UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Caio Oliveira Carneloz

Auxílio no diagnóstico da doença de Alzheimer a partir de imagens de ressonância magnética utilizando competição e cooperação entre partículas

São José do Rio Preto
2020

Caio Oliveira Carneloz

Auxílio no diagnóstico da doença de Alzheimer a partir de imagens de ressonância magnética utilizando competição e cooperação entre partículas

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: CAPES – Proc..1743481

Orientador: Prof. Dr. Fabricio Aparecido Breve

São José do Rio Preto
2020

C289a

Carneloz, Caio Oliveira

Auxílio no diagnóstico da doença de Alzheimer a partir de imagens de ressonância magnética utilizando competição e cooperação entre partículas / Caio Oliveira Carneloz. -- São José do Rio Preto, 2020

90 f. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Fabricio Aparecido Breve

1. Computação - Matemática. 2. Inteligência Artificial. 3. Aprendizado de Computador. 4. Diagnóstico por Imagem. 5. Doença de Alzheimer. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Caio Oliveira Carneloz

Auxílio no diagnóstico da doença de Alzheimer a partir de imagens de ressonância magnética utilizando competição e cooperação entre partículas

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: CAPES – Proc..1743481

Comissão Examinadora

Profª. Drª. Fabricio Aparecido Breve
UNESP – Câmpus de Rio Claro
Orientador

Prof. Dr. João Roberto Bertini Junior
UNICAMP – Campus de Campinas

Prof. Dr. Danilo Medeiros Eler
UNESP – Câmpus de Presidente Prudente

Rio Claro
06 de Setembro de 2019

AGRADECIMENTOS

Primeiramente, aos meus amados pai, mãe e irmã por todo amor e carinho que sempre me deram. Agradeço a compreensão de vocês em todos os momentos que me fiz ausente para atingir este objetivo. Sei que sempre acreditaram em mim, mesmo quando ninguém acreditava. Eu não seria nada sem vocês. Muito obrigado. Espero um dia recompensar um décimo de tudo que vocês fizeram por mim.

Agradeço também ao meu orientador por todo suporte durante este período e pela oportunidade que me foi dada. Sou muito grato por todo o aprendizado e pela confiança em minha pessoa.

Também gostaria de dedicar um agradecimento especial aos meus professores Danillo Pereira e Danilo Bosso. Ao meu professor da faculdade, Danillo Pereira, por guiar, com absoluta precisão, praticamente todos os meus passos na vida acadêmica e profissional. Também agradeço por todo aprendizado sobre a vida. Ao meu professor do ensino médio, Danilo Bosso, pelo incrível empenho em exercer sua profissão. Graças ao senhor, enxerguei em mim uma pessoa que eu não sabia que existia. Tudo que me ocorreu academicamente começa depois disso. Serei eternamente grato!

Muito obrigado a todos que direta ou indiretamente participaram desta minha conquista. À minha namorada, que me apoiou do início ao fim, sendo sempre a melhor pessoa possível, me ajudando em todos os momentos que precisei. Aos meus amigos Alan, Bruno, Jorge e Wesley por todos os momentos vividos durante este período. Aos alunos da UNESP de Prudente, pela incrível recepção que tive na universidade. A todos os professores que tive desde ensino fundamental até o fim do mestrado. Aos meus amigos da faculdade e do ensino médio, que sempre estiveram presentes em todos os momentos. E, por fim, aos manos da minha quebrada por todos os churrascos, pelas peladas e principalmente pelo Cartola FC nosso de cada dia. Muito obrigado a todos!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, processo 1743481.

RESUMO

O diagnóstico precoce da doença de Alzheimer é algo crucial para que seu tratamento seja efetivo. No entanto, detectar a doença em estágio inicial não é uma tarefa trivial, uma vez que seus sintomas podem ser confundidos com consequências naturais do envelhecimento. Para que métodos de aprendizado supervisionado identifiquem diferenças entre cérebros saudáveis e com Alzheimer, é necessária uma grande base de dados rotulados, o que é caro e leva tempo. Em contrapartida, o uso de métodos não-supervisionados elimina o impasse dos dados rotulados, mas ao mesmo tempo não lida bem com problemas onde as amostras possuem diferenças sutis entre si. Neste trabalho, será apresentada uma abordagem semi-supervisionada para classificar pacientes com a doença de Alzheimer a partir de imagens cerebrais, através do uso do modelo de competição e cooperação entre partículas. Para que seja possível lidar com imagens, foram utilizados extratores de características de descritores de propósito geral e redes pré-treinadas de aprendizagem profunda. Os experimentos foram realizados fazendo uso da base *Open Access Series of Imaging Studies*. Como forma de avaliar o desempenho do modelo de partículas, foram aplicados os algoritmos SVM, kNN, Naive Bayes, Label Propagation e Label Spreading.

De modo geral, através da métrica *F1-Score*, foi possível observar que os extratores de características baseados em aprendizagem profunda trouxeram uma melhor representação das imagens em relação aos descritores de uso geral. A segmentação de regiões cerebrais trouxe ganhos de até 131% no valor *F1-Score* para determinados algoritmos. Os experimentos também demonstraram um melhor desempenho médio do modelo de partículas em comparação aos demais métodos utilizados.

Palavras-chave: Competição e Cooperação de Partículas. Aprendizado de Máquina. Aprendizado Semi-Supervisionado. Doença de Alzheimer. Imagens de Ressonância Magnética.

ABSTRACT

The early detection of Alzheimer's disease is a crucial step to have effective treatment of it. However, early identifying this disease is not an easy task, since their symptoms can be confused with natural aging consequences. To have supervised machine learning methods doing this well, a large labeled dataset is needed, which is expensive and time-consuming. On the other hand, the use of unsupervised methods solves the dataset question, but at the same time, it doesn't handle well with samples that have a small difference between each other. In this paper, we will present a semi-supervised approach for classifying patients with Alzheimer's disease from brain imaging, using the particle competition and cooperation model. In order to deal with images, feature extractors from general-purpose descriptors and pre-trained deep learning networks were used. The experiments were performed using the Open Access Series of Imaging Studies database. To evaluate the performance of particle competition and cooperation model, the SVM, kNN, Naive Bayes, Label Propagation and Label Spreading algorithms were applied. In a general manner, through the *F1-Score* metric, it is possible to see that the pre-trained deep learning networks have brought better image descriptions regarding pre-built image descriptors. With the segmentation of brain areas, gains around 131% were observed for some algorithms. The experiments also show that the particle model had a higher average score when compared to other tested algorithms.

Keywords: Particle Competition and Cooperation. Machine Learning. Semi-Supervised Learning. Alzheimer's Disease. Magnetic Resonance Imaging.

Lista de Figuras

Lista de Tabelas

Sumário

1 Introdução

A doença de Alzheimer, chamada internacionalmente de *Alzheimer's Disease* ou abreviada por DA, é uma disfunção cerebral neurodegenerativa associada à perda de conexões cerebrais (Tufail et al., 2012). Ela é o tipo mais comum de demência (Yang et al., 2013) e atinge principalmente pessoas com mais de 65 anos de idade (Hebert et al., 2003). De acordo com um estudo do ano de 2015, cerca de 35 milhões de pessoas sofrem da doença em um escopo global (Guzior et al., 2015). As consequências que este mal traz são bem conhecidas pelo senso comum. Em suma, há uma grande alteração, ou perda, das funções cognitivas. Dentre os principais problemas causados, a perda da memória de curto prazo e possíveis tombos fazem com que a pessoa afetada corra vários riscos enquanto sozinha, tornando-se extremamente dependente de outras pessoas (Christofoletti et al., 2006; Nagahara et al., 2009). Atualmente, não são conhecidos tratamentos suficientemente eficazes para os pacientes da doença. Em outras palavras, não existem métodos que permitam a recuperação ou, no mínimo, a desaceleração do progresso da mesma (Weller e Budson, 2018; Guzior et al., 2015).

Além disso, diagnosticar a doença é algo extremamente difícil (D'Errico e Divieto, 2018) e isso se deve ao fato de que os danos causados por ela são muito semelhantes aos que são causados pelo processo natural de envelhecimento (Klöppel et al., 2008b). Entre os danos causados naturalmente e o surgimento de demência, há o comprometimento cognitivo leve (CCL ou *MCI*), que pode ser categorizado como um estágio intermediário (Petersen et al., 2014). Ainda que nem todo caso de *MCI* evolua para um quadro de demência, a possibilidade do surgimento da doença de Alzheimer nesses casos é superior.

O diagnóstico precoce da doença de Alzheimer permite com que pacientes sejam tratados da forma adequada, utilizando medicamentos que permitam uma melhor convivência com a mesma. Desta forma, é importante que o diagnóstico seja feito o quanto antes e de forma correta. Consequentemente, esforços devem e vêm sendo dedicados a isso (Weller e Budson, 2018; Neto et al., 2005).

Atualmente, sabe-se que a doença de Alzheimer está diretamente relacionada à perda da massa cinzenta presente no cérebro humano (Karas et al., 2003) e também à

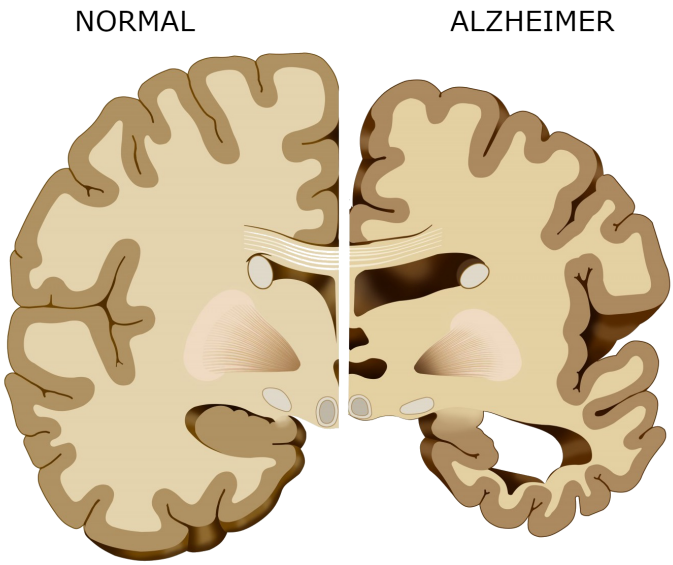


Figura 1: Ilustração das possíveis diferenças entre o cérebro de uma pessoa comum e de uma pessoa com DA.
Adaptado de: alzheimersnewstoday.com/2015/03/11/

algumas outras mudanças cerebrais que poderiam ter seus padrões identificados através de Imagens de Ressonância Magnética (*Magnetic Resonance Imaging* ou *MRI*). A Figura 1 ilustra as diferenças entre um cérebro saudável e o cérebro de um paciente com Alzheimer. Partindo desse pressuposto, diversos autores utilizaram métodos computacionais para classificarem automaticamente, através de *MRI's*, imagens que pertenciam a pacientes com a doença (Desikan et al., 2009; Zhang et al., 2011; Davatzikos et al., 2008; Vemuri et al., 2008; Klöppel et al., 2008a).

O uso de máquinas para realização deste tipo de tarefa se mantém em constante evolução e cada vez mais áreas se beneficiam disso. A área médica é um exemplo. Um retrato disso é a possibilidade de utilizar métodos computacionais para reconhecer padrões, como Pereira et al. (2016) demonstra em seu trabalho na análise de movimentos caligráficos para o auxílio na identificação da doença de Parkinson. Tal feito pode ser alcançado utilizando métodos de aprendizado de máquina. Resumidamente, a área de aprendizado de máquina estuda como computadores podem aprender a realizar tarefas específicas, como reconhecer caracteres manuscritos, identificar patologias de doenças, classificar categorias de vinhos e etc.

Comumente, o aprendizado de máquina é dividido em duas sub-categorias, sendo

o aprendizado supervisionado e aprendizado não-supervisionado (de Mello e Ponti, 2018). Cada sub-categoria opera de um modo. O aprendizado supervisionado recebe amostras rotuladas como entrada e, a partir do conhecimento adquirido, tenta prever o rótulo de amostras desconhecidas. O aprendizado não-supervisionado recebe amostras não-rotuladas, sendo assim, é criada uma relação entre as amostras partindo da semelhança das suas características (Grira et al., 2004). Em meio a ambas as técnicas, há também o aprendizado semi-supervisionado, que mistura características das duas abordagens anteriores. Este tipo de estratégia consiste em utilizar amostras rotuladas e não-rotuladas. Pode-se dizer que a abordagem semi-supervisionada faz um uso mais oportuno dos dados, uma vez que dados não-rotulados podem não trazer informações suficientes sobre como os dados se comportam, e dados rotulados tendem a ser mais caros e complicados de se obter em consequência da necessidade de especialistas para a rotulagem (Kunze et al., 2018).

De forma geral, para utilizar métodos de aprendizado de máquina no processamento de dados em formato multimídia, é necessário um pré-processamento. Sendo assim, é possível citar os Sistemas de Recuperação de Imagens Baseada em Conteúdo (*Content Based Image Retrieval* ou *CBIR*), que se baseiam neste pré-processamento para operar. Os sistemas de *CBIR* aparecem em diversas aplicações onde a base de dados é formada por arquivos de imagem (Caicedo et al., 2008). Através de uma imagem $\hat{I}$, é possível buscar n imagens com conteúdos semelhantes à ela, como demonstra a Figura 2. Basicamente, os sistemas de *CBIR* podem se comportar como classificadores para dividirem a base de dados em duas classes, sendo elas a classe de imagens relevantes e a classe de imagens não-relevantes (Dharani e Aroquiaraj, 2013). Além disso, também é justificável associar tais sistemas à regressores, uma vez que estimam numericamente a relevância ou semelhança de imagens. As imagens relevantes são as que melhor se assemelham às imagem usada na busca. Para que este processo de busca ocorra, é necessário que haja a extração de características dessas imagens. Esta mesma etapa deve ocorrer ao utilizar métodos de aprendizado de máquina. Cor, textura, formato e intensidade são algumas das possíveis características a serem extraídas (Torres e Falcão, 2008). Neste trabalho,

serão abordadas duas formas de extração de caraterísticas: via descritores de imagens e via transferência de aprendizado com redes profundas pré-treinadas.

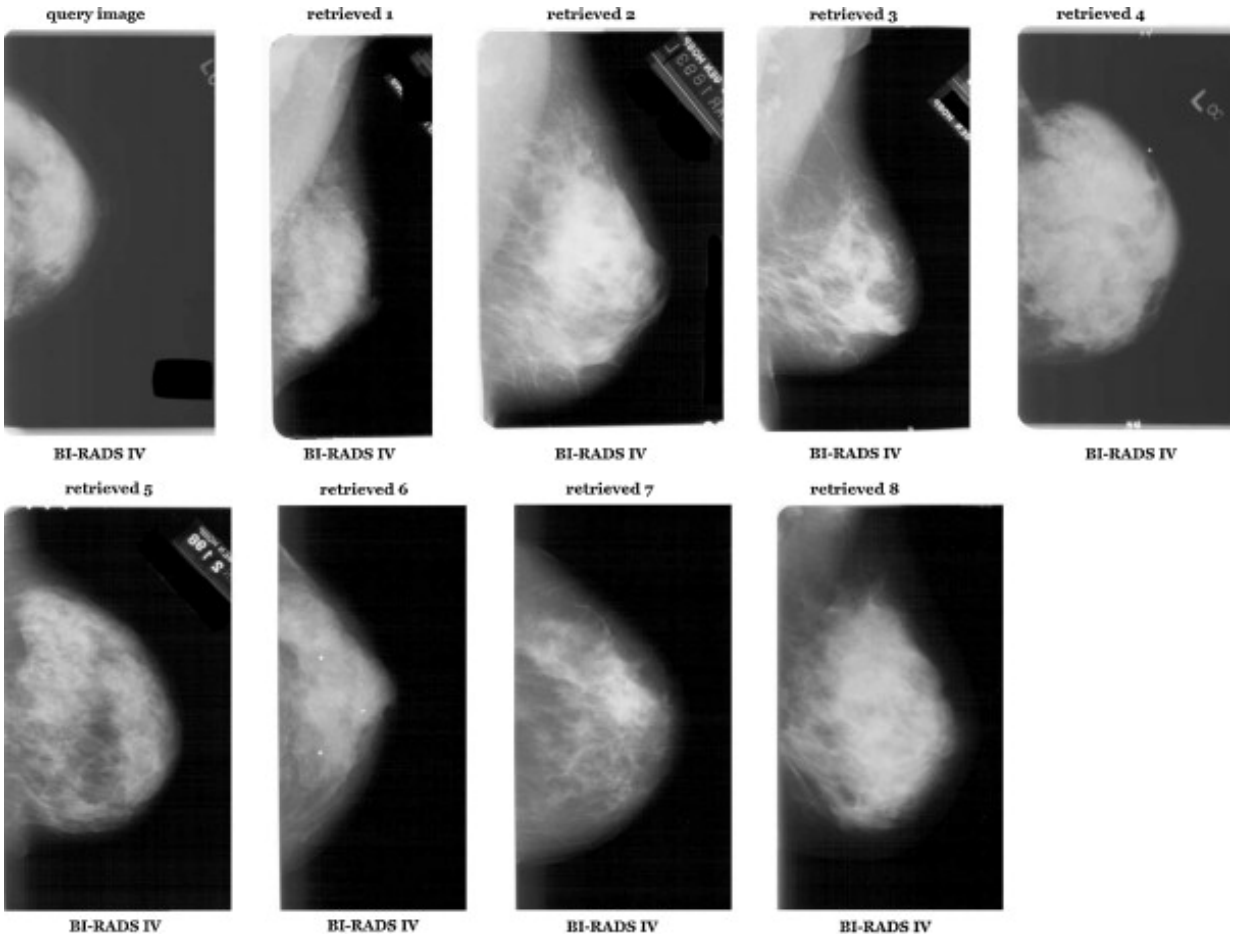


Figura 2: Exemplo de recuperação de imagens baseada em conteúdo, onde, o principal fator utilizado na busca é a densidade mamária.

Fonte: De Oliveira et al. (2010)

Baseado nas premissas anteriores, foi possível observar a oportunidade de aplicação do método de Competição e Cooperação entre Partículas para a classificação de imagens de ressonância magnética, sendo possível auxiliar o reconhecimento da doença de Alzheimer através da análise de padrões. O método a ser utilizado neste trabalho é categorizado como um método de aprendizado semi-supervisionado que tem sua estrutura baseada em um grafo (Breve et al., 2015).

