meu desenvolvimento, de forma ainda incompreensível, sobre tantas coisas na vida, assim como esta etapa.

Ao meu orientador, João Paulo Papa, pela amizade, pelas diversas orientações profissionais e pessoais, e pela oportunidade de trabalhar ao seu lado.

À FUSP, que em grande parte do meu doutorado me auxiliou financeiramente, como agência de fomento.

Agradeço aos meus amigos, que estiveram e estão ao meu lado, total ou parcialmente, durante o desenvolvimento desta e de outras pesquisas/atividades.

Aos meus pais, que sempre torceram pela minha trilha acadêmica, e me ajudaram à forma deles.

À minha amada esposa, Ana Carolina, que acompanhou atenta e ativamente essa jornada, e me apoiou da melhor forma possível em todas as decisões que tomei.

Por fim, agradeço também a mim, por não desistir da academia, e por abraçar causas que podem salvar vidas.

*Disse-lhe Jesus: Eu sou o caminho, a verdade e a vida; ninguém vem ao Pai, senão por mim.*  
*(João 14:6)*

# Resumo

Na última década, técnicas de aprendizado profundo (do inglês, *Deep Learning* - DL) têm sido amplamente pesquisadas e aplicadas a diversos problemas, abrangendo uma vasta gama que vai desde sistemas de recomendação de produtos para consumidores, até aplicações na área médica, como a classificação de acidente vascular cerebral (AVC), e do pré-processamento de imagens médicas. No entanto, mesmo com excelentes resultados nas tarefas de visão computacional mencionadas, ainda há áreas a serem exploradas à medida que os problemas tornam-se cada vez mais complexos, ou as aplicações do mundo real exigem novas restrições que prejudicam/limitam o desempenho das técnicas estado-da-arte. Com relação à classificação de acidente vascular cerebral, um problema de suprimento sanguíneo cerebral, pode-se observar algumas dificuldades nas pesquisas acadêmicas realizadas, como a falta de grandes conjuntos de dados para treinar redes neurais profundas, falta de imagens anotadas para utilizar o paradigma de aprendizado supervisionado, e a dificuldade em encontrar trabalhos com redes neurais de baixa complexidade, como as Máquinas de Boltzmann Restritas. Deste modo, o presente trabalho atua no estudo e desenvolvimento de modelos baseados em Máquinas de Boltzmann aplicados no contexto da classificação de acidente vascular cerebral, empregando redes neurais artificiais treinadas através dos paradigmas de aprendizado não-supervisionado e supervisionado, utilizando um banco de dados público com baixo volume de dados rotulados (principalmente para o ajuste-fino supervisionado), concomitante à transformada de Fourier para adicionar informação relevante aos modelos. Os resultados atingidos foram expressivos e apontam para um novo estado-da-arte do conjunto de dados empregado nesta tese, com taxas de acerto médias iguais a 99,94%. Além disso, a metodologia proposta que adiciona informação multimodal por meio da transformada de Fourier impactou positivamente no desempenho preditivo dos modelos. Por fim, a abordagem empregada utilizando Máquinas de Boltzmann Restritas e Redes de Crença Profundas, bem como suas variantes convolucionais, são mais simples que as estado-da-arte baseadas em redes neurais convolucionais padrão.

**Palavras-chave:** AVC. Máquinas de Boltzmann Restritas. Redes Neurais Convolucionais. Aprendizado Profundo

# Abstract

In the last decade, deep learning techniques (DL) have been widely researched and applied to several problems, covering a wide range that goes from product recommendation systems for consumers to noble applications in the medical field, such as stroke classification and medical image pre-processing. However, even with excellent results in the mentioned computer vision tasks, there are still areas to explore as the problems become more and more complex, or real-world applications prevent new restrictions that harm/limit the performance of the current techniques' state-of-the-art. Regarding the classification of stroke, a problem of cerebral blood supply, some difficulties can be observed in the academic research carried out, such as the lack of large datasets to train Deep Neural Networks, lack of annotated images to use the supervised learning paradigm, and the difficulty in finding works with low complexity neural networks, such as Restricted Boltzmann Machines. In this way, the present thesis acts in the study and development of models based on Boltzmann Machines applied in the context of stroke classification, using artificial neural networks trained through the unsupervised and supervised learning paradigms, employing a public database with a low volume of labeled data (mainly for supervised fine-tuning), concomitantly with Fourier transform to add relevant information to the models. The achieved results were expressive and pointed to a new state-of-the-art for the dataset employed in this thesis, with an average accuracy of 99.94%. In addition, adding multimodal information through the Fourier transform, positively impacting the predictive performance of the models. Finally, the proposed approach with Restricted Boltzmann Machines and Deep Belief Networks, as well as their convolutional variants, is simpler than the state-of-the-art based on standard convolutional neural networks.

**Key-words:** Stroke. Restricted Boltzmann Machines. Convolutional Neural Networks. Deep Learning.

# Lista de ilustrações

|                                                                                                                                                               |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 – Representação dos tipos de AVC. . . . .                                                                                                            | 20 |
| Figura 2 – Representação do fluxo de identificação e classificação do AVC, adaptado de Yew e Cheng (YEW; CHENG, 2015). . . . .                                | 22 |
| Figura 3 – Tomografia computadorizada referente a um acidente vascular cerebral hemorrágico (região destacada). . . . .                                       | 23 |
| Figura 4 – Arquitetura de uma RBM. . . . .                                                                                                                    | 34 |
| Figura 5 – Ilustração do método baseado em divergência contrastiva. . . . .                                                                                   | 36 |
| Figura 6 – Arquitetura padrão de uma DBN. . . . .                                                                                                             | 39 |
| Figura 7 – Exemplo de aplicação de um <i>kernel</i> convolucional de dimensão $3 \times 3$ . . . .                                                            | 40 |
| Figura 8 – Exemplo de uma arquitetura padrão para uma CNN. . . . .                                                                                            | 41 |
| Figura 9 – Exemplo de uma arquitetura CDBN com duas camadas ocultas. . . . .                                                                                  | 43 |
| Figura 10 – Arquitetura MultFRRBM-P com dois tipos de entradas, isto é, a imagem original e o ângulo de fase correspondente a esta a partir da DFT. . . . .   | 46 |
| Figura 11 – Esquema da transformação aplicada para a entrada nos modelos multimodais. . . . .                                                                 | 47 |
| Figura 12 – Arquitetura FCDBN-P com dois tipos de dados de entrada, isto é, as imagens originais e a representação dos ângulos de fase da DFT. . . . .        | 48 |
| Figura 13 – Modelos multimodais resultantes da abordagem proposta. . . . .                                                                                    | 49 |
| Figura 14 – Tomografias do cérebro: saudável, com acidente vascular cerebral hemorrágico e com acidente vascular cerebral isquêmico, respectivamente. . . . . | 51 |
| Figura 15 – RBM alimentada com a fase. . . . .                                                                                                                | 55 |
| Figura 16 – RBM alimentada com a magnitude. . . . .                                                                                                           | 55 |
| Figura 17 – RBM alimentada com dados Gaussianos. . . . .                                                                                                      | 56 |
| Figura 18 – FCDBN alimentada com a fase. . . . .                                                                                                              | 57 |
| Figura 19 – FCDBN alimentada com a magnitude. . . . .                                                                                                         | 58 |
| Figura 20 – CDBN alimentada com os dados originais (CT). . . . .                                                                                              | 59 |
| Figura 21 – Acurácia média sob o conjunto de teste considerando a separação 50/50. . . .                                                                      | 61 |
| Figura 22 – Acurácia média sob o conjunto de teste considerando a separação 75/25. . . .                                                                      | 61 |
| Figura 23 – Acurácia média sob o conjunto de teste considerando a separação 25/75. . . .                                                                      | 63 |
| Figura 24 – Acurácia média sob o conjunto de teste considerando a separação 50/50. . . .                                                                      | 63 |
| Figura 25 – Acurácia média sob o conjunto de teste considerando a separação 75/25. . . .                                                                      | 64 |
| Figura 26 – Acurácia média final no conjunto de teste considerando as três partições empregadas. . . . .                                                      | 64 |

# Lista de tabelas

|                                                                                                                                                                                                                                                                                                                                                                                                                                          |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabela 1 – Resumo dos bancos de dados empregados nos trabalhos. Nome é auto-explicativo (caso possua); Problema (tipo) representa qual a tarefa do trabalho, ou seja, classificação (Class), segmentação (Segm), ou ambas (Segm-Class); Volume representa o volume de dados (Pequeno; Médio; Grande); Específico para AVC significa se a base de dados é específica para AVC ou não; Disponibilidade e Ano são auto-explicáveis. . . . . | 31 |
| Tabela 2 – Configuração de hiper-parâmetros para os modelos baseados em RBMs. . .                                                                                                                                                                                                                                                                                                                                                        | 51 |
| Tabela 3 – Configuração dos hiper-parâmetros para os modelos CDBN com multimodalidade (FCDBN-). . . . .                                                                                                                                                                                                                                                                                                                                  | 52 |
| Tabela 4 – Taxas de acerto médias e seus respectivos desvios padrão para o conjunto de teste considerando a divisão de dados proposta pelos autores do banco de dados. . . . .                                                                                                                                                                                                                                                           | 59 |
| Tabela 5 – Taxas de acerto médias e seus respectivos desvios padrão para o conjunto de teste considerando as divisões de dados propostas. . . . .                                                                                                                                                                                                                                                                                        | 62 |
| Tabela 6 – Análise de parâmetros para os modelos empregados. . . . .                                                                                                                                                                                                                                                                                                                                                                     | 65 |

# Lista de abreviaturas e siglas

|             |                                                                                 |
|-------------|---------------------------------------------------------------------------------|
| AVC         | Acidente Vascular Cerebral                                                      |
| CD          | <i>Contrastive Divergence</i>                                                   |
| CDBN        | <i>Convolutional Deep Belief Networks</i>                                       |
| CT          | <i>Computed Tomography</i>                                                      |
| CNN         | <i>Convolutional Neural Networks</i>                                            |
| CRBM        | <i>Convolutional Restricted Boltzmann Machines</i>                              |
| DBM         | <i>Deep Boltzmann Machines</i>                                                  |
| DBN         | <i>Deep Belief Networks</i>                                                     |
| DFT         | <i>Discrete Fourier Transform</i>                                               |
| DICOM       | <i>Digital Imaging and Communications in Medicine</i>                           |
| DL          | <i>Deep Learning</i>                                                            |
| FC          | <i>Fully-Connected</i>                                                          |
| FCDBN       | <i>Fourier-based Convolutional Deep Belief Network</i>                          |
| FCDBN-M     | <i>Fourier-based Convolutional Deep Belief Network with Magnitude</i>           |
| FCDBN-P     | <i>Fourier-based Convolutional Deep Belief Network with Phase</i>               |
| FCDBN-PM    | <i>Fourier-based Convolutional Deep Belief Network with Phase and Magnitude</i> |
| GaussianRBM | <i>Gaussian Restricted Boltzmann Machines</i>                                   |
| IA          | Inteligência Artificial                                                         |
| ISLES       | <i>Ischemic Stroke Lesion Segmentation</i>                                      |
| MCMC        | <i>Markov Chain Monte Carlo</i>                                                 |
| MRI         | <i>Magnetic Resonance Imaging</i>                                               |
| MultFRRBM   | RBM Multimodal baseada em Fourier                                               |
| MultFRRBM-M | RBM Multimodal com a magnitude da DFT                                           |

MultFRRBM-P    RBM Multimodal com a fase da DFT

MultFRRBM-PM    RBM Multimodal com fase e magnitude da DFT

NIfTI            *Neuroimaging Informatics Technology*

NIHSS           *National Institute of Health Stroke Scale*

RBM            *Restricted Boltzmann Machines*

SGD            *Stochastic Gradient Descent*

# Sumário

|                |                                                                                    |           |
|----------------|------------------------------------------------------------------------------------|-----------|
| <b>1</b>       | <b>INTRODUÇÃO</b>                                                                  | <b>14</b> |
| <b>1.1</b>     | <b>Problema</b>                                                                    | <b>17</b> |
| <b>1.2</b>     | <b>Hipótese e Questões da Pesquisa</b>                                             | <b>18</b> |
| <b>1.3</b>     | <b>Objetivo e Contribuições</b>                                                    | <b>18</b> |
| <b>1.4</b>     | <b>Estrutura da Tese</b>                                                           | <b>19</b> |
| <b>2</b>       | <b>ACIDENTE VASCULAR CEREBRAL E REVISÃO DA LITERATURA</b>                          | <b>20</b> |
| <b>2.1</b>     | <b>Acidente Vascular Cerebral - AVC</b>                                            | <b>20</b> |
| <b>2.2</b>     | <b>Classificação do AVC</b>                                                        | <b>21</b> |
| <b>2.3</b>     | <b>Trabalhos Correlatos à Classificação de AVC utilizando Aprendizado Profundo</b> | <b>23</b> |
| <b>2.4</b>     | <b>Banco(s) de Dados sobre AVC</b>                                                 | <b>28</b> |
| <b>2.4.1</b>   | <b>Sumarização</b>                                                                 | <b>30</b> |
| <b>3</b>       | <b>FUNDAMENTAÇÃO TEÓRICA</b>                                                       | <b>32</b> |
| <b>3.1</b>     | <b>Modelos Baseados em Energia</b>                                                 | <b>33</b> |
| <b>3.1.1</b>   | <b>Máquinas de Boltzmann Restritas</b>                                             | <b>33</b> |
| <b>3.1.1.1</b> | <b>RBM para Dados Contínuos</b>                                                    | <b>37</b> |
| <b>3.1.2</b>   | <b>Redes de Crença em Profundidade</b>                                             | <b>38</b> |
| <b>3.2</b>     | <b>Modelos Convolucionais</b>                                                      | <b>39</b> |
| <b>3.2.1</b>   | <b>Redes Neurais Convolucionais</b>                                                | <b>39</b> |
| <b>3.2.2</b>   | <b>RBM e DBNs Convolucionais</b>                                                   | <b>41</b> |
| <b>3.3</b>     | <b>Transformada de Fourier</b>                                                     | <b>43</b> |
| <b>4</b>       | <b>PROPOSTA</b>                                                                    | <b>45</b> |
| <b>4.1</b>     | <b>RBM e DBNs Multimodais baseadas em Fourier</b>                                  | <b>45</b> |
| <b>5</b>       | <b>METODOLOGIA</b>                                                                 | <b>50</b> |
| <b>5.1</b>     | <b>O Banco de Dados</b>                                                            | <b>50</b> |
| <b>5.2</b>     | <b>Configuração Experimental</b>                                                   | <b>51</b> |
| <b>6</b>       | <b>RESULTADOS EXPERIMENTAIS</b>                                                    | <b>54</b> |
| <b>6.1</b>     | <b>Aprendizado Não-Supervisionado</b>                                              | <b>54</b> |
| <b>6.1.1</b>   | <b>Análise Visual Não-Convolucional</b>                                            | <b>54</b> |
| <b>6.1.2</b>   | <b>Análise Visual Convolucional</b>                                                | <b>56</b> |
| <b>6.2</b>     | <b>Avaliação dos Modelos para Classificação de AVC</b>                             | <b>58</b> |
| <b>6.2.1</b>   | <b>Classificação com Modelos Não-Convolucionais</b>                                | <b>58</b> |

|            |                                                     |           |
|------------|-----------------------------------------------------|-----------|
| 6.2.2      | Classificação com Modelos Convolucionais . . . . .  | 60        |
| <b>6.3</b> | <b>Complexidade dos Modelos Propostos . . . . .</b> | <b>65</b> |
| <b>7</b>   | <b>CONCLUSÃO E PRÓXIMOS PASSOS . . . . .</b>        | <b>67</b> |
| <b>7.1</b> | <b>Publicações Derivadas da Tese . . . . .</b>      | <b>68</b> |
| <b>7.2</b> | <b>Próximos Passos . . . . .</b>                    | <b>68</b> |
|            | <b>REFERÊNCIAS . . . . .</b>                        | <b>70</b> |

# 1 Introdução

Diante dos grandes avanços tecnológicos, a nova era da informação passa a modificar a sociedade em diversos aspectos, fortemente apoiada pela 4<sup>a</sup> Revolução Industrial e pela Inteligência Artificial (IA), caracterizadas pela elevada conectividade entre sistemas ciber-físicos em diversos ambientes cotidianos. Deste modo, diferentes tecnologias são constantemente criadas e integradas na busca pela ampliação do uso destas na sociedade para facilitar tarefas e auxiliar os profissionais especialistas, como sistemas de visão computacional baseados em inteligência artificial, beneficiados pela grande quantidade de dados gerada com o aumento da conectividade entre dispositivos (LI; HOU; WU, 2017).

A visão computacional é uma área que fornece ferramentas importantes para a tomada de decisão baseada em imagens, buscando representações de alto nível do mundo real, possibilitando o uso de características intrínsecas para a execução de tarefas como detecção e/ou classificação de objetos por técnicas de aprendizado de máquina (BISHOP, 1995). Uma vez que o mundo real não é tão simples de ser parametrizado, tais tarefas tornam-se mais complexas quando existem variações ambientais de luminosidade, diferentes perspectivas e/ou planos de captura de imagens, e variações na resolução, dentre outros fatores (LECUN; KAVUKCUOGLU; FARABET, 2010).

Atualmente, a IA abrange majoritariamente técnicas de aprendizado profundo, do inglês *Deep Learning* (DL), possuindo grande destaque e foco de estudo da comunidade científica (LECUN et al., 1998; HINTON; OSINDERO; TEH, 2006; SALAKHUTDINOV; HINTON, 2012; GAUTAM; RAMAN, 2021; FATIMAH; SINGHAL; SINGH, 2022). Estas técnicas baseiam-se no aprendizado hierárquico de características, similar ao processamento visual humano, por meio de abstrações em diferentes níveis (camadas) que auxiliam na extração de características. Algumas das principais técnicas de DL utilizadas são as Redes Neurais Convolucionais, do inglês *Convolutional Neural Networks* (CNNs) (LECUN et al., 1998), e as Máquinas de Boltzmann Profundas, do inglês *Deep Boltzmann Machines* (DBMs) (SALAKHUTDINOV; HINTON, 2012). As CNNs são capazes de modelar as informações hierárquicas, em basicamente três etapas. A primeira, corresponde à aplicação de convoluções no sinal de entrada e diferentes filtros, seguida por uma amostragem do sinal, e por fim um processo de normalização. Porém, sofrem com a necessidade de uma grande quantidade de dados rotulados quando o processo de treinamento é puramente supervisionado. De forma semelhante, as DBMs buscam aprender de forma hierárquica, com ou sem operações de convolução, porém, diferindo no processo de aprendizado com o emprego do paradigma não-supervisionado junto de técnicas como *mean-field* (SALAKHUTDINOV; HINTON, 2012) para possibilitar a inferência entre camadas de neurônios interdependentes.

Como dito anteriormente, os avanços em inteligência artificial e visão computacional concomitante ao desenvolvimento de *hardwares* altamente eficientes, possibilitaram o emprego das técnicas de aprendizado profundo em diversas áreas, como no monitoramento por câmeras de segurança (MOHAMMADI et al., 2016; AFONSO et al., 2018), e na medicina (WONG; FORTINO; ABBOTT, 2020; HU; WHITNEY; GIGER, 2020; BHATT et al., 2020), por exemplo.

Tratando-se de aplicações médicas, um problema de grande importância diz respeito à detecção e classificação de acidente vascular cerebral (AVC). Esta, é uma lesão que atinge abruptamente o tecido cerebral, sendo causada principalmente por alterações no suprimento sanguíneo para uma determinada região do cérebro, resultando em perda ou redução de suas funções. Esta doença cerebrovascular afeta milhões de pessoas em todo o mundo a cada ano, das quais uma quantidade expressiva deve ir a óbito (VIRANI et al., 2020). No entanto, o principal problema relacionado ao AVC não é apenas a mortalidade, pois é acompanhado pelas sequelas em muitos sobreviventes ao longo da vida pós AVC, que são diferentes e complexas. Belasco e Dicini (RANGEL; BELASCO; DICCINI, 2013), concluíram que as sequelas após o AVC levavam os pacientes a dificuldades em atividades triviais. Os autores observaram que 49,6% dos pacientes apresentavam dependência moderada a grave e 49,7% apresentavam sintomas disfóricos ou depressivos. Além disso, o tratamento dessas pessoas representa um custo de bilhões de dólares por ano (VIRANI et al., 2020).

Na tentativa de auxiliar computacionalmente o problema, alguns trabalhos aplicaram técnicas de aprendizado profundo para classificar imagens de pacientes que sofreram um acidente vascular cerebral, uma vez que seu diagnóstico está ligado diretamente às neuroimagens geradas por tomografia computadorizada ou ressonância magnética. Porém, antes da solicitação do exame de neuroimagem, os médicos primeiro analisam os sintomas do paciente, seu histórico (e o familiar) para casos de acidente vascular cerebral, o que desempenha um papel crucial no fluxo de diagnóstico. Uma vez que o paciente foi acometido por um acidente vascular cerebral anteriormente, há uma maior probabilidade de sofrer o acidente novamente. Caso contrário, é realizada uma busca por outros sintomas a partir da família; se corresponderem aos sintomas do AVC, então o médico solicita aos outros membros da equipe (geralmente radiologistas) para gerarem os exames de neuroimagem do paciente.

Uma vez que o diagnóstico e a classificação do AVC estão altamente relacionados com a análise de neuroimagens, diversos autores e equipes se esforçam para auxiliar este processo com o uso de inteligência artificial. Um exemplo é o trabalho de Pereira et al. (PEREIRA et al., 2018), em que os autores realizaram um estudo clínico com pacientes, acometidos ou não pelo AVC, empregando a coleta de imagens de tomografia computadorizada (do inglês, *Computed Tomography* - CT), para desenvolver modelos de aprendizado profundo capazes de classificar se uma imagem de tomografia representa um paciente com AVC (isquêmico ou hemorrágico), ou não. Os autores atingiram altas taxas de acerto para a classificação de AVC (entre 77,33% e 97,20%), entretanto, o tamanho do banco de dados é relativamente

pequeno (aproximadamente 300 imagens), representando uma das grandes dificuldades na área, agravada pelo fato de que a maioria dos pacientes que passam por triagens em centros de saúde não estão sofrendo da lesão cerebral.

Em concordância, alguns autores buscam apoiar o diagnóstico do AVC com a segmentação da lesão cerebral, ou seja, empregam modelos de aprendizado profundo para destacar/segmentar a região de um possível AVC para fornecer informações mais detalhadas a médicos e radiologistas de forma visual, agregada ao exame, para que a tomada de decisão seja mais rápida e assertiva. Pedemonte et al. (PEDEMONTE et al., 2018), por exemplo, propuseram uma rede convolucional modificada (YNet) que utiliza imagens de ressonância magnética como dados de treinamento para segmentar lesões, a partir do aprendizado fracamente supervisionado. O objetivo do estudo foi segmentar as imagens e detectar AVC em pacientes para o correto direcionamento de terapias de tratamento. O trabalho logrou êxito na tarefa, porém, o modelo proposto consome um grande tempo de treinamento, uma vez que a arquitetura empregada possui alta complexidade, o que pode dificultar a aplicação em dispositivos não tão robustos, como celulares e *tablets*, por exemplo. Além disso, foi treinado em um grande banco de dados privado, fato que inviabiliza outros pesquisadores a melhorarem o estado-da-arte desta pesquisa.

Devido ao crescente número de aplicações de aprendizado profundo para o diagnóstico de acidente vascular cerebral, diversos trabalhos foram desenvolvidos nos últimos anos (SARAGIH et al., 2020; RAJINIKANTH; ASLAM; KADRY, 2021; PAN et al., 2021; GAUTAM; RAMAN, 2021; BRIDGE et al., 2022) e, além disso, alguns autores forneceram estudos que resumem diferentes trabalhos e suas principais realizações da área. Por exemplo, Castillo et al. (CASTILLO; LAKSHMINARAYANAN; RODRÍGUEZ-ÁLVAREZ, 2021) forneceram uma revisão robusta da literatura publicada sobre sistemas e algoritmos que permitem a classificação, identificação e detecção de “hiper-intensidades de matéria branca” (do inglês, *white matter hyperintensities*) de imagens de ressonância magnética cerebral. No entanto, os autores se concentram apenas em casos de acidente vascular cerebral isquêmico e compreendem uma análise até o ano de 2020.

Outro exemplo que destaca a importância de aplicações médicas diz respeito ao trabalho de Peng et al. (PENG et al., 2022), que abordaram o problema de segmentação e classificação de AVC. Os autores propuseram dividir as imagens de tomografia computadorizada de baixo contraste (banco de dados privado) em sub-imagens (*patches*) de  $16 \times 16$  *pixels*, e transformá-las em uma matriz *t-score*, que são os novos dados para alimentar uma rede neural convolucional. Os resultados foram promissores, porém as consecutivas transformações aplicadas podem tornar o processo custoso, além de fazer com que as redes convolucionais empregadas requeiram uma maior quantidade de dados para serem treinadas. Por fim, os autores empregaram um banco de dados prioritário, dificultando aprofundamentos no problema.