1.1 Desafios de Pesquisa

Baseando-se no que foi estudado, certos desafios são discutidos no decorrer da pesquisa, sendo alguns deles:

- A que ponto o entendimento patológico da doença é capaz de auxiliar nas maneiras de detectá-la via métodos computacionais?
- É possível obter uma boa extração de características em imagens de ressonância magnética a partir da transferência de aprendizagem de outro domínio?
- Etapas de pré-processamento de imagem e pré-processamento de características de imagens podem auxiliar na etapa de classificação?
- A quantidade de informação obtida por meio de dados rotulados é suficiente para que, através do método proposto, se possa obter uma boa inferência sob os dados não-rotulados?

Como consequência dos desafios apresentados acima, a seguinte seção retrata os objetivos do trabalho, visando quais tipos de contribuição espera-se obter com o desenvolvimento do mesmo.

1.2 Objetivos

O presente projeto de pesquisa tem como principais objetivos:

- Estudo e entendimento das características da doença que possam ser relevantes na etapa de análise computacional.
- Avaliação do uso de descritores de propósito geral e transferência de aprendizagem para extração de informações cerebrais através de imagens de ressonância magnética.
- Análise do uso de aprendizado semi-supervisionado para a classificação de imagens de ressonância magnética via extração de características.
- Desenvolvimento de um framework capaz de auxiliar os profissionais da área médica à minimizarem os erros presentes no diagnóstico da doença de Alzheimer.

1.3 Organização do Documento

Este trabalho está dividido em 10 capítulos, organizados da seguinte maneira: o Capítulo 2 visa dar uma visão geral do tipo de dado a ser trabalhado, trazendo o conceito de imagens de ressonância magnética, assim como propriedades da doença de Alzheimer nesse tipo de imagem. No Capítulo 3 é introduzido o conceito do procedimento de extração de características, onde são demonstrados alguns exemplos de aplicações. Seguidamente, nos subcapítulos, é explicado o funcionamento de alguns algoritmos relevantes dessa área e alguns atributos que imagens digitais possuem. O 4º Capítulo traz uma introdução sobre aprendizado de máquina, explicando os objetivos desta área assim como seus conceitos principais. Em seguida, são abordados dois grandes setores deste domínio, sendo eles o aprendizado supervisionado e aprendizado não-supervisionado, onde, em cada um deles, é discorrido sobre um algoritmo relevante. O Capítulo 5 aborda com mais especificidade o aprendizado de máquina com a abordagem semi-supervisionada, a qual será utilizada neste trabalho. Neste capítulo, além de uma introdução sobre o aprendizado semi-supervisionado, estão presentes subcapítulos dedicados a explicar diferentes técnicas para tal. O Capítulo 6 é dedicado a demonstrar dois trabalhos relacionados, evidenciando a maneira pela qual os autores abordaram o problema e seus respectivos resultados. No 7º Capítulo é demonstrado, detalhadamente, o funcionamento do algoritmo de Competição e Cooperação entre Partículas que será utilizado neste projeto de pesquisa. Em seguida, o Capítulo 8, apresenta a metodologia e plano de trabalho. No penúltimo Capítulo, de número 9, é apresentado o protocolo experimental seguido dos experimentos realizados. É abordado o desempenho dos descritores de imagens e de algoritmos de aprendizado supervisionado e semi-supervisionado. Por fim, no 10º e último Capítulo, são relatadas as considerações finais.

2 Imagens de Ressonância Magnética no Diagnóstico da Doença de Alzheimer

Imagens de ressonância magnética têm como fundamento o uso de um intenso campo magnético para que seja possível o envio de pulsos de radiofrequência que, posteriormente, são coletados trazendo informações de regiões inacessíveis visualmente (Mazzola, 2009) (Giedd, 2004). Além disso, as chamadas sequências de pulso definem um conjunto de parâmetros presentes em uma seção de ressonância magnética que, quando modificadas, possibilitam que as imagens demonstrem características distintas entre si (observável na Figura 3). Isto se deve aos princípios físicos do processo, onde, através da divergência entre tempos de relaxação (Kwong et al., 1992), é viável alcançar diferentes contrastes entre as estruturas do corpo humano, sendo possível destacar certas regiões (Stanisz et al., 2005). Por este motivo, existem diversas categorias de imagens de ressonância magnética. Abaixo estão imagens ponderadas em sinal T1 e sinal T2, que representam diferentes tempos de relaxação.

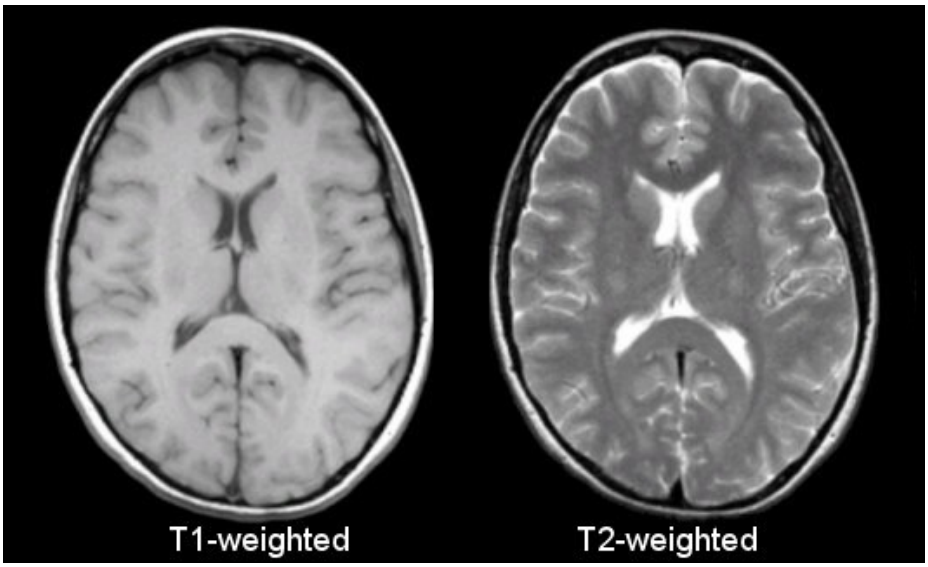


Figura 3: Imagens de ressonância magnética das duas categorias citadas sendo, à esquerda, a imagem de ponderação em sinal T1 e, à direita, a imagem de ponderação em sinal T2. Adaptado de: http://casemed.case.edu/clerkships/neurology/Web%20Neurorad/MRI%20Basics.htm

As imagens resultantes da ressonância magnética podem ser úteis na visualização de órgãos em diversos ângulos e profundidades. Além disso, a técnica permite a detecção de

múltiplas propriedades desses órgãos. Possíveis danos que pacientes possam ter sofrido também são evidenciados, assim como deformidades. No caso do Alzheimer, pode-se observar perdas de tecido, diferenças na espessura de determinadas áreas do cérebro, enfraquecimento de fibras e outros (Iyappan et al., 2017). A Figura 4 demonstra com clareza a característica que pode ser chamada de "perda de massa cinzenta", citada anteriormente, onde fica visível a diferença entre um paciente saudável e um paciente com a doença. Também é possível perceber um leve alargamento dos ventrículos, ao centro do cérebro do paciente com Alzheimer (em vermelho), além de outras diferenças, semelhantes a ilustração da Figura 1.

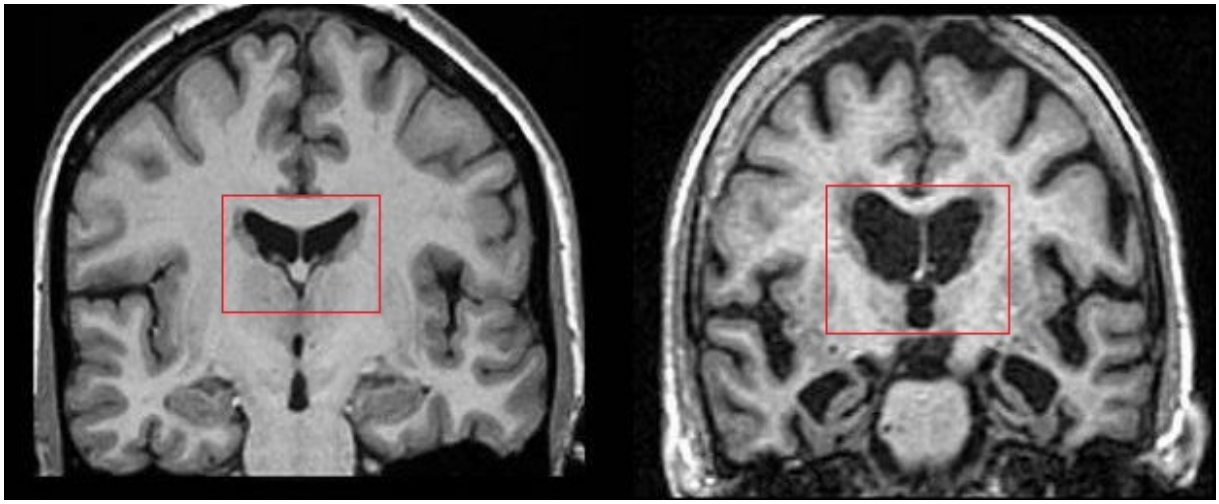


Figura 4: Exemplo de Imagem de Ressonância Magnética, onde, à esquerda tem-se o cérebro de um paciente saudável e à direita, o cérebro de um paciente com Alzheimer.
Fonte: bbc.com/news/health-31807961

Com o uso de imagens de ressonância magnética é viável obter, via extração de características, dados que ligam o estado atual do cérebro do paciente com os possíveis diagnósticos e, consequentemente, o estágio em que a doença se encontra (Iyappan et al., 2017). De acordo com Cuingnet et al. (2011) as abordagens para a extração de características de *MRI's* podem ser agrupadas em três categorias. A primeira, e mais comum, se baseia nas proporções de massa cinzenta (*gray matter* ou *GM*), massa branca (*white matter* ou *WM*) e líquido cefalorraquidiano (*cerebrospinal fluid* ou *CSF*) presentes no cérebro do paciente. Todas essas características podem ser obtidas através da análise dos *voxels* que compõem uma imagem volumétrica obtida através de ressonância magnética.

Voxel é o termo usado para representar um *pixel* volumétrico, que compõe três dimensões (Koketsu et al., 2004). A segunda abordagem faz uso apenas de dados do córtex cerebral. E, por fim, a terceira abordagem se baseia totalmente no hipocampo do cérebro.

Ainda de acordo com o trabalho de Cuingnet et al. (2011), métodos que fazem uso de toda área do cérebro possuem vantagem sobre métodos que se baseiam em regiões específicas. Desta forma, foi adotado para o presente projeto a abordagem que faz uso dos nos níveis de *GM*, *WM* e *CSF* considerando toda a área do cérebro. Para que seja possível fazer uso deste tipo de informação, são necessários métodos computacionais que recuperam informações de imagens digitais. Portanto, serão apresentadas na próxima seção, conceitos e maneiras de extrair características de imagens digitais.

3 Extração de Características

Extração de características é o processo de retirar informações relevantes de imagens digitais. Uma vez que os métodos convencionais de aprendizado de máquina nem sempre são capazes de lidar com imagens em sua forma crua (LeCun et al., 2015), a extração de características se torna uma etapa importante no processo de classificá-las (Ponti et al., 2016). O objetivo é fazer com que o amontoado de valores avulsos que compõem uma imagem torne-se uma representação de seus principais atributos visuais. Isso pode ser feito baseando-se em alguns fatores como formato, cores, textura, volume e afins. A estrutura que configura essa representação de atributos é comumente chamada de vetor de características (Gudivada e Raghavan, 1995) e, para uma imagem $\hat{I}$, é definida por Torres e Falcão (2008) como um ponto no espaço $\mathbb{R}^n : \vec{v}_{\hat{I}} = (v_1, v_2, ..., v_n)$ onde n é a dimensão deste vetor. A Figura 5 ilustra um exemplo de vetor de características baseado em bordas. Neste caso, a imagem será descrita com base em quantas vezes cada tipo de borda aparece.

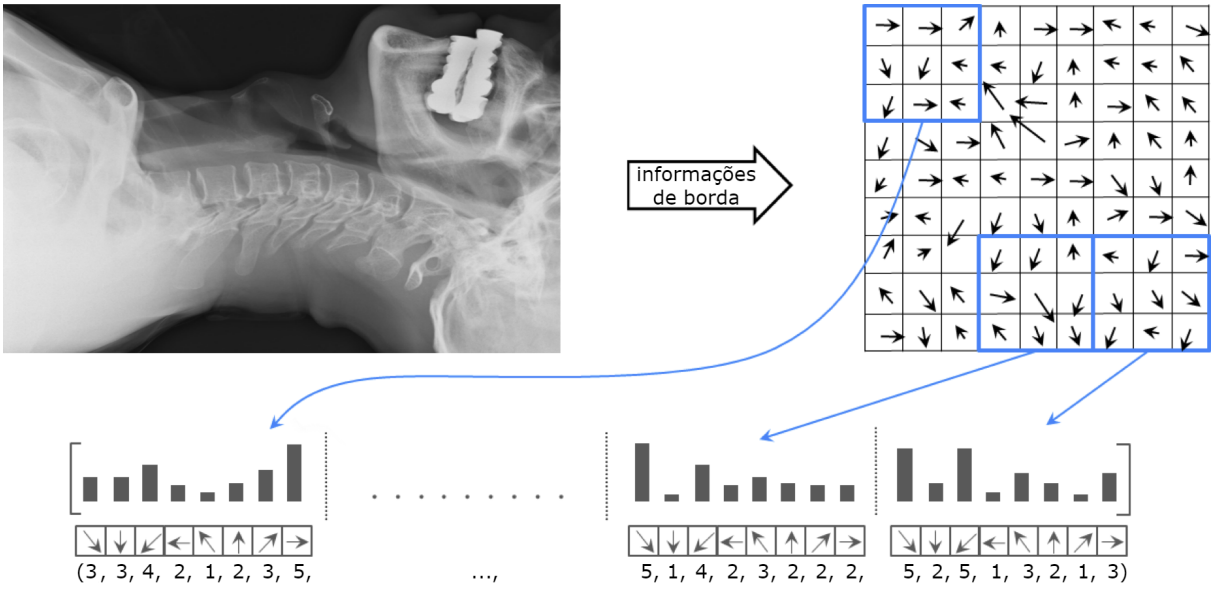


Figura 5: No exemplo, é gerada uma matriz de dimensão 9x9 contendo informações sobre a direção das bordas da imagem. Para cada submatriz 3x3 (sem que haja sobreposição) é gerado um histograma. Ao todo, serão criados 9 histogramas. No fim, é gerado o vetor de características a partir da concatenação dos histogramas.
Adaptado de: https://jermwatt.github.io/machine_learning_refined

É importante que os fatores da imagem a serem considerados sejam bem escolhidos

de acordo com a tarefa para que, assim, seja possível obter uma representação coerente para o problema. Isto é, informações extraídas devem ter relação com o contexto no qual as imagens estão inseridas. O impacto que a falta de coerência causa pode ser descrito, por exemplo, pelo ato de utilizar características de textura para identificar o alargamento dos ventrículos no cérebro humano. Neste cenário, as mudanças na textura podem não ser muito bem representadas, portanto informações sobre forma seriam melhor aproveitadas. As informações extraídas podem ser utilizadas tanto para analisar singularmente uma imagem, quanto para relacionar imagens entre si. Para isto, são apresentados os descritores de imagem que, de acordo com Torres e Falcão (2008), podem ser formalmente definidos como um par finito $(\epsilon D, \delta D)$ onde $\epsilon D : \hat{\mathrm{I}} \rightarrow \mathbb{R}^n$ é uma função que extrai o vetor de características de uma imagem $\hat{\mathrm{I}}$ e $\delta D : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função de similaridade que calcula o quão parecidas são duas imagens a partir da distância entre seus vetores.

Uma outra forma de construir extratores de características de imagens é através do treinamento de algoritmos de aprendizagem profunda com imagens diversas. A aprendizagem profunda pode ser definida como uma sub-área de aprendizado de máquina que explora várias camadas de processamento não-linear para extração de características, transformação de dados, análise de padrões e classificação (Deng et al., 2014). Nesta abordagem, algoritmos são treinados com uma densa quantidade de dados. Durante o processo, os filtros aplicados na extração de características de imagens são otimizados para extraírem informações relevantes. Em alguns casos, é possível assumir que o conhecimento adquirido em um certo domínio pode ser utilizado em algum domínio distinto. Sendo assim, quando uma imagem jamais vista for submetida, a rede irá extrair informações que bem representavam as imagens vistas durante o treinamento. Este conceito é conhecido por transferência de aprendizagem (Pan e Yang, 2009) e será descrito adiante.

A extração de características é um problema antigo (Hong, 1991) e diversos descritores, foram, e vêm sendo desenvolvidos para que, cada vez mais, haja uma recuperação efetiva das informações presentes em imagens (Rui et al., 1999). Além disso, o uso de aprendizagem profunda se mostra relevante neste âmbito (Long et al., 2013). Abaixo, são trazidas informações sobre algumas características de imagens e extratores.

3.1 Características de Imagem

Nas seguintes sub-seções será discorrido sobre as principais características presentes em imagens digitais que podem ser utilizadas na extração e posterior comparação dos vetores de características.

3.1.1 Cor

Descritores baseados em cor são extremamente comuns em diversas aplicações, uma vez que tal característica pode dizer muito sobre uma amostra. A informação de cor é normalmente representada em três canais, como RGB, HSV, YIQ e outros.