Diante dos diversos trabalhos que estudam tarefas relacionadas ao diagnóstico de AVC,

e da gravidade do problema relacionado, bem como a ausência de bancos de dados públicos, confirma-se a importância do desenvolvimento de técnicas alternativas às apresentadas, que possam auxiliar o diagnóstico médico e o posterior aprofundamento da pesquisa por outros pesquisadores de forma pública.

## 1.1 Problema

Os esforços para desenvolver uma ferramenta de apoio à tomada de decisão na classificação de AVC são constantes, com bons e promissores trabalhos na literatura. Porém, dois pontos de grande relevância são levantados: a ausência de bancos de dados públicos anotados para a classificação de AVC, e o uso extensivo de CNNs profundas e complexas.

Para mitigar o primeiro ponto, não apenas para o problema de classificação de AVC, muitos autores utilizam abordagens de transferência de aprendizado (RAINA; NG; KOLLER, 2006), em que um modelo de DL complexo é treinado em um banco de dados grande o suficiente, diferente do problema original, e depois é realizado um ajuste-fino para a tarefa alvo. Outra abordagem diz respeito ao aumento artificial de dados, em que diferentes e sucessivas transformações são realizadas nos dados originais para diferenciá-los e aumentar o volume do banco de dados. É importante salientar que mesmo funcionando bem, essas abordagens possuem suas limitações, como por exemplo alguns vieses imprecisos que podem ser inseridos por meio de domínios de dados diferentes e/ou transformações errôneas no aumento artificial.

Com relação ao segundo ponto, observa-se na literatura o aumento da utilização de redes neurais cada vez mais profundas e complexas, ou que demandam diferentes tipos de pré-processamentos dos dados. O aumento da complexidade não é um ponto negativo, a não ser quando o acesso a grandes quantidades de dados para treinar esses modelos é restrito, ou quando existe limitação de *hardware* para treinamento e inferência. Ademais, o estado-da-arte para classificação de AVC, empregando DL, diz respeito à utilização de técnicas de alta complexidade (redes neurais convolucionais muito profundas), com os paradigmas de aprendizado puramente supervisionado e/ou fracamente supervisionado, não havendo por exemplo, a exploração de técnicas menos complexas como as Máquinas de Boltzmann Restritas Convolucionais (do inglês, *Convolutional Restricted Boltzmann Machines* - CRBMs) (LEE et al., 2009), ou suas variantes em profundidade, as Redes de Crença Profundas (do inglês, *Deep Belief Networks* - DBNs) e Redes de Crença Profundas Convolucionais (do inglês, *Convolutional Deep Belief Networks* - CDBNs) (LEE et al., 2009).

Além disso, outro ponto importante relacionado ao segundo ponto é a aplicação de técnicas de processamento de imagens amplamente estudadas e bem estabelecidas, como a transformada de Fourier (GONZALEZ; WOODS, 2000). Neste sentido, a técnica se mostra promissora de ser estudada (MASUTANI; BAHRAMI; HSIAO, 2020; FATIMAH; SINGHAL; SINGH, 2022; XU et al., 2023), uma vez que no domínio das frequências podem emergir padrões

relevantes e serem empregados como forma de agregação de informações. Além disso, não há registros de trabalhos que empreguem as referidas redes neurais utilizando a transformada de Fourier como forma de agregar informação, além das aqui apresentadas, mesmo a técnica sendo amplamente aplicada em processamento de sinais nos mais diversos domínios, abrindo uma linha de estudo e aplicação.

Desta forma, o principal problema a ser investigado por este trabalho é a classificação de AVC, utilizando redes neurais menos complexas, como as RBMs e DBNs (convolucionais e não-convolucionais), bem como a aplicação da transformada de Fourier no domínio original dos dados, preenchendo essas lacunas das atuais pesquisas na área com a metodologia que será apresentada para o problema.

## 1.2 Hipótese e Questões da Pesquisa

Diante dos problemas citados na seção anterior, é apresentada a hipótese desta pesquisa: É viável, em termos de desenvolvimento e desempenho preditivo, aplicar técnicas de aprendizado profundo com baixa complexidade como as RBMs, DBNs, e até suas variantes convolucionais (CRBMs, e CDBNs) pouco profundas, e agregar informação por meio da transformada de Fourier responsável pela multimodalidade dos dados, para a tarefa de classificação de AVC a partir de imagens de tomografia computadorizada.

Frente à hipótese supracitada, são levantadas algumas questões de pesquisa a serem respondidas com o desenvolvimento desta tese, sendo apresentadas a seguir:

- (1) É possível desenvolver redes neurais robustas e menos complexas (em profundidade) para a classificação com êxito de AVC utilizando, por exemplo, Máquinas de Boltzmann Restritas não-convolucionais e/ou convolucionais, bem como as DBNs?
- (2) A fusão das imagens originais com as informações obtidas da transformada de Fourier melhora a eficácia na classificação de AVC?
- (3) É possível visualizar e entender o que está sendo aprendido pelos modelos empregados para a classificação de AVC?

Para validar a hipótese, ou rejeitá-la, estas questões serão consideradas e respondidas frente aos resultados atingidos com a metodologia proposta.

## 1.3 Objetivo e Contribuições

O objetivo desta tese parte da hipótese de que é possível empregar técnicas de aprendizado profundo com baixa complexidade para a de classificação de AVC, ou seja, propor uma metodologia eficiente utilizando as redes neurais mencionadas com a agregação de informação

pela transformada de Fourier para o problema em questão. Além disso, como um subproduto da pesquisa, objetiva-se criar um *framework* com redes neurais menos profundas que as CNNs estado-da-arte. Para validar o estudo, serão empregados os paradigmas de aprendizado não-supervisionado e supervisionado.

Arelado ao objetivo principal desta pesquisa, toda modelagem e experimentação foi desenvolvida em linguagem Python, disponibilizada em um repositório do GitHub para tornar o *framework* desenvolvido público e disponível para toda a comunidade científica. O *framework* desenvolvido é um subproduto desta pesquisa para apoiar o desenvolvimento da área.

Uma vez definidos todos os pontos citados anteriormente, desde o problema até o objetivo, este trabalho possui cinco principais contribuições, sendo estas elencadas como segue:

1. Desenvolvimento de redes neurais de complexidade menor (RBMs/DBNs convolucionais ou não) que as estado-da-arte (CNNs);
2. Introdução das redes de Boltzmann restritas para o problema de classificação de AVC;
3. Introdução da transformada de Fourier junto das RBMs e DBNs;
4. Propor e disponibilizar um *framework* com uma metodologia para resolução do problema de classificação de AVC;
5. Atualização da literatura acadêmica com a aplicação da nova metodologia empregada.

## 1.4 Estrutura da Tese

A tese está organizada da seguinte maneira:

- O Capítulo 2 apresenta um panorama geral sobre o acidente vascular cerebral e o problema de classificação do mesmo;
- O Capítulo 3 apresenta conceitos e definições sobre o aprendizado em profundidade;
- O Capítulo 4 apresenta abordagens propostas para mitigar os problemas levantados, bem como a proposta desenvolvida e aplicada para parte destes;
- O Capítulo 5 trata do banco de dados utilizado e da metodologia para a abordagem proposta;
- O Capítulo 6 mostra os resultados alcançados com a metodologia adotada, tanto para aprendizado não-supervisionado quanto para o aprendizado supervisionado;
- O Capítulo 7 apresenta as conclusões sobre o que foi desenvolvido, bem como pontos em aberto para futuros estudos.

## 2 Acidente Vascular Cerebral e Revisão da Literatura

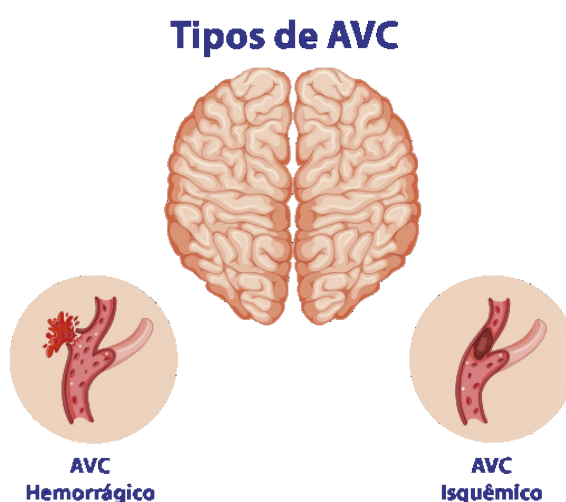
Nesta seção são apresentados os conceitos teóricos necessários para o entendimento geral sobre o AVC, bem como a introdução de conceitos para a classificação do mesmo, junto de alguns conceitos e referenciais médicos para tal.

### 2.1 Acidente Vascular Cerebral - AVC

AVC é o nome comum para a anomalia cerebral chamada acidente vascular cerebral. É caracterizado por uma lesão que atinge abruptamente o tecido cerebral, causada principalmente por alterações no fornecimento de sangue a uma determinada região do cérebro. O evento cerebral mencionado resulta na perda ou redução das funções do cérebro. O AVC é usualmente classificado em isquêmico (o mais comum) e hemorrágico. O primeiro, podendo ser chamado de agudo, é caracterizado pelo bloqueio de um vaso responsável pela irrigação cerebral, ou seja, aterosclerose de grandes artérias, embolia cardíaca, oclusão de pequenos vasos, ou outras causas indeterminadas (YEW; CHENG, 2009). Já o hemorrágico representa a ruptura de um vaso sanguíneo dentro ou ao redor do cérebro. A Figura 1 retrata tais diferenças.

Vários fatores de risco influenciam a ocorrência precoce do AVC, sendo categorizados como não modificáveis e modificáveis. O primeiro representa a idade, sexo, raça e etnia, enquanto a segunda categorização diz respeito a fatores que dependem de condições clínicas, tais como doenças cardíacas e diabetes mellitus. Além disso, os fatores modificáveis também se referem ao estilo de vida, como inatividade física, obesidade, desnutrição, uso de tabaco, e

Figura 1 – Representação dos tipos de AVC.



Fonte: Associação Brasil AVC <<https://abavc.org.br/index.php/sobre-o-avc/>>

álcool.

Para prevenir o AVC, e o subsequente desenvolvimento de sequelas, é crucial que os médicos apliquem com sucesso uma rápida identificação e se possível apoiem o controle dos fatores de risco modificáveis, uma vez que estes desempenhem um papel significativo no diagnóstico. O diagnóstico envolve uma análise profunda da história médica, exames físicos e neurológicos, além de análises de imagem cerebral (SIRSAT; FERMÉ; CÂMARA, 2020), o que dificulta diagnósticos precoces.

Ano após ano, muitas pessoas (milhões) em todo o mundo são atingidas por um AVC, em qualquer grau de gravidade, em que uma quantidade elevada e expressiva acaba por morrer (VIRANI et al., 2020). No entanto, o maior problema relacionado às anomalias do AVC vai para além da mortalidade, ou seja, muitos sobreviventes desenvolvem sequelas pós-AVC ao longo da vida. Alguns exemplos de tais sequelas podem ser dificuldades em atividades cotidianas triviais com graus de dependência moderada ou grave para estas tarefas (RANGEL; BELASCO; DICCINI, 2013).

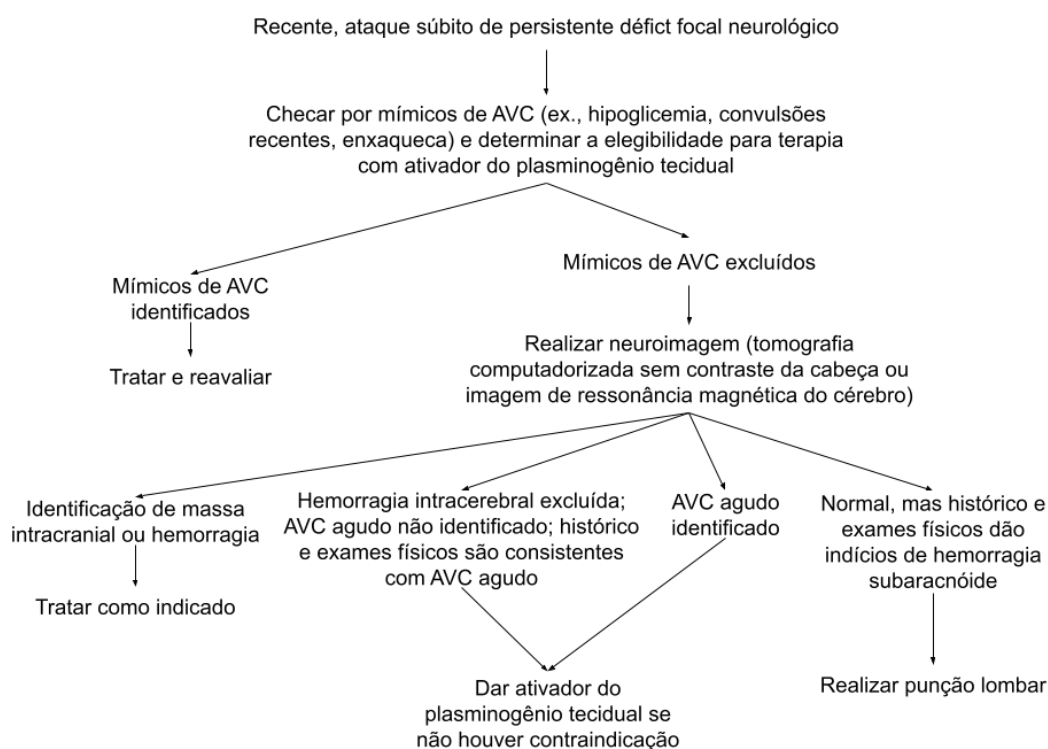
Para ajudar no diagnóstico e na detecção de anomalias cerebrais por imagens, tais como hemorragias, trombozes, pequenas fraturas cranianas e AVC, a medicina tem tido o apoio da inteligência artificial. Tais aplicações utilizam, geralmente, dois tipos de imagens, provenientes de tomografia computadorizada (comumente chamadas de *CT Scans*), ou ressonância magnética (MRI), do inglês *Magnetic Resonance Imaging*.

## 2.2 Classificação do AVC

Com relação à tarefa de classificação do AVC, uma diretriz é frequentemente empregada pelos médicos. Tal guia parte da história e do exame físico do paciente, no qual é possível observar alguns sintomas como distúrbio de fala, fraqueza subjetiva de perna, fraqueza subjetiva de braço e ataque agudo. Além disso, os médicos podem empregar o NIHSS (*National Institute of Health Stroke Scale*), uma das classificações mais comuns de gravidade inicial do AVC que fornece um exame neurológico estruturado com diagnóstico (YEW; CHENG, 2009).

Em um estudo na comunidade médica, os médicos de cuidados primários que atuam em um cenário de emergência tiveram uma sensibilidade de 92% para o diagnóstico de AVC com base no histórico e observação (YEW; CHENG, 2015). Além disso, a confiabilidade geral do diagnóstico de AVC de um clínico é de moderada a boa, com menor confiabilidade em examinadores menos experientes ou menos confiantes, naturalmente. Os autores Yew e Cheng (YEW; CHENG, 2009) também destacam que a característica histórica mais comum de um acidente vascular cerebral isquêmico é o despertar com início ou ataque agudo dos sintomas. Por outro lado, as evidências físicas mais comuns são fraqueza unilateral e distúrbios da fala (YEW; CHENG, 2015). A Figura 2 ilustra o fluxo para detectar AVC em pacientes.

Figura 2 – Representação do fluxo de identificação e classificação do AVC, adaptado de Yew e Cheng (YEW; CHENG, 2015).



Da Figura 2, pode-se notar que é unânime o emprego da neuroimagem, uma vez que marcadores mímicos são descartados. Aqui, destacam-se a tomografia computadorizada sem contraste (CT) e a ressonância magnética por imagem (MRI), e também, suas principais diferenças. A CT sem contraste é considerada suficientemente sensível no campo de detecção de lesão de massa, como uma hemorragia aguda. No entanto, menos de dois terços dos AVCs são detectados por CT sem contraste três horas após o infarto (LATCHAW et al., 2009) e têm sensibilidade ainda menor para AVCs pequenos (YEW; CHENG, 2015).

Por outro lado, a MRI multimodal, particularmente as imagens ponderadas em difusão, têm melhor resolução do que a CT sem contraste. Além disso, tais ressonâncias magnéticas têm maior sensibilidade para detectar acidente vascular cerebral isquêmico agudo (LATCHAW et al., 2009). Outro ponto é que as sequências de MRI são tão sensíveis quanto a CT sem contraste para detectar acidente vascular cerebral hemorrágico intracerebral (WINTERMARK et al., 2013), (YEW; CHENG, 2015).

Com relação às duas técnicas de neuroimagem, a MRI apresenta melhor resolução que a CT sem contraste, porém a CT sem contraste é mais rápida, tem maior disponibilidade, menos dispendiosa e pode ser aplicada em pessoas com dispositivos implantados (por exemplo, marcapasso). Mesmo com essas diferenças de alto nível, o ponto principal é que para realizar a classificação do AVC ou confirmar o diagnóstico, uma das técnicas de imagem deve ser

empregada pelo médico (YEW; CHENG, 2015). Por fim, a Figura 3 apresenta um exemplo de acidente vascular cerebral hemorrágico.

Figura 3 – Tomografia computadorizada referente a um acidente vascular cerebral hemorrágico (região destacada).



## 2.3 Trabalhos Correlatos à Classificação de AVC utilizando Aprendizado Profundo

Tratando-se da classificação de AVC utilizando aprendizado profundo, temos alguns trabalhos recentes que abordam a problemática.

Com relação ao diagnóstico médico baseado em IA, Pereira et al. (2018) empregaram redes neurais convolucionais, como a AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2017), por exemplo, para detectar e classificar imagens de pacientes com/sem AVC. Os autores obtiveram taxas de acerto relativamente altas com relação à classificação de AVC usando imagens de tomografia, com 93% e 97%, para duas divisões dos dados. Entretanto, imagens médicas são difíceis de se obter, e mais difíceis de rotular (devido à alta demanda dos médicos), o que pode levar a bancos de dados pequenos, como apontou Pereira et al. (2018). Assim, os autores também afirmaram que seu conjunto de dados (300 imagens) era relativamente pequeno para aproveitar o poder das CNNs empregadas, principalmente observando a alta variação dos resultados demonstrada pelo desvio padrão.

Pedemonte et al. (PEDEMONTE et al., 2018) propuseram uma rede convolucional modificada (YNet) que utilizou imagens de ressonância magnética como dados de treinamento baseados em aprendizado fracamente supervisionado. O estudo teve como objetivo segmentar as imagens e detectar acidente vascular cerebral isquêmico em pacientes para o correto direcionamento das terapias de tratamento. O trabalho foi bem sucedido na tarefa, no entanto, o modelo proposto possui uma arquitetura complexa e consumiu muito tempo de treinamento,

principalmente devido a grande quantidade de dados utilizados, aproximadamente 10.000 estudos provenientes de uma extensa base de dados privada.

Chilamkurthy et al. (CHILAMKURTHY et al., 2018) estudaram extensivamente tomografias computadorizadas cerebrais sem contraste, lidando com uma ampla gama de anomalias cerebrais, como hemorragia intracraniana (ICH) e seus tipos, fraturas cerebrais, desvio da linha média e efeito de massa. O principal objetivo dos autores não era detectar o AVC; no entanto, o ICH compreende um subtipo de acidente vascular cerebral hemorrágico. O objetivo era detectar e classificar as anormalidades anteriores a partir de um grande volume de tomografias computadorizadas de vários centros médicos da Índia. Os autores empregaram o modelo CNN ResNet18 (HE et al., 2015) como base, ajustado para cada tipo de anormalidade cerebral, pois os autores treinaram um modelo para cada anormalidade e então encaminharam a saída dos modelos para diferentes florestas aleatórias (BREIMAN, 2001) para executar classificação. Os autores também empregaram uma U-Net (originalmente proposta por Ronneberger et al. (RONNEBERGER; FISCHER; BROX, 2015)) para fazer a segmentação das amostras classificadas em ICH, anotando a região delimitadora com o apoio de especialistas (radiologistas).

Lee et al. (LEE et al., 2019) também realizaram um excelente trabalho na detecção e classificação de hemorragia intracraniana. Os autores diferem de Chilamkurthy et al. (CHILAMKURTHY et al., 2018) ao focar na segmentação de ICH e sua classificação a partir de um conjunto de dados pequeno e desbalanceado, composto por menos de 1.000 amostras de tomografias computadorizadas sem contraste. Os autores também empregaram o mapeamento de ativação de classe (método CAM) (SELVARAJU et al., 2017) para adicionar explicabilidade aos modelos. Para tratar o desequilíbrio de classes, Lee et al. (LEE et al., 2019) selecionaram quatro arquiteturas baseadas em CNN, todas pré-treinadas no ImageNet<sup>1</sup>, com o emprego do aumento artificial de dados, a criação de um atlas (como um banco de informações sobre imagens de treinamento) e uma conversão de várias janelas de visualização, seguida de uma interpolação de corte para imitar o fluxo de trabalho dos radiologistas. Após um extensivo processo de otimização, os autores alcançaram desempenho notável em métricas como área sob a curva (AUC), sensibilidade e especificidade, que variaram de 0,70 a 0,99. Eles empregaram uma composição de redes neurais (*ensemble*) para segmentar hemorragias e classificar a ICH e seus subtipos, com a explicabilidade das redes profundas fornecidas pelo método CAM.

Junior et al. (JUNIOR et al., 2019) realizaram um estudo sobre a detecção e classificação de AVC, se concentrando no desenvolvimento de um sistema de internet das coisas (IoT), capaz de receber imagens de CT a partir de dispositivos móveis para extrair informações dos exames com modelos baseados em CNN pré-treinadas e realizar a classificação com alguns algoritmos clássicos de aprendizado de máquina. Os autores empregaram dois bancos de dados particulares e relativamente pequenos, compreendendo 420 imagens de CT em cada um, distribuídos igualmente nas três classes (cérebro normal, AVC hemorrágico, e AVC isquêmico). No entanto,

---

<sup>1</sup> <<https://www.image-net.org/>>

a metodologia mostra que ambas as bases de dados são semelhantes, diferindo apenas no formato da imagem, ou seja, uma é em DICOM e a outra em escala de cinza, levando a resultados experimentais muito próximos nas métricas avaliadas. Porém, o trabalho é, de fato, um marco significativo na tarefa de detecção e classificação de AVC devido à implementação de um sistema IoT.

Nishio et al. (NISHIO et al., 2020) conduziram um importante estudo para auxiliar os radiologistas a melhorar sua sensibilidade para a tarefa de detecção e classificação de AVC isquêmico agudo (AIS). Como o problema é a detecção de AIS ou uma imagem de paciente falso positivo (FP), os autores coletaram 238 exames de dois hospitais distintos no Japão, onde os pacientes foram classificados como tendo AIS ou não, para treinar um fluxo robusto de aprendizado profundo. Os autores não disponibilizam o banco de dados publicamente. Eles desenvolveram um sistema de detecção e classificação de dois estágios empregando as arquiteturas YOLOv3 (You Only Look Once v3) (REDMON; FARHADI, 2018) e VGG16 (Visual Geometry Group 16) (SIMONYAN; ZISSERMAN, 2014). De modo geral, a metodologia empregada compreende duas etapas. A primeira etapa consiste em alimentar as imagens CT na rede de detecção de objetos (YOLOv3) e, a partir de suas probabilidades de classe de saída, os autores criaram um novo subconjunto do conjunto de dados original com amostras que podem ser FP, dado um limite. Depois disso, tais imagens alimentam o modelo de classificação (VGG16) para classificá-las em AIS ou FP. Os resultados experimentais mostraram uma melhora significativa no desempenho da sensibilidade dos radiologistas.

Em uma direção semelhante, Federau et al. (FEDERAU et al., 2020) propôs melhorar a segmentação e a sensibilidade de detecção de lesões de AVC ponderadas por difusão com técnicas de aprendizagem profunda aprimoradas sinteticamente. Eles usaram dois bancos de dados de “tamanho médio” (quase 3.000 exames) que foram adquiridos de um sistema de comunicação e arquivamento de imagens (PACS) do University Hospital Basel (privado). Os autores geraram sinteticamente imagens relacionadas ao AVC a partir das amostras normais ponderadas por difusão e, aleatoriamente, combinaram as sintéticas com as normais, aplicando a orientação necessária e a correção de localização no parênquima cerebral para gerar as novas imagens finais. Eles obtiveram bons resultados em relação à sensibilidade com uma 3D U-Net, superando a abordagem padrão (sem aumento de dados sintéticos); no entanto, tal abordagem pode ser tratada como, até certo ponto, injusta, uma vez que o melhor modelo empregou 40.000 amostras geradas sinteticamente, o que é várias vezes maior do que o banco de dados original. Os autores argumentam que tal metodologia é um ponto de partida para um aprimoramento posterior, tratando-a como uma opção para imagens geradas sinteticamente a partir de GANs (Generative Adversarial Networks) (GOODFELLOW et al., 2020), por exemplo, o que pode ser positivo.

Herzog et al. (HERZOG et al., 2020) estudaram especificamente o problema de classificar uma imagem de ressonância magnética como contendo ou não um AVC isquêmico. Os autores

fornece uma estrutura completa para diagnosticar pacientes com AVC isquêmico a partir de imagens de ressonância magnética incorporando a incerteza Bayesiana em um procedimento de aprendizado de uma CNN semelhante a uma VGG16 (SIMONYAN; ZISSERMAN, 2014). Com relação ao banco de dados, Herzog et al. (HERZOG et al., 2020) coletaram ressonâncias magnéticas ponderadas em difusão de 511 pacientes com acidente vascular cerebral isquêmico ou ataque isquêmico transitório internados no Departamento de Neurologia do Hospital Universitário de Zurique. Os autores estudaram extensivamente a decisão e a incerteza das redes, combinando medidas como entropia e informação mútua, e um número variável de previsões de imagem para o diagnóstico. Por fim, as abordagens propostas melhoraram a confiança e o desempenho da classificação da rede neural, superando as linhas de base empregadas, ou seja, CNN sem agregação de incerteza.

Por outro lado, Saragih et al. (SARAGIH et al., 2020) combinaram aprendizado profundo com uma técnica clássica de aprendizado de máquina, as florestas aleatórias (RF) (BREIMAN, 2001), para classificar AVC isquêmico em 92 imagens de CT fornecidas por um hospital na Indonésia (banco de dados privado). Neste estudo, a CNN foi empregada como um extrator de características com cinco camadas convolucionais para alimentar o classificador RF, provavelmente devido ao baixíssimo volume de imagens. Os autores obtiveram bons resultados quanto à precisão da classificação do AVC com a RF. O estudo é bem fundamentado, porém apresenta algumas fragilidades, como a falta de explicação sobre a metodologia de pré-treinamento da CNN (por exemplo, pré-treinamento no ImageNet ou banco de dados próprio, bem como a configuração experimental) antes da fase de classificação empregando a RF com características extraídas da CNN, por exemplo.

Rajinikanth et al. (RAJINIKANTH; ASLAM; KADRY, 2021) estudaram a segmentação e classificação (binária) de AVC isquêmico em diferentes modalidades de MRI. Eles usaram o conjunto de dados do desafio *Ischemic Stroke Lesion Segmentation* (ISLES2015), um banco de dados de tamanho médio, pré-processado. A abordagem proposta foi uma combinação de diferentes técnicas para lidar com a classificação do AVC. Os autores empregaram uma rede baseada em VGG para segmentar as imagens com AVC isquêmico e extrair suas características, em combinação com uma segunda fase de extração de características utilizando informações como momento Hu (HU, 1962), e padrão binário local (LBP) (OJALA; PIETIKAINEN; HARWOOD, 1994), para serem concatenadas com as características da CNN. Após a concatenação, diferentes classificadores, como máquinas de vetores de suporte (SVMs) (CORTES; VAPNIK, 1995) e florestas aleatórias (BREIMAN, 2001), foram empregados para fins de classificação. Os experimentos mostraram resultados satisfatórios e bem fundamentados. Apesar de não atingirem o estado-da-arte, os autores empregaram diferentes métricas para medir o desempenho da classificação, como acurácia, precisão, revocação e sensibilidade, dando robustez à análise, uma vez que a base de dados empregada é desbalanceada.

Pan et al. (PAN et al., 2021) abordaram uma tarefa semelhante em seu trabalho, ou seja,

a segmentação e posterior classificação do AVC isquêmico agudo. Eles empregaram imagens de CT e dados de MRI ponderados por difusão em vez de usar apenas um tipo de exame. Os autores propuseram detectar e classificar o AIS com uma arquitetura baseada em convolução, a ResNet18 (HE et al., 2015), bem como uma abordagem de *patch* em que o tamanho do *patch* da imagem varia em três escalas. A base de dados é privada e com menos de 150 amostras. A abordagem proposta foi treinar a ResNet18 para identificar cada *patch* de imagem extraído das imagens e, em seguida, tomar o resultado da identificação como o rótulo do pixel central. Este método baseado na classificação do *patch* permite que os autores obtenham a classificação do AVC e sua localização. Os autores alcançaram precisões, sensibilidades, e especificidades satisfatórias, com quase todas as métricas atingindo 90% no conjunto de teste. No entanto, os autores poderiam ter fornecido mais detalhes e explicações sobre a agregação dos modelos para segmentação e classificação, uma vez que três ResNet18 foram empregados para processar os diferentes tamanhos de *patch*.

Em relação à tarefa de classificação pura do AVC, Cetinoglu et al. (CETINOGLU et al., 2021) desenvolveram um estudo no qual os autores empregaram arquiteturas CNN para classificar imagens de MRI ponderada por difusão em AVC ou não, e também para classificar o território vascular quando a imagem é AVC positiva. A segunda tarefa representa a indicação da região arterial quando o AVC estava presente, por exemplo, na região da artéria cerebral média ou na região da circulação posterior. A base de dados (privada) utilizada foi desenvolvida pelos autores e dividida em duas, uma para a classificação do AVC e outra para a classificação do território vascular. Cetinoglu et al. (CETINOGLU et al., 2021) mostraram que com a transferência de aprendizado, é possível alcançar resultados promissores para ambas as tarefas de classificação medindo o desempenho dos modelos sob várias métricas, como sensibilidade, especificidade, e precisão. O melhor modelo alcançou mais de 80% em todas as métricas para ambas as tarefas. Os autores destacaram a importância de grandes conjuntos de dados para problemas tão desafiadores como os relacionados ao AVC, além de uma validação extra com dados externos.

Gautam e Raman (GAUTAM; RAMAN, 2021) também trabalharam com classificação de AVC; seu banco de dados privado inclui 900 imagens de CT com cérebros normais, derrames isquêmicos e hemorrágicos, coletadas do Instituto de Ciências Médicas do Himalaia (Índia). Eles produziram uma abordagem de fusão de imagem para melhorar a detecção de AVC utilizando uma CNN própria, em que contraste e filtro da mediana são aplicados nas imagens, além de um método multi-foco para reconstruir as imagens modificadas com um leve destaque nas áreas de AVC. Esta etapa de pré-processamento é um conjunto de operações complexas baseadas em gradiente e laplaciano, que requerem uma quantidade significativa de processamento computacional, corroborado pelo hardware empregado. Em relação aos resultados experimentais, os autores elaboraram duas estratégias. Uma delas é para o caso de alimentar o sistema apenas com AVC isquêmico e hemorrágico (classificação binária). O outro adiciona cérebros saudáveis, levando a uma classificação de três classes. Considerando as duas abordagens com a mesma

estrutura, a CNN proposta alcançou resultados significativos em várias métricas, como acurácia, precisão, e os verdadeiro-positivos, por exemplo, contra linhas de base de ResNet50 (HE et al., 2015) e AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Tal resultado é apropriado, uma vez que o pré-processamento proposto foi projetado para melhorar a região do AVC, e CNNs mais complexas precisam de mais dados para aprender.

Ural (URAL, 2022) estudou as áreas anômalas em tomografias cerebrais de pacientes com AVC utilizando metodologias de processamento de imagem (operações de limiar, segmentação, etc), e avaliou as alterações cerebrais causadas pelo AVC nos tecidos cerebrais entre pacientes, utilizando modelos de aprendizado profundo. Para isso, o autor empregou tanto redes neurais rasas quanto profundas para a classificação entre pacientes sem AVC e com AVC. Os resultados se mostraram interessantes, com AUC de 98,5, para um banco de dados de tamanho médio, porém, com a ressalva de que o sistema de classificação era alimentado apenas com regiões de interesse nas imagens de tomografia extraídas da segmentação, realizada por uma sequência de operações matemáticas para essa finalidade.

Ainda em 2022, Peng et al. (PENG et al., 2022) desenvolveram um método automatizado baseado em CNNs para facilitar uma detecção fácil e rápida (segundo os autores) de AVC, utilizando imagens privadas de tomografia do cérebro. Eles empregaram abordagens específicas de pré-processamento de imagens para então alimentar as CNNs, compostas basicamente por três camadas convolucionais e camadas de normalização. Parte do pré-processamento é composto por análises estatísticas e aumento artificial de dados para tentar melhorar o desempenho de detecção e classificação de AVC pela CNN desenvolvida.

Finalmente, Bridge et al. (BRIDGE et al., 2022) desenvolveram um modelo de aprendizado profundo (3D U-Net) que empregou o coeficiente de difusão aparente junto da série de imagens ponderadas por difusão, ou seja, multimodalidade de dados. O banco de dados (privado), foi rotulado como positivos ou negativos para AVC, com 377 amostras contendo a região de interesse delineada (anotação de segmentação). Os diferentes tipos de anotações auxiliaram o treinamento, sem exigir muito tempo para segmentar manualmente cada estudo. Os autores atingiram bons resultados, como sensibilidade de 89,3% e especificidade de 94,8%, mas infelizmente não é possível utilizar o mesmo banco de dados para possíveis melhorias por outros autores.

## 2.4 Banco(s) de Dados sobre AVC

Ao se tratar de classificação de AVC por imagens, a literatura carece de bancos de dados públicos disponíveis, mesmo com uma grande quantidade de trabalhos na área. Isto representa uma das dificuldades apresentadas na Seção 1.1, em que emergem algumas alternativas para mitigar esse problema. A seguir são apresentados alguns bancos de dados como alternativa para o problema, e o único banco de dados de classificação de AVC encontrado.

Em 2015 surgiu uma competição de segmentação de lesão cerebral causada por AVC isquêmico, chamada ISLES <sup>2</sup> (*Ischemic Stroke Lesion Segmentation*). A competição ocorreu até 2018 e, atualmente, está em andamento. Especificamente, o ISLES 2018 objetivou encontrar métodos que permitam a segmentação de lesões de acidente vascular cerebral agudo com base em dados de perfusão de tomografia computadorizada. Para isso, foi proposto um conjunto de dados a partir de tomografias de 103 pacientes com AVC com suas respectivas máscaras de segmentação criadas por especialistas. Os dados de treinamento do ISLES 2018 compreendem 94 conjuntos de dados individuais de 63 pacientes. Alguns casos de indivíduos têm duas fatias de leitura de tomografia para cobrir a lesão pelo AVC. Estas são regiões cerebrais não, ou parcialmente, sobrepostas. Já os casos individuais consistem em apenas fatias ao redor das lesões de AVC, com o número de fatias por caso variando de 2 a 16. Para cada caso, foram fornecidas imagens de CT nativas, bem como imagens de CT quadridimensionais aprimoradas por meio de contraste. A partir destas últimas, os mapas de perfusão de fluxo sanguíneo cerebral (do inglês, *cerebral blood flow*), volume sanguíneo cerebral (do inglês, *cerebral blood volume*), tempo médio de trânsito (do inglês, *mean transit time*) e tempo para fluxo máximo foram calculados e também fornecidos.

Em 2020 Hssayeni et al. (HSSAYENI et al., 2020), publicaram um banco de dados para detecção e segmentação de hemorragia intracraniana, um tipo de AVC que pode ocorrer. Os autores coletaram um conjunto de dados de 82 tomografias computadorizadas de pacientes com lesão cerebral traumática. As regiões de hemorragia intracraniana nesses exames foram delineadas/annotadas em cada corte por dois radiologistas. Eles também anotaram cada corte de CT quanto à presença de diferentes tipos de hemorragia intracraniana e fraturas. As tomografias computadorizadas são disponibilizadas no formato NIfTI (*Neuroimaging Informatics Technology Initiative*).

No estudo, Hssayeni et al. (HSSAYENI et al., 2020) coletaram 82 exames de CT, incluindo 36 exames de pacientes diagnosticados com hemorragia intracraniana. Cada tomografia computadorizada de cada paciente inclui cerca de 30 cortes com espessura de corte de 5 mm. A média e o desvio padrão de idade dos pacientes foram 27,8 e 19,5, respectivamente, em que 46 dos pacientes eram do sexo masculino e 36 deles eram do sexo feminino. Cada corte das tomografias sem contraste foi feito por dois radiologistas que registraram os tipos de hemorragia, e se ocorreu hemorragia ou fratura. Os radiologistas também delinearam as regiões hemorrágicas em cada fatia. Houve um consenso entre os radiologistas. Além disso, para facilitar, os autores processaram as imagens de CT de formato DICOM (*Digital Imaging and Communications in Medicine*) usando a janela cerebral (nível=40, largura=120) e salvos novamente em arquivos NIfTI para 75 pacientes.

Como pode ser observado, os bancos de dados apresentados anteriormente não representam dados específicos para a classificação de AVC, mas podem ser utilizados de alguma

<sup>2</sup> <http://www.isles-challenge.org/ISLES2018/>

forma, como por exemplo para o pré-treinamento de redes neurais, para posterior ajuste-fino em um cenário de classificação de AVC com outro banco de dados.

Por fim, foi encontrado na literatura o trabalho de Pereira et al. (PEREIRA et al., 2018), em que os autores disponibilizaram o banco de dados para a classificação de AVC utilizando imagens de tomografia computadorizada. Segundo Pereira et al. (PEREIRA et al., 2018), o banco de dados é composto por 25 exames de tomografia computadorizada cerebral, que compreendem pacientes com (a) cérebro saudável, (b) acidente vascular cerebral isquêmico e (c) acidente vascular cerebral hemorrágico. As imagens foram obtidas com o apoio da Clínica Trajano Almeida - Diagnóstico por Imagem, e criadas especificamente para este trabalho. As 300 imagens de tomografias cerebrais são divididas igualmente, sendo 100 imagens de cérebros saudáveis, e 200 com AVC (100 hemorrágicas e 100 isquêmicas). Além disso, as imagens foram armazenadas usando o formato DICOM, que é um padrão de imagens médicas digitais amplamente utilizado, criado pelo *American College of Radiology* (ACR) e pela *National Electrical Manufacturers Association* (NEMA). Os autores (PEREIRA et al., 2018) utilizaram a biblioteca de código aberto DCMTK (DICOM Toolkit), fornecida pela OFFIS <sup>3</sup>, para ler as imagens e convertê-las para uma representação de 8 bits.

### 2.4.1 Sumarização

Para facilitar a visualização dos trabalhos correlatos previamente citados, e os respectivos bancos de dados empregados, estes foram sumarizados na Tabela 2.4.1. É importante ressaltar que o tamanho da base de dados foi padronizado na análise, adotando o termo pequeno para bancos de até 1.000 amostras, médio de 1.000 a 100.000, e grande para mais de 100.000 amostras.

---

<sup>3</sup> <<https://dicom.offis.de/index.php.en>>

| Trabalho                          | Nome      | Problema (tipo) | Volume  | Específico para AVC | Disponibilidade            | Ano  |
|-----------------------------------|-----------|-----------------|---------|---------------------|----------------------------|------|
| (PEREIRA et al., 2018)            | –         | Class           | Pequeno | AVC                 | Público                    | 2018 |
| (PEDEMONTE et al., 2018)          | –         | Segm            | Médio   | AVC                 | Privado                    | 2018 |
| (CHILAMKURTHY et al., 2018)       | Qure25k   | Segm-Class      | Grande  | Não                 | Privado                    | 2018 |
| (CHILAMKURTHY et al., 2018)       | CQ500     | Segm-Class      | Grande  | Não                 | Público (SR <sup>1</sup> ) | 2018 |
| (LEE et al., 2019)                | –         | Segm-Class      | Pequeno | AVC                 | Privado                    | 2019 |
| (JUNIOR et al., 2019)             | –         | Class           | Pequeno | AVC                 | Privado                    | 2019 |
| (HSSAYENI et al., 2020)           | –         | Segm            | Médio   | AVC                 | Público                    | 2020 |
| (NISHIO et al., 2020)             | –         | Segm-Class      | Médio   | AVC                 | Privado                    | 2020 |
| (FEDERAU et al., 2020)            | –         | Segm-Class      | Médio   | AVC                 | Privado                    | 2020 |
| (HERZOG et al., 2020)             | –         | Class           | Pequeno | AVC                 | Privado                    | 2020 |
| (SARAGIH et al., 2020)            | –         | Class           | Pequeno | AVC                 | Privado                    | 2020 |
| (RAJINIKANTH; ASLAM; KADRY, 2021) | ISLES2015 | Segm-Class      | Médio   | AVC                 | Público                    | 2021 |
| (PAN et al., 2021)                | –         | Segm-Class      | Pequeno | AVC                 | Privado                    | 2021 |
| (CETINOGLU et al., 2021)          | –         | Class           | Médio   | AVC                 | Privado                    | 2021 |
| (GAUTAM; RAMAN, 2021)             | –         | Class           | Médio   | AVC                 | Privado                    | 2021 |
| (URAL, 2022)                      | –         | Class           | Médio   | AVC                 | Privado                    | 2022 |
| (PENG et al., 2022)               | –         | Segm-Class      | Pequeno | AVC                 | Privado                    | 2022 |

<sup>1</sup> Sob Requisição.

Tabela 1 – Resumo dos bancos de dados empregados nos trabalhos. Nome é auto-explicativo (caso possua); Problema (tipo) representa qual a tarefa do trabalho, ou seja, classificação (Class), segmentação (Segm), ou ambas (Segm-Class); Volume representa o volume de dados (Pequeno; Médio; Grande); Específico para AVC significa se a base de dados é específica para AVC ou não; Disponibilidade e Ano são auto-explicáveis.

### 3 Fundamentação Teórica

Com o desenvolvimento tecnológico acelerado, houve (e há) a necessidade de serem criados métodos automáticos para o processamento e análise de dados, como a transformada de Fourier por exemplo, bem como algoritmos que apoiam a tomada de decisão ou realizam tarefas específicas de forma automatizada, como as redes neurais artificiais. Este fato impulsionou o desenvolvimento da área de aprendizado de máquina. Esta, é responsável por modelar algoritmos adaptativos independentes que aprendem padrões ocultos nos dados. Esta característica fez com que as técnicas desenvolvidas ganhassem destaque rapidamente, concomitante ao fato de gerarem resultados promissores (BISHOP, 1995), (CHARU, 2018).

Entretanto, as aplicações, tanto acadêmicas (ATWOOD; TOWSLEY, 2016; CARON et al., 2018) quanto aplicadas no dia-a-dia (JUNIOR et al., 2019), têm se tornado cada vez mais complexas, por diferentes fatores como a diversidade de dispositivos geradores de dados, ou pela natureza dos próprios dados, com relacionamentos complexos de serem estabelecidos e modelados. Um exemplo de complexidade diz respeito à análise de imagens RGB, em que cada componente de cor (*red*, *green*, *blue*) representa um plano bidimensional que compõe a imagem  $2D$ , bem como as mais diversas aplicações que tangem desde a medicina até ambientes industriais complexos.

Ao considerarmos imagens, alguns pontos fazem com que a aplicação seja analisada com cautela, como por exemplo a luminosidade do ambiente e fatores como rotação da imagem ou objetos, que alteram a intensidade ou posição de um determinado *pixel*. Com isso, começou o desenvolvimento de métodos para extrair características relevantes de um domínio altamente variável (BISHOP, 1995). A partir da necessidade de embutir o próprio processo de caracterização dos dados às técnicas de aprendizado de máquina, emergiu a sub-área denominada aprendizado em profundidade ou profundo (do inglês, *Deep Learning* - DL). As técnicas de DL possuem a particularidade de utilizar os dados brutos como entrada, no caso de imagens, os *pixels*, para extrair informações relevantes sem necessariamente utilizar algum pré-processamento.

Em 2006, Geoffrey Hinton deu início a uma “revolução” na inteligência artificial, estabelecendo o aprendizado profundo. Este, diz respeito a uma maior quantidade de camadas ocultas (profundidade) das redes neurais artificiais, concomitante ao aprendizado de padrões/características profundas em relação aos dados. O termo DL teve sua introdução com as *Deep Belief Network* (DBN), ou rede de crença em profundidade (HINTON; OSINDERO; TEH, 2006), que são compostas por máquinas de Boltzmann restritas (RBMs). Esse fato, popularizou o termo “aprendizado profundo”, uma vez que possibilitou o treinamento de redes neurais artificiais com uma ou mais camadas de neurônios, e elevou o patamar dentro da

inteligência artificial com o aprendizado de padrões ocultos (profundos).

No aprendizado profundo, ou em profundidade, busca-se representar informações sofisticadas a partir da combinação de representações mais simples de forma hierárquica, fazendo com que o aprendizado de relacionamentos, ou eventos complexos, seja atingido por meio de uma modelagem eficiente. Essa abordagem se aproxima ao aprendizado do ser humano, em que o cérebro torna as informações hierárquicas a partir de entradas simples, e é capaz de construir conhecimentos robustos e complexos.

A partir de então, as inovações escalaram de maneira rápida e eficiente, com o auxílio de uma grande quantidade de trabalhos e do desenvolvimento de *hardwares* cada vez mais robustos para a alta carga computacional requerida, cenário em que o aprendizado de máquina começou a ganhar destaque na última década, bem como técnicas eficientes para processar dados complexos, como a transformada de Fourier discreta (GONZALEZ; WOODS, 2000).

## 3.1 Modelos Baseados em Energia

Nesta seção são apresentados dois modelos baseados em energia, amplamente utilizados em diversas tarefas, tais como pré-treinamento de redes profundas, classificação de padrões, extração de características e geração de dados.

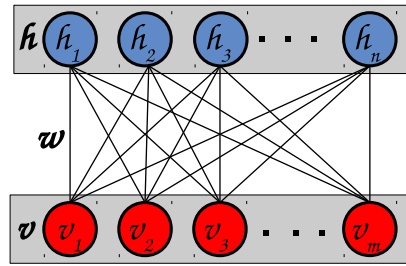
### 3.1.1 Máquinas de Boltzmann Restritas

Conhecidas por sua versatilidade e inspiração física, as máquinas de Boltzmann restritas, ou *Restricted Boltzmann Machines* - RBMs, são redes neurais estocásticas modeladas a partir da distribuição de Boltzmann e princípios de energia, entropia e temperatura. As RBMs são compostas por duas camadas de neurônios, a visível e a oculta, capazes de modelar problemas sob os paradigmas de aprendizado não-supervisionado (HINTON, 2002) ou supervisionado (LAROCHELLE; BENGIO, 2008). As RBMs são uma variação da clássica Máquina de Boltzmann (ACKLEY; HINTON; SEJNOWSKI, 1988), com o diferencial no tocante das conexões serem restritas apenas entre camadas, e não intra-camada.

Sendo derivada de uma máquina de Boltzmann, as RBMs são representadas como um grafo bipartido de conexões não direcionais. A Figura 4 descreve a arquitetura de uma máquina de Boltzmann restrita, com a camada visível  $\mathbf{v}$  possuindo  $m$  unidades, e a camada oculta  $\mathbf{h}$  com  $n$  neurônios. A matriz  $\mathbf{w}$ , com valores reais, modela os pesos (conexão neural) entre os neurônios visíveis e ocultos, possuindo dimensão  $m \times n$ .

No início do seu desenvolvimento, os neurônios (visíveis e ocultos) das RBMs foram modelados para estados binários, utilizando a distribuição de Bernoulli para amostrar os estados visíveis e ocultos. Posteriormente, Welling et al. Welling, Rosen-zvi e Hinton (2005) e Hinton (HINTON, 2012) apresentaram variações para as unidades que podem ser usadas

Figura 4 – Arquitetura de uma RBM.



Fonte: Elaborado pelo autor.

em uma RBM, como as binomiais, as unidades lineares retificadas (ReLU) e as gaussianas. Essas e quaisquer outras variações de RBMs são generalizações da RBM binária, portanto, os conceitos relacionados a esta são apresentados.

Sejam  $\mathbf{v}$  e  $\mathbf{h}$  as unidades visíveis e ocultas (binárias), respectivamente, ou seja,  $\mathbf{v} \in \{0, 1\}^m$  e  $\mathbf{h} \in \{0, 1\}^n$ . A energia de uma RBM é modelada como segue:

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{i=1}^m a_i v_i - \sum_{j=1}^n b_j h_j - \sum_{i=1}^m \sum_{j=1}^n v_i h_j w_{ij}, \quad (1)$$

em que  $\mathbf{a}$  e  $\mathbf{b}$  são os valores dos vieses (*biases*) das unidades visíveis e ocultas, respectivamente. A probabilidade de uma configuração conjunta  $(\mathbf{v}, \mathbf{h})$  é calculada como segue:

$$P(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}}, \quad (2)$$

onde o denominador da equação é a função de partição (fator de normalização), que leva em conta todas as possíveis configurações envolvendo as unidades visíveis e ocultas, sendo intratável para espaços de alta dimensão como imagens, por exemplo.