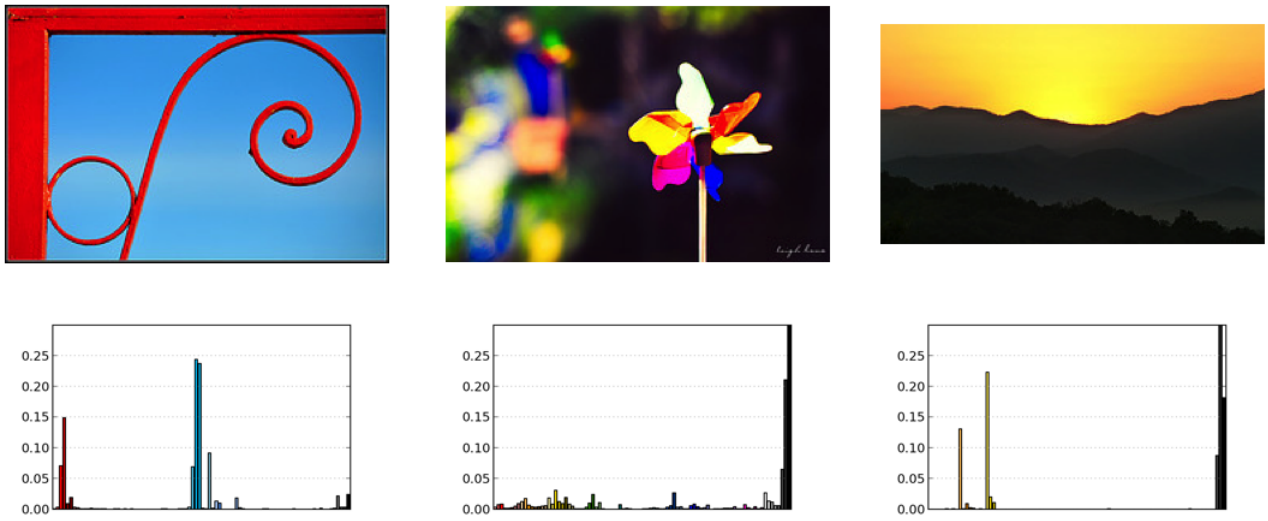


Figura 6: Acima, imagens digitais com a presença de cores e, abaixo, o histograma de cada imagem representando as cores e sua frequência de aparição.
Adaptado de: http://sergeykarayev.com/rayleigh/doc/images/examples1.png

O mais comum exemplo de um descritor de cor é o histograma, que traz a informação global de uma imagem através da quantidade de pixels que representam cada cor (Torres e Falcão, 2008). Analisando graficamente em um plano com duas dimensões, x e y, sendo x a cor e y a frequência de aparição daquela cor em uma imagem z_1, uma outra imagem, z_2, pode ser comparada com z_1 através de medidas de similaridade entre o histograma de ambas, utilizando a distância Euclidiana, por exemplo. Um exemplo de histogramas de cores pode ser visto na Figura 6. No entanto, pelo fato de *MRI's* serem geradas em tons de cinza, histogramas representando cores não seriam viáveis.

Sendo assim, uma alternativa seria um histograma de intensidade, presente na Figura 7, demonstrado por Despotović et al. (2015) em seu trabalho voltado a segmentação de estruturas cerebrais em imagens de ressonância magnética.

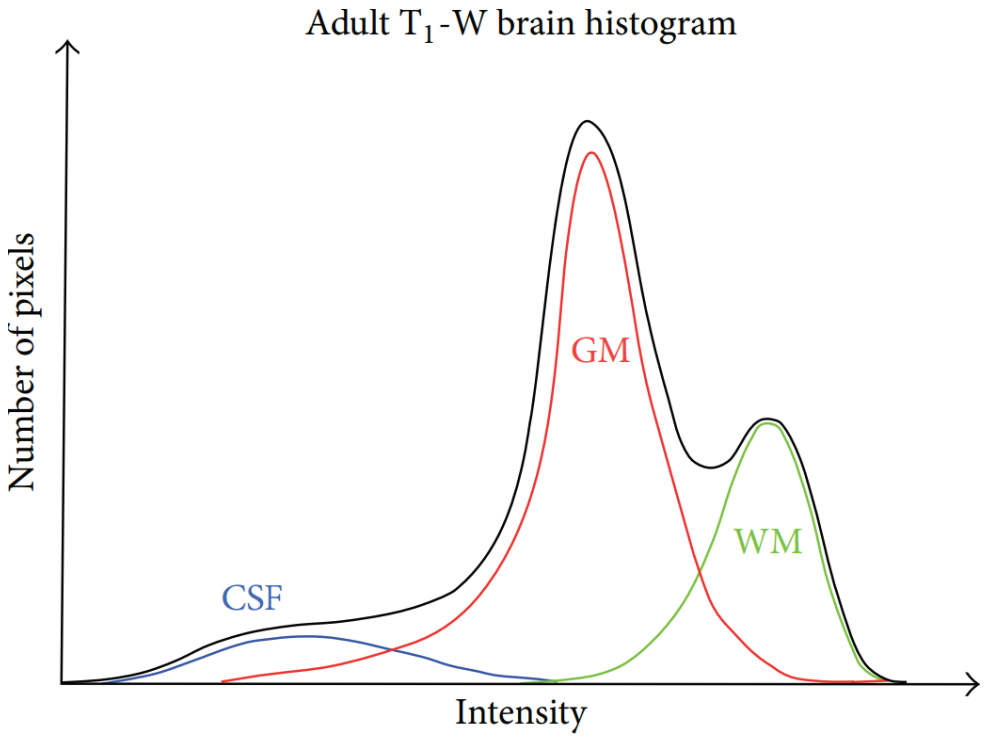


Figura 7: Histograma do cérebro de um adulto obtido através de ressonância magnética ponderada em T1, onde, a intensidade dos tons de cinza descrevem a densidade de cada região a ser segmentada, sendo elas líquido cefalorraquidiano (CSF), massa cinzenta (GM) e massa branca (WM), explicadas no Capítulo 2.
Fonte: Despotović et al. (2015)

3.1.2 Textura

Textura é uma propriedade presente em imagens digitais que representam a superfície e estrutura de uma imagem. De modo geral, pode ser definida como uma repetição regular de um elemento ou padrão em uma superfície, muitas vezes repetitivos, que descrevem diversas informações. É possível, através dela, medir níveis de cinza, distribuição espacial, granularidade, orientação (direção), suavidade e assim por diante (Torres e Falcão, 2008). Devido ao que foi descrito, o uso de descritores que se baseiam em textura são aplicados em diversos sistemas de visão computacional. A classificação de regiões em imagens obtidas por satélites são um bom exemplo disto (Guo et al., 2011).

Na Figura 8 é possível observar alguns tipos de texturas nas quais Hafiane et al.

(2008) utilizaram para treinar classificadores de forma à reconhecerem certas texturas em variados ângulos de rotação. Nas imagens é possível perceber parte das características citadas anteriormente, sendo as mais evidentes, orientação, granularidade e suavidade, por exemplo.

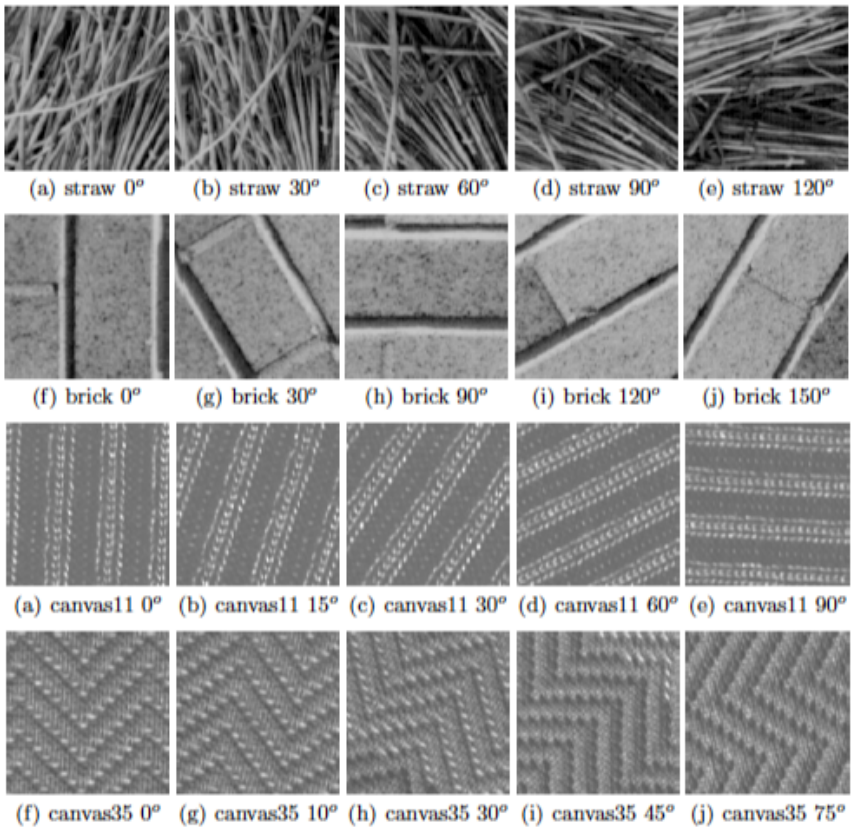


Figura 8: Diferentes tipos de texturas representadas em orientações diversas.
Fonte: Hafiane et al. (2008)

Dependendo do método utilizado para a obtenção de imagens cerebrais, é possível perceber que os tecidos do corpo humano também possuem certos padrões de texturas, sendo assim, tal característica também é de grande utilidade neste setor.

3.1.3 Forma

A forma de um objeto costuma ser uma característica muito singular e é evidente que classificar diferenças (ou semelhanças) através dela é algo totalmente factível. Este atributo traz informações tão importantes que muitas vezes nós, seres humanos, reconhecemos características de objetos unicamente através de tal (Pedronette e da Silva Torres, 2010). De acordo com Latecki et al. (2000), descritores baseados em forma podem ser

divididos em três categorias, sendo elas, baseados em contorno, baseados em imagem e baseados em esqueletização. Cada uma delas visa um tipo de problema, podendo oferecer diferentes formas de robustez, ou seja, uma vez que o principal objetivo é analisar a forma de objetos, diferenças de ângulo, escala e outras alterações (que não interferem diretamente no formato), não devem dificultar a classificação das imagens (Latecki et al., 2000).

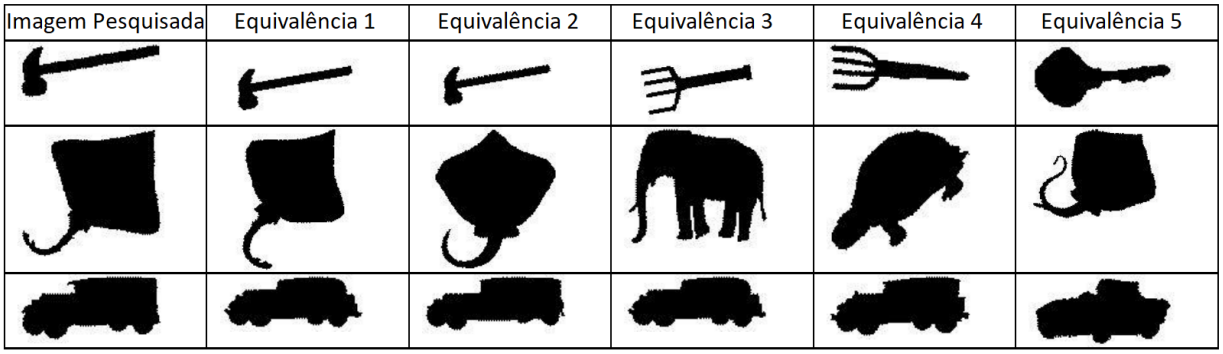


Figura 9: Resultados de uma pesquisa utilizando, em cada uma das linhas, a imagem da primeira coluna para buscar, em ordem de proximidade (esquerda para direita), outras imagens com formatos semelhantes.
Adaptado de: http://jonvolkmar.com/portfolio/image-retrieval-earth-movers-distance/

A Figura 9 demonstra a recuperação de imagens baseando-se no formato das mesmas. Na segunda linha é possível perceber que diferenças no ângulo e, inclusive, escala, não impedem que os dois primeiros resultados estejam como os mais semelhantes.

3.2 Descritores

Nesta sub-seção serão abordados alguns dos descritores mais comuns em recuperação de imagens baseada em conteúdo e também parte do funcionamento de cada um deles.

3.2.1 Fuzzy Color and Texture Histogram

FCTH ou *Fuzzy Color and Texture Histogram* (Chatzichristofis e Boutalis, 2008b) é um descritor que faz uso da lógica Fuzzy para a extração de características em imagens. Dentre as possíveis informações a serem extraídas, o método faz uso dos dados de cor e textura para alcançar os resultados. Primeiramente, a imagem é segmentada em blocos,

onde, cada um desses blocos passa por todas as três unidades de defuzzificação. No primeiro bloco, é gerado um histograma de 8 intervalos, onde cada intervalo representa uma cor predefinida. Após, o histograma inicial é expandido em 24 e, posteriormente, 192 intervalos. Com o histograma final, é utilizado um classificador difuso (Gustafson e Kessel, 1979) que quantifica os valores obtidos. Ao fim, é aplicada uma métrica de similaridade para que seja possível o cálculo de distância entre as imagens submetidas. A informação recuperada através do método não ocupa mais do que 72 bytes.

3.2.2 Color and Edge Directivity Descriptor

O descritor chamado *Color and Edge Directivity Descriptor* (Chatzichristofis e Boutalis, 2008a), ou CEDD, proposto pelo mesmo autor do FCTH, utiliza dados de cor e textura, assim como é feito no método anterior. No entanto, informações de bordas são utilizadas em conjunto com as demais características. O uso de apenas uma característica não fornece um resultado satisfatório e, por isso, a maioria dos métodos fazem uso de duas ou mais. Novamente semelhante ao FCTH, o método separa a imagem em blocos e as submete, primeiramente, em duas unidades de defuzzificação. Na primeira unidade, é gerado um histograma quantificado de 10 classes. Mais uma vez, cada classe corresponde a uma cor predefinida. Em seguida, o histograma é expandido em 24 classes. O próximo passo é utilizar os 5 filtros digitais propostos no *MPEG-7 Edge Histogram Descriptor* (Won et al., 2002) para recuperar informações relacionadas a textura da imagem, classificando cada bloco de forma a gerar o histograma de 144 intervalos. Por fim, é utilizado o mesmo classificador difuso do método FCTH para quantificação dos valores.

3.2.3 Scalable Color Descriptor

Assim como os descritores FCTH e CEDD, a abordagem do SCD (ou *Scalable Color Descriptor*) traz um histograma de cores da imagem. Através do padrão HSV (que se baseia na matiz, saturação e brilho), são definidos espaços de cores fixas. Durante o processo, é utilizada a transformada de Haar para que seja possível obter diferentes

"perfis" no uso deste descritor, como por exemplo mudar a escala de complexidade e de outras características. Tal adaptabilidade ocorre devido a possibilidade de variar a quantidade de coeficientes do algoritmo, ou seja, diminuir o número de compartimentos, e também do tamanho (em bits) de tais compartimentos.

3.2.4 Edge Histogram Descriptor

O descritor conhecido como *Edge Histogram Descriptor* também faz uso de histogramas, mas não envolve cores no processo. O histograma gerado é baseado apenas em informações de bordas presentes na imagem e seu principal foco é em bordas locais, não sendo uma boa alternativa para recuperação de imagens (Won et al., 2002). Uma vez que o foco é em bordas locais, o método divide a imagem em sub-imagens (também chamadas de blocos) para analisar singularmente cada uma delas. Sendo assim, um histograma é gerado para cada bloco da imagem principal. Um histograma do EHD pode ser divido em cinco categorias, sendo elas: vertical, horizontal, diagonal de 45 graus, diagonal de 135 graus e não-diagonal (Prajapati et al., 2016). Desta forma, as características de um bloco são dadas através da frequência de aparição de um determinado tipo de borda.

3.3 Transferência de Aprendizagem

A maioria dos métodos de aprendizado de máquina partem do pretexto de que os conjuntos de treino e teste devem compartilhar do mesmo espaço de características e mesma distribuição. No entanto, seria caro ou até mesmo inviável retreinar modelos quando algum desses fatores mudam. Para isso, foi apresentado o conceito de transferência de aprendizagem. Fundamentalmente, o conceito de transferência de aprendizagem pode ser definido como, dado um domínio fonte D_S, uma tarefa de aprendizado T_S, um domínio alvo D_T e uma tarefa de aprendizado T_T, busca-se aprimorar o aprendizado da função de predição $f_T(\cdot)$ no domínio D_T usando o conhecimento adquirido em D_S e T_S, sendo $D_S \neq D_T$ ou $T_S \neq T_T$ (Pan e Yang, 2009). Apoiada nisso, a área de transferência de aprendizagem busca solucionar as seguintes questões: o que transferir, como transferir e quando transferir. A primeira questão diz respeito a quais tipos de conhecimentos

podem ser transferidos entre domínios distintos. Alguns domínios possuem informações muito singulares, que podem não ser úteis em outros domínios. No entanto, também existem domínios que compartilham informações em comum. Um exemplo disso seriam imagens, que independente da área de domínio, possuem formas, bordas, texturas, cores e etc. (Quattoni et al., 2008). A questão de "quando transferir" lida apenas com as situações onde deve ser feita a transferência de aprendizagem ou não, como já discutido ao falar sobre áreas de domínio. Por fim, existem diversas maneiras de realizar a transferência de aprendizagem. Uma delas é através do uso de aprendizagem profunda.

Aprendizagem profunda é uma categoria de aprendizado de máquina que explora diversas camadas de processamento de informações não-lineares, semelhante às redes neurais artificiais. Esta área tem como objetivo oferecer funções como extração de características supervisionada e não-supervisionada, transformações de dados, análise de padrões e classificação de dados (Deng et al., 2014). No âmbito de imagens digitais, podem ser destacadas as redes neurais convolucionais (Krizhevsky et al., 2012), que lidam com imagens digitais indo dos seus *pixels* até a etapa de classificação, sem o uso de nenhum descritor de imagem pré-definido (Sermanet et al., 2012).

Basicamente, a arquitetura desta rede é composta por repetidas camadas de extração de características, onde cada uma possui um módulo convolucional (Figura 10) seguido de um módulo de *pooling* (Figura 11) e um módulo de normalização. O módulo convolucional aplica filtros de imagem de modo a reconhecer características básicas, como curvas, cores, retas e etc. Como saída, obtém-se um valor que representa o quão próximo uma certa região da imagem está do filtro aplicado. Uma vez que se tem uma enorme quantidade de filtros, é possível detectar atributos de mais alto nível, como quadrados e outras formas geométricas. A camada de *pooling*, também chamada de camada de sub-amostragem, é responsável pela compactação dos dados. No *maxpooling*, que é uma categoria de *pooling*, são selecionadas as sub-amostras que obtiveram a maior resposta a um dado filtro em uma determinada região da imagem.