Em uma RBM o processo de treinamento tem por objetivo maximizar as probabilidades observadas de uma configuração  $P(\mathbf{v})$ , ao passo que é necessário estimar e ajustar os valores de  $\mathbf{w}$ ,  $\mathbf{a}$  e  $\mathbf{b}$ . Portanto, os parâmetros da RBM podem ser otimizados através da técnica da subida do gradiente estocástico, método dual para a descida do gradiente estocástico (*Stochastic Gradient Descent*- SGD). A técnica de otimização é comumente aplicada no logaritmo da verossimilhança (*Log-Likelihood*) dos dados de treinamento. Esta, é calculada para uma amostra apresentada às unidades visíveis, e sua probabilidade é obtida como segue:

$$P(\mathbf{v}) = \frac{\sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}}. \quad (3)$$

Isto posto, para atualizar os pesos e vieses, é necessário computar as seguintes derivadas:

$$\frac{\partial \log P(\mathbf{v})}{\partial w_{ij}} = E[h_j v_i]_{data} - E[h_j v_i]_{modelo}, \quad (4)$$

$$\frac{\partial \log P(\mathbf{v})}{\partial a_i} = E[v_i]_{data} - E[v_i]_{modelo}, \quad (5)$$

$$\frac{\partial \log P(\mathbf{v})}{\partial b_j} = E[h_j]_{data} - E[h_j]_{modelo}, \quad (6)$$

em que  $E[\cdot]$  representa a esperança estatística sob uma distribuição, e  $E[\cdot]_{data}$  e  $E[\cdot]_{modelo}$  representam as probabilidades dos dados originais e reconstruídos, respectivamente.

Em termos práticos, podemos computar  $E[h_j v_i]_{data}$  considerando  $\mathbf{h}$  e  $\mathbf{v}$  como segue:

$$E[\mathbf{h}\mathbf{v}]_{data} = P(\mathbf{h}|\mathbf{v})\mathbf{v}^T, \quad (7)$$

aqui,  $P(\mathbf{h}|\mathbf{v})$  é a probabilidade associada ao espaço latente, representado pelas unidades ocultas  $\mathbf{h}$ , dada uma observação no espaço visível  $\mathbf{v}$  (dado de treinamento):

$$P(h_j = 1|\mathbf{v}) = \sigma \left( \sum_{i=1}^m w_{ij} v_i + b_j \right), \quad (8)$$

onde  $\sigma(\cdot)$  é a função sigmoide-logística<sup>4</sup>. Deste modo,  $E[\mathbf{h}\mathbf{v}]_{data}$  é obtida da seguinte maneira: dada uma amostra de treinamento  $\mathbf{x} \in X$ , onde  $X$  é um conjunto de treinamento, precisa-se ajustar  $\mathbf{v} \leftarrow \mathbf{x}$ , e então utilizar a Equação 8 para obter  $P(\mathbf{h}|\mathbf{v})$  e, através da Equação 7, obtém-se  $E[\mathbf{h}\mathbf{v}]_{data}$ .

Uma vez obtidas as estatísticas a partir dos dados ( $E[\mathbf{h}\mathbf{v}]_{data}$ ), é possível obter  $E[\mathbf{h}\mathbf{v}]_{modelo}$ , o qual representa a modelagem da distribuição aprendida pelo modelo. Da mecânica estatística, uma estratégia é obter a esperança do modelo por meio da técnica de amostragem de Gibbs, baseada no método de Monte Carlo com Cadeias de Markov (MCMC, do inglês *Markov Chain Monte Carlo*). Esta inicia as unidades visíveis com valores aleatórios, e atualiza as unidades ocultas utilizando a Equação 8, seguida pela atualização das unidades visíveis usando  $P(\mathbf{v}|\mathbf{h})$ , dada por:

$$P(v_i = 1|\mathbf{h}) = \sigma \left( \sum_{j=1}^n w_{ij} h_j + a_i \right), \quad (9)$$

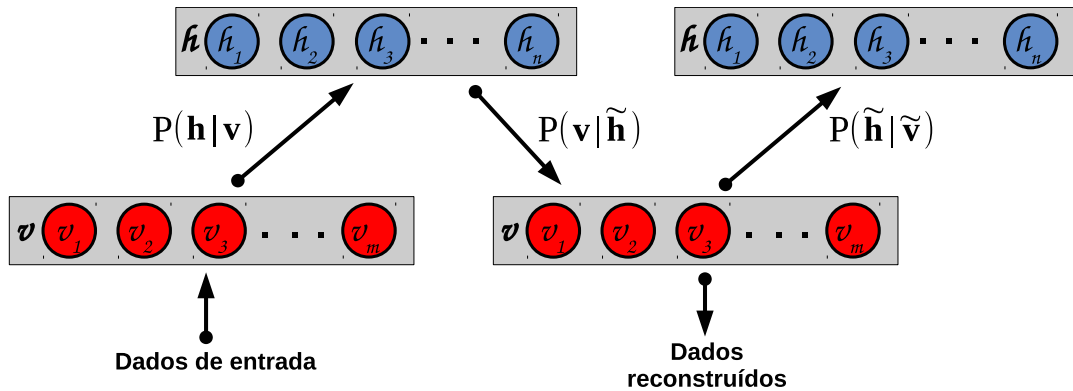
posteriormente, atualizam-se novamente as unidades ocultas utilizando a Equação 8 até um critério de convergência da cadeia ser atingido, como por exemplo,  $k$  iterações. Em resumo, é possível estimar  $E[\mathbf{h}\mathbf{v}]_{modelo}$  a partir de valores aleatórios com a técnica citada.

<sup>4</sup> A função sigmoide-logística é computada pela seguinte equação:  $\sigma(x) = 1/(1 + \exp(-x))$ .

Entretanto, iterar uma cadeia Markov com o método de Monte Carlo é computacionalmente custoso, e se torna inviável para espaços de alta dimensão, uma vez que a convergência da cadeia é garantida quando as iterações tendem ao infinito  $k \rightarrow +\infty$ . Por conta desta dificuldade, alguns trabalhos buscam e apresentaram alternativas à amostragem de Gibbs original, sendo a principal delas a Divergência Contrastiva (*Contrastive Divergence* - CD) (HINTON, 2002).

Hinton (HINTON, 2002) tomou como base a amostragem de Gibbs para propor a divergência contrastiva, introduzindo uma metodologia mais simples, eficaz e rápida para o cálculo de  $E[\mathbf{h}\mathbf{v}]_{\text{modelo}}$ . Em suma, a simplicidade dá-se pelo fato da inicialização das unidades visíveis com amostras dos dados de treinamento, para inferir os estados latentes utilizando as probabilidade da Equação 8. Uma vez calculadas estas probabilidades pode-se calcular os estados das unidades visíveis, ou seja, a reconstrução dos dados, por meio da Equação 9. Este procedimento é equivalente a execução da amostragem de Gibbs usando  $k = 1$ , com os valores da cadeia inicializados com amostras de treinamento. A Figura 5 ilustra essa abordagem.

Figura 5 – Ilustração do método baseado em divergência contrastiva.



Fonte: Elaborado pelo autor.

O procedimento apresentado na Figura 5 pode ser iterado por  $k$  vezes, porém, na maioria das aplicações  $k = 1$  é suficiente para gerar boas aproximações, se, e somente se, a cadeia possui uma alta taxa de mistura, isto é, consegue convergir rapidamente ao passo que reduz as dependências anteriores.

Diante do exposto anteriormente, pode-se calcular  $E[\mathbf{h}\mathbf{v}]_{\text{modelo}}$  como segue:

$$E[\mathbf{h}\mathbf{v}]_{\text{modelo}} = P(\tilde{\mathbf{h}}|\tilde{\mathbf{v}})\tilde{\mathbf{v}}^T. \quad (10)$$

Portanto, o problema de obtenção das estatísticas do modelo, e consequentemente do aprendizado de seus parâmetros, é sanado. Isso possibilita atualizar a matriz de pesos  $\mathbf{W}$ , como segue:

$$\begin{aligned}
\mathbf{w}^{t+1} &= \mathbf{w}^t + \eta(E[\mathbf{h}\mathbf{v}]_{data} - E[\mathbf{h}\mathbf{v}]_{modelo}) \\
&= \mathbf{w}^t + \eta(P(\mathbf{h}|\mathbf{v})\mathbf{v}^T - P(\tilde{\mathbf{h}}|\tilde{\mathbf{v}})\tilde{\mathbf{v}}^T),
\end{aligned} \tag{11}$$

onde  $\mathbf{w}^t$  é a matriz de pesos no instante  $t$ , e  $\eta$  corresponde à taxa de aprendizado. Adicionalmente, os vieses das unidades visíveis e ocultas são atualizados seguindo as formulações:

$$\begin{aligned}
\mathbf{a}^{t+1} &= \mathbf{a}^t + \eta(\mathbf{v} - E[\mathbf{v}]_{modelo}) \\
&= \mathbf{a}^t + \eta(\mathbf{v} - \tilde{\mathbf{v}}),
\end{aligned} \tag{12}$$

e

$$\begin{aligned}
\mathbf{b}^{t+1} &= \mathbf{b}^t + \eta(E[\mathbf{h}]_{data} - E[\mathbf{h}]_{modelo}) \\
&= \mathbf{b}^t + \eta(P(\mathbf{h}|\mathbf{v}) - P(\tilde{\mathbf{h}}|\tilde{\mathbf{v}})),
\end{aligned} \tag{13}$$

em que  $\mathbf{a}^t$  e  $\mathbf{b}^t$  são os valores de *bias* das unidades visíveis e ocultas no momento  $t$ , respectivamente. Em resumo, com as Equações 11, 12 e 13 pode-se atualizar os parâmetros da RBM.

### 3.1.1.1 RBMs para Dados Contínuos

Dada a necessidade de trabalhar com dados não binários, houve um esforço dos pesquisadores para generalizar as RBMs de modo que acomodassem dados contínuos, principalmente na camada visível. Essa modelagem é útil para modelar diferentes tipos de entradas ou sinais que variam em largas faixas de valores. As mudanças que possibilitam esta utilização ocorrem na função de energia, como segue:

$$E(\mathbf{v}, \mathbf{h}) = \sum_{i=1}^m \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^n b_j h_j - \sum_{i=1}^m \sum_{j=1}^n \frac{v_i}{\sigma_i} h_j w_{ij}, \tag{14}$$

em que  $\sigma_i$  representa o desvio padrão dos dados de entrada, e  $\sigma_i^2$  a variância, para cada neurônio  $i$ . Levando em consideração as derivadas analogamente às RBMs binárias, as probabilidades condicionais dos neurônios da camada visível se tornam:

$$P(v_i = 1|\mathbf{h}) \sim \mathcal{N}\left(\sum_{j=1}^n w_{ij} h_j + a_i, \sigma_i^2\right). \tag{15}$$

Da Equação 15, é simples notar que dados normalizados em uma Gaussiana padrão, ou seja, com média zero e desvio padrão unitário ( $\sigma_i = 1$ ), o procedimento de aprendizado mantém-se o mesmo, sem maiores dificuldades.

### 3.1.2 Redes de Crença em Profundidade

Com o aumento da complexidade dos problemas, surgiu a necessidade de melhorar a representação dos dados, e aprender características profundas e intrínsecas do domínio. Esse fato fomentou o desenvolvimento das redes de crença em profundidade (*Deep Belief Networks*) (HINTON; OSINDERO; TEH, 2006), compostas por uma ou mais RBMs empilhadas que formam uma rede híbrida com conexões direcionais após o treinamento.

A intuição das DBNs é tal que, o empilhamento de RBMs auxilia no processo de extração das características de alto nível, isto é, são capazes de modelar a hierarquia dos dados apresentados à primeira camada visível. Com isso, espera-se que no decorrer das passagens pelas camadas intermediárias, a função *Log-Likelihood* tenha seu limite assintótico (*lower bound*) aumentado, representando uma melhor modelagem da distribuição dos dados.

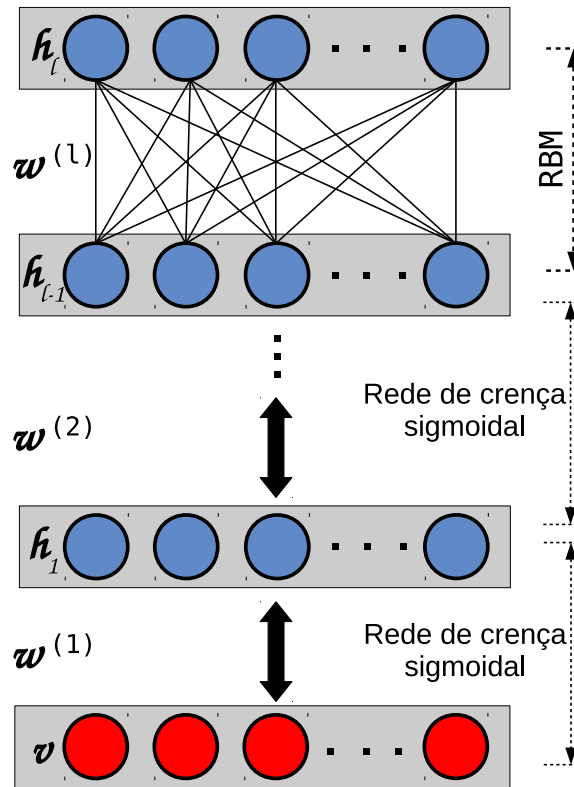
O treinamento de uma DBN é relativamente simples, cada RBM da arquitetura é treinada “isoladamente”, ou seja, a rede final é formada bloco a bloco. Esse procedimento é chamado de *Greedy Layer-wise training*, e utiliza o processo de aprendizado apresentado na Seção 3.1.1, cuja RBM que está sendo treinada em uma camada arbitrária não considera outras durante o seu treinamento. Após o treinamento de um bloco, os anteriores tornam-se redes de crença sigmoidais, uma vez que suas ativações são ditadas pela função logística, além da possibilidade de realizar tanto geração quanto inferência de dados.

A Figura 6 ilustra a arquitetura de uma DBN, em que cada RBM de uma dada camada é representada conforme ilustrado na Figura 4, novamente com os respectivos vieses ocultos. Nesse caso, temos uma DBN composta por  $l$  camadas, sendo  $\mathbf{w}^l$  as conexões neurais entre as RBMs da camada  $l$ . É importante ressaltar que as unidades ocultas da camada  $l$  tornam-se as visíveis da  $l + 1$ .

Após a etapa previamente descrita, também conhecida como pré-treinamento, Hinton et al. (HINTON; OSINDERO; TEH, 2006) propõem a realização de um ajuste adicional e final dos parâmetros da rede (*fine-tuning*), o qual pode ser realizado sob os paradigmas não-supervisionado ou supervisionado. O primeiro é realizado pelo algoritmo *Wake-Sleep* (HINTON et al., 1995), cuja intuição é apresentar os dados à camada de entrada  $\mathbf{v}$ , propagar o sinal pela rede e, atualizar as conexões de acordo com o resultado obtido. Posteriormente, é realizada a propagação “para baixo”, que representa a geração de amostras para os dados iniciais, seguida pelo ajuste dos pesos frente a geração obtida.

Já o ajuste supervisionado utiliza o algoritmo de retro-propagação, a fim de ajustar as matrizes de peso  $\mathbf{w}^l$ ,  $l = 1, 2, \dots, l$ . O algoritmo de otimização trabalha minimizando uma medida de erro, geralmente o de classificação, a partir da saída de uma camada extra adicionada ao topo da DBN após o *Greedy Layer-wise training*. Essa camada geralmente é composta por unidades logísticas ou do tipo *softmax*, podendo também ser substituída por uma técnica de classificação de padrões supervisionada, como máquinas de vetores de suporte (CORTES;

Figura 6 – Arquitetura padrão de uma DBN.



Fonte: Elaborado pelo autor.

VAPNIK, 1995).

## 3.2 Modelos Convolucionais

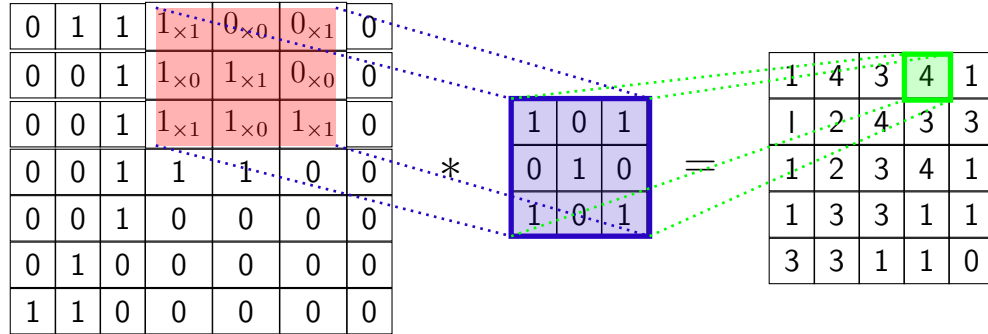
Para desenvolver o equacionamento das redes de Boltzmann restritas convolucionais - CRBMs, e consequentemente, as redes de crença profundas convolucionais - CDBNs, é importante explicar os conceitos básicos da operação de convolução e como foi o desenvolvimento das primeiras redes neurais convolucionais - CNNs, portanto, abaixo serão abordados os conceitos necessários.

### 3.2.1 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (LECUN et al., 1998) atingiram excepcional popularidade desde o início da década de 2010, tornando-se fundamentais para resolver problemas relacionados ao processamento de imagens, como classificação de imagens (TAN; LE, 2019) e segmentação (ZOPH et al., 2020). Como o nome sugere, os blocos básicos das CNNs são compostos por neurônios convolucionais, que representam *kernels* (núcleos), geralmente de dimensão  $3 \times 3$  ou  $5 \times 5$ , responsáveis por realizar operações de convolução sobre os dados de entrada. Ao aplicar os núcleos de convolução sobre os dados, essa transformação gera um

novo conjunto de matrizes, sendo estas empregadas como dados de entrada para as camadas subsequentes do modelo. No processamento de sinais, a convolução é descrita como uma multiplicação de dois sinais para gerar um terceiro (OPPENHEIM; BUCK; SCHAFER, 2001). A Figura 7 demonstra esse processo.

Figura 7 – Exemplo de aplicação de um *kernel* convolucional de dimensão  $3 \times 3$ .



A partir da Figura 7, pode-se descrever melhor o processo de convolução, assumindo as três matrizes representadas como **dados**, **kernel**, e **saída**, respectivamente da esquerda para a direita. Neste processo, o *kernel* é aplicado sobre os dados por meio da multiplicação matricial, cujo resultado é um elemento da matriz de saída, como indicado no *pixel* marcado em verde. Toda a imagem passa por esse processamento, ocorrendo o deslocamento do *kernel* de acordo com o passo (*stride*) escolhido, formando a saída completa, de tamanho reduzido ou não. Aqui, de acordo com o tamanho do passo de deslocamento do *kernel* sobre os dados, e da sua dimensão, pode ser trabalhada a dimensão da saída para posteriores camadas. Matematicamente, a convolução discreta é definida pela Equação 16 (GONZALEZ; WOODS, 2000) como segue:

$$\begin{aligned}
 (f * g)(k) &= h(k) \\
 &= \sum_{j=0}^k f(j) \cdot g(k - j),
 \end{aligned} \tag{16}$$

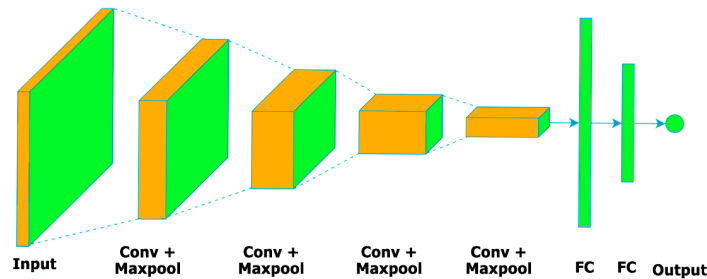
em que,  $f$  e  $g$  são sequências de tamanho  $s$ , e a equação fornece o valor do  $k$ -ésimo elemento do resultado. Deve-se notar que o tamanho de  $h$  é  $2s - 1$ , isto é,  $k$  varia de 0 a  $2s - 1$ . As sequências  $f$  e  $g$  não precisam necessariamente ter o mesmo tamanho, supondo  $f$  com  $s_1$  elementos e  $g$  com  $s_2$  elementos, o tamanho de  $k$  será dado por  $s_1 + s_2 - 1$ .

As CNNs foram propostas com uma sequência base de operações, as convoluções, a aplicação da função de ativação na saída destas, e, opcionalmente, a amostragem (*pooling*) (LECUN; KAVUKCUOGLU; FARABET, 2010). Como mencionado anteriormente, o processo de convolução nos dados representa a multiplicação matricial da janela dos dados pelo *kernel*. Posteriormente, é feita a passagem dos dados transformados por uma função de ativação, cujas opções são inúmeras, como a sigmoide, tangente hiperbólica, e ReLU, por

exemplo. Neste passo ocorre a quebra da linearidade, por exemplo, com a aplicação da função ReLU na Equação 16,  $\text{ReLU}[h(k)]$ , e naturalmente pode ocorrer a redução da dimensão dos dados, uma vez que a aplicação do *kernel* possui essa característica.

Por fim, o processo pode ser seguido pela redução da dimensão de saída por meio da camada de *pooling*, usualmente escolhendo uma janela menor que a dimensão do *kernel*. Neste passo, a maioria das aplicações utilizam a amostragem semelhante a um filtro passa-alta, deixando passar apenas os máximos valores de cada janela (*max-pooling*). Todos os passos descritos, e o processo de aprendizado em uma CNN foi amplamente discutido por Yamashita et al. (2018). Esta sequência representa um bloco convolucional, como mostrado na Figura 8, em que a rede mostrada é inspirada na arquitetura LeNet-5.

Figura 8 – Exemplo de uma arquitetura padrão para uma CNN.



Portanto, uma CNN pode ser interpretada como uma pilha de *kernels* convolucionais gerando a entrada das camadas sucessivas. Observe que algumas camadas intermediárias de agrupamento podem ser incluídas no processo e uma camada totalmente conectada (*Fully-Connected* - FC) é acoplada na parte final da arquitetura para fins de classificação. O processo de aprendizagem é realizado pelo algoritmo de retro-propagação (*backpropagation*) de erro, trabalhando com o cálculo do gradiente descendente (YAMASHITA et al., 2018).

### 3.2.2 RBMs e DBNs Convolucionais

Lee et al. (LEE et al., 2009) propuseram a RBM baseada em convolução, e consequentemente, a generalização deste modelo para sua versão profunda, a DBN convolucional. Como mostrado pelos experimentos dos autores, esses modelos atingiram resultados extremamente positivos na época, fato que representou uma quebra de paradigmas no que diz respeito às RBMs, uma vez que eram limitadas a entrada de dados em formato vetorizado, o que dificultava a aplicação em tarefas de visão computacional.

Para modelar a CDBN, primeiro é necessário definir as formulações da CRBM, uma vez que o conceito das redes de crença continua o mesmo, ou seja, empilhar várias redes de Boltzmann e treiná-las de forma independente por pares. Sejam  $\mathbf{V}$  e  $\mathbf{H}$  as camadas de neurônios visíveis e ocultos, tal que  $\mathbf{V}$  é uma matriz de  $C \times N_V \times N_V$  neurônios visíveis, com  $C$  canais de entrada (tipicamente os canais de cores de uma imagem) e dimensão  $(N_V, N_V)$ . Já  $\mathbf{H}$  consiste

em  $K$  filtros (*kernels*) convolucionais, que possuem  $N_H \times N_H$  neurônios ocultos, de dimensão  $(N_V - N_w + 1, N_V - N_w + 1)$ , em que  $N_w$  representa o tamanho do filtro convolucional. As camadas visível e oculta são conectadas pela matriz  $\mathbf{W}$  sobre todos os filtros convolucionais, de dimensão  $(K, C, N_w, N_w)$ , com vieses  $\mathbf{b}$  e  $\mathbf{c}$  associados, respectivamente, às camadas visível e oculta. A energia de uma CRBM é descrita como segue:

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{k=1}^K h^k \bullet (W^k * v) - \sum_{k=1}^K b_k \sum_{i,j} h_{ij}^k - \mathbf{c} \sum_{i,j} v_{ij}, \quad (17)$$

em que  $b_k$  e  $\mathbf{c}$  são os valores do vieses (*bias*) dos neurônios ocultos e visíveis, respectivamente. O operador  $\bullet$  é o produto dos elementos seguido pela soma da matriz resultante, enquanto  $*$  é o operador de convolução.

Como todas as variantes de RBMs são grafos bipartidos, e a probabilidade de uma configuração conjunta depende da energia fornecida, pode-se calcular as probabilidades visíveis e ocultas como na RBM padrão. Portanto, as Equações 18 e 19 descrevem as probabilidades condicionais para cada camada de uma CRBM:

$$P(h_{ij}^k = 1 | \mathbf{v}) = \sigma \left( (W^k * v)_{i,j} + b_k \right), \quad (18)$$

e,

$$P(v_{ij} = 1 | \mathbf{h}) = \sigma \left( \left( \sum_k \tilde{W}^k * h^k \right)_{i,j} + c \right), \quad (19)$$

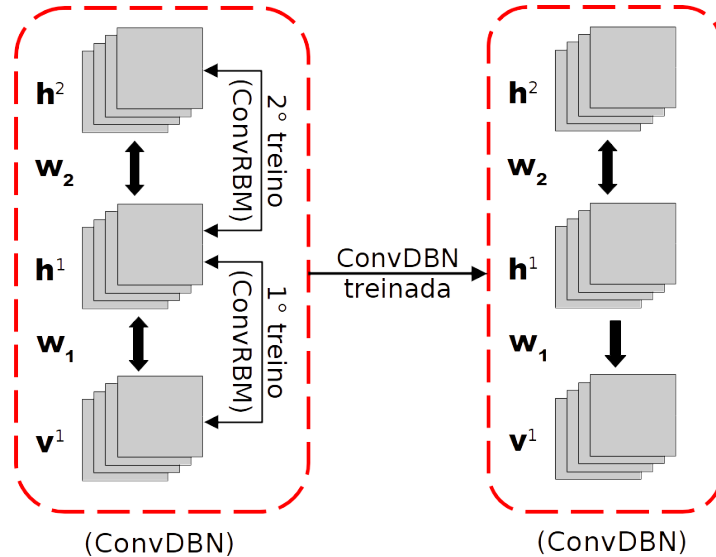
em que  $\sigma(\cdot)$  é a função logística. A Equação 18 representa a probabilidade de um neurônio oculto  $h_{ij}^k$  para um filtro convolucional  $k$  no espaço  $(N_H, N_H)$ . Já o  $\tilde{W}$  representa a matriz de pesos transposta para o emprego da convolução transposta, que faz o mapeamento para a camada visível.

Em suma, o treinamento de uma CRBM visa maximizar o logaritmo da verossimilhança sob os parâmetros convolucionais mencionados ( $\mathbf{W}$ ,  $\mathbf{b}$  e  $\mathbf{c}$ ). Para resolver este problema de otimização, a Divergência Contrastiva (HINTON, 2002) é aplicada para aproximar os gradientes do modelo e atualizar seus parâmetros segundo o gradiente estocástico ascendente, conforme apontado por Lee et al. (LEE et al., 2009). Deste modo, torna-se simples mudar a ativação dos neurônios visíveis para distribuição gaussiana, ou ainda aplicar a função de ativação ReLU, como Lee et al. (LEE et al., 2009) apontaram.

Uma vez que a CRBM foi definida, é fácil extrapolar seus conceitos para versões mais profundas, como a CDBN. Este modelo representa uma pilha de duas ou mais CRBMs treinadas de forma gulosa, ou seja, um bloco de CRBM de cada vez, que compõe o modelo CDBN final com  $L$  camadas ocultas (cada camada oculta é aproximada por uma CRBM). Para conectar cada bloco de CRBM, basta utilizar a ativação do mapa de características dos kernels convolucionais da camada  $l - 1$  para alimentar a camada  $l$ , e treinar essa CRBM com a amostragem Gibbs utilizando o método CD. Além disso, pode-se aplicar o *max-pooling*

probabilístico (LEE et al., 2009) ou o *max-pooling* padrão (BOUREAU; PONCE; LECUN, 2010). A Figura 9 ilustra uma arquitetura CDBN com duas camadas ocultas.

Figura 9 – Exemplo de uma arquitetura CDBN com duas camadas ocultas.



### 3.3 Transformada de Fourier

A transformada de Fourier é uma poderosa ferramenta matemática datada de meados do século XIX. Sua postulação indica que qualquer função periódica pode ser expressa como uma soma de senos e cossenos de diferentes frequências, cada uma ponderada por um coeficiente diferente, independentemente da complexidade da função. Se tal função não for periódica, ainda é possível aplicar a decomposição em senos e cossenos, mas a partir da integral sob as curvas. Com o advento da computação e os avanços tecnológicos, a transformada rápida de Fourier (do inglês, *Fast Fourier Transform* - FFT) foi desenvolvida (GONZALEZ; WOODS, 2000).

A transformada de Fourier pode ser aplicada em domínios contínuos ou discretos, onde esta última permite aproximações eficientes computacionalmente rápidas para cálculos. A FFT permite diversas manipulações de sinais digitais, principalmente em imagens, como a aplicação de diversos filtros no domínio da frequência (GONZALEZ; WOODS, 2000). Em suma, a transformada discreta de Fourier (do inglês, *Discrete Fourier Transform* - DFT) é aplicada a um determinado arranjo  $D$ -dimensional produzindo um arranjo complexo com a mesma dimensão, além da magnitude (ou módulo) e ângulo de fase. Matematicamente, a DFT é definida para duas dimensões, por exemplo, como segue (GONZALEZ; WOODS, 2000):

$$\begin{aligned} F(u, v) &= R(u, v) + jI(u, v) \\ &= |F(u, v)|e^{j\phi(u, v)}, \end{aligned} \tag{20}$$

onde  $R$  e  $I$  representam as componentes real e imaginária, respectivamente,  $u$  e  $v$  representam as coordenadas da matriz bidimensional e  $\phi$  é o ângulo de fase. A magnitude e o ângulo de fase do sinal de entrada são definidos pelas Equações 21 e 22, respectivamente:

$$|F(u, v)| = [R^2(u, v) + I^2(u, v)]^{1/2} \quad (21)$$

e

$$\phi(u, v) = \arctan \left[ \frac{I(u, v)}{R(u, v)} \right]. \quad (22)$$

As transformadas de Fourier têm uma ampla gama de aplicações, desde métodos clássicos de processamento de imagens até métodos baseados em aprendizado profundo. Lotfollahi et al. (LOTFOLLAHI et al., 2019), por exemplo, empregaram a transformada de Fourier na coloração de imagens digitais infravermelhas de alta definição por meio de aprendizado profundo. Os autores exploraram um domínio expressivamente difícil de adquirir dados categorizados/rotulados, as manchas histológicas. Eles mostraram que é possível empregar técnicas de aprendizado profundo não-supervisionado juntamente com a transformada de Fourier para mapear imagens do espectro infravermelho para imagens de alta resolução de manchas histológicas.

Chen et al. (CHEN et al., 2022) introduziram um modelo de rede neural profunda *end-to-end*, denominado *Fourier Image Network* (FIN). A estrutura de funcionamento não requer etapas de pré-processamento e utiliza duas ou mais imagens holográficas brutas de tipos diferentes. A FIN usa informações de frequência espacial global processadas por módulos treinados de transformada espacial de Fourier (SPAF), que usa a DFT em um de seus módulos, semelhante ao que ocorre em (LI et al., 2020). A eficácia da abordagem proposta foi demonstrada experimentalmente treinando o modelo com amostras de imagens de tecido pulmonar humano e testando cegamente em imagens de outros tecidos do corpo humano, como amostras de próstata, e glândula salivar. De acordo com os autores (CHEN et al., 2022), a velocidade do FIN é 27 vezes mais rápida que algoritmos iterativos baseados em propagação de onda para reconstrução de imagem holográfica. Além disso, em comparação com os modelos existentes baseados em CNNs, a FIN exibe grande desempenho de generalização e é mais rápida na fase de inferência.

## 4 Proposta

A partir do levantamento bibliográfico e de estudos práticos tangentes à linha principal da tese, ou seja, a classificação de AVC utilizando redes de Boltzmann e suas variantes, foram observados três pontos principais: (i) a ausência de bancos de dados volumosos e anotados para o problema de classificação de AVC; (ii) a necessidade de técnicas menos complexas que as estado-da-arte em aprendizado profundo (CNNs); e, (iii) a ausência de trabalhos que empregam RBMs/DBNs não-convolucionais e convolucionais neste domínio de trabalho.

Portanto, para mitigar os problemas supracitados, foi desenvolvido um *framework* utilizando as máquinas de Boltzmann restritas e a transformada de Fourier, empregando a multimodalidade para os dados de entrada, utilizando uma base com baixo volume de dados. É importante ressaltar que a abordagem proposta neste *framework* possui flexibilidade para empregar o aprendizado profundo por meio das DBNs e CDBNs. O *framework* está disponível nos respectivos repositórios do GitHub: <[https://github.com/MateusRoder/fourier-rbm\\_avc](https://github.com/MateusRoder/fourier-rbm_avc)>, e <<https://github.com/gugarosa/learnergy>>. Por fim, a abordagem será descrita adiante.

### 4.1 RBMs e DBNs Multimodais baseadas em Fourier

Nesta seção serão abordados os temas a respeito da abordagem proposta empregada, ou seja, a utilização da multimodalidade para as RBMs e DBNs não-convolucionais e convolucionais no problema de classificação de AVC.

Esta abordagem se propõe a representar informações em diferentes domínios através da transformada de Fourier. Deste modo, é esperado que o uso da fusão de informações para o treinamento de RBMs e DBNs seja bem sucedido no problema de classificação de AVC em seus tipos (isquêmico e hemorrágico). Além disso, essas redes foram escolhidas devido a sua baixa complexidade e ausência de trabalhos envolvendo os temas pesquisados.

A fusão de informações agrega dados de diferentes domínios, e muitas vezes é denotada como multimodal devido à diferença nas distribuições agregadas. A multimodalidade pode ser usada com  $N$  tipos de dados, como bi e tri-modais. No que diz respeito a este trabalho, os dados são compostos por imagens originais de tomografia computadorizada e suas respectivas transformadas de Fourier, onde esta última é composta por duas componentes diferentes: a magnitude (Equação 21) e o ângulo de fase (Equação 22).

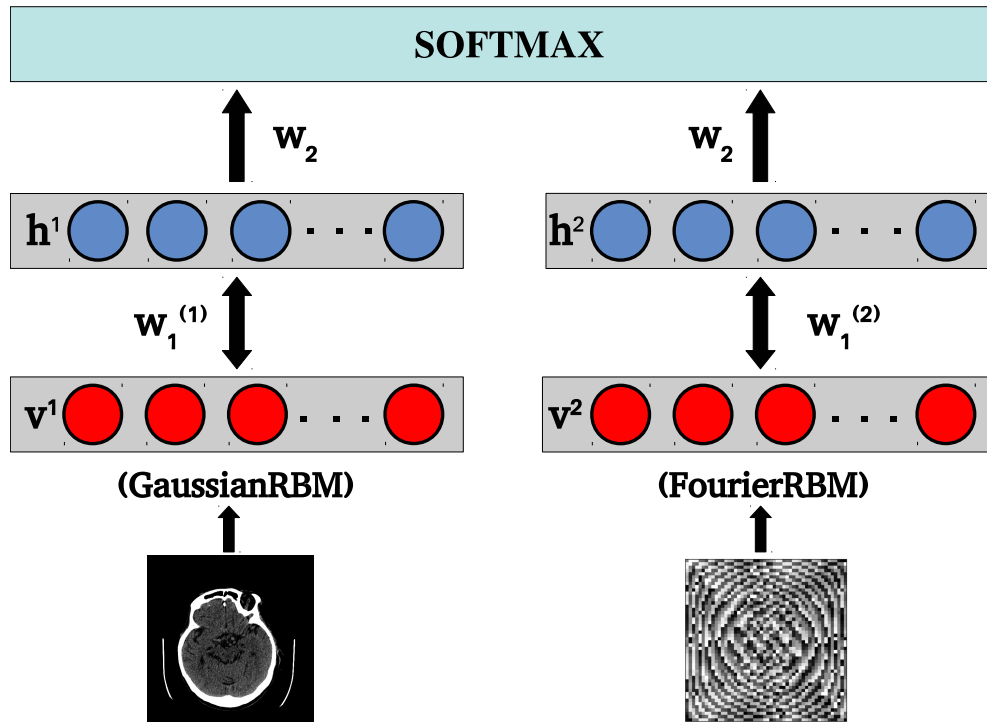
Seja a abordagem proposta denotada por MultFRRBM (RBM Multimodal baseada em Fourier), e suas variações MultFRRBM-P, MultFRRBM-M, e MultFRRBM-PM, denotando quais componentes são empregadas na multimodalidade<sup>5</sup>. Formalmente, as representações dos

<sup>5</sup> P representa a fase (do inglês, *phase*), M representa a magnitude (do inglês, *magnitude*), e PM significa

dados de entrada podem ser definidas da seguinte forma: seja  $l$  uma imagem com componentes espaciais  $(x_1, x_2) \in \mathbb{R}^2$ . Seja sua transformada de Fourier 2-D denotada por  $F$ , com módulo  $|F(x_1, x_2)|$  e ângulo de fase  $\phi(x_1, x_2)$  definido pelas Equações 21 e 22, respectivamente. De forma complementar, seja FCDBN (Fourier-CDBN, ou seja, CDBN multimodal baseada em Fourier) a abordagem proposta para os modelos CDBN, com as variantes: FCDBN-P, FCDBN-M, e FCDBN-PM, representando as componentes empregadas na multimodalidade, assim como as MultFRRBM.

Para facilitar o entendimento, serão apresentadas separadamente as versões não-convolucionais e convolucionais, iniciando pelos modelos MultFRRBM. Dadas as três arquiteturas MultFRRBM possíveis, elas são treinadas sob a abordagem padrão de aprendizado não-supervisionado (HINTON, 2002) e ajustadas para a tarefa de classificação por meio de uma camada *Softmax* adicional, que recebe as ativações das camadas ocultas  $\mathbf{h}^1$  e  $\mathbf{h}^2$  para dois tipos de dados, ou  $\mathbf{h}^1$ ,  $\mathbf{h}^2$  e  $\mathbf{h}^3$  para três tipos de dados. A Figura 10 ilustra uma MultFRRBM-P com duas entradas, onde a primeira RBM (GaussianRBM) recebe a imagem original da tomografia e a segunda (FourierRBM) recebe o ângulo de fase da DFT aplicada na imagem.

Figura 10 – Arquitetura MultFRRBM-P com dois tipos de entradas, isto é, a imagem original e o ângulo de fase correspondente a esta a partir da DFT.

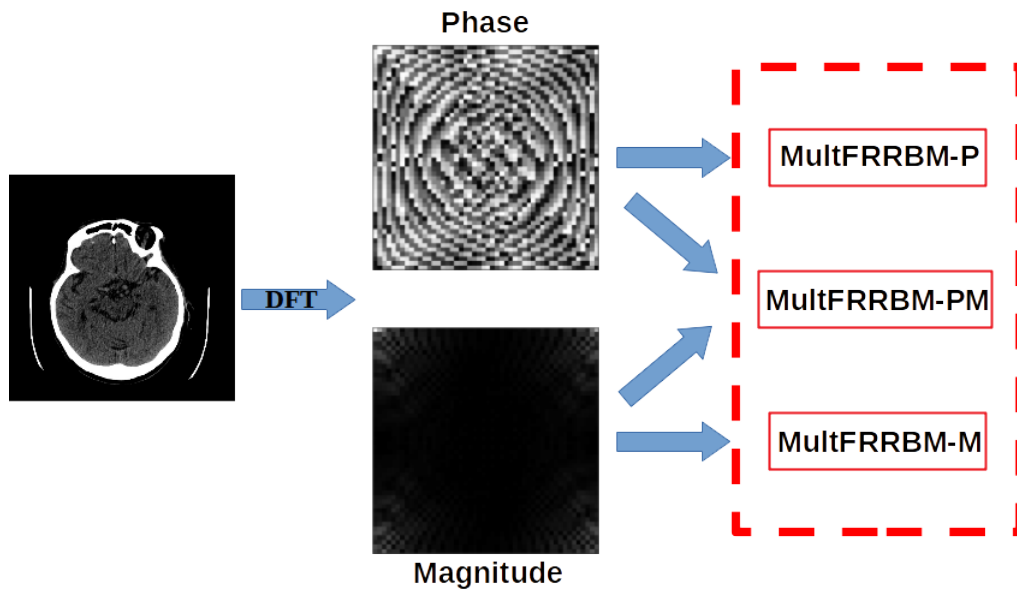


Nesta abordagem, as RBMs multimodais abordam o problema empregando unidades visíveis Gaussianas, ou seja, *pixels* de imagem como valores contínuos (normalizados), a partir das imagens originais em escala de cinza. Inicialmente, uma ou mais RBMs Gaussianas são empregadas para processar os dados transformados pela DFT, ou seja, se a arquitetura final usa

ambos.

o ângulo de fase (MultFRRBM-P), os dados são normalizados para uma Gaussiana padrão e alimentados para as RBMs. O mesmo processo ocorre para RBMs com magnitude (MultFRRBM-M), onde a magnitude do espectro de frequência é normalizada em uma Gaussiana padrão para servir outra RBM Gaussiana. Finalmente, quando ambas componentes da DFT são empregadas (MultFRRBM-PM), fase e magnitude são normalizadas e alimentadas em duas RBMs Gaussianas (simultaneamente). A Figura 11 ilustra o processamento descrito.

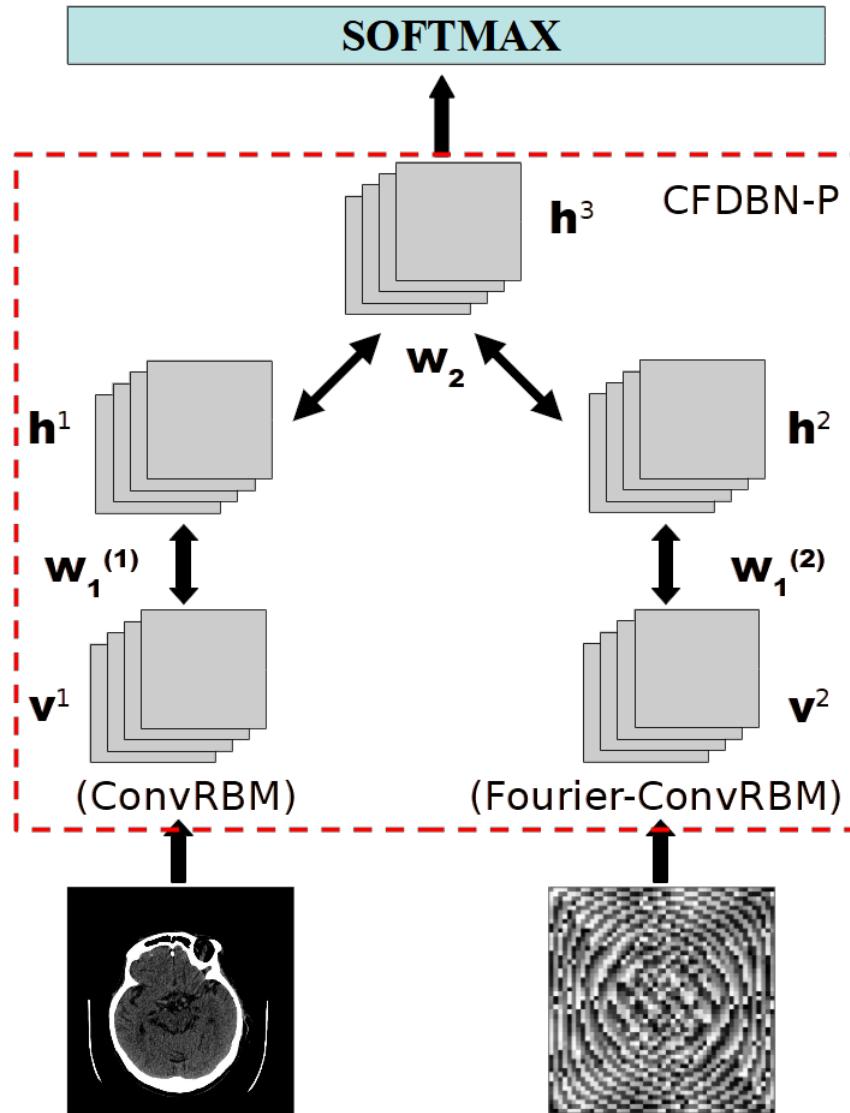
Figura 11 – Esquema da transformação aplicada para a entrada nos modelos multimodais.



Como mencionado anteriormente, a FCDBN possui três arquiteturas possíveis com as componentes de Fourier, e elas também são treinadas sob o paradigma de aprendizado não-supervisionado (HINTON; OSINDERO; TEH, 2006; LEE et al., 2009), em que a segunda camada oculta ( $\mathbf{h}^3$ ) recebe as ativações dos filtros convolucionais de todas as CRBMs ( $\mathbf{h}^1$  e  $\mathbf{h}^2$ ) que processam os dados de entrada, destacando que as ativações dos filtros convolucionais são amostradas com o *max-pooling* de tamanho  $2 \times 2$  após o treinamento de cada CRBM. Esta abordagem é representada na Figura 12, em que  $\mathbf{h}^1$  e  $\mathbf{h}^2$  (primeira camada oculta) representam os filtros convolucionais dos dois modelos de entrada. Depois disso, o modelo FCDBN é ajustado para a tarefa de classificação por meio de uma camada *Softmax* adicional, que recebe as ativações dos filtros convolucionais da última camada oculta,  $\mathbf{h}^3$  (ou  $\mathbf{h}^L$  para mais camadas ocultas). A Figura 12 ilustra uma FCDBN-P com duas camadas ocultas e duas entradas, onde a primeira CRBM recebe a imagem CT original e a segunda (Fourier-ConvRBM) recebe o ângulo de fase da DFT aplicada à imagem.

Neste trabalho, as FCDBNs multimodais abordam o problema com neurônios visíveis gaussianos para as operações de convolução, ou seja, *pixels* da imagem como valores contínuos de uma distribuição Gaussiana das imagens originais em tons de cinza. Para a abordagem proposta, CRBMs gaussianas são empregadas para processar os dados da DFT, ou seja, se uma determinada arquitetura usa o ângulo de fase (FCDBN-P), as duas modalidades de dados são

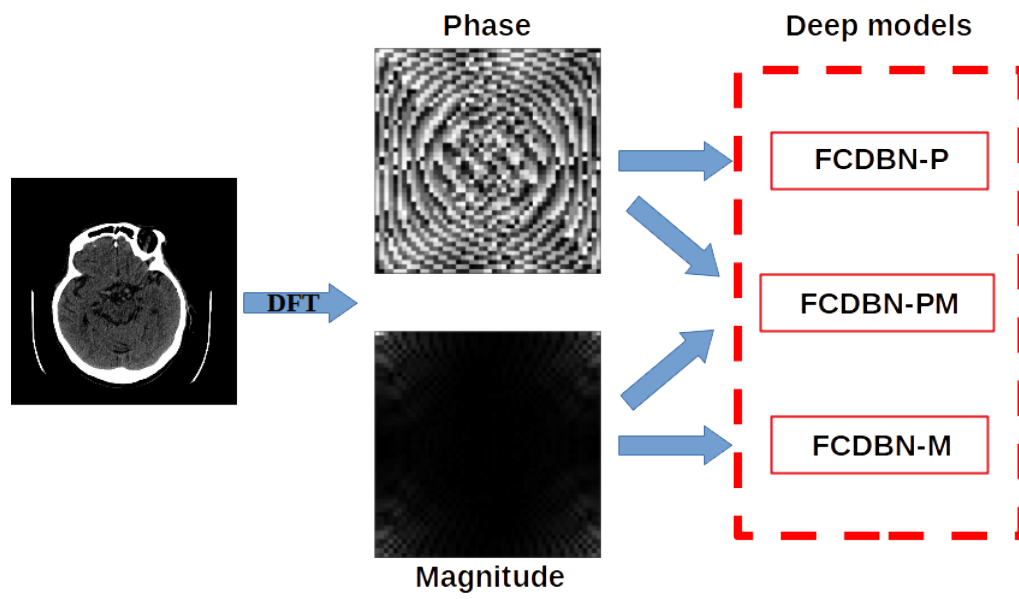
Figura 12 – Arquitetura FCDBN-P com dois tipos de dados de entrada, isto é, as imagens originais e a representação dos ângulos de fase da DFT.



normalizadas em uma Gaussiana padrão (média zero e variância unitária) e fornecidas para as duas CRBMs. O mesmo processo ocorre para FCDBNs com a magnitude (FCDBN-M), onde a magnitude do espectro de frequência é normalizada para uma Gaussiana padrão e alimenta outra CRBM. Finalmente, quando ambas as componentes da DFT são empregadas (FCDBN-PM), a fase e a magnitude são normalizadas e fornecidas a duas CRBMs, simultaneamente, ao passo que a terceira CRBM recebe as imagens de CT originais. A Figura 13 ilustra os modelos gerados pela abordagem proposta para as FCDBNs.

Portanto, a abordagem proposta agrega informações de diferentes domínios, ou seja, o componente espacial originado das imagens de tomografia computadorizada, e a frequência originada das componentes da transformada de Fourier, tanto para os modelos não-convolucionais quanto para os convolucionais. Adicionalmente, a multimodalidade empregada possibilita a análise de qual configuração é mais adequada para o problema em questão e confere maior

Figura 13 – Modelos multimodais resultantes da abordagem proposta.



flexibilidade/escolha à aplicação.

## 5 Metodologia

Neste capítulo é descrita a metodologia da abordagem proposta nesta tese. Essencialmente, são abordados os tópicos relacionados aos problemas (ii) e (iii) citados no Capítulo 4, que se referem às questões de pesquisa, Seção 1.2, endereçando o problema de classificação de AVC com redes neurais menos complexas que as CNNs estado-da-arte, uso de multimodalidade e visualização do aprendizado dos modelos.