Analisando a Figura 10, tendo como entrada de uma rede neural convolucional uma imagem $\hat{I}$ com dimensões 5×5 (em verde, representada como uma matriz binária) e uma

camada convolucional com um filtro F, de dimensão 3×3 (com valores em vermelho), é calculada a matriz de convolução a partir da multiplicação da matriz de filtro com os valores da imagem. Esta etapa é realizada com diversos filtros, que podem ser inicializados aleatoriamente ou de acordo com alguma política (Glorot e Bengio, 2010).

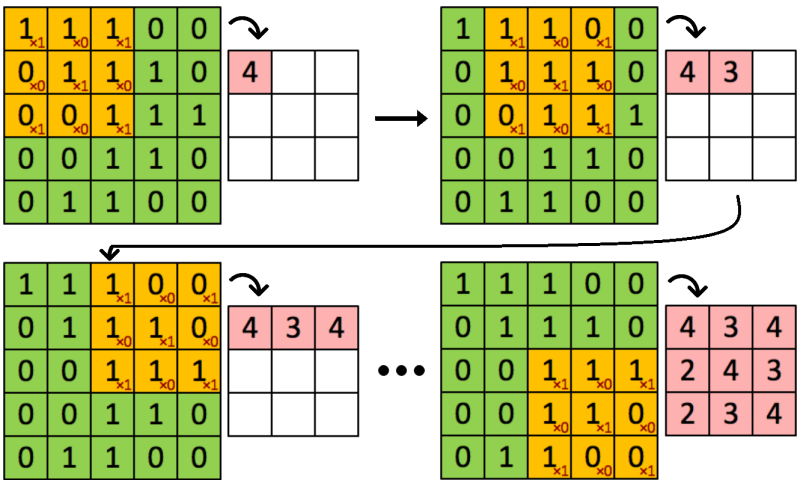


Figura 10: Funcionamento da camada convolucional em uma rede neural convolucional.
Adaptado de: icecreamlabs.com/wp-content/uploads/2018/08/33-con.gif

Na camada de *pooling*, é realizada outra filtragem. Para melhor entendimento, foi considerada uma diferente matriz de convolução, tendo dimensão 6×6. Na Figura 11 a matriz de convolução é reduzida ao ser extraído o maior valor de cada região não sobreposta de dimensão 3×3. Como já dito, este processo é conhecido por *maxpooling*. Outra forma de realizar esta compactação seria extraindo o valor médio de uma determinada região, sendo este, chamado de *averagepooling*.

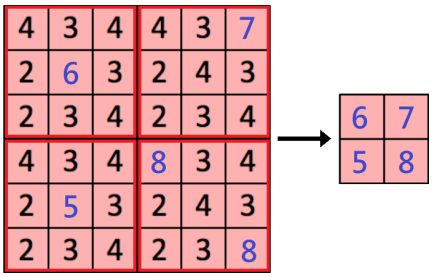


Figura 11: Funcionamento da camada de *pooling* em uma rede neural convolucional.
Adaptado de: icecreamlabs.com/wp-content/uploads/2018/08/33-con.gif

Por fim, é realizada a normalização dos dados de saída (Krizhevsky et al., 2012)

e o resultado obtido pode servir como entrada para uma outra camada convolucional, possuindo o mesmo funcionamento da explicada anteriormente, porém, com filtros distintos. Para a classificação dos dados, ao fim das camadas convolucionais é adicionada uma camada totalmente conectada, semelhante a camada oculta de redes perceptron. O ajuste dos pesos das camadas é realizado através do algoritmo *backpropagation* (Hinton et al., 2012) que, ao comparar o resultado esperado com resultado obtido na saída, calcula o erro e o propaga para as camadas anteriores visando ajustar os pesos de forma a minimizar o erro da rede (Zhou et al., 2008).

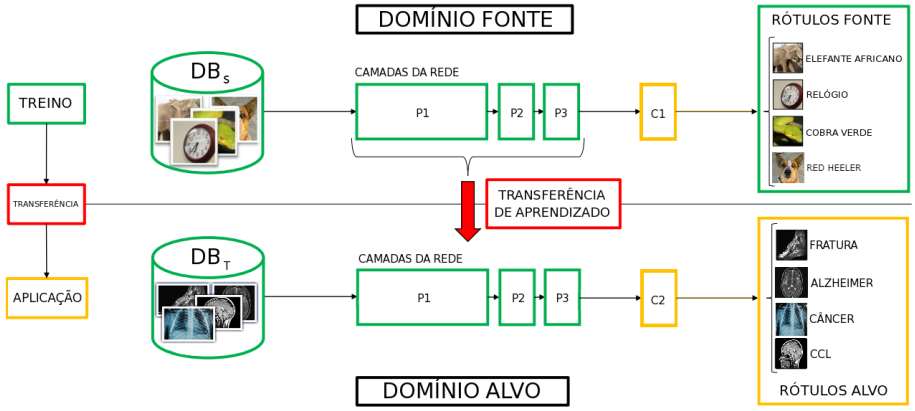


Figura 12: Funcionamento da transferência de aprendizado em redes convolucionais.

Adaptado de: Oquab et al. (2014)

Desta forma, a transferência de aprendizado em redes profundas pode ser exemplificada pela Figura 12. Considerando que o domínio fonte seja treinado com uma base de dados DB_S a partir do classificador $C1$ (que representa a camada totalmente conectada) e que seus pesos na camada de convolução ($P1$, $P2$ e $P3$) sejam devidamente ajustados, é possível utilizar os mesmos pesos da camada de convolução para que seja feita a extração de características em imagens do domínio alvo, presentes no banco DB_T. Assim, $C2$ é treinado a partir dos rótulos alvo, usando o conhecimento de extração adquirido por $C1$.

4 Aprendizado de Máquina

Uma das propostas da inteligência artificial é criar mecanismos que sejam capazes de imitar não só a nossa mente (de Fernandes Teixeira, 2014), mas a de diversos outros seres vivos. Embutido nesta concepção, o aprendizado de máquina trata especificamente do conceito de aprendizado, buscando técnicas que permitem que máquinas se aprimorem automaticamente.

Basicamente, o conceito de aprendizado pode ser dito como o ato de, a partir de experiências passadas, obter conhecimento suficiente para executar alguma tarefa com precisão (Jordan e Mitchell, 2015). Tal evento é explicitado por de Mello e Ponti (2018) ao descreverem qual o processo até que uma criança aprenda a reconhecer um objeto o qual jamais havia visto. Imaginando que este objeto seja uma cadeira, quando a criança tiver o seu primeiro contato, como ela associará o objeto físico "cadeira" com o conceito de uma cadeira? O fato é que além dos pais da criança estarem dizendo a todo momento o que é aquele objeto e, além disso, corrigirem quando a criança erra, é possível que a criança também foque em características daquele objeto para associá-lo a uma cadeira. Podem existir uma gama de características dependendo da situação, mas, neste caso, o padrão de formato da cadeira seria uma boa forma de diferenciá-la de outros objetos. Da mesma forma, o aprendizado de máquina lida com algoritmos que melhoram seu desempenho para resolver problemas a partir de experiências a quais foi submetido (Blum, 2007), além de se basear em padrões presentes nos dados de entrada.

Como já mencionado, o aprendizado de máquina é geralmente dividido em duas principais categorias, sendo elas, o aprendizado supervisionado e aprendizado não-supervisionado (de Mello e Ponti, 2018). Nas próximas subseções, serão descritas tais abordagens e, posteriormente, a interseção entre ambas, que trata da abordagem semi-supervisionada, utilizada no presente projeto de pesquisa.

4.1 Aprendizado Supervisionado

Uma das ramificações mais comuns do aprendizado supervisionado pode ser exemplificada pela criação de um classificador capaz de prever a classe de novas amostras a partir dos

exemplos de amostras já vistas. Assim, dada uma entrada desconhecida E, é esperado que o classificador retorne uma saída S que corresponda a classe daquela entrada (Dietterich, 2002). O objetivo é que o algoritmo, durante o treinamento, aprenda o padrão de cada uma das classes existentes. Desta forma, quando uma amostra sem nenhum gabarito atrelado for dada como entrada, o método retorna como saída a provável classe na qual a entrada pertence.

A abordagem supervisionada é comumente utilizada no treinamento de redes neurais artificiais (Riedmiller, 1994). Uma abordagem clássica deste processo pode ser exemplificada com o uso de redes *perceptron*. Tal abordagem tem seu funcionamento descrito a seguir. Primeiramente, é montada uma rede neural artificial de forma que, a quantidade de neurônios da camada de entrada seja compatível com a quantidade de características que a informação a ser submetida possui. Seguindo o mesmo raciocínio, a quantidade de neurônios na camada de saída deve ser capaz de representar a quantidade de classes que o problema demonstra. Sendo assim, a partir de uma entrada, um neurônio de saída é ativado (ou não) indicando qual a classe que esta entrada será associada (Haykin et al., 2009).

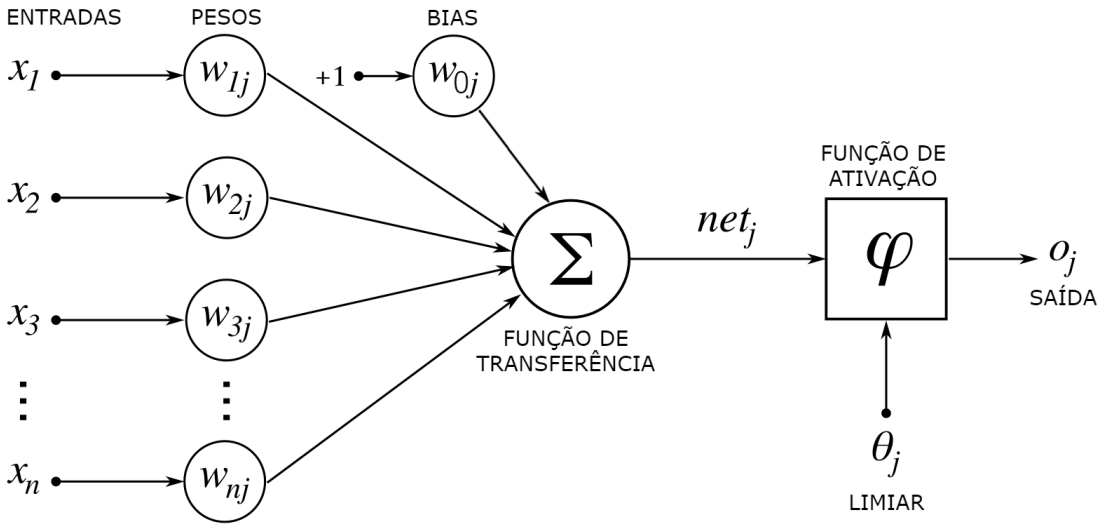


Figura 13: Representação gráfica da estrutura da rede neural artificial perceptron de uma única camada.
Adaptado de: innoarchitech.com/assets/images/ArtificialNeuronModel_english.png

Na Figura 13, há uma representação onde existem n neurônios de entrada que podem ser classificados através de *um* neurônio de saída, ou seja, é um problema com apenas duas

classes (representadas pela ativação, ou não, do neurônio de saída). O treinamento da rede funciona da forma explicada a seguir. As entradas do treinamento já possuem uma classe atrelada, no entanto, esta informação só é utilizada durante a obtenção do erro da rede. Os valores dessas entradas, que são representados de x_1 a x_n, são multiplicados por pesos que, comumente, são inicializados com valores aleatórios próximos de zero. Além da entrada advinda da amostra a ser submetida, um valor *bias* também é utilizado, possuindo valor fixo e também sendo multiplicado pelo peso atrelado a si. Após esta etapa, a função de transferência realiza a soma do produto de cada uma das multiplicações, gerando o valor conhecido como *net*. O próximo passo é utilizar a função de ativação para definir se o neurônio de saída será ou não ativado. A função de ativação representada neste problema é definida através do limiar θ_j, sendo assim, define-se que caso net_j ultrapasse tal limiar, o neurônio de saída é ativado. Caso o valor net_j seja igual ou inferior ao valor de limiar estipulado, o neurônio de saída não é ativado. Todo o procedimento pode ser descrito pela seguinte equação:

$$\boldsymbol{o_j = \varphi(net) = \sum_{i=1}^{n} x_{ij} w_{ij} + bias_j w_{0j}} \qquad \boldsymbol{\varphi(net)} = \begin{cases} \mathbf{1}, & \text{se } \boldsymbol{net > \theta} \\ \mathbf{0}, & \text{senão} \end{cases} \tag{1}$$

Se o valor do neurônio de saída não for compatível com o valor da classe da entrada, é necessário que os pesos sejam corrigidos. A correção ocorre a partir do cálculo do erro da rede, que é dado por: $\boldsymbol{e_j = c_j - o_j}$ onde, c_j representa o rótulo real da amostra j. o_j representa o rótulo estipulado pela rede neural para a amostra j. e_j define o valor do erro obtido na amostra j. Por fim, todos os pesos são ajustados à partir da multiplicação do erro com a entrada x_{ij}:

$$\boldsymbol{w_{ij} = w_{ij} + e_j x_{ij}} \tag{2}$$

4.2 Aprendizado Não-Supervisionado

A abordagem não-supervisionada não recebe amostras rotuladas, portanto, a partir da semelhança entre os dados de entrada, é feita a separação destes em diferentes grupos. Dentre as diversas formas de separação de dados onde não se tem o conhecimento prévio do rótulo, a mais comum é a conhecida por agrupamento. Resumidamente, é possível afirmar que dados que fazem parte de um mesmo grupo, são semelhantes entre si. Consequentemente, dados de diferentes grupos, possuem diferenças. Este comportamento pode ser observado na Figura 14 (Murty et al., 1999).

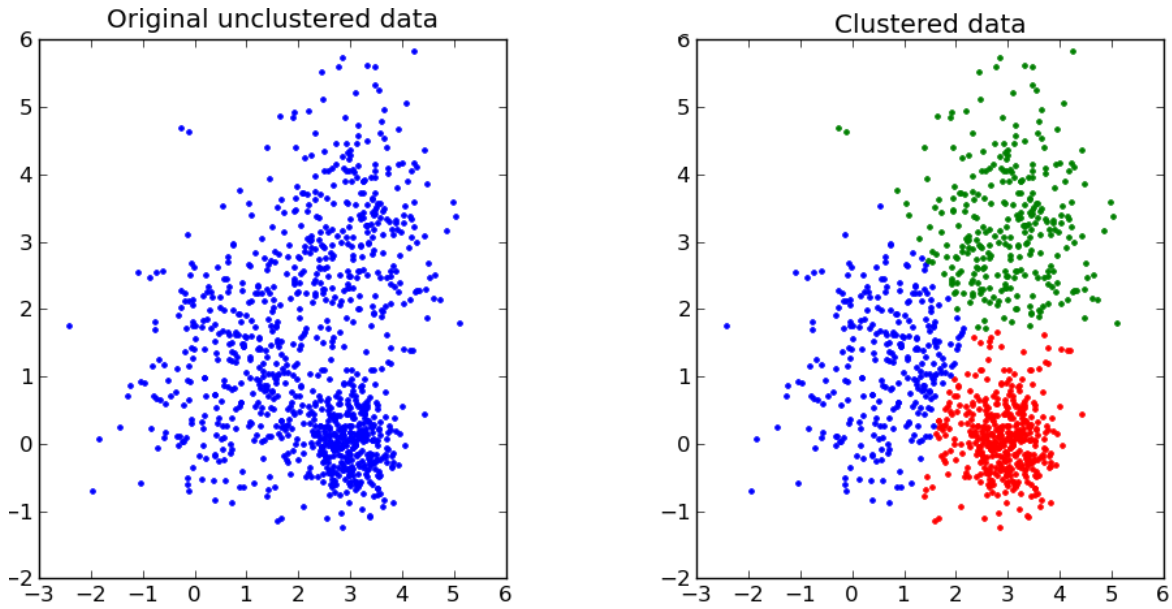


Figura 14: Exemplo do método de agrupamento conhecido como *K-Means* ou K-Médias realizando a separação de grupos (*clusters*) em um plano de duas dimensões.
Fonte: sourceforge.net/_images/kmeans_2d.png

Basicamente, o processo de agrupamento tem como objetivo agrupar conjuntos de dados baseando-se em suas similaridades. O termo "*cluster*" pode ser utilizado para denominar grupos gerados no processo. Novamente, na Figura 14, é possível ver que, o aglomerado de amostras presente no primeiro gráfico pôde ser separado através do algoritmo *k-means* em três diferentes *clusters*, simbolizados pelas cores azul, verde e vermelho. Para que seja possível tal separação, é necessário que existam métricas predefinidas para a medição de similaridade, para que assim, seja viável relacionar os dados entre si. O algoritmo *k-means* é um dos mais populares e pode ser implementado de maneira trivial. Seus passos são descritos a seguir. Considerando um plano bidimensional, tem-se

n pontos que representam amostras de um determinado conjunto de dados. De início, é necessário estipular a quantidade de *clusters* em que esses pontos podem ser agrupados. O número de *clusters* pode ser representado pelo valor k. O primeiro passo do algoritmo é estipular, aleatoriamente, a posição dos k pontos que irão ser colocados no plano para representarem um determinado *cluster*. Baseando-se no primeiro passo da Figura 15, tais pontos são representados por um X nas cores azul, vermelho e verde.