### 5.1 O Banco de Dados

Pereira et al. (PEREIRA et al., 2018) propuseram o conjunto de dados empregado, que consiste em tomografias cerebrais, incluindo pacientes<sup>6</sup> com (a) cérebro saudável, (b) acidente vascular cerebral isquêmico e (c) acidente vascular cerebral hemorrágico. O número total de imagens é de 300, em que 100 representam cérebros saudáveis e 200 com acidente vascular cerebral (100 hemorrágico e 100 isquêmico). Além disso, as imagens são compostas por padrões irregulares, iluminação irregular e características estruturais diferentes.

Além disso, Pereira et al. (PEREIRA et al., 2018) utilizaram duas metodologias adicionais de pré-processamento para geração de imagens, sendo elas, segmentação de imagens para extrair apenas o cérebro, e imagens cerebrais geradas a partir dos mapas de densidade radiológicos (mapa de atributos de densidade tecidual); no entanto, reduzir etapas de pré-processamento pode evitar possíveis atrasos no sistema em aplicações do mundo real. Um ponto interessante é que nos resultados experimentais, os autores mostraram que essas metodologias de pré-processamento não melhoraram o desempenho das CNNs empregadas. Assim, o presente trabalho considera apenas as imagens originais em escala de cinza.

Cada imagem original tem uma resolução de  $512 \times 512$  *pixels*; porém, foram redimensionadas para  $50 \times 50$  visando diminuir a utilização de recursos computacionais necessários e o tempo de treinamento das RBMs. Já para os modelos baseados em convolução, as imagens foram redimensionadas para  $150 \times 150$  devido à possibilidade de tratar os dados em seu formato original bidimensional. As imagens em escala de cinza foram normalizadas de acordo com uma distribuição Gaussiana padrão, ou seja, média zero e variância unitária. A Figura 14 ilustra algumas amostras de exames baseados em tomografia de um cérebro saudável, acidente vascular cerebral hemorrágico e acidente vascular cerebral isquêmico, respectivamente, com destaque nas regiões afetadas.

Da Figura 14, nota-se que o cérebro saudável não possui qualquer indicativo visual de anomalia, o que não ocorre nos casos hemorrágico e isquêmico. Para o AVC hemorrágico, há a

---

<sup>6</sup> A identificação dos pacientes foi omitida por razões éticas.

Figura 14 – Tomografias do cérebro: saudável, com acidente vascular cerebral hemorrágico e com acidente vascular cerebral isquêmico, respectivamente.



presença de uma massa hiper-densa, com alta intensidade de cor, ou seja, a radiação (ondas de raio-x) tem grande dificuldade para atravessar o material. Por outro lado, a lesão do AVC isquêmico gera pouca resistência para a passagem das ondas de raio-x, fazendo com que a região fique escura e sem destaque de intensidade.

## 5.2 Configuração Experimental

Os experimentos foram separados em duas frentes, ou seja, primeiro para os modelos RBM multimodais (não-convolucionais), e posteriormente para as CDBNs multimodais. Sobre a abordagem proposta, isto é, a multimodalidade em RBMs utilizando a transformada de Fourier, esta compreende quatro diferentes arquiteturas de RBMs, porém com o mesmo processo de treinamento, ou seja, redes treinadas com o método CD e com igual número de épocas para cada RBM que compõe o modelo multimodal. As arquiteturas de referência são as redes neurais convolucionais,  $CNN_1$ , e  $CNN_2$ , propostas por Pereira et al. (PEREIRA et al., 2018), enquanto as RBMs correspondem a uma puramente Gaussiana, e as variantes multimodais empregadas, ou seja, MultFRRBM-P, MultFRRBM-M e MultFRRBM-PM. A configuração de hiper-parâmetros dessas arquiteturas é representada pela Tabela 2.

| Arquitetura  | # Neurônios Ocultos | Taxa de Aprendizado  | Momento       |
|--------------|---------------------|----------------------|---------------|
| GaussianRBM  | 2.000               | 0,0001               | 0,5           |
| MultFRRBM-P  | 500; 1.500          | 0,001; 0,0001        | 0,9; 0,5      |
| MultFRRBM-M  | 500; 1.500          | 0,001; 0,0001        | 0,9; 0,5      |
| MultFRRBM-PM | 250; 250; 1.500     | 0,001; 0,001; 0,0001 | 0,9; 0,9; 0,5 |

Tabela 2 – Configuração de hiper-parâmetros para os modelos baseados em RBMs.

Os valores descritos na Tabela 2 correspondem a hiper-parâmetros escolhidos empi-

ricamente, selecionados em experimentos preliminares. Além disso, o número de neurônios ocultos é baseado no modelo âncora (GaussianRBM), que apresentou um bom desempenho com 2.000 unidades. Tal valor é utilizado como a quantidade máxima de neurônios para todas as arquiteturas propostas, proporcionando assim uma comparação justa entre modelos que utilizam uma ou mais RBMs. Por outro lado, tanto as taxas de aprendizado quanto os momentos são escolhidos para fornecer convergências mais suaves (valores mais baixos), enquanto as 50 épocas de treinamento são escolhidas a partir de análises prévias para que os modelos tenham tempo para a estabilização do aprendizado, e mini-lotes de tamanho igual a 10, complementam os hiper-parâmetros dos modelos.

Com relação à última etapa de ajuste (ajuste-fino), é adicionada uma camada *Softmax* na parte superior das RBMs, treinada com o otimizador Adam (KINGMA; BA, 2015) usando uma taxa de aprendizado de 0,00001 para as RBMs e 0,001 para a *Softmax*. Essa camada de classificação adicional também tem seus hiper-parâmetros escolhidos empiricamente. O ajuste-fino compreende 50 épocas usando mini-lotes com tamanho igual a 60. Por fim, a medida de avaliação empregada é a mesma utilizada por Pereira et al. (PEREIRA et al., 2018), ou seja, é a acurácia padrão.

Isto posto, adiante é descrita a configuração experimental para os modelos CDBNs multimodais. Da mesma forma, as arquiteturas propostas e empregadas são: FCDBN-P, FCDBN-M, e FCDBN-PM; com a CDBN sendo o modelo base. A configuração dos hiper-parâmetros dessas arquiteturas é apresentada abaixo, na Tabela 3.

| Arquitetura | # Formato do filtro | # Filtros  | Taxa de aprendizado                                        | Momentum | Camadas |
|-------------|---------------------|------------|------------------------------------------------------------|----------|---------|
| CDBN        | (5, 5)              | 32         | $10^{-6}$ ; $10^{-7}$                                      | 0.5      | 2       |
| FCDBN-P     | (5, 5)              | 32, 32     | $10^{-5}$ ; $10^{-6}$ ; $10^{-6}$                          | 0.5      | 2       |
| FCDBN-M     | (5, 5)              | 32, 32     | $10^{-5}$ ; $10^{-6}$ ; $10^{-6}$                          | 0.5      | 2       |
| FCDBN-PM    | (5, 5)              | 32, 32, 32 | $10^{-6}$ ; $10^{-6}$ ; $10^{-6}$ ; $10^{-7}$ <sup>7</sup> | 0.5      | 2       |

Tabela 3 – Configuração dos hiper-parâmetros para os modelos CDBN com multimodalidade (FCDBN-).

A partir da Tabela 3, é importante ressaltar que todos os hiper-parâmetros também foram escolhidos empiricamente nos experimentos exploratórios, levando às melhores configurações em termos de convergência e estabilidade. A forma dos filtros convolucionais é igual para todas as camadas e dados de entrada, para tornar justa a análise comparativa, e o mesmo raciocínio é empregado no número do filtros, mantendo o mesmo que o modelo de âncora (CDBN) para cada CRBM adicionada. Além disso, a taxa de aprendizado representa os valores que proporcionaram melhor estabilidade no processo de aprendizado. Adicionalmente, todos os modelos têm *momentum* de 0,5 na descida do gradiente estocástico para todas as CRBMs. Por fim, foram empregadas apenas duas camadas ocultas para todos os modelos CDBN e

<sup>7</sup> Este modelo requereu taxa de aprendizado mais baixa para a convergência.

FCDBN-, uma vez que experimentos preliminares não indicaram melhorias de desempenho com a adição de muitas camadas.

Com relação ao fluxo do processo de treinamento, todos os modelos são inicialmente treinados de maneira não-supervisionada (fase de pré-treinamento), como as CDBNs foram propostas (LEE et al., 2009), durante 50 épocas, com um mini-lote de 50 amostras. Após o pré-treinamento, é adicionada uma camada *Softmax* sobre todas as CDBNs (multimodais ou não) para compor de forma eficiente os modelos de classificação, nos quais é empregado o otimizador Adam (KINGMA; BA, 2015) com taxa de aprendizado de  $10^{-4}$  para a camada *Softmax*, por 50 épocas de ajuste-fino, e 60 amostras por mini-lote. Como o conjunto de dados tem distribuição de classe igual, ou seja, é balanceado, a acurácia padrão é mantida como medida de desempenho, conforme Pereira et al. (PEDEMONTE et al., 2018).

Considerando o conjunto de dados empregado, optou-se por seguir a metodologia proposta por Pereira et al. (PEREIRA et al., 2018), onde os autores forneceram duas configurações de dados distintas: uma divisão meio a meio entre conjuntos de treinamento e teste (50/50) e três quartos dos dados para treinamento (75/25). Porém, para os modelos convolucionais propostos foi empregada uma partição adicional, 25/75, a fim de avaliar se os modelos baseados nas máquinas de Boltzmann também sofrem com uma menor quantidade de dados, como é visto em CNNs. Além disso, os experimentos foram avaliados sob 15 execuções independentes para reduzir a estocasticidade dos modelos, remover qualquer viés de dados baseado na divisão destes, e fornecer execuções suficientes para realizar um teste estatístico. Para tal, foi empregado o teste ranqueado de Wilcoxon (WILCOXON, 1945) com 5% de significância.

Por fim, para executar os experimentos citados, foi utilizada uma máquina com processador AMD Ryzen 5 3600X contendo 12 núcleos de até 3.800MHz, 16Gb de memória RAM, e uma placa de vídeo GeForce GTX 1600 com 6Gb de memória. Todos os experimentos foram realizados utilizando a linguagem Python, a partir da biblioteca Learnergy (RODER; ROSA; PAPA, 2020), onde serão inseridos os modelos desenvolvidos que estão disponíveis no repositório: <[https://github.com/MateusRoder/fourier-rbm\\_avc](https://github.com/MateusRoder/fourier-rbm_avc)>.

## 6 Resultados Experimentais

Este capítulo é dedicado à apresentação dos resultados experimentais abordando a resolução do problema (ii) e (iii), citados no Capítulo 4, utilizando a multimodalidade de dados para as Redes de Boltzmann, não-convolucionais e convolucionais, com a transformada de Fourier, no problema de classificação de AVC com o banco de dados fornecido por Pereira et al. (PEREIRA et al., 2018). São apresentadas três seções, a primeira (Seção 6.1) diz respeito às análises visuais do aprendizado não-supervisionado, a segunda (Seção 6.2) aborda a avaliação dos modelos para a classificação de AVC, enquanto a última (Seção 6.3) analisa a complexidade das arquiteturas empregadas.

### 6.1 Aprendizado Não-Supervisionado

A análise visual dos pesos aprendidos por uma técnica de aprendizado de máquina, ou de aprendizado profundo, sob o paradigma não-supervisionado, é frequentemente uma boa ferramenta para analisar os padrões que foram aprendidos e fornecer interpretabilidade ao modelo. Com essa abordagem é possível observar como as informações de baixo e alto nível são capturadas, se aproximando do processamento no córtex visual humano. Portanto, as análises referentes aos modelos não-convolucionais e convolucionais são feitas nas seguintes subseções.

#### 6.1.1 Análise Visual Não-Convolutional

Frente aos modelos empregados neste estudo, é possível analisar as conexões das RBMs, diferindo em relação à modelagem de dados para cada uma, bem como na quantidade de neurônios em cada arquitetura. As Figuras 15, 16, e 17 mostram as visualizações referentes às conexões da variante MultFRRBM-PM, em que a Figura 15 mostra a RBM que recebe a fase da DFT, a Figura 16 representa a RBM que recebe a magnitude da DFT, e Figura 17 recebe as imagens normalizadas pela Gaussiana. As RBMs possuem, respectivamente, 250, 250, e 1.500 neurônios ocultos, como explicado na Tabela 2.

Da Figura 15, nota-se que foram aprendidos padrões previamente visualizados com o exemplo da imagem transformada, na Figura 11, em que a componente fase da DFT apresenta formas onduladas que percorrem do centro para as bordas da imagem. Essas ondulações representam justamente o comportamento do ângulo de fase (cíclico). Deste modo, acredita-se que a RBM conseguiu extrair informações representativas, e possivelmente discriminantes, sobre os padrões comportamentais da fase para este banco de dados, e, na etapa de classificação esta conclusão será validada, positiva ou negativamente.

A respeito da Figura 16, nota-se que foram aprendidas conexões (pesos) que parecem

Figura 15 – RBM alimentada com a fase.

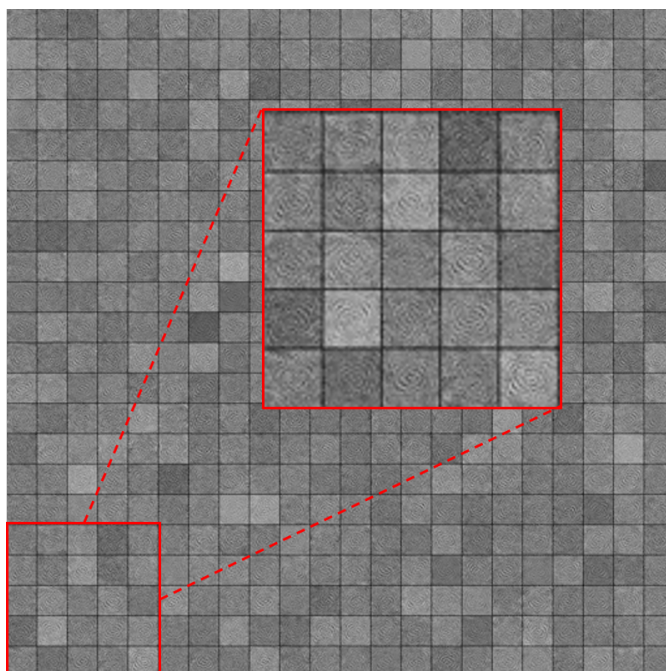
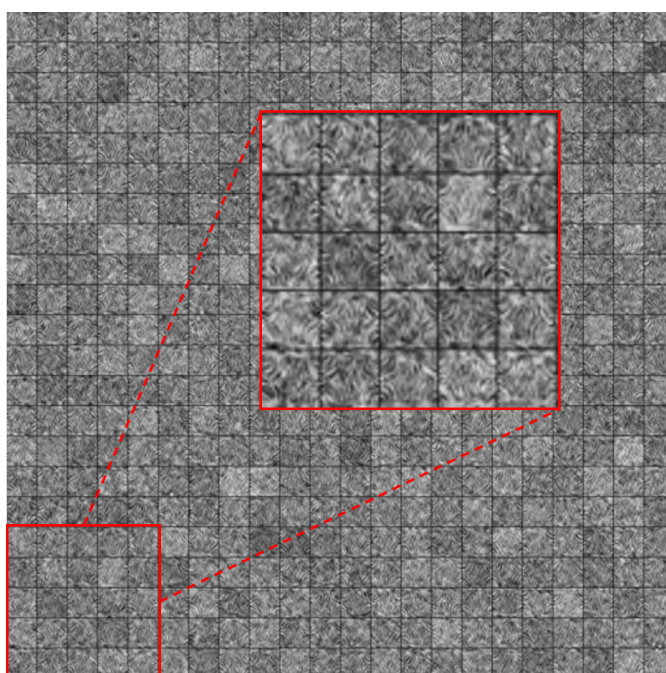
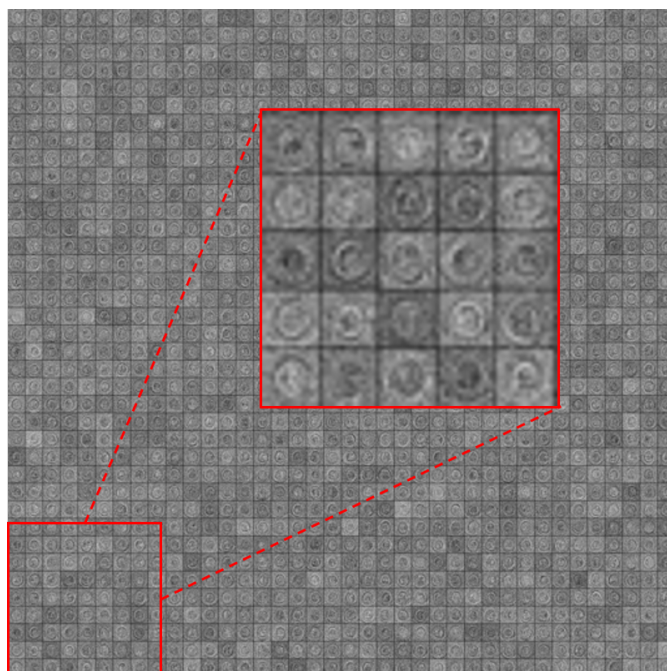


Figura 16 – RBM alimentada com a magnitude.



ressaltar o contraste da componente magnitude da DFT, principalmente ao compararmos com o exemplo da imagem transformada na Figura 11, que mostra, com certa dificuldade, alguns comportamentos ondulados de borda para o centro da imagem. A RBM responsável por essa componente conseguiu ressaltar informações sobre os padrões comportamentais da magnitude, porém, qualitativamente, os pesos aprendidos parecem não representar padrões discriminativos para diferenciar os tipos de AVC, o que será validado com a análise de classificação.

Figura 17 – RBM alimentada com dados Gaussianos.



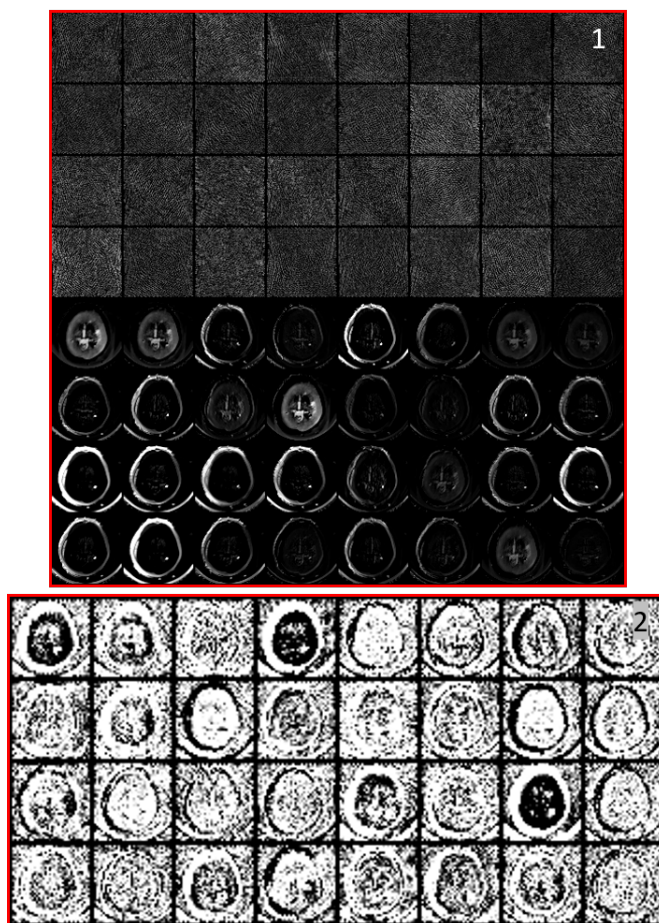
Por fim, a Figura 17, ilustra padrões cerebrais bem definidos, principalmente no que diz respeito a contornos e formas das imagens que foram apresentadas à rede no treinamento. Também não foram encontradas representações diretas de lesões cerebrais (AVC) nos pesos da rede, porém, a combinação das conexões com diferentes anatomias cerebrais, e tons de cinza, pode vir a caracterizar de forma mais discriminativa as diferentes classes do problema após o ajuste-fino, o que será validado com a análise de classificação.

### 6.1.2 Análise Visual Convolutacional

Tratando-se dos modelos convolucionais empregados, é possível analisar o mapa de ativações dos filtros convolucionais de cada camada oculta das CRBMs, diferindo em relação à modelagem de dados para cada uma. Portanto, as Figuras 18, 19, e 20 mostram as visualizações referentes às ativações das camadas ocultas 1 (imagem do quadro superior) e 2 (imagem do quadro inferior), para uma dada imagem de tomografia apresentada, utilizando as variantes FCDBN-P, FCDBN-M, e CDBN, respectivamente, como exemplo visual. O modelo FCDBN-P é alimentado com a imagem original e a componente de fase da transformada de Fourier, a FCDBN-M recebe a imagem original e a magnitude de sua transformada, enquanto a CDBN recebe apenas os dados originais.

Da Figura 18, nota-se que os filtros convolucionais ativados na camada 1 conseguem trazer uma boa representação dos dados de entrada, ou seja, é possível vê-los como uma combinação dos padrões presentes nas imagens de tomografia apresentadas, com a fase sendo bem representada na parte superior da imagem 1, e características cerebrais como contorno e tecido interno, bem representados na parte inferior da imagem 1. Além disso, na camada 2,

Figura 18 – FCDBN alimentada com a fase.

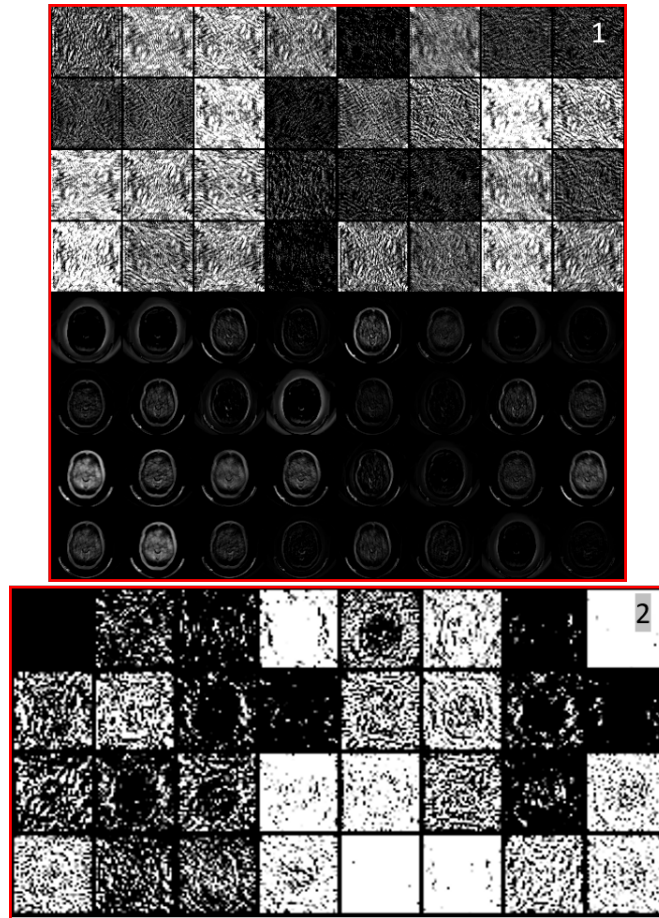


nota-se que as ativações realizaram uma sobreposição entre os padrões da camada anterior, o que é interessante uma vez que regiões internas do cérebro ganharam destaque, sendo possivelmente uma combinação (a arquitetura) eficiente para a classificação de AVC.

A partir da Figura 19, observa-se que os filtros convolucionais ativados na camada 1 trazem boa representação dos dados de entrada, como o comportamento da magnitude na parte superior da imagem, enquanto na parte inferior da mesma notam-se variações do cérebro, e características como contorno e tecido interno. Porém, na camada 2, as ativações se sobrepueram de tal forma que é difícil encontrar padrões cerebrais nítidos para descrever o domínio, indicando que talvez não seja uma combinação tão eficiente para a classificação. Esse fato possui relação com as imagens geradas pela componente magnitude da transformada de Fourier, uma vez que possuem características como baixo contraste entre os tons de cinza, e uma grande quantidade dos *pixels* são pretos.

Por fim, da Figura 20, observa-se que os filtros convolucionais na camada 1 trazem uma representação nítida do cérebro, com regiões de contorno bem definidas, e tecidos cerebrais em diferentes tonalidades de cinza. Entretanto, ao analisar as ativações da camada 2 a partir da 1, vê-se que os padrões já não são tão aparentes, e passam a ter destaque para a região tecidual interna do cérebro, concomitante a regiões muito escuras, fato que pode não ser tão

Figura 19 – FCDBN alimentada com a magnitude.



discriminativo quanto os padrões observados na Figura 18, com a FCDBN-P. Mesmo assim, a CDBN parece ser superior em representatividade e extração de características que a FCDBN-M (Figura 19).