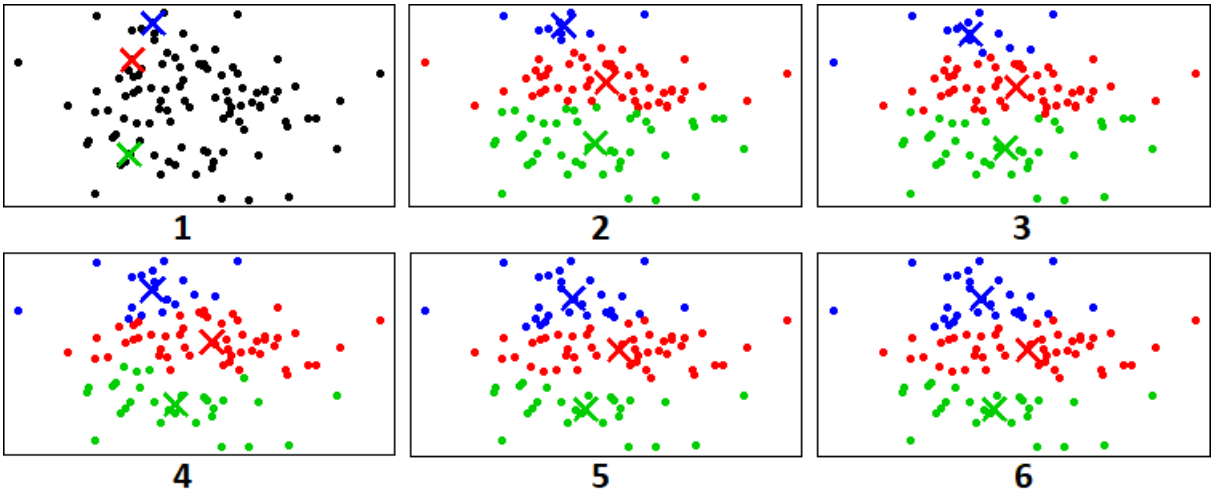


Figura 15: Demonstração do funcionamento do algoritmo *K-Means* até sua convergência.
Adaptado de: learnbymarketing.com/wp-content/uploads/2015/01/method-k-means-steps-example.png

O próximo passo é atrelar cada ponto do plano ao representante mais próximo, criando um *cluster* por centroide. Centroide é o ponto médio de um determinado *cluster*, ou seja, seu possível centro. Esta etapa pode ser realizada através do cálculo de distância, como a distância Euclidiana, por exemplo, que é medida a partir da soma dos quadrados da diferença de cada ponto. Após, será calculada a posição média de cada *cluster* e os k pontos representantes serão deslocados para lá. A medida que os pontos de referência se deslocam, sua distância para as amostras mudam, podendo existir uma mudança na composição dos *clusters*. No segundo passo da Figura 15, é possível ver as amostras do plano sendo atreladas a um grupo e, também, o deslocamento dos representantes para seu atual centroide. O posterior funcionamento do algoritmo baseia-se em recalcular os centroides e então, deslocar os representantes para o novo centroide (demais passos da Figura 15). Esta etapa é repetida até que seja alcançado algum critério de parada.

5 Aprendizado Semi-Supervisionado

O aprendizado semi-supervisionado faz uso tanto de dados rotulados quanto de dados não-rotulados (Zhu, 2011), ou seja, traz características presentes no aprendizado supervisionado e não-supervisionado (Chapelle et al., 2009). Uma vez que métodos de aprendizado supervisionado utilizam apenas dados rotulados, em algum momento é necessário o auxílio de especialistas para que a rotulagem seja feita. No entanto, devido a massiva quantidade de dados disponível, não é possível obter conjuntos rotulados em qualquer ocasião. Uma alternativa para esses casos é o uso dos métodos não-supervisionados. Porém, ao utilizar métodos não-supervisionados, outros obstáculos acabam surgindo. Devido a falta de informação prévia sobre as amostras, métodos não-supervisionados precisam assumir algumas premissas que não necessariamente representam a verdade. Um exemplo disso é o fato de partir da proposição de que dados com características similares pertencem à mesma classe sendo que, em alguns casos, as classes podem estar sobrepostas no plano (Jain et al., 1999).

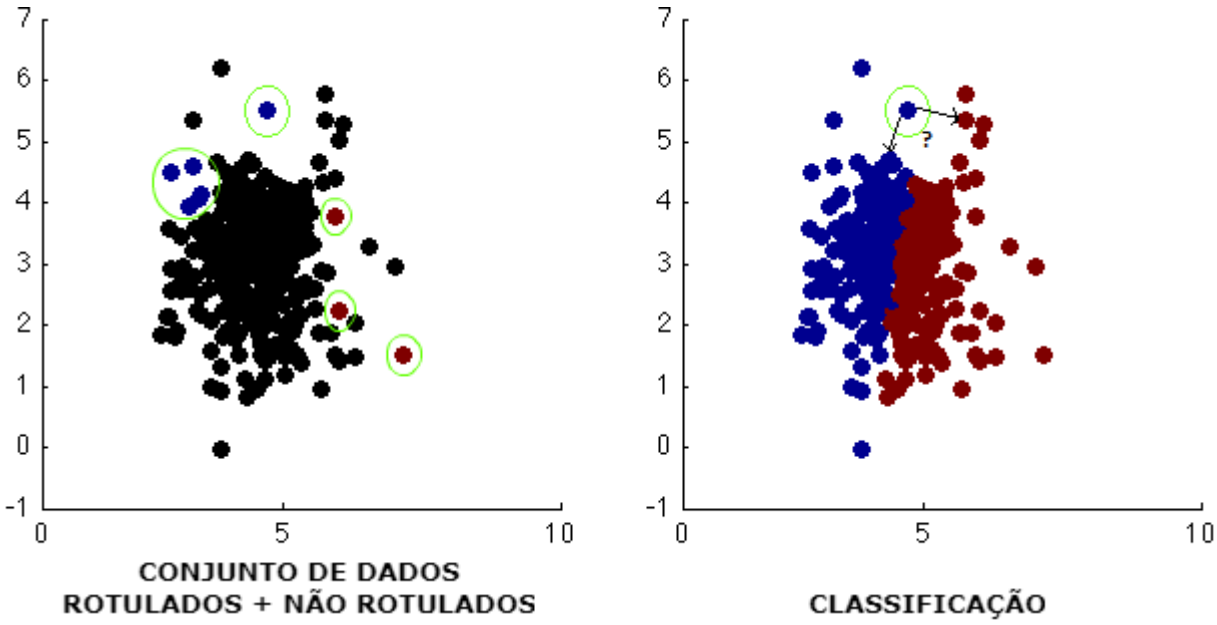


Figura 16: Representação de um conjunto de amostras que exemplifica a possível relevância da existência de dados rotulados em meio a outros durante uma classificação.
Adaptado de: matlab-cookbook.com/recipes/0100_Statistics/kmeans_groups_uni.png

As principais vantagens da ideia estão na redução do uso de dados rotulados, melhor aproveitando a quantidade de dados disponível. Fazendo uso de dados rotulados para

inferirem em não-rotulados (como exemplificado da Figura 16), economiza-se tempo e recursos. No entanto, é necessário que os métodos de aprendizado semi-supervisionado invistam esforços de forma a obter um bom aproveitamento dos dados não-rotulados, que constituem a maior parte do conjunto a ser utilizado (Zhu, 2005).

Parte da ideia do funcionamento de um método de aprendizado semi-supervisionado está na comparação entre pares de dados, classificando-os com base em alguma medida de distância (Silva e Breve, 2015). Atualmente, as abordagens semi-supervisionadas com maior destaque são as que utilizam métodos baseados em grafos, no entanto, existem diversos outros tipos de implementação de tal técnica. Nas subseções seguintes, serão apresentados os conceitos e funcionamento de algumas.

5.1 Auto-Treinamento

Na abordagem conhecida por auto-treinamento, dados rotulados compõem a etapa mais relevante do processo. O método consiste em utilizar a confiabilidade desses dados para que seja possível predizer a classe dos demais, evitando fazer previsões de risco. Os passos do processo são descritos a seguir.

Primeiramente, tem-se na base de dados D, os subconjuntos D_r e D_n que representam, respectivamente, os dados rotulados e não-rotulados (Roli e Marcialis, 2006). Com as amostras divididas desta forma, um classificador é treinado com todos os dados do subconjunto D_r. Após esta etapa de treinamento, o conhecimento adquirido pelo classificador é então utilizado para fornecer rótulos à amostras do subconjunto D_n, transformando dados presentes neste subconjunto em dados rotulados, que agora passam a fazer parte do subconjunto D_r. Os dados não-rotulados, que recebem sua classe durante este processo, são escolhidos através de métricas que medem o quão confiável aquela escolha é, ou seja, o quão alta é a chance daquela amostra ter sido classificada corretamente, fazendo com que o subconjunto D_r possua informações fundamentadas. Com a base D_r de dados rotulados alterada, tendo agora amostras adicionais, a etapa de treinamento com dados rotulados é realizada novamente, assim como a etapa de fornecer rótulos à amostras do subconjunto D_n. Portanto, a iteração do algoritmo ocorre em torno desses dois estágios. O ciclo conti-

nua até que a convergência seja atingida (onde todas as amostras se tornam rotuladas) ou até que algum critério de parada seja satisfeito, uma vez que a convergência nem sempre é garantida.

5.2 Co-Treinamento

Proposto inicialmente por Blum e Mitchell (1998), a ideia do método é fundamentada no treinamento de dois classificadores distintos. Em tal proposta, cada classificador faz uso de características distintas do dado na etapa de treinamento. Posteriormente, o aprendizado de ambos é combinado para que os dados sejam classificados. Este processo é conhecido como co-treinamento de duas visões. Goldman e Zhou (2000) apresentam uma forma diferente de implementar tal método, onde ambos classificadores se baseiam no mesmo conjunto de características para a etapa de treinamento (co-treinamento de visão única). O funcionamento desta abordagem é descrito na sequência.

Como citado anteriormente, dois classificadores são responsáveis por realizarem a rotulagem dos dados. Cada um deles faz uso de abordagens estatísticas para selecionarem amostras não-rotuladas e, posteriormente, classificá-las para o outro classificador. Considerando os classificadores como A e B onde, A é um algoritmo de aprendizado supervisionado e B é outro algoritmo de aprendizado supervisionado diferente de A, ambos são treinados de forma a rotularem amostras para que o outro classificador possa utilizar tais amostras rotuladas durante sua etapa de treinamento. O fator que define a confiabilidade dessas escolhas é chamado de hipótese, e para cada classificador deve haver uma hipótese previamente definida. Esta etapa se repete até não ocorrerem mudanças nos conjuntos que estão sendo rotulados. Por fim, assim que o conjunto rotulado de B fornecido por A e o conjunto rotulado de A fornecido por B estejam suficientemente definidos, são combinadas as hipóteses de cada algoritmo para que se tenha a hipótese final, fornecendo então, a classificação definitiva dos dados.

5.3 Co-Treinamento Duas Visões

O co-treinamento duas visões funciona de forma muito semelhante ao de visão única. No entanto, como já mencionado na subseção anterior, as amostras tem seus atributos divididos em duas partições, e cada classificador irá inferir apenas sob sua visão. Para melhor exemplificar tal comportamento, as tabelas a seguir ilustram uma estrutura hipotética dos dados.

Amostra	A_1	A_2	A_3	A_4	*Classe*
1	457	0,54	0,64	0,081	B
2	789	0,23	0,77	0,051	C
3	338	0,57	0,65	0,050	?
4	127	0,44	0,57	0,057	A
5	399	0,66	0,72	0,087	?

Tabela 1: Conjunto hipotético de amostras que possuem quatro atributos, referenciados por A_i.

Amostra	A_1	A_2	*Classe*
1	457	0,54	B
2	789	0,23	C
3	338	0,57	?
4	127	0,44	A
5	399	0,66	?

Amostra	A_3	A_4	*Classe*
1	0,64	0,081	B
2	0,77	0,051	C
3	0,65	0,050	?
4	0,57	0,057	A
5	0,72	0,087	?

Tabela 2: À esquerda, visão do classificador A. À direita, visão do classificador B.

Como é possível verificar, o conjunto de dados inicial é dividido em dois subconjuntos, onde, cada um deles leva uma parcela dos atributos. Desta forma, os classificadores realizam o mesmo processo descrito na subseção 5.2. A vantagem desta abordagem está no fato de que, a partir da visão do classificador A, B pode tomar decisões sob seu conjunto com mais confiabilidade, uma vez que a decisão de um, infere na amostra correspondente do outro, embora os atributos estejam separados (Sanches, 2003). O mesmo ocorre considerando a situação contrária, onde B infere sua visão em A.

5.4 Modelos Generativos

Esta abordagem adota uma forma de aprendizado semi-supervisionado que toma como base algum modelo estatístico como, por exemplo, a cadeia oculta de Markov em situações onde existem árvores ou sequências. A partir de parâmetros observáveis, determina-se o valor de parâmetros ocultos. Modelos generativos são, talvez, o método mais antigo de aprendizado semi-supervisionado e podem ser definidos pela fórmula $p(x, y) = p(y)p(x|y)$, onde, $p(x|y)$ é uma distribuição com mistura identificável, podendo ser um modelo de mistura Gaussiano, por exemplo (Zhu, 2005). Em conjuntos com uma enorme quantia de dados não-rotulados, os componentes da mistura podem possuir apenas um dado rotulado, enquanto o restante das amostras (não-rotuladas) seria presumida.

Uma das abordagens em modelos de mistura Gaussianos é o algoritmo de Maximização de Expectativa (*Expectative Maximization* ou *EM*) que foi utilizado para classificação de textos e trouxe resultados melhores que algoritmos treinados apenas com dados rotulados (Nigam et al., 2000). No entanto, é necessário que a presunção dos modelos de mistura estejam corretas, caso contrário, dados não-rotulados podem prejudicar a acurácia da classificação ao invés de a melhorarem (Castelli e Cover, 1995).

5.5 Separação de Baixa Densidade

Neste modelo é assumido que os limites, ou divisões das classes, estão em regiões de baixa densidade. Logo, a separação (ou classificação) dos dados ocorre a partir de tais regiões. Este tipo de abordagem pode ser diretamente relacionada a máquinas de vetor de suporte, no entanto, por se tratar de uma abordagem semi-supervisionada, há a existência tanto de dados rotulados como também de dados não-rotulados.

A Figura 17 ilustra a separação das classes baseando-se em regiões de baixa densidade através do algoritmo *TSVM* ou *Transductive Support Vector Machine*. Assim como no algoritmo *SVM*, tal abordagem semi-supervisionada também busca traçar uma separação de forma que a reta fique o mais distante possível de ambas as classes.

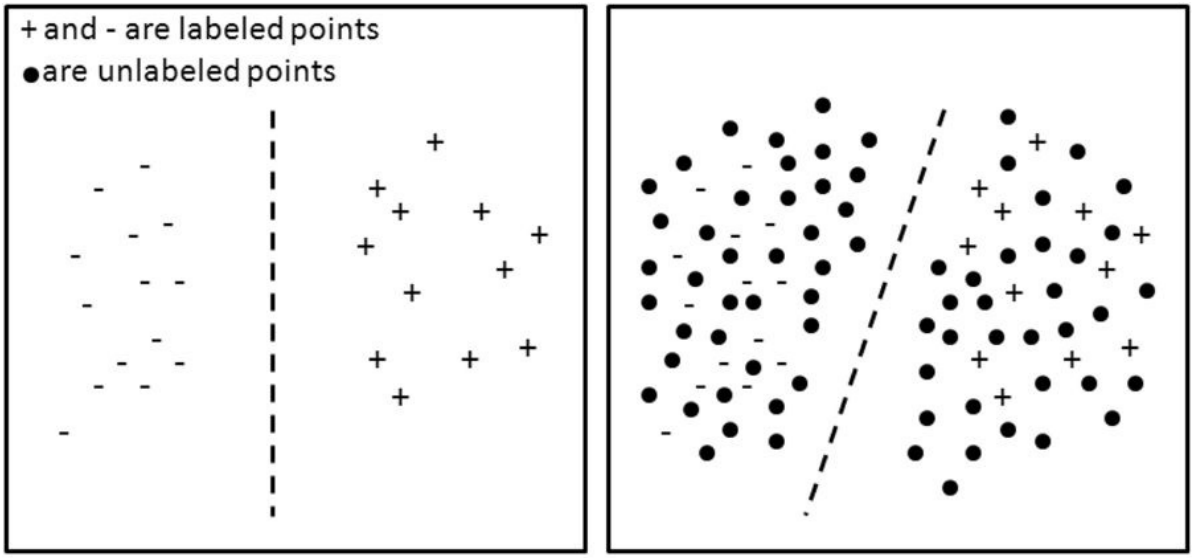


Figura 17: Ilustração da separação de baixa densidade, onde tem-se dados não-rotulados (representados por •), em meio a dados rotulados (representados por sinais de mais e menos). Linhas tracejadas indicam a separação das classes. À esquerda, tem-se a presença apenas dos dados rotulados. À direita, dos dados não-rotulados em meio aos rotulados.

Fonte: Peikari et al. (2018).

5.6 Métodos Baseados em Grafos

Métodos baseados em grafos, como o próprio nome já diz, se sustentam na estrutura de grafos para realizarem a classificação de dados. A abordagem semi-supervisionada deste modelo tem suas características definidas na sequência.

De forma simplificada e geral, grafos utilizados em tal método possuem seus nós representando amostras que, por sua vez, podem ser divididas em rotuladas e não-rotuladas. As arestas que ligam os nós do grafo também podem possuir pesos que indicam a similaridade entre duas amostras distintas. Uma outra forma de definir a similaridade é com a existência de arestas (sem pesos) apenas entre os k vizinhos mais próximos de cada nó. A grande maioria dos métodos funciona de forma iterativa e, desta maneira, a função dos nós rotulados é propagar informações de rótulo para os demais nós. Este passo é feito a cada iteração até que seja alcançada a convergência (Breve, 2010).

Ainda que métodos baseados em grafos sejam uma das áreas mais ativas de pesquisa

em aprendizado semi-supervisionado (Chapelle et al., 2006), a grande maioria dos algoritmos ainda possui uma complexidade computacional muito alta ($O(n^3)$), inviabilizando o uso de bases de dados muito densas. O algoritmo de competição e cooperação entre partículas que será utilizado no presente projeto de pesquisa, descrito na seção 7, se destaca neste quesito, possuindo complexidade linear ($O(n)$) no laço principal (Breve et al., 2012).

No trabalho de Park et al. (2014) para análise de recorrência de câncer utilizando aprendizado semi-supervisionado, a abordagem baseada em grafos é utilizada. Através da Figura 18, gerada pelos autores para ilustrar o funcionamento do método, é possível observar como podem funcionar as técnicas desta categoria. Inicialmente tem-se um conjunto de três dados diferentes sendo, rotulados pertencentes a classe branca (ou classe 1), rotulados pertencentes a classe preta (ou classe 2) e não rotulados. O grafo da esquerda representa o estado inicial e o da direita a convergência, ou seja, a completa rotulagem dos dados que não possuíam classe.