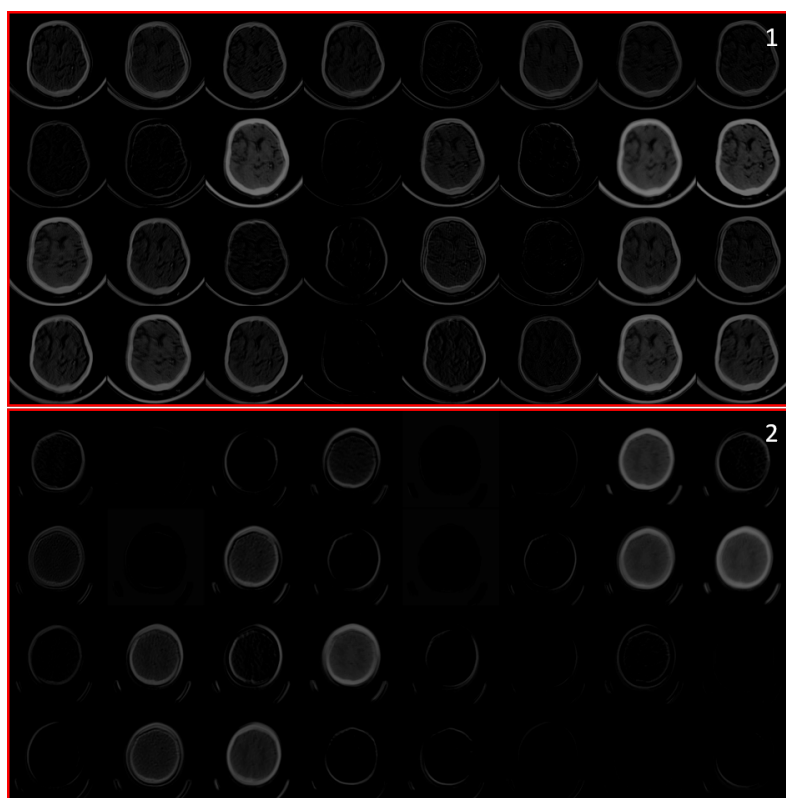
## 6.2 Avaliação dos Modelos para Classificação de AVC

Nesta seção são apresentados os resultados de classificação dos modelos propostos, separadamente em duas subseções para melhor entendimento do impacto da fusão de informação nos modelos não-convolucionais e convolucionais.

### 6.2.1 Classificação com Modelos Não-Convolutionais

A Tabela 4 apresenta os resultados médios de taxa de acerto, e seus respectivos desvios padrão, obtidos das 15 execuções independentes, considerando ambas divisões de dados, ou seja, 50/50 e 75/25. Além disso, o teste Wilcoxon com um valor de  $p$  de 5% foi aplicado aos modelos propostos seguindo o raciocínio: selecionar a melhor acurácia média em uma divisão de dados e aplicar o teste contra todos os outros valores disponíveis para testar (um-contra-todos), permitindo marcar com ( $=$ ) os resultados estatisticamente semelhantes, e com ( $\neq$ )

Figura 20 – CDBN alimentada com os dados originais (CT).



os diferentes. Por fim, a maior média alcançada por um determinado modelo em cada divisão é sublinhada.

| Arquitetura                             | 50/50               | 75/25                   |
|-----------------------------------------|---------------------|-------------------------|
| GaussianRBM                             | <u>99,66 ± 0,52</u> | 99,66 ± 0,52 (=)        |
| MultFRRBM-P                             | 98,76 ± 1,58 (! =)  | <u>99,72 ± 0,40 (=)</u> |
| MultFRRBM-M                             | 98,53 ± 1,70 (! =)  | 99,49 ± 0,60 (=)        |
| MultFRRBM-PM                            | 97,94 ± 1,66 (! =)  | 99,32 ± 0,71 (=)        |
| CNN <sub>1</sub> (PEREIRA et al., 2018) | 93,46 ± 16,54       | 97,20 ± 2,45            |
| CNN <sub>2</sub> (PEREIRA et al., 2018) | 83,55 ± 13,09       | 77,33 ± 22,24           |

Tabela 4 – Taxas de acerto médias e seus respectivos desvios padrão para o conjunto de teste considerando a divisão de dados proposta pelos autores do banco de dados.

Da Tabela 4, pode-se observar que todos os modelos RBM alcançaram maior taxa de acerto que as CNNs, considerando ambas as divisões de dados. Tais resultados são notáveis, pois modelos não-convolucionais obtiveram uma melhor taxa de acerto em uma tarefa tão importante. Além disso, os modelos baseados em RBMs empregados têm um pré-processamento

mínimo de imagem, e baixa complexidade de arquitetura, com no máximo 2.000 de neurônios ocultos (sendo uma camada visível e uma oculta).

De um modo geral, os melhores resultados foram com a GaussianRBM para a divisão 50/50, alcançando um desempenho consideravelmente maior, aproximadamente 7%, que o melhor modelo convolucional ( $CNN_1$ ). Considerando a mesma divisão, os modelos multimodais não tiveram um desempenho tão bom quanto o unimodal, indicando que a multimodalidade pode precisar de mais dados de treinamento; no entanto, seus resultados ainda foram superiores aos obtidos pela  $CNN_1$  e  $CNN_2$ .

Com relação à segunda divisão (75/25), houve uma melhora nos modelos multimodais, mas, ainda assim, estatisticamente iguais ao modelo unimodal, e todas as abordagens obtiveram uma acurácia média superior a 99%. Dentre esses modelos, a MultFRRBM-P se destacou, alcançando 99,72% e superando tanto o modelo unimodal quanto o modelo de linha de base. Além disso, obteve o menor desvio padrão, enquanto todos os outros modelos baseados em RBMs tiveram melhor acurácia média e menor desvio padrão do que as CNNs.

Vale ressaltar que os modelos MultFRRBM-PM obtiveram a menor acurácia média entre os modelos baseados em RBM, indicando a componente de magnitude agrega menos informação que a componente de fase, como também mostrado pelo professor Kundur em suas notas de aula (KUNDUR, 2013). Além disso, observa-se na imagem da magnitude uma grande quantidade de *pixels* pretos presentes (Figura 12), e a combinação de componentes de magnitude e fase gera modelos que podem exigir mais dados para treinamento.

Pode-se também analisar a evolução de desempenho (acurácia) dos modelos no ajuste-fino. As Figuras 21 e 22 ilustram este processo, em que podemos destacar certa instabilidade da MultFRRBM-PM quando comparada a outros modelos, conforme mostrado nas regiões com o *zoom* no final das épocas de ajuste-fino.

## 6.2.2 Classificação com Modelos Convolucionais

A Tabela 5 apresenta os resultados médios de taxa de acerto, e seus respectivos desvios padrão, obtidos das 15 execuções independentes, considerando as divisões de dados propostas por Pereira et al. (PEREIRA et al., 2018), ou seja, 50/50 e 75/25, com a adição de uma nova partição, 25/75, para avaliar a dificuldade dos modelos convolucionais em aprender sobre o domínio, uma vez que é conhecida a dependência de grandes volumes de dados para tais. Além disso, o teste Wilcoxon foi aplicado com o mesmo racional dos modelos não-convolucionais, marcando com (=) quando os resultados são estatisticamente semelhantes, e com (!=) os diferentes. Por fim, a maior média alcançada por um determinado modelo em cada divisão é sublinhada.

Da Tabela 5, e a primeira divisão de dados (25/75), pode-se observar que todos os modelos baseados em CDBN alcançaram uma acurácia média decente, em torno de 97%,

Figura 21 – Acurácia média sob o conjunto de teste considerando a separação 50/50.

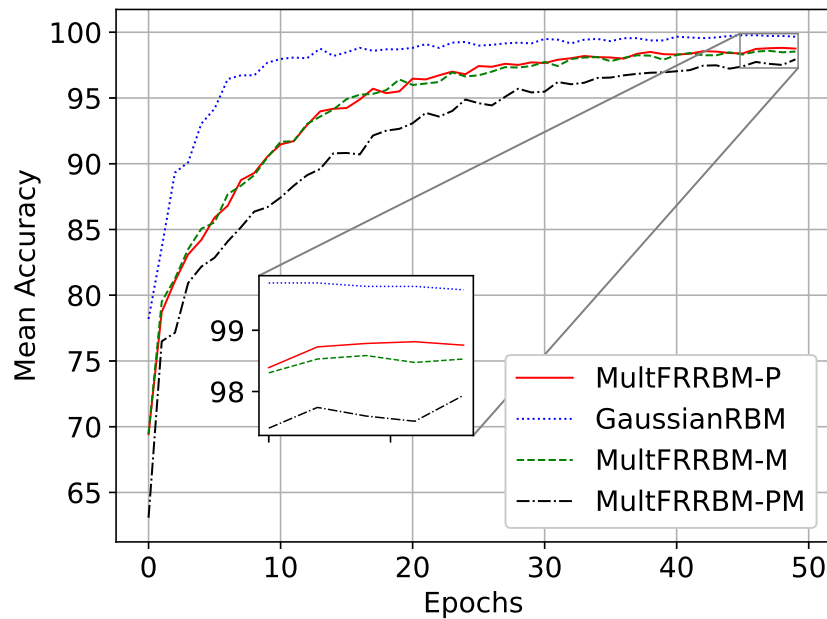
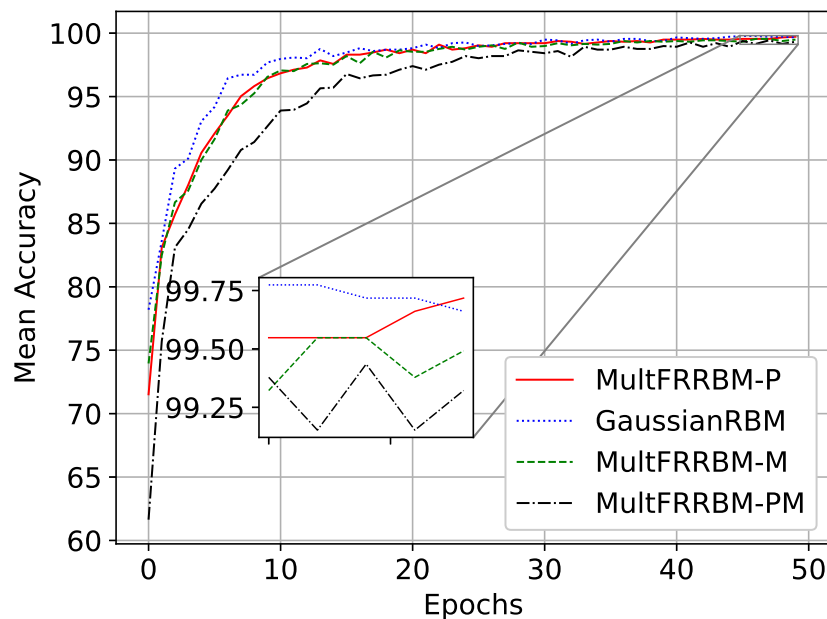


Figura 22 – Acurácia média sob o conjunto de teste considerando a separação 75/25.



melhor que o esperado, pois o número de amostras de treinamento foi bastante baixo e modelos convolucionais costumam requerer um grande volume de dados para um treinamento eficiente. No entanto, o desvio padrão foi alto, mostrando que os modelos propostos ainda podem necessitar de mais dados para se tornarem mais estáveis. Além disso, o melhor modelo até agora foi o FCDBN-P, sendo estatisticamente diferente do FCDBN-PM, a versão com mais variedade de dados derivados das imagens originais.

Com relação à divisão 50/50, o melhor modelo foi o FCDBN-P, alcançando uma acurácia média de 99,71% superando os modelos multimodais não-convolucionais e as CNNs de Pereira et

| Arquitetura                             | 25/75               | 50/50               | 75/25               |
|-----------------------------------------|---------------------|---------------------|---------------------|
| CDBN                                    | 97,86 ± 2,41 (=)    | 99,41 ± 0,50 (!=)   | 99,83 ± 0,33 (=)    |
| FCDBN-P                                 | <u>97,99 ± 2,19</u> | <u>99,71 ± 0,60</u> | <u>99,94 ± 0,21</u> |
| FCDBN-M                                 | 97,88 ± 2,38 (=)    | 99,48 ± 0,65 (!=)   | 99,83 ± 0,34 (=)    |
| FCDBN-PM                                | 97,41 ± 2,21 (!=)   | 99,57 ± 0,79 (=)    | 99,72 ± 0,50 (=)    |
| GaussianRBM                             | -                   | 99.66 ± 0.52 (=)    | 99.66 ± 0.52 (!=)   |
| MultFRBM-P                              | -                   | 98.76 ± 1.58 (=)    | 99.72 ± 0.40 (=)    |
| MultFRBM-M                              | -                   | 98.53 ± 1.70 (=)    | 99.49 ± 0.60 (!=)   |
| MultFRBM-PM                             | -                   | 97.94 ± 1.66 (!=)   | 99.32 ± 0.71 (!=)   |
| CNN <sub>1</sub> (PEREIRA et al., 2018) | -                   | 93,46 ± 16,54       | 97,20 ± 2,45        |
| CNN <sub>2</sub> (PEREIRA et al., 2018) | -                   | 83,55 ± 13,09       | 77,33 ± 22,24       |

Tabela 5 – Taxas de acerto médias e seus respectivos desvios padrão para o conjunto de teste considerando as divisões de dados propostas.

al. (PEREIRA et al., 2018). Além disso, este modelo foi estatisticamente diferente da CDBN, da FCDBN-M, e da MultFRBM-PM. Olhando ainda para as abordagens convolucionais propostas, todas elas superam as RBMs multimodais não-convolucionais em mais de 1 ponto percentual em acurácia média e baixo desvio padrão, porém, apenas com o modelo MultFRBM-PM houve diferença estatística, enquanto com os demais não-convolucionais não houve provavelmente pelo elevado desvio padrão. Mesmo assim, a abordagem multimodal convolucional se mostra robusta com baixo volume de dados, haja visto os resultados de acurácia e desvio padrão.

Finalmente, em relação à terceira divisão (75/25), houve uma melhora interessante nos modelos CDBN multimodais, onde todas as abordagens obtiveram uma acurácia média superior a 99,72%. Entre os modelos propostos, a FCDBN-P se destacou novamente, atingindo 99,94% de acerto, e superando tanto os modelos unimodais (CDBN), com significância estatística para todos, exceto a MultFRBM-P, quanto os de base comparativa (CNNs). Além disso, a FCDBN-P apresentou o menor desvio padrão (0,21) em comparação a todos os modelos. Entretanto, entre os modelos multimodais convolucionais não houve diferença estatística significativa. Tais resultados corroboram os efeitos positivos em relação à abordagem multimodal proposta, em que todos os modelos superam os de base comparativa. Além disso, a melhoria de desempenho com mais dados de treinamento indica que a metodologia pode se beneficiar em aplicações onde mais dados estão disponíveis para treinar.

Adicionalmente, as Figuras 23, 24 e 25 ilustram o procedimento de ajuste-fino (aprendizado supervisionado), onde pode-se destacar uma instabilidade da FCDBN-PM quando comparada a outros modelos, como mostrado de forma proeminente no início dos gráficos, e

nas regiões finais com a ampliação.

Figura 23 – Acurácia média sob o conjunto de teste considerando a separação 25/75.

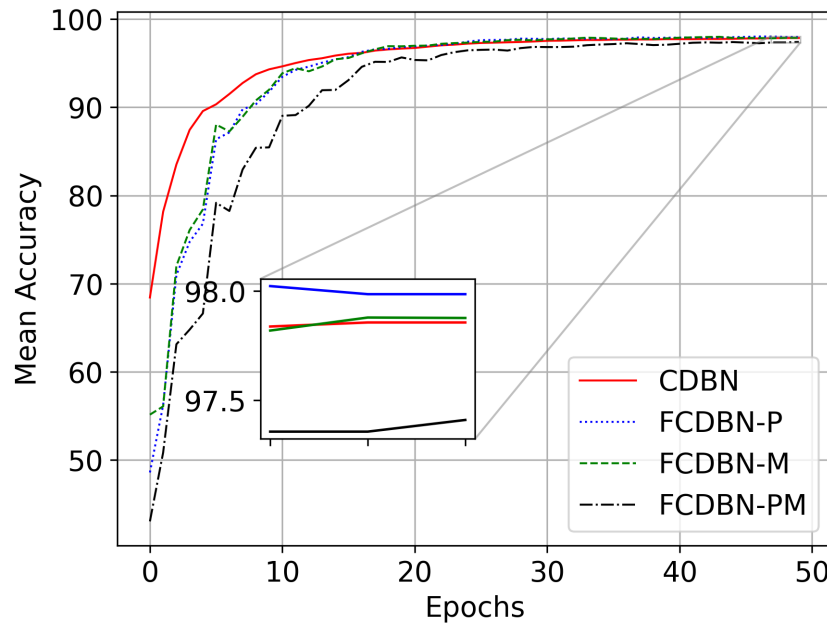
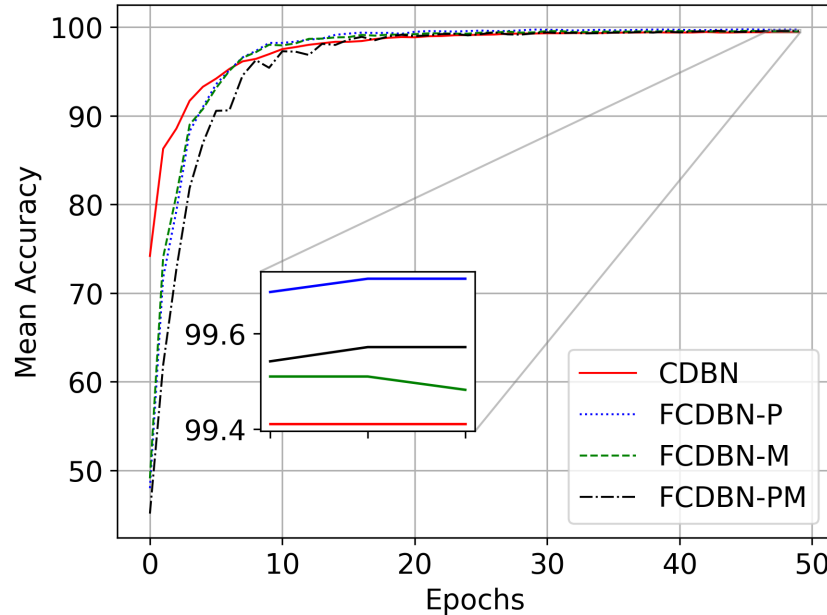


Figura 24 – Acurácia média sob o conjunto de teste considerando a separação 50/50.



Para ilustrar, e facilitar a comparação visual dos resultados finais, estes também foram resumidos na Figura 26, onde são representadas apenas as acurácias médias na época final (50) para as três partições empregadas. Da Figura 26, vê-se de forma clara a superioridade do modelo FCDBN-P frente aos demais da mesma “família”, ou seja, FCDBN-, para todas as partições de dados. Além disso, fica claro que a fase da transformada de Fourier contribuiu significativamente para as melhores acurácias, principalmente em situações com menor volume de dados.

Figura 25 – Acurácia média sob o conjunto de teste considerando a separação 75/25.

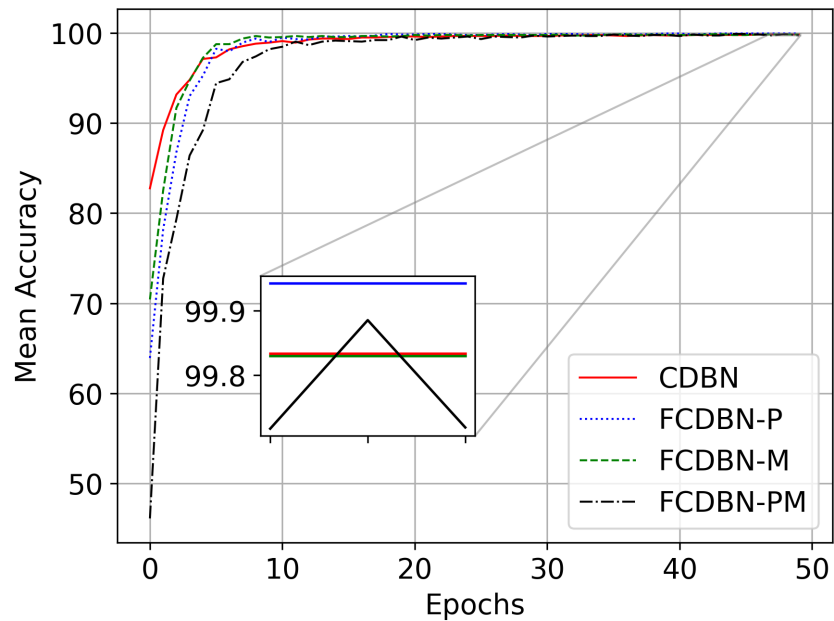
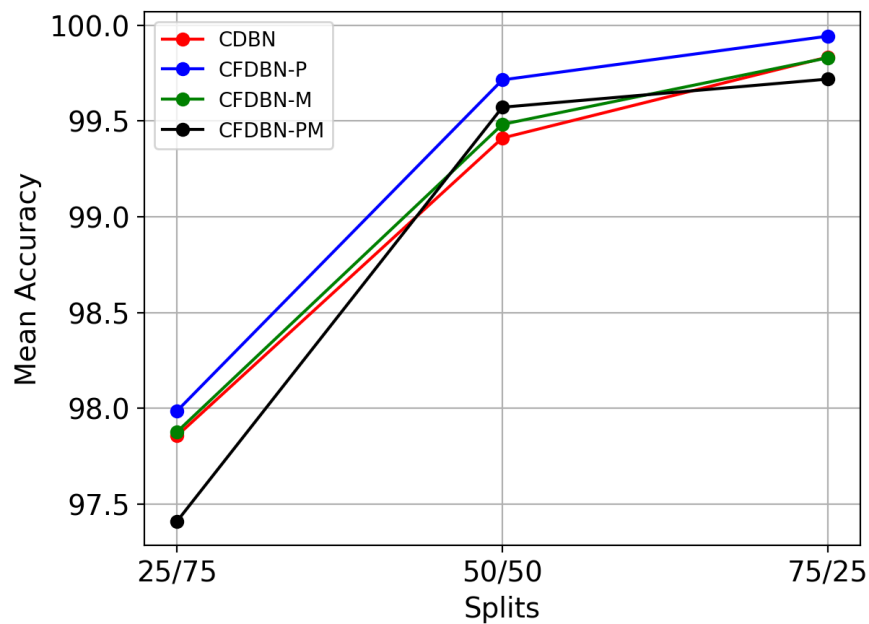


Figura 26 – Acurácia média final no conjunto de teste considerando as três partições empregadas.



Por fim, dentre todos os modelos multimodais propostos, não-convolucionais e convolucionais, para todas as partições empregadas, o melhor foi o FCDBN-P com 75% de dados de treinamento e acurácia média de 99,94%. Este fato foi igualmente notado para todas as partições utilizadas (vide Figura 26), mesmo nos modelos não-convolucionais, indicando que a componente de fase da transformada de Fourier é mais informativa e discriminativa para esta abordagem, neste problema. De fato, ao observar as representações visuais, como nas Figuras 11 e 13, vemos que a fase traz consigo imagens mais expressivas que as obtidas da componente magnitude.

### 6.3 Complexidade dos Modelos Propostos

Adicionalmente, foi analisado o número de parâmetros das redes para complementar a análise, fornecendo uma comparação direta entre os modelos propostos, não-convolucionais e convolucionais, com os utilizados por Pereira et al. (PEREIRA et al., 2018) como base comparativa. Tais valores foram calculados com base na descrição das arquiteturas dos modelos (Tabelas 2 e 3), e com base em Pereira et al. (PEREIRA et al., 2018) para as CNNs. A Tabela 6 mostra o número de parâmetros para as abordagens propostas e as arquiteturas de linha de base, bem como o número aproximado de tais parâmetros (para simplificar a visualização).

|                                         | Parâmetros | Aproximação    |
|-----------------------------------------|------------|----------------|
| RBM/MultFRRBM                           | 5.006.000  | $\approx 5M$   |
| CDBN                                    | 26.497     | $\approx 26K$  |
| FCDBN-P/M                               | 52.962     | $\approx 53K$  |
| FCDBN-PM                                | 79.427     | $\approx 80K$  |
| CNN <sub>1</sub> (PEREIRA et al., 2018) | 116.874    | $\approx 117K$ |
| CNN <sub>2</sub> (PEREIRA et al., 2018) | 60.954.656 | $\approx 61M$  |

Tabela 6 – Análise de parâmetros para os modelos empregados.

Da Tabela 6, pode-se notar que os modelos baseados em RBM, ou seja, a GaussianRBM e suas versões multimodais (MultFRRBM), possuem significativamente mais parâmetros que os modelos convolucionais propostos (CDBN e FCDBN-), e que a CNN<sub>1</sub>, possuindo quase  $43\times$  (aqui, “ $\times$ ” significa vezes) mais parâmetros que esta. No entanto, esse número maior não foi uma barreira para a excelente classificação de AVC, uma vez que todos os modelos baseados em RBM superaram os resultados da CNN<sub>1</sub>. Por outro lado, os baseados em RBM têm  $\approx 92\times$  menos parâmetros que a CNN<sub>2</sub>, que é um valor impressionante.