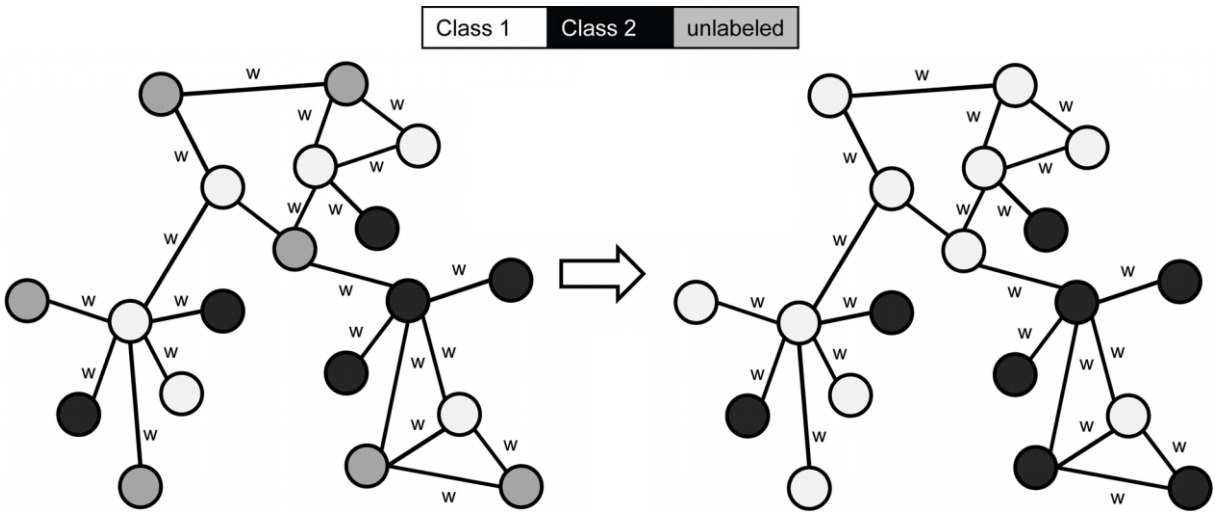


Figura 18: Demonstração ilustrativa da convergência de um método semi-supervisionado baseado em grafos. No topo da imagem está presente a legenda indicando o significado de cada nó. A letra w representa a existência de um peso w em uma aresta.

Adaptado de: Park et al. (2014).

6 Trabalhos Relacionados

Com o objetivo de entender melhor as abordagens utilizadas na classificação da doença de Alzheimer, este capítulo discorre sobre alguns trabalhos relacionados. São trazidos os passos envolvidos no processo de classificação, assim como algumas conclusões nas quais os autores chegaram. Por fim, é apresentada uma breve discussão.

6.1 Discriminação entre Alzheimer e CCL utilizando SOM e PSO-SVM

Yang et al. (2013) propuseram um *framework* baseado em imagens de ressonância magnética para diferenciar pacientes que possuem a doença de Alzheimer ou comprometimento cognitivo leve, de pessoas saudáveis. Foram utilizadas diferentes características e classificadores para tal. A etapa de extração de características das imagens de ressonância magnética fez uso de informações presentes no formato e volume das imagens para a criação das amostras a serem utilizadas nos algoritmos. Os dados obtidos na extração, são submetidos ao algoritmo de Análise de Componentes Principais (do inglês *Principal Component Analysis*, ou *PCA*), que teve como função, de acordo com os autores, reduzir a dimensionalidade dos dados. Esta etapa foi nomeada como "redução de características" e trouxe uma significante melhora nos resultados.

Com os dados devidamente extraídos e pré-processados, os autores desenvolveram um *framework* que combinava o algoritmo de Máquinas de Vetor de Suporte (SVM) com o algoritmo de Otimização por Enxame de Partículas (PSO) para chegar a classificação das amostras. Também foram utilizados, separadamente, mapas auto-organizáveis (SOM) e, novamente, Máquinas de Vetor de Suporte, para que houvesse a classificação da base de dados como uma maneira de comparar as abordagens. A Figura 19 demonstra os passos descritos acima. Os resultados obtidos na classificação focada em diferenciar pacientes saudáveis, de pacientes com a doença de Alzheimer, são apresentados na Tabela 3.

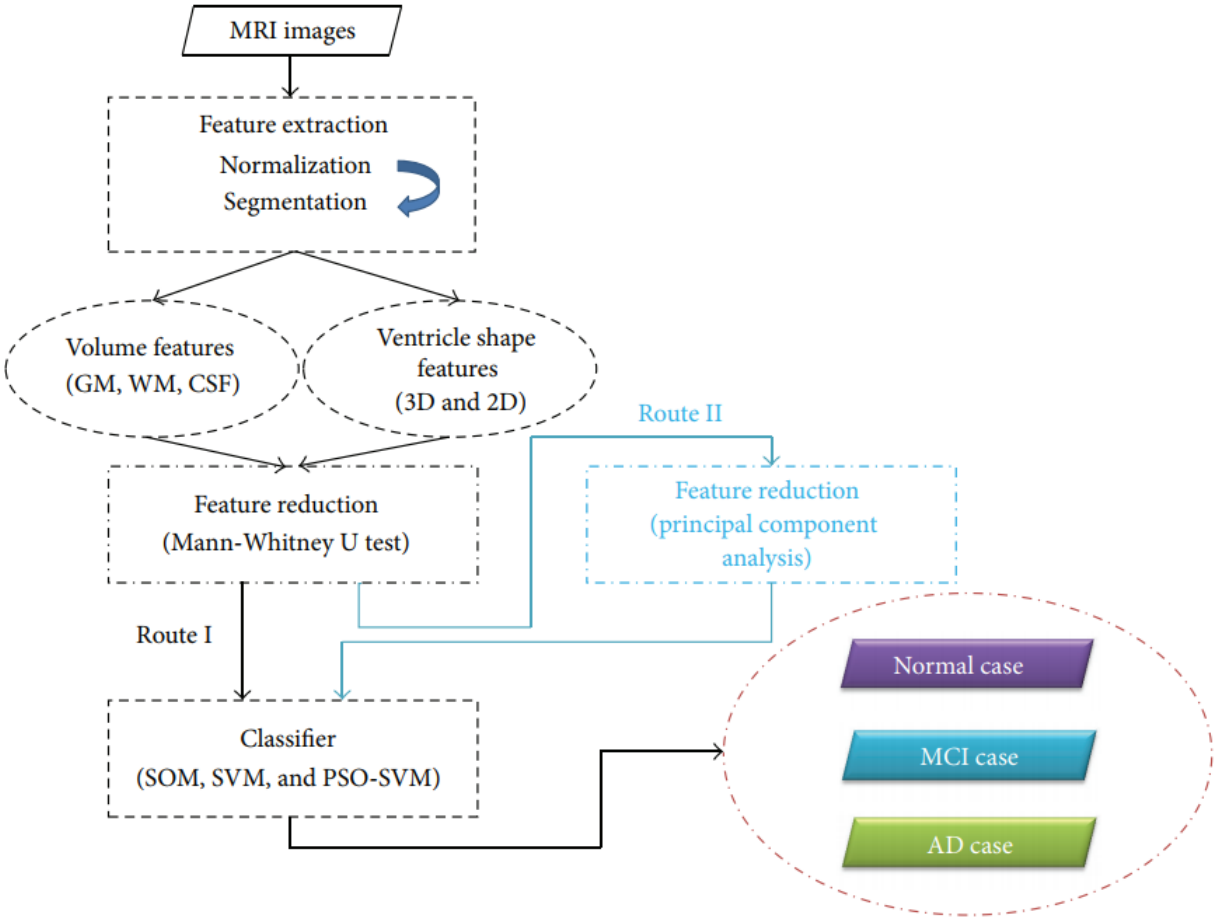


Figura 19: Diagrama representando os passos até a classificação dos dados.

Fonte: Yang et al. (2013).

Relação	Volume	Volume + PCA	Formato	Formato + PCA	Volume + Formato	Volume + Formato + PCA
Acurácia	76,47%	82,35%	64,71%	70,59%	76,47%	88,24%
Sensibilidade	81,25%	87,50%	68,75%	70,59%	76,47%	88,24%
Especificidade	77,78%	83,33%	66,67%	70,59%	76,47%	88,24%

Tabela 3: Resultados combinando as diferentes características e o uso do método PCA.

Os autores concluem que a combinação de diferentes características extraídas das imagens resultam em melhores resultados. Isto é possível de ser verificado através da Tabela 3, onde, o ganho entre as características de volume e formato, quando juntas, se sobressaem quando comparadas às demais. O mesmo pode ser concluído quando há o uso da análise estatística do PCA, onde há um significante aumento na acurácia em todos os casos.

6.2 Classificação automática de estágios iniciais da doença de Alzheimer com PSVM, kNN e ANN

Tufail et al. (2012) também apresentaram uma abordagem baseada no uso de imagens de ressonância magnética para classificar pacientes em três estados diferentes, sendo eles, estado de comprometimento cognitivo leve, doença de Alzheimer e estado normal, assim como Yang et al. (2013) o fez um ano mais tarde. Além disso, uma outra semelhança é o uso do método *PCA*, que também está presente em ambos os trabalhos. A solução proposta por Tufail et al. (2012) é descrita por seis estágios, ilustrados pela Figura 20, sendo possível agrupá-los em duas fases. Na fase de pré-processamento dos dados, estão a formação do conjunto de imagens, o redimensionamento das mesmas através de uma interpolação linear de três dimensões e, por último, o uso do *PCA* para redução de dimensionalidade, onde o autor menciona que o espaço é reduzido para cem características. Uma vez que os dados estão devidamente trabalhados, ocorre a extração de características através do método de Análise de Componentes Independentes (*Independent Component Analysis* ou *ICA*), seguido da seleção das características a serem utilizadas. Por fim, as amostras são submetidas aos métodos *PSVM* (*Proximal Support Vector Machines*), que é uma variante das máquinas de vetor de suporte, *kNN*, que é um método de aprendizado supervisionado e também redes neurais artificiais (*Artificial Neural Networks*).

Figura 20: Diagrama representando os passos até a classificação dos dados.

Fonte: Tufail et al. (2012).

O método *ICA* foi utilizado devido ao objetivo de encontrar características não-gaussianas que possam proporcionar componentes originais, destacando-os de outras pro-

priedades possivelmente compartilhadas entre todas as amostras.

Os resultados obtidos pelos autores foram divididos em nove tabelas, que podem ser separadas em três análises (Tabelas 4,5,6). Em todas as tabelas o autor traz a acurácia dos métodos dividida por classes, ou seja, é informada a porcentagem que a classe c foi classificada corretamente como classe c. A primeira mostra o desempenho dos três métodos quando o propósito é classificar as amostras no estado de comprometimento cognitivo leve (MCI) ou estado normal (NC).

	MCI	NC
kNN	72,86%	43,21%
ANN	42,86%	64,29%
PSVM	26,53%	67,86%

Tabela 4: Resultados de Tufail et al. (2012) para MCI vs NC.

O segundo conjunto de tabelas demonstra resultados provenientes da classificação entre estado normal (NC) e doença de Alzheimer (AD).

	AD	NC
kNN	82,35%	47,06%
ANN	33,33%	47,06%
PSVM	66,15%	51,19%

Tabela 5: Resultados de Tufail et al. (2012) para AD vs NC.

A terceira e última análise traz tabelas da acurácia de cada método, tendo comprometimento cognitivo leve (MCI) e doença de Alzheimer (AD) como classes.

	AD	MCI
kNN	97,92%	38,19%
ANN	33,33%	50,00%
PSVM	65,08%	56,22%

Tabela 6: Resultados de Tufail et al. (2012) para MCI vs AD.

A conclusão dos autores foi que os resultados foram satisfatórios em termos de acurácia e tempo de execução para os algoritmos kNN e PSVM. Eles ainda complementam mencionando que seria algo interessante analisar resultados provenientes de outros métodos de extração de características, assim como avaliar a melhora utilizando uma base de dados mais extensa.

6.3 Discussão

Nos trabalhos citados, ainda que o algoritmo SVM lide bem com a alta dimensionalidade dos dados (Colas et al., 2009), foi possível observar que ambos os autores optaram pela utilização do método de análise de componentes principais (*PCA*) para a redução de dimensionalidade. O *PCA* analisa o conjunto de dados em busca de variáveis correlacionadas e, a partir delas, extrai a informação contida representando-a em um novo conjunto de variáveis chamadas de componentes principais (Abdi e Williams, 2010). Basicamente, a proposta é reduzir a dimensionalidade de um certo dado e, ao mesmo tempo, preservar a variabilidade das suas características (Jolliffe e Cadima, 2016). A partir do trabalho de Yang et al. (2013), é possível observar que não apenas o *PCA* aprimora o resultado como, a presença de mais categorias de características (volume e formato) também o fazem. No trabalho de Tufail (Tufail et al., 2012), a extração de características é feita a partir do método da análise de componentes independentes (*ICA*), que separa características independentes em meio a mistura presente na imagem (Hoyer e Hyvärinen, 2000). Ainda que de modo geral o resultado do algoritmo kNN tenha sido superior, o algoritmo PSVM se destaca por classificar bem as amostras de pacientes saudáveis, reduzindo a quantidade de falsos positivos (melhor explicado no Capítulo 9.1), que seriam pacientes diagnosticados, e posteriormente tratados como sendo portadores de Alzheimer, sem tê-lo. Tendo em vista os aspectos observados, a seção de experimentos (Capítulo 9) abordará novamente alguns pontos aqui colocados de forma a avaliar os estudos dos autores.

7 Competição e Cooperação entre Partículas

O modelo de competição e cooperação entre partículas a ser aplicado neste trabalho é classificado como um algoritmo de aprendizado semi-supervisionado e, mais especificamente, é baseado em uma busca por território em um grafo.

Tal abordagem é derivada da combinação do modelo de competição de partículas (Quiles et al., 2008), que, basicamente, adota que partículas caminham no grafo e competem entre si de modo que cada uma visa possuir o maior território possível (Figura 21), e, do conceito de cooperação entre partículas, onde partículas de mesma classe se movimentam de forma cooperativa para buscar esse território (Breve et al., 2012).

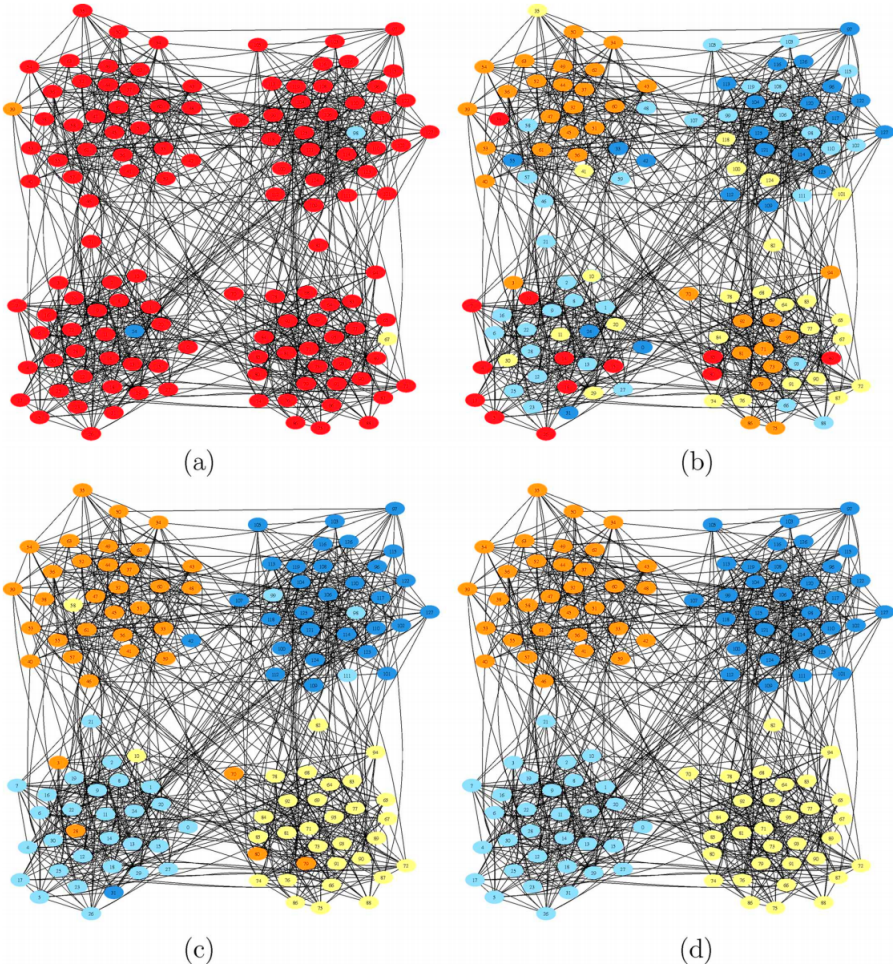


Figura 21: Detecção de 4 comunidades distintas em uma rede com 128 amostras através do algoritmo de competição de partículas. A Figura 21a ilustra a rede em seu estado inicial. As ilustrações seguintes (10b, 10c, 10d) demonstram o estado da rede após, 250, 3500 e 7000 iterações, respectivamente. A cor vermelha simboliza amostras não-rotuladas, enquanto o restante das cores correspondem a cada uma das 4 comunidades.
Fonte: Quiles et al. (2008).

A partir dos conceitos abordados, obteve-se um novo classificador semi-supervisionado que faz uso da estrutura de grafos para a detecção de comunidades em redes utilizando dados rotulados para inferirem sobre dados não-rotulados.

Além disso, diferente da maioria dos modelos baseados em grafo que possuem complexidade cúbica ($O(n^3)$) (Zhu, 2005), o modelo de competição e cooperação entre partículas leva vantagem sobre os demais pelo fato de ter complexidade linear ($O(n)$) no laço principal (Breve et al., 2012), como já mencionado na seção 5.6. Isto se deve ao fato de que a propagação de rótulos deste modelo é feita de maneira local, ou seja, apenas um nó é inferido a cada movimento, diferente da grande maioria dos métodos, onde a propagação ocorre de forma global. Demais conceitos e o funcionamento do algoritmo são descritos a seguir.

Inicialmente, é necessário que a base de dados a ser utilizada seja transformada em uma rede (ou grafo). A conversão ocorre da seguinte forma: no primeiro passo, cada amostra do conjunto de dados é transformada em um nó. Cada nó pode ser rotulado ou não-rotulado. Seguidamente, cria-se a relação inicial entre todas as amostras, ou seja, ocorre a ligação dos nós através de arestas não-direcionadas que não possuem pesos, como ilustra a Figura 22.