Ainda da Tabela 6, vê-se que os modelos CDBN possuem a menor quantidade de parâmetros frente aos demais, aproximadamente 26K. Este fato não impediu o modelo de alcançar melhores resultados de classificação que as CNN<sub>1</sub> e CNN<sub>2</sub>. Já os modelos FCDBN-P/M ficam com aproximadamente 53K parâmetros, sendo o modelo baseado na fase da transformada de Fourier (-P) o melhor sobre todos no que diz respeito à classificação. Por fim, o modelo multimodal convolucional mais complexo é o FCDBN-PM, com 80K parâmetros, fato que traz luz ao desempenho não tão bom quanto os atingidos pelos demais da abordagem proposta. Além disso, todos os modelos CDBN são muito mais leves que os propostos não-convolucionais (MultFRRBM) e os linha de base de Pereira et al. (PEREIRA et al., 2018).

Finalmente, pode-se destacar que as redes padrão baseadas em convolução (CNN<sub>1</sub> e CNN<sub>2</sub>) são muito mais sensíveis ao volume de dados, e os modelos baseados em RBM podem

ser mais robustos quando menos dados estão disponíveis para treinamento, como corroborado com os modelos FCDBN-P e MultFRRBM-P. Entretanto, mesmo os multimodais propostos possuem certa afinidade com maiores volumes de dados, como foi observado na Figura 26, onde as taxas de acerto são crescentes conforme o volume de dados aumenta, porém, a baixa quantidade de parâmetros destes modelos é extremamente positiva para a aplicação.

## 7 Conclusão e Próximos Passos

A presente tese explorou o uso de dados multimodais no contexto das Máquinas de Boltzmann Restritas não-convolucionais e convolucionais aplicadas à tarefa de classificação de acidente vascular cerebral. A multimodalidade empregada utilizou a transformada de Fourier para fornecer duas distribuições de dados adicionais, ou seja, a magnitude e o ângulo de fase dos componentes da transformada.

Os resultados experimentais mostraram que a abordagem proposta pode alcançar medidas de desempenho mais expressivas do que algumas técnicas, atingindo a marca de 99,94% de acurácia média para o modelo FCDBN-P, superando as abordagens convolucionais apresentadas por Pereira et al. (PEREIRA et al., 2018), indicando que o presente banco de dados não requer modelos altamente complexos, e maiores volumes de dados podem auxiliar, dados os melhores resultados com a partição 75/25.

Utilizando a metodologia proposta, todos os modelos baseados em Máquinas de Boltzmann, convolucionais ou não, atingiram resultados superiores aos comparativos do banco de dados, corroborando a hipótese de pesquisa levantada nesta tese, bem como a validação positiva da aplicação de multimodalidade com a transformada de Fourier, mesmo que essa possa requerer uma quantidade de dados levemente maior para treinamento em alguns casos. Os resultados mostram o sucesso da aplicação, porém é de grande valia possuir outro banco de dados, com maior volume de dados, para testar a generalização da abordagem proposta no domínio estudado.

Em concordância com as conclusões anteriores, é importante destacar o potencial dos dados multimodais em modelos baseados em RBMs, uma vez que a combinação com a transformada de Fourier trouxe resultados promissores ao estudo, como visto nas Tabelas 4 e 5, em que os modelos multimodais atingiram resultados significativamente melhores. Além disso, pelas Figuras 15 e 18, fica evidente que o ângulo de fase codifica informações relevantes, e os resultados quantitativos da taxa de acerto corroboram que a fase contribui mais para o poder discriminativo do modelo que a magnitude, aumentando a viabilidade da abordagem, bem como seu potencial de exploração. Além disso, ao adicionar convolução nas arquiteturas as possibilidades de desenvolvimento e melhoria aumentam, haja visto que foram os modelos de menor complexidade dentre todo o desenvolvimento da pesquisa, o que é interessante para a área.

Com isso, alguns pontos positivos (+) e negativos (-) podem ser destacados a pesquisa:

- (+) Os modelos utilizados para o problema são mais simples que os de base, ou seja, os modelos convolucionais aplicados possuem menos parâmetros que as CNNs de base;

- (+) As Máquinas de Boltzmann propostas com a transformada de Fourier atingiram êxito de forma surpreendente para a classificação de AVC;
- (+) A área de estudo de classificação de AVC tem a introdução de Modelos de Boltzmann de maneira efetiva;
- (+) Foi gerada uma biblioteca em Python junto do trabalho, disponível no respectivo link: [https://github.com/MateusRoder/fourier-rbm\\_avc](https://github.com/MateusRoder/fourier-rbm_avc);
- (-) O banco de dados de Pereira et al. (PEREIRA et al., 2018) possui pouco volume de dados, por isso seria interessante dispor de outro banco de dados para a classificação de AVC visando uma “dupla validação” da abordagem atual.

Por fim, revisitando e analisando a hipótese de pesquisa (Seção 1.2), houve a validação positiva desta, ou seja, é viável, em termos de desenvolvimento e desempenho preditivo, aplicar RBMs, CRBMs, e CDBNs pouco profundas, e agregar informação multimodal por meio da transformada de Fourier, para a tarefa de classificação de AVC a partir de imagens de tomografia computadorizada. A validação da hipótese veio através das questões de pesquisa respondidas, ou seja, sucesso no desenvolvimento de RBMs para a tarefa de classificação de AVC, sucesso na fusão de informação com a transformada de Fourier, e possibilidade de visualização do aprendizado dos modelos propostos.

## 7.1 Publicações Derivadas da Tese

Com o desenvolvimento de uma nova metodologia multimodal para o problema de classificação de AVC utilizando Máquinas de Boltzmann Restritas, e os resultados atingidos, três trabalhos foram derivados desta tese. Estes, são citados a seguir com os nomes dos locais de submissão e os devidos estados de publicação até a defesa da tese.

- Enhancing shallow neural networks through fourier-based information fusion for stroke classification. (RODER et al., 2021);
- Deep Learning for Stroke Classification: A Survey. (ACM Computing Survey - Em revisão);
- Multimodal Convolutional Deep Belief Networks for Stroke Classification with Fourier Transform. (SIBGRAPI 2023 - Em revisão)

## 7.2 Próximos Passos

Por fim, com relação aos próximos passos, pretende-se aplicar as CRBMs/CDBNs em ao menos um dos bancos de dados de segmentação de AVC apresentados na Seção 2.4, como o

ISLES18 por exemplo, para entender as limitações e possibilidades das técnicas aqui utilizadas em tarefas de segmentação, uma vez que esta também possui aplicações importantes no apoio ao diagnóstico médico, e representa uma tarefa de alta complexidade a depender do tipo de neuroimagem utilizada (tomografia ou ressonância).

Outro ponto diz respeito à aplicação de RBMs convolucionais utilizando a abordagem proposta por Roder et al. (RODER et al., 2020), em que os autores geraram imagens sintéticas a partir de amostras originais com RBMs para aumentar o volume de dados, mostrando-se uma abordagem promissora para melhoria de desempenho em bancos de dados desbalanceados. Esta aplicação pode ser analisada como uma forma de aumento artificial de dados fiel ao domínio.

Um último ponto a ser ressaltado é implementação e aplicação dos modelos DBMs e CDBMs para classificação de AVC, uma vez que não foram abordados nesta tese. São modelos promissores, porém mais complexos de serem treinados, o que pode deixar o processo mais custoso quando comparado às CDBNs, por exemplo. Entretanto, esses modelos possuem características interessantes que podem ser exploradas, como a interdependência entre camadas ocultas para a ativação de neurônios.

# Referências

- ACKLEY, D.; HINTON, G.; SEJNOWSKI, T. J. A learning algorithm for boltzmann machines. In: WALTZ, D.; FELDMAN, J. (Ed.). *Connectionist Models and Their Implications: Readings from Cognitive Science*. Norwood, NJ, USA: Ablex Publishing Corp., 1988. p. 285–307. ISBN 0-89391-456-8.
- AFONSO, B. M.; CINELLI, L. P.; THOMAZ, L. A.; SILVA, A. F. da; SILVA, E. A. da; NETTO, S. L. Moving-camera video surveillance in cluttered environments using deep features. In: IEEE. *2018 25th IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2018. p. 2296–2300.
- ATWOOD, J.; TOWSLEY, D. Diffusion-convolutional neural networks. *Advances in neural information processing systems*, v. 29, 2016.
- BHATT, C.; KUMAR, I.; VIJAYAKUMAR, V.; SINGH, K. U.; KUMAR, A. The state of the art of deep learning models in medical science and their challenges. *Multimedia Systems*, Springer, p. 1–15, 2020.
- BISHOP, C. *Neural networks for pattern recognition*. [S.l.]: Oxford University Press, 1995.
- BOUREAU, Y.-L.; PONCE, J.; LECUN, Y. A theoretical analysis of feature pooling in visual recognition. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. [S.l.: s.n.], 2010. p. 111–118.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.
- BRIDGE, C. P.; BIZZO, B. C.; HILLIS, J. M.; CHIN, J. K.; COMEAU, D. S.; GAURIAU, R.; MACRUZ, F.; PAWAR, J.; NORO, F. T.; SHARAF, E. et al. Development and clinical application of a deep learning model to identify acute infarct on magnetic resonance imaging. *Scientific reports*, Nature Publishing Group, v. 12, n. 1, p. 1–11, 2022.
- CARON, M.; BOJANOWSKI, P.; JOULIN, A.; DOUZE, M. Deep clustering for unsupervised learning of visual features. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. [S.l.: s.n.], 2018. p. 132–149.
- CASTILLO, D.; LAKSHMINARAYANAN, V.; RODRÍGUEZ-ÁLVAREZ, M. J. Mr images, brain lesions, and deep learning. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 11, n. 4, p. 1675, 2021.
- CETINOGLU, Y. K.; KOSKA, I. O.; ULUC, M. E.; GELAL, M. F. Detection and vascular territorial classification of stroke on diffusion-weighted mri by deep learning. *European Journal of Radiology*, Elsevier, v. 145, p. 110050, 2021.
- CHARU, C. A. *Neural networks and deep learning: a textbook*. [S.l.]: Springer, 2018.
- CHEN, H.; HUANG, L.; LIU, T.; OZCAN, A. Fourier imager network (fin): A deep neural network for hologram reconstruction with superior external generalization. *Light: Science & Applications*, Nature Publishing Group UK London, v. 11, n. 1, p. 254, 2022.

- CHILAMKURTHY, S.; GHOSH, R.; TANAMALA, S.; BIVIJI, M.; CAMPEAU, N. G.; VENUGOPAL, V. K.; MAHAJAN, V.; RAO, P.; WARIER, P. Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study. *The Lancet*, Elsevier, v. 392, n. 10162, p. 2388–2396, 2018.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.
- FATIMAH, B.; SINGHAL, A.; SINGH, P. A multi-modal assessment of sleep stages using adaptive fourier decomposition and machine learning. *Computers in Biology and Medicine*, Elsevier, v. 148, p. 105877, 2022.
- FEDERAU, C.; CHRISTENSEN, S.; SCHERRER, N.; OSPEL, J. M.; SCHULZE-ZACHAU, V.; SCHMIDT, N.; BREIT, H.-C.; MACLAREN, J.; LANSBERG, M.; KOZERKE, S. Improved segmentation and detection sensitivity of diffusion-weighted stroke lesions with synthetically enhanced deep learning. *Radiology: Artificial Intelligence*, Radiological Society of North America, v. 2, n. 5, 2020.
- GAUTAM, A.; RAMAN, B. Towards effective classification of brain hemorrhagic and ischemic stroke using cnn. *Biomedical Signal Processing and Control*, Elsevier, v. 63, p. 102178, 2021.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento de imagens digitais*. [S.l.]: Editora Blucher, 2000.
- GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial networks. *Communications of the ACM*, ACM New York, NY, USA, v. 63, n. 11, p. 139–144, 2020.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. Disponível em: <<http://arxiv.org/abs/1512.03385>>.
- HERZOG, L.; MURINA, E.; DÜRR, O.; WEGENER, S.; SICK, B. Integrating uncertainty in deep neural networks for mri based stroke analysis. *Medical Image Analysis*, Elsevier, v. 65, p. 101790, 2020.
- HINTON, G. Training products of experts by minimizing contrastive divergence. *Neural Computation*, MIT Press, Cambridge, MA, USA, v. 14, n. 8, p. 1771–1800, 2002. ISSN 0899-7667.
- HINTON, G. A practical guide to training restricted boltzmann machines. In: MONTAVON, G.; ORR, G.; MÜLLER, K.-R. (Ed.). *Neural Networks: Tricks of the Trade*. [S.l.]: Springer Berlin Heidelberg, 2012, (Lecture Notes in Computer Science, v. 7700). p. 599–619.
- HINTON, G. E.; DAYAN, P.; FREY, B. J.; NEAL, R. M. The "wake-sleep" algorithm for unsupervised neural networks. *Science*, v. 268 5214, p. 1158–61, 1995.
- HINTON, G. E.; OSINDERO, S.; TEH, Y.-W. A fast learning algorithm for deep belief nets. *Neural Computation*, MIT Press, Cambridge, MA, USA, v. 18, n. 7, p. 1527–1554, 2006. ISSN 0899-7667.
- HSSAYENI, M. D.; CROOCK, M. S.; SALMAN, A. D.; AL-KHAFAJI, H. F.; YAHYA, Z. A.; GHORAANI, B. Intracranial hemorrhage segmentation using a deep convolutional model. *Data*, v. 5, n. 1, 2020. ISSN 2306-5729. Disponível em: <<https://www.mdpi.com/2306-5729/5/1/14>>.

HU, M.-K. Visual pattern recognition by moment invariants. *IRE transactions on information theory*, IEEE, v. 8, n. 2, p. 179–187, 1962.

HU, Q.; WHITNEY, H. M.; GIGER, M. L. A deep learning methodology for improved breast cancer diagnosis using multiparametric mri. *Scientific reports*, Nature Publishing Group, v. 10, n. 1, p. 1–11, 2020.

JUNIOR, C. M. J. M. D.; SILVA, S. P. P. da; NOBREGA, R. V. M. da; BARROS, A. C. d. S.; FILHO, P. P. R.; ALBUQUERQUE, V. H. C. de. Deep learning iot system for online stroke detection in skull computed tomography images. *Computer Networks*, Elsevier, v. 152, p. 25–39, 2019.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. In: *3rd International Conference on Learning Representations, ICLR*. [S.l.: s.n.], 2015.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, AcM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017.

KUNDUR, P. D. *Fourier Magnitude and Phase*. 2013. Disponível em: <[https://www.comm.utoronto.ca/~dkundur/course\\_info/signals/notes/Kundur\\_FourierMagPhase.pdf](https://www.comm.utoronto.ca/~dkundur/course_info/signals/notes/Kundur_FourierMagPhase.pdf)>.

LAROCHELLE, H.; BENGIO, Y. Classification using discriminative restricted boltzmann machines. In: ACM. *Proceedings of the 25th international conference on Machine learning*. [S.l.], 2008. p. 536–543.

LATCHAW, R. E.; ALBERTS, M. J.; LEV, M. H.; CONNORS, J. J.; HARBAUGH, R. E.; HIGASHIDA, R. T.; HOBSON, R.; KIDWELL, C. S.; KOROSHETZ, W. J.; MATHEWS, V. et al. Recommendations for imaging of acute ischemic stroke: a scientific statement from the american heart association. *Stroke*, Am Heart Assoc, v. 40, n. 11, p. 3646–3678, 2009.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. [S.l.: s.n.], 2010. p. 253–256.

LEE, H.; GROSSE, R.; RANGANATH, R.; NG, A. Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *Proceedings of the 26th annual international conference on machine learning*. [S.l.: s.n.], 2009. p. 609–616.

LEE, H.; YUNE, S.; MANSOURI, M.; KIM, M.; TAJMIR, S. H.; GUERRIER, C. E.; EBERT, S. A.; POMERANTZ, S. R.; ROMERO, J. M.; KAMALIAN, S. et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature biomedical engineering*, Nature Publishing Group, v. 3, n. 3, p. 173–182, 2019.

LI, G.; HOU, Y.; WU, A. Fourth industrial revolution: technological drivers, impacts and coping methods. *Chinese Geographical Science*, v. 27, n. 4, p. 626–637, 2017. ISSN 1993-064X.

LI, Z.; KOVACHKI, N.; AZIZZADENESHELI, K.; LIU, B.; BHATTACHARYA, K.; STUART, A.; ANANDKUMAR, A. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

LOTFOLLAHI, M.; BERISHA, S.; DAEINEJAD, D.; MAYERICH, D. Digital staining of high-definition fourier transform infrared (ft-ir) images using deep learning. *Applied spectroscopy*, SAGE Publications Sage UK: London, England, v. 73, n. 5, p. 556–564, 2019.

MAIER, O.; MENZE, B. H.; GABLENTZ, J. von der; HÄNI, L.; HEINRICH, M. P.; LIEBRAND, M.; WINZECK, S.; BASIT, A.; BENTLEY, P.; CHEN, L. et al. Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical image analysis*, Elsevier, v. 35, p. 250–269, 2017.

MASUTANI, E. M.; BAHRAMI, N.; HSIAO, A. Deep learning single-frame and multiframe super-resolution for cardiac mri. *Radiology*, Radiological Society of North America, v. 295, n. 3, p. 552–561, 2020.

MOHAMMADI, S.; PERINA, A.; KIANI, H.; MURINO, V. Angry crowds: Detecting violent events in videos. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision – ECCV 2016*. [S.l.]: Springer International Publishing, 2016. p. 3–18. ISBN 978-3-319-46478-7.

NISHIO, M.; KOYASU, S.; NOGUCHI, S.; KIGUCHI, T.; NAKATSU, K.; AKASAKA, T.; YAMADA, H.; ITOH, K. Automatic detection of acute ischemic stroke using non-contrast computed tomography and two-stage deep learning model. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 196, p. 105711, 2020.

OJALA, T.; PIETIKAINEN, M.; HARWOOD, D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: IEEE. *Proceedings of 12th international conference on pattern recognition*. [S.l.], 1994. v. 1, p. 582–585.

OPPENHEIM, A. V.; BUCK, J. R.; SCHAFER, R. W. *Discrete-time signal processing*. Vol. 2. [S.l.]: Upper Saddle River, NJ: Prentice Hall, 2001.

PAN, J.; WU, G.; YU, J.; GENG, D.; ZHANG, J.; WANG, Y. Detecting the early infarct core on non-contrast ct images with a deep learning residual network. *Journal of Stroke and Cerebrovascular Diseases*, Elsevier, v. 30, n. 6, p. 105752, 2021.

PEDEMONTE, S.; BIZZO, B.; POMERANTZ, S.; TENENHOLTZ, N.; WRIGHT, B.; WALTERS, M.; DOYLE, S.; MCCARTHY, A.; ALMEIDA, R. R. D.; ANDRIOLE, K. et al. Detection and delineation of acute cerebral infarct on dwi using weakly supervised machine learning. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2018. p. 81–88.

PENG, S.-J.; CHEN, Y.-W.; YANG, J.-Y.; WANG, K.-W.; TSAI, J.-Z. Automated cerebral infarct detection on computed tomography images based on deep learning. *Biomedicines*, v. 10, n. 1, 2022. ISSN 2227-9059. Disponível em: <<https://www.mdpi.com/2227-9059/10/1/122>>.

PEREIRA, D. R.; FILHO, P. P. R.; ROSA, G. H. de; PAPA, J. P.; ALBUQUERQUE, V. H. C. de. Stroke lesion detection using convolutional neural networks. In: IEEE. *2018 International joint conference on neural networks (IJCNN)*. [S.l.], 2018. p. 1–6.

RAINA, R.; NG, A. Y.; KOLLER, D. Constructing informative priors using transfer learning. In: *Proceedings of the 23rd International Conference on Machine Learning*. [S.l.]: Association for Computing Machinery, 2006. (ICML '06), p. 713–720. ISBN 1595933832.

RAJINIKANTH, V.; ASLAM, S. M.; KADRY, S. Deep learning framework to detect ischemic stroke lesion in brain mri slices of flair/dw/t1 modalities. *Symmetry*, MDPI, v. 13, n. 11, p. 2080, 2021.

RANGEL, E. S. S.; BELASCO, A. G. S.; DICCINI, S. Qualidade de vida de pacientes com acidente vascular cerebral em reabilitação. *Acta paulista de enfermagem*, SciELO Brasil, v. 26, n. 2, p. 205–212, 2013.

REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

RODER, M.; PASSOS, L. A.; RIBEIRO, L. C. F.; BENATO, B. C.; FALCÃO, A. L.; PAPA, J. P. Intestinal parasites classification using deep belief networks. In: IEEE. *The 19th International Conference on Artificial Intelligence and Soft Computing (ICAISC)*. [S.l.], 2020.

RODER, M.; ROSA, G. H.; PAPA, J. P.; PEDRONETTE, D. C. G. In: IEEE. *2021 34th SIBGRAP Conference on Graphics, Patterns and Images (SIBGRAP)*. [S.l.], 2021. p. 378–385.

RODER, M.; ROSA, G. H. de; PAPA, J. P. Learnergy: Energy-based machine learners. *arXiv preprint arXiv:2003.07443*, 2020.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241.

SALAKHUTDINOV, R.; HINTON, G. E. An efficient learning procedure for deep boltzmann machines. *Neural Computation*, v. 24, n. 8, p. 1967–2006, 2012.

SARAGIH, G. S.; RUSTAM, Z.; ALDILA, D.; HIDAYAT, R.; YUNUS, R. E.; PANDELAKI, J. Ischemic stroke classification using random forests based on feature extraction of convolutional neural networks. *International Journal on Advanced Science, Engineering and Information Technology*, INSIGHT-Indonesian Society for Knowledge and Human Development, v. 10, n. 5, p. 2177–2182, 2020.

SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017. p. 618–626.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SIRSAT, M. S.; FERMÉ, E.; CÂMARA, J. Machine learning for brain stroke: A review. *Journal of Stroke and Cerebrovascular Diseases*, Elsevier, v. 29, n. 10, p. 105162, 2020.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2019. p. 6105–6114.

URAL, A. B. Computer aided deep learning based assessment of stroke from brain radiological ct images. *Avrupa Bilim ve Teknoloji Dergisi*, n. 34, p. 42–52, 2022.

VIRANI, S. S.; ALONSO, A.; BENJAMIN, E. J.; BITTENCOURT, M. S.; CALLAWAY, C. W.; CARSON, A. P.; CHAMBERLAIN, A. M.; CHANG, A. R.; CHENG, S.; DELLING, F. N. et al. Heart disease and stroke statistics—2020 update: a report from the american heart association. *Circulation*, Am Heart Assoc, v. 141, n. 9, p. e139–e596, 2020.

WELLING, M.; ROSEN-ZVI, M.; HINTON, G. Exponential family harmoniums with an application to information retrieval. In: SAUL, L.; WEISS, Y.; BOTTOU, L. (Ed.). *Advances in Neural Information Processing Systems 17*. [S.l.]: MIT Press, 2005. p. 1481–1488.

WILCOXON, F. Individual comparisons by ranking methods. *Biometrics Bulletin*, v. 1, n. 6, p. 80–83, 1945.

WINTERMARK, M.; SANELLI, P. C.; ALBERS, G. W.; BELLO, J.; DERDEYN, C.; HETTS, S. W.; JOHNSON, M. H.; KIDWELL, C.; LEV, M. H.; LIEBESKIND, D. S. et al. Imaging recommendations for acute stroke and transient ischemic attack patients: a joint statement by the american society of neuroradiology, the american college of radiology, and the society of neurointerventional surgery. *American Journal of Neuroradiology*, Am Soc Neuroradiology, v. 34, n. 11, p. E117–E127, 2013.

WONG, K. K.; FORTINO, G.; ABBOTT, D. Deep learning-based cardiovascular image diagnosis: A promising challenge. *Future Generation Computer Systems*, v. 110, p. 802–811, 2020. ISSN 0167-739X.

XU, Q.; ZHANG, R.; FAN, Z.; WANG, Y.; WU, Y.-Y.; ZHANG, Y. Fourier-based augmentation with applications to domain generalization. *Pattern Recognition*, Elsevier, v. 139, p. 109474, 2023.

YAMASHITA, R.; NISHIO, M.; DO, R. K. G.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, Springer, v. 9, n. 4, p. 611–629, 2018.

YEW, K. S.; CHENG, E. Acute stroke diagnosis. *American family physician*, NIH Public Access, v. 80, n. 1, p. 33, 2009.

YEW, K. S.; CHENG, E. Diagnosis of acute stroke. *Am Fam Physician*, NIH Public Access, v. 91, n. 8, p. 528–536, 2015.

ZOPH, B.; GHIASI, G.; LIN, T.-Y.; CUI, Y.; LIU, H.; CUBUK, E. D.; LE, Q. V. Rethinking pre-training and self-training. *arXiv preprint arXiv:2006.06882*, 2020.