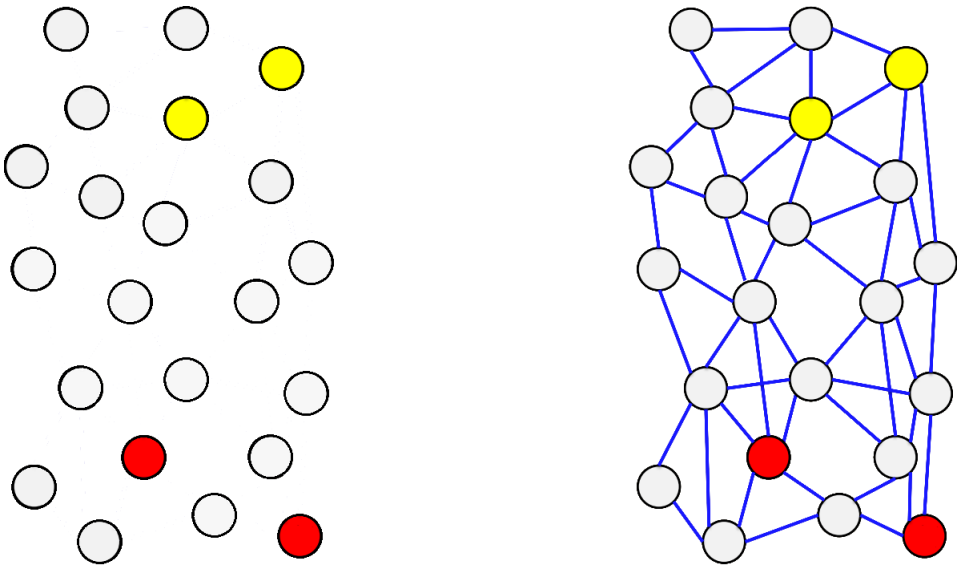


Figura 22: Ilustração do grafo no modelo de partículas. Nós brancos representam as amostras não-rotuladas. Nós vermelhos e amarelos representam as amostras rotuladas.

A ligação dos nós pode ser feita de várias formas. Uma das formas é adotar que cada

nó será conectado aos vizinhos que possuírem uma distância menor que um dado limiar, ou seja, a aresta E_{ij}, que liga os nós v_i e v_j, existirá apenas se a amostra v_i estiver em uma distância menor ou igual a σ do nó v_j. Uma outra forma adota que cada nó será conectado aos seus k vizinhos mais próximos, ou seja, todo nó terá no mínimo k arestas para k nós distintos.

Com os nós devidamente conectados entre si, o próximo passo é gerar as partículas que caminharão na rede. A Figura 23 ilustra o procedimento que pode ser explicado da seguinte forma: cada nó do grafo que já possui um rótulo irá receber uma partícula. Este nó rotulado será o nó *casa* da partícula que foi ali gerada. Nós *casa* contam com propriedades diferentes dos demais nós.

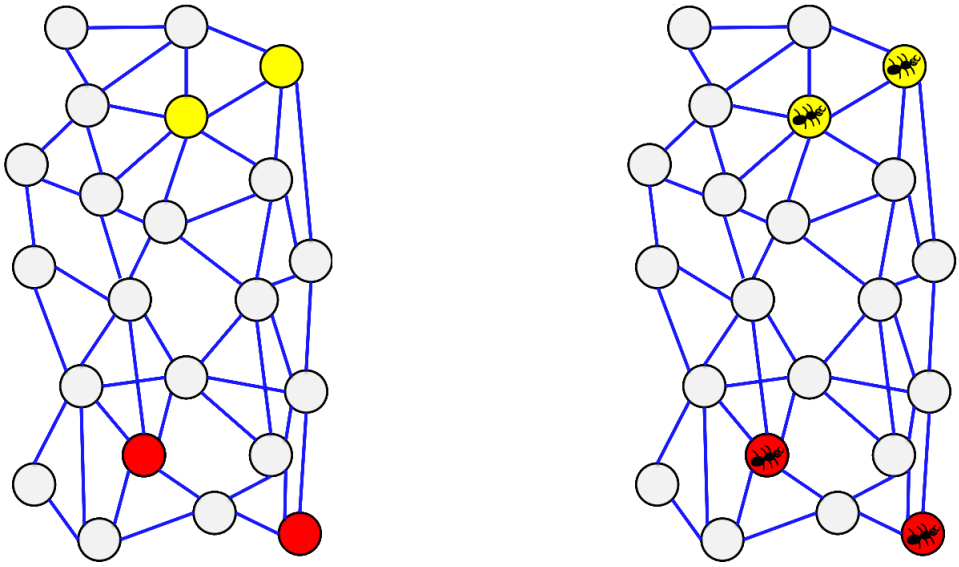


Figura 23: Ilustração do surgimento de partículas no grafo.

A dinâmica dos nós é dada por uma estrutura chamada de "vetor de domínio". Vetores de domínio indicam o quão dominado por uma classe aquele nó está. Em um exemplo com 4 classes, um vetor de domínio não-rotulado, inicialmente, estaria dominado igualitariamente por todas as classes, ou seja, cada classe dominaria 25% daquele território (Figura 24B). Enquanto isso, vetores de domínio de nós rotulados seriam 100% dominados pela classe de seu rótulo (Figura 24A). Com os valores inicialmente pré-definidos, as partículas irão caminhar no grafo buscando aumentar o domínio de sua classe em todos os nós. Cada vez que uma partícula visita um nó não-rotulado, a mesma aumenta o nível de domínio do seu rótulo e diminui o nível de domínio dos demais rótulos (Figura 25A).

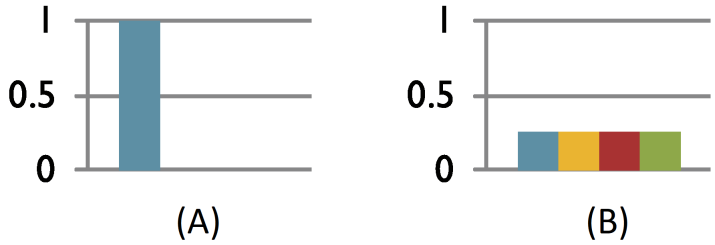


Figura 24: Comparação entre vetores de domínio. À esquerda, vetor de domínio de um nó rotulado. À direita, vetor de domínio de um nó não-rotulado.
Fonte: Breve et al. (2012).

Desta forma, considere um conjunto de rótulos L onde $y_i \in L$ configura uma amostra rotulada e $y_i = \emptyset$ uma amostra não rotulada. Sendo $\rho_j^{\omega}(t)$ a força de uma partícula ρ_j com rótulo ρ_j^f em uma iteração t, cada nó v_i tem seu vetor de domínio $v_i^{\omega_\ell}(t)$ atualizado da seguinte maneira:

$$v_i^{\omega_\ell}(t+1) = \begin{cases} \max\{0, v_i^{\omega_\ell}(t) - \frac{\Delta_v \rho_j^{\omega}(t)}{c-1}\} \\ \quad \text{se } y_i = \emptyset \text{ e } \ell \neq \rho_j^f \\ v_i^{\omega_\ell}(t) + \sum_{q \neq \ell} v_i^{\omega_q}(t) - v_i^{\omega_q}(t+1) \quad , \\ \quad \text{se } y_i = \emptyset \text{ e } \ell = \rho_j^f \\ v_i^{\omega_\ell}(t) \quad \text{se } y_i \in L \end{cases} \tag{3}$$

Onde ℓ representa uma determinada classe, Δ_v é um parâmetro que controla a taxa de alteração nos níveis de domínio e c é a quantidade de classes. Uma vez que o vetor de domínio é atualizado, a partícula tem sua força ditada pela equação abaixo:

$$\rho_j^{\omega}(t+1) = v_i^{\omega_\ell}(t+1), \tag{4}$$

Desta forma, percebe-se que partículas que caminham em território inimigo vão perdendo força gradativamente. Para evitar esta situação, o grafo é atualizado automaticamente a cada passo para que a partícula saiba o quão longe está do seu nó *casa*. Sendo assim, a partícula prioriza proteger sua própria vizinhança. As Figuras 25B e 25C demonstram visualmente a força de uma partícula de acordo com o território.

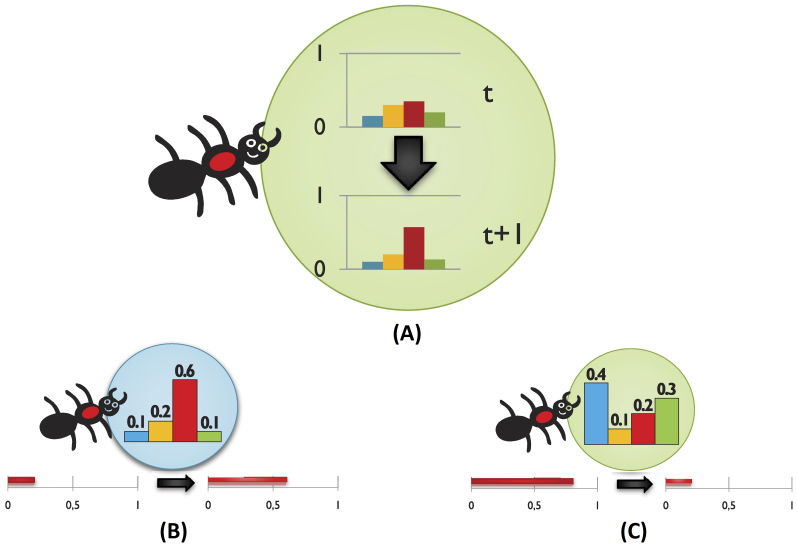


Figura 25: (A) Ilustração de uma partícula aplicando seu domínio a um nó (antes (t) e depois ($t+1$)). (B) e (C), demonstração da força de uma partícula (abaixo do círculo, em vermelho) antes, e depois de visitar um certo território.

Fonte: Breve et al. (2012).

A movimentação das partículas pode ocorrer de duas formas. Uma delas é aleatória, onde a partícula escolhe o próximo nó sem se basear em níveis de domínio. A outra forma de movimentação é gulosa, onde a preferência da partícula é por nós onde já existe um certo domínio do seu rótulo. Ambas as formas de movimentação ocorrem, para que assim, a partícula equilibre seu comportamento de explorar e ao mesmo tempo, defender. As duas formas de movimentação são ilustradas pela Figura 26. O cálculo que descreve cada forma de movimento é explicado na sequência.

Para o movimento aleatório, sendo q o índice do nó atual da partícula, a probabilidade do nó v_i ser visitado pela partícula ρ_j pode ser descrita pela equação abaixo:

$$p(v_i|\rho_j) = \frac{W_{qi}}{\sum_{\mu=1}^{n} W_{q\mu}}, \tag{5}$$

É atribuído o valor 1 para W_{qi} caso exista uma aresta entre o nó atual e qualquer outro nó. Caso contrário, W_{qi} possui valor 0. No movimento guloso, com $\rho_j^{d_i}$ representando o

inverso da distância entre o nó v_i e o nó casa da partícula (v_j), tem-se:

$$p(v_i|\rho_j) = \frac{W_{qi} v_i^{\omega_\ell} (1 + \rho_j^{d_i})^{-2}}{\sum_{\mu=1}^{n} W_{q\mu} v_\mu^{\omega_\ell} (1 + \rho_j^{d_\mu})^{-2}}. \tag{6}$$

A cada iteração, cada partícula tem a probabilidade P_{grd} de realizar o movimento guloso. Desta forma, a partícula tem $1 - P_{grd}$ de probabilidade de realizar o movimento aleatório ($0 \leq p_{\text{grd}} \leq 1.$).

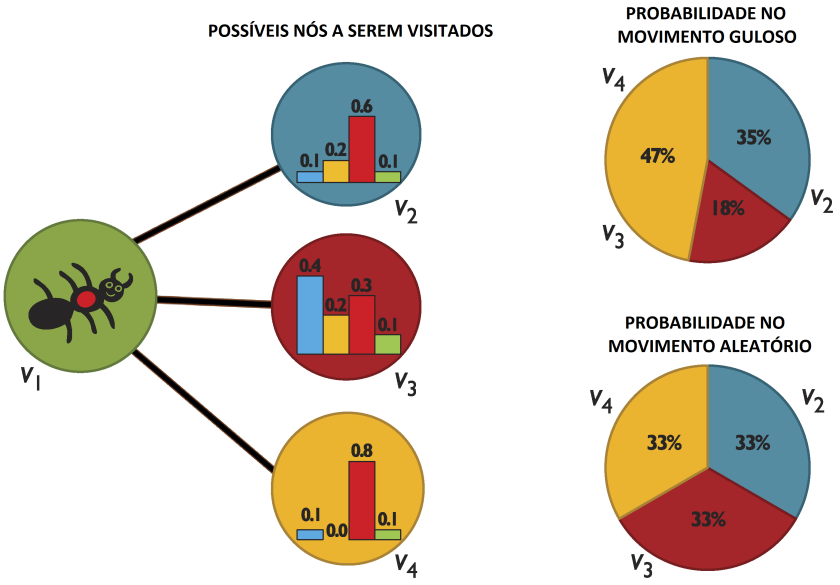


Figura 26: Ilustração da probabilidade de movimentação para os nós azul, vermelho e laranja quando adotada a estratégia aleatória e gulosa.

Adaptado de: Breve et al. (2012).

Por fim, as partículas se movimentam pelo grafo até que algum critério de parada seja atingido. Um dos critérios de parada que podem ser adotados levam em consideração o nível de mudanças que a rede está sofrendo e, caso ele seja mínimo, entende-se que as classes já foram definidas. Quando isto ocorre, os nós do grafo que não possuem rótulos são rotulados com a classe que possui o maior domínio.

8 Metodologia e Plano de Trabalho

Nesta seção serão abordados temas referentes à forma de execução da abordagem proposta, ferramentas utilizadas e descrição da base de dados.

8.1 Abordagem Proposta

Como já mencionado anteriormente, o uso do método de competição e cooperação entre partículas para auxílio no diagnóstico da doença de Alzheimer se faz interessante pelo fato de permitir um bom aproveitamento dos dados não-rotulados, sem deixar de lado a informação presente nos dados rotulados. No entanto, para que seja possível utilizar imagens de ressonância magnética, é necessária a extração das características presentes em tais.

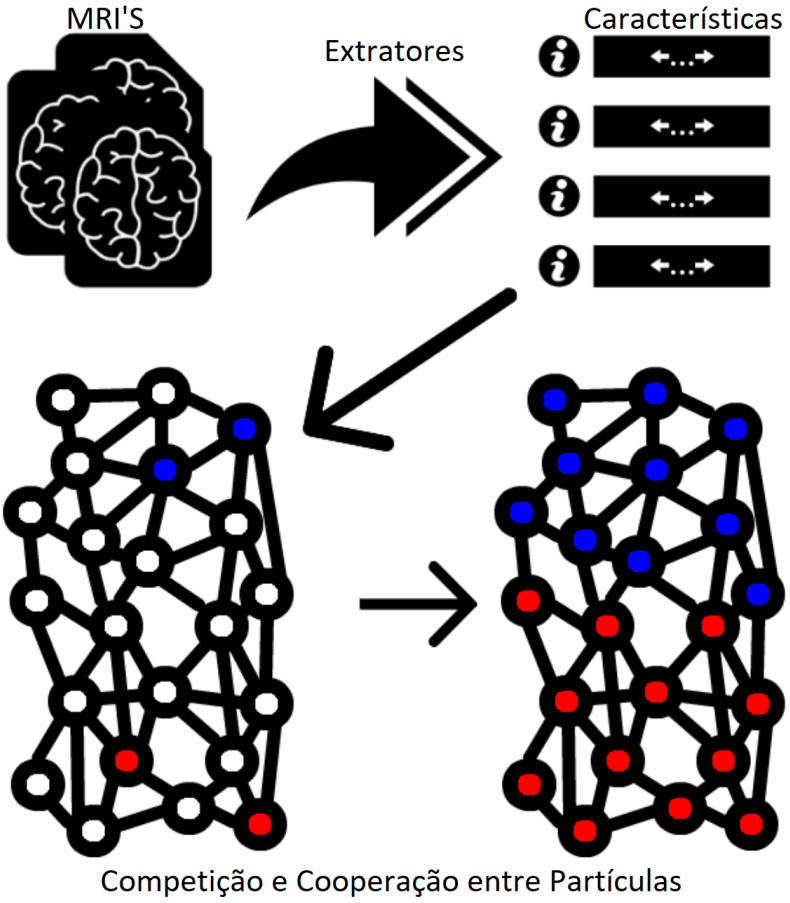


Figura 27: Ilustração simplificada da metodologia de trabalho.

Fonte: Elaborada pelo autor.

Simplificados pela Figura 27, os passos envolvidos na abordagem proposta iniciam-se na separação das imagens de ressonância magnética a serem utilizadas, retiradas da base de dados OASIS (Marcus et al., 2007), descrita na próxima seção. Em seguida, as imagens serão submetidas à etapa de extração de características, através da biblioteca de recuperação de imagens baseada em conteúdo LIRe (Lux e Chatzichristofis, 2008) e biblioteca de aprendizagem profunda Keras (Chollet et al., 2015). Por fim, as amostras em formato de vetores de características servirão de entrada para o método de competição e cooperação entre partículas e outros métodos descritos no Capítulo 9.

A etapa de escolha das imagens a serem utilizadas se baseia na forma a qual o trabalho será conduzido, considerando as três possíveis descritas por Cuingnet et al. (2011), presentes no Capítulo 2. A ideia é utilizar imagens retiradas do topo do cérebro do paciente, para que assim, se faça uso da primeira abordagem descrita por Cuingnet, analisando as proporções de massa cinzenta, massa branca e líquido cefalorraquidiano.

Uma vez definida a forma de abordar os sinais da doença, serão escolhidos os descritores que melhor evidenciam estes sinais. Prioritariamente, as deformações cerebrais relacionadas à *WM*, *GM* e *CSF* alteram dados como intensidade de cinza e formato, induzindo à preferência de descritores que explorem essas propriedades.

Com as amostras devidamente convertidas em vetores de características, serão efetuados testes de forma à avaliar o desempenho do método de competição e cooperação entre partículas. Em virtude de todos os dados da base OASIS possuírem rótulos, também será possível mensurar a proporção de dados rotulados que trazem o melhor aproveitamento, sendo esta, a principal medida para ponderar o impacto do trabalho.

8.1.1 Base de Dados OASIS

A base de dados OASIS (sigla para *Open Access Series of Imaging Studies*) (Marcus et al., 2007) fornece imagens cerebrais gratuitas, livres para uso e distribuição da comunidade científica. A coleção traz três tipos de conjuntos, sendo eles o conjunto OASIS-3, que agrupa dados longitudinais, trazendo informações clínicas e cognitivas de 1098 pacientes entre 45 e 95 anos de idade com o uso tanto de ressonância magnética quanto de

tomografia por emissão de positrões (popularmente conhecida por *PET Scan*). O conjunto OASIS-2, que também traz dados longitudinais, mas utilizando apenas imagens de ressonância magnética, onde estão presentes dados de 150 pacientes tendo entre 60 e 96 anos. E, por fim, o conjunto OASIS-1, que traz *MRI's* de 416 indivíduos, com idade entre 18 e 96 anos. Em todos os conjuntos, os indivíduos podem ou não possuir algum tipo de demência (entre elas, a doença de Alzheimer em qualquer estágio).

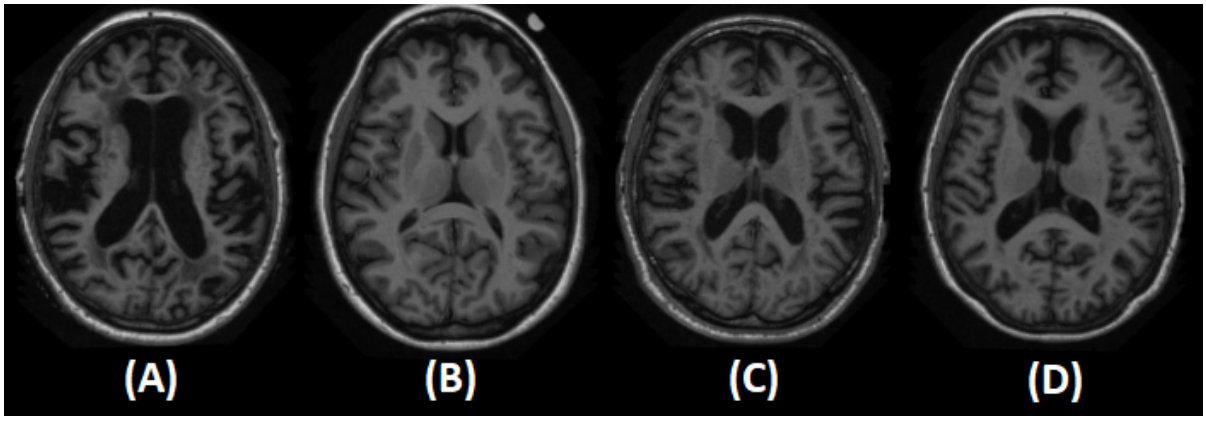


Figura 28: Imagem trazendo 4 diferentes sessões de *MRI's* da base OASIS-1, onde, cada uma possui um diagnóstico diferente. A Figura 28A é de um indivíduo que foi diagnosticado com a doença de Alzheimer em estágio leve, a Figura 28B é de um paciente que não foi diagnosticado com nenhum tipo de demência, a Figura 28C é de um dos dois pacientes que foram diagnosticados com Alzheimer em estágio moderado e, por fim, a A Figura 28D é de um paciente diagnosticado com doença de Alzheimer em estágio muito leve.

Fonte: Marcus et al. (2007).

Para este projeto de pesquisa, foi utilizado o conjunto OASIS-1 que, dentre os três, traz informações mais robustas e de maior viabilidade para a extração de características e posterior classificação automática. A apresentação mais detalhada de tal conjunto é traçada a seguir.

Para cada um dos 416 indivíduos, 3 ou 4 imagens de ressonância magnética ponderadas em T1 (explicadas no Capítulo 2) foram obtidas. A base é composta tanto por pacientes do sexo masculino quanto do sexo feminino. Todos os indivíduos são destros. Dos 416 indivíduos presentes na base, 98 deles foram diagnosticados com DA leve ou muito leve

e apenas 2 possuem diagnóstico de DA moderada, todos com 60 ou mais anos de idade. Os restantes 316 indivíduos não foram diagnosticados com nenhum tipo de demência. A Figura 28 traz sessões de ressonância magnética de indivíduos com as quatro classes presentes na base de dados, sendo elas as sessões 0031, 0002, 0308 e 0041 respectivamente.

A base também conta com imagens já segmentadas através do método *k-means* (que tem seu funcionamento descrito no Capítulo 4.2) diferenciando tecidos brancos e cinzentos, assim como a área onde há presença de líquido cefalorraquidiano. Abaixo, a Figura 29 traz algumas imagens que expõem o resultado deste processamento.

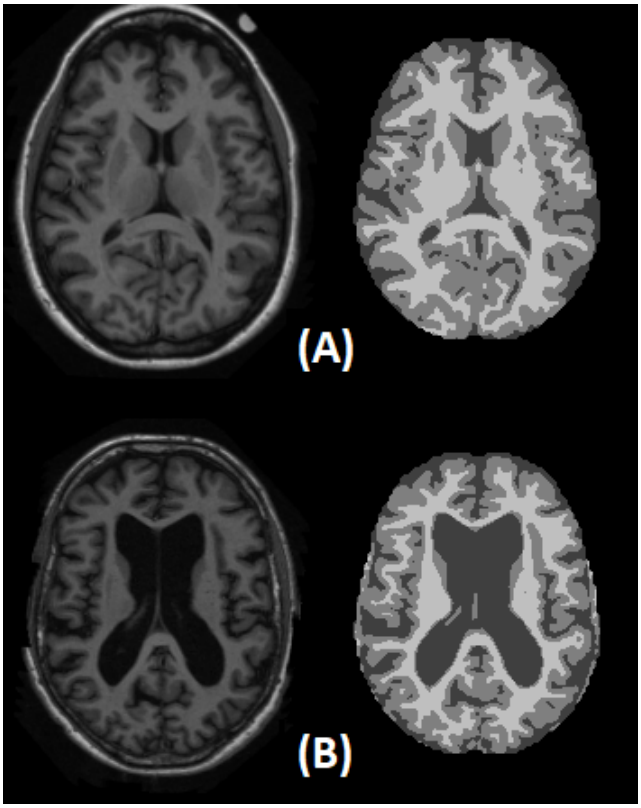


Figura 29: A Figura 29A, pertencente a um paciente saudável, traz, em sua esquerda, uma imagem sem nenhum tipo de segmentação, enquanto, à direita, tem-se a mesma imagem, porém, segmentada através do método *k-means*. O mesmo acontece com a Figura 29B, de um indivíduo com DA leve.

Fonte: Marcus et al. (2007).

8.1.2 Biblioteca LIRe

LIRe (Lux e Chatzichristofis, 2008), abreviação de *Lucene Image Retrieval* é uma biblioteca *open source* programada em Java que fornece aplicações para sistemas de recuperação de imagens baseada em conteúdo. Como principal contribuição para este trabalho, a biblioteca oferece diversos métodos de extração de características, baseando-se em atributos como cor, textura, forma (abordados no Capítulo 3) e outros. Além disso, a biblioteca também viabiliza o armazenamento de tais características em índices para posterior recuperação.

Fazendo uma comparação dos recursos inicialmente disponibilizados por Lux e Chatzichristofis com o estado atual da biblioteca, houve um significante aumento no número de descritores assim como outras funcionalidades, porém, entre os principais recursos disponíveis estão os métodos como CEDD, FCTH, EHD e SCD (também abordados no Capítulo 3). À vista disso, pretende-se fazer uso das técnicas oferecidas pela biblioteca para que seja possível alcançar uma boa descrição das imagens de ressonância magnética, trazendo um impacto direto na acurácia de classificação das mesmas pelo método de competição e cooperação entre partículas.

8.1.3 Biblioteca Keras

Keras (Chollet et al., 2015) é uma biblioteca de aprendizagem profunda que foi inicialmente projetada para pesquisadores com o propósito de fornecer rápidas experimentações. Com o uso desta biblioteca, é possível definir e treinar diversos modelos de aprendizagem profunda. Outros recursos relevantes serão destacados na sequência.

Com o uso da biblioteca Keras, é possível executar códigos tanto na unidade central de processamento (também conhecida como *CPU*), como na unidade de processamento gráfico (*GPU*) sem que haja qualquer alteração no código. Desta forma, unidades de processamento gráfico que oferecem um alto grau de paralelismo (Buck, 2007) podem ser utilizadas para execução de um código inicialmente projetado para ser executado em uma *CPU*. A biblioteca também oferece suporte para o desenvolvimento e uso de redes neurais convolucionais (descritas no Capítulo 3.3). Além disso, também são suportadas re-

des profundas com diversas arquiteturas, possuindo múltiplas saídas, múltiplas entradas, compartilhamento de camadas e etc. No entanto, a construção de tais modelos através da biblioteca Keras é oferecida em alto nível, não servindo para operações de baixo nível, como manipulação de tensores (Duchenne et al., 2011). Sendo assim, Keras oferece recursos de forma modular, ainda que faça uso de bibliotecas de baixo nível como TensorFlow (Abadi et al., 2015), CNTK (Seide e Agarwal, 2016) e Theano (Al-Rfou et al., 2016).

No presente projeto de pesquisa, foram utilizadas redes convolucionais pré-treinadas oferecidas pela biblioteca Keras. A diferença entre cada rede utilizada está na arquitetura, ou seja, nos componentes de baixo nível utilizados na construção da rede. As redes pré-treinadas tem como foco aplicações com imagens e podem ser utilizadas para extração de características. Este processo também é descrito no Capítulo 3.3. Assim como justificado no capítulo anterior, pretende-se fazer uso dos recursos aqui descritos para obtenção de uma boa descrição das imagens de ressonância magnética no contexto da doença de Alzheimer. Além disso, no contexto de extração de características, espera-se avaliar as diferenças entre descritores *feitos à mão*, oferecidos pela biblioteca LIRe, e redes pré-treinadas, oferecidas pela biblioteca Keras.

9 Avaliação Experimental

Neste capítulo, serão apresentadas as configurações dos experimentos, métricas de avaliação utilizadas, resultados considerando diferentes extratores de características, comparação do método de Competição e Cooperação entre partículas com demais algoritmos de aprendizado supervisionado e semi-supervisionado e, por fim, discussão sobre os resultados obtidos em cada parte do experimento.

9.1 Protocolo Experimental

A fim de garantir a reprodutibilidade do presente trabalho, serão descritos a seguir os detalhes sobre as bases de dados utilizadas, implementação dos algoritmos, escolha de parâmetros, métricas de avaliação e outros fatores relevantes.

Como descrito na seção 8.1.1, as imagens da base de dados OASIS são divididas em duas categorias, segmentadas e não-segmentadas. Em busca de avaliar os impactos que a segmentação das áreas de massa cinzenta, massa branca e líquido cefalorraquidiano podem causar, foram realizados experimentos considerando ambas categorias.

Descritores	Características
CEDD	145
FCTH	193
EHD	81
SCD	65

Modelos Pré-Treinados	Características
VGG16	6227
NASNetLarge	80493
ResNet50	47685
Xception	61527

Tabela 7: Lista de descritores e modelos pré-treinados usados durante os experimentos.

Para a extração de características das imagens foram utilizadas duas bibliotecas, sendo a biblioteca LIRe, com a linguagem Java e a biblioteca Keras, com a linguagem Python. Na biblioteca LIRe, apoiando-se no trabalho de Padovese (Padovese, 2017), foram escolhidos alguns descritores de propósito geral para extração de características, enquanto na biblioteca Keras, a extração de características foi realizada por modelos pré-treinados de aprendizagem profunda. A Tabela 7 lista os descritores e modelos pré-treinados que foram utilizados durante a etapa experimental. Baseado nos trabalhos de Tufail et al. (2012) e Yang et al. (2013), foi utilizado o algoritmo PCA considerando 100 componentes

para os valores obtidos em modelos pré-treinados.

Com exceção do modelo de partículas (Breve et al., 2012), implementado pelo autor[1], os algoritmos utilizados nos experimentos aqui presentes são oriundos da biblioteca scikit-learn (Pedregosa et al., 2011) em sua versão 0.20.1. A Tabela 8 traz informações sobre os métodos, com a categoria e o módulo da biblioteca em que estão implementados. Mais detalhes serão encontrados nas seções seguintes. A fim de validar os resultados dos experimentos, foi utilizado o método *holdout*. O método *holdout* separa a base de dados em dois conjuntos mutuamente exclusivos, sendo o conjunto de treino e de teste ou, conjunto rotulado e não-rotulado. Em algoritmos de aprendizado supervisionado, o conjunto rotulado é utilizado para treino do modelo, enquanto o não-rotulado, para teste (Kohavi et al., 1995). Com o propósito de comparar o resultado de cada algoritmo, o método *holdout* é executado n vezes e o cálculo do desempenho dos algoritmos é feito a partir da média do resultado de todas execuções.

Algoritmo	Categoria	Módulo	Acrônimo
Support Vector Machine	Supervisionado	sklearn.svm	SVM
k-Nearest Neighbors	Supervisionado	sklearn.neighbors	kNN
Naive Bayes	Supervisionado	sklearn.naive_bayes	NB
Label Spreading	Semi-Supervisionado	sklearn.semi_supervised	LS
Label Propagation	Semi-Supervisionado	sklearn.semi_supervised	LP

Tabela 8: Lista de algoritmos da biblioteca scikit-learn utilizados durante os experimentos.

Em razão da existência de parâmetros variáveis nos algoritmos presentes na experimentação, foi necessário definir alguns valores iniciais, assim como realizar testes que demonstram o impacto da variação de tais parâmetros. Com base no trabalho de Breve et al. (2011), para o modelo de partículas foram definidos $\Delta_v = 0,35$ e Euclidiana como métrica de distância. Os valores P_{grd} e k foram definidos arbitrariamente. Sendo $P_{grd} = 0,5$, $k = 64$ para os algoritmos semi-supervisionados e $k = 16$ para o algoritmo kNN. Para o restante dos algoritmos, serão utilizados os parâmetros iniciais definidos pela biblioteca scikit-learn.

Para a avaliação dos resultados, a principal métrica utilizada foi a *F1 Score*, que mesmo em casos onde existe um grande desbalanceamento entre as classes (como acontece na

[1] https://github.com/caiocarneloz/pycc

base OASIS), é possível obter um valor representativo acerca da taxa de acerto do modelo (Fujino et al., 2008; da Costa et al., 2016). A *F1 Score* é a combinação de duas outras métricas: precisão e revocação, que serão descritas na sequência.

Primeiramente, será apresentado o conceito de verdadeiros positivos e negativos, falsos positivos e negativos. No contexto atual, pode-se considerar como verdadeiro positivo (ou VP) a quantidade de amostras pertencentes à classe Alzheimer que foram classificadas corretamente. Desta forma, entende-se que os falsos positivos (FP) são a quantidade de amostras pertencentes à classe saudável que foram classificadas como Alzheimer. Da mesma forma, ao analisar a classe das amostras saudáveis, falsos negativos são amostras de cérebros com Alzheimer que foram ditas pelo modelo como sendo saudáveis, enquanto os verdadeiros negativos são os pacientes saudáveis classificados corretamente (Olson e Delen, 2008). Para melhor entendimento, a Figura 30 ilustra este conceito.

	DIAGNOSTICADO COM ALZHEIMER	DIAGNOSTICADO COMO SAUDÁVEL
Previsão de Positivos	VERDADEIRO POSITIVO (VP)	FALSO POSITIVO (FP)
Previsão de Negativos	FALSO NEGATIVO (FN)	VERDADEIRO NEGATIVO (VN)

Figura 30: Tabela demonstrando verdadeiros positivos e negativos, falsos positivos e negativos.

Adaptado de: Olson e Delen (2008)

Formalizadas abaixo, a métrica de precisão indica quantas das amostras classificadas como positivas realmente eram tal, enquanto a revocação indica quantas amostras positivas foram corretamente classificadas no total (Tan, 2005).

$$\text{Precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}} \quad \text{Revocação} = \frac{\text{VP}}{\text{VP} + \text{FN}} \tag{7}$$

Para obter o *F1-Score*, é calculada a média harmônica a partir dos resultados adqui-

ridos no cálculo de precisão e revocação, podendo ser descrita como:

$$F1\text{-}Score = \frac{2}{\frac{1}{\text{Precisão}} + \frac{1}{\text{Revocação}}} \tag{8}$$

Levando em consideração todos os aspectos apresentados, as próximas seções trazem os resultados obtidos para a base de dados OASIS, assim como a discussão sobre os mesmos.

9.2 Resultados Experimentais

Nos experimentos seguintes, foram utilizadas as imagens retiradas em ângulo transversal dos 416 exames disponibilizadas pelo conjunto OASIS-1, descrito anteriormente no Capítulo 8. Alguns pontos a serem considerados são: todas as amostras da base de dados possuem rótulo, sendo assim, foi possível mensurar o valor *F1-Score* de cada método. Uma vez que o estágio de demência interfere na facilidade de sua detecção, a Figura 31 traz informações sobre a parcela dos indivíduos da base de dados que se encontram em cada um desses estágios.

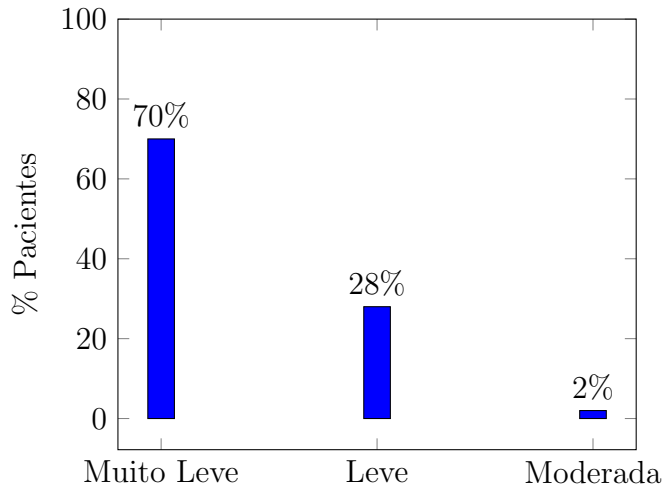


Figura 31: Proporção de pacientes da base OASIS por estágio de demência.

Fonte: Elaborada pelo autor.

Com o objetivo de avaliar o desempenho dos extratores de características utilizados, foi realizado um experimento inicial considerando 5%, 10% e 15% dos dados rotulados.

9.2.1 Avaliação dos Extratores

Para esta análise, os parâmetros dos algoritmos utilizados serão os iniciais descritos no Capítulo 9.1. O número n de execuções do método *holdout* com separação aleatória dos conjuntos foi estabelecido em 50. Serão apresentados média e desvio padrão dos valores *F1-Score* obtidos nas 50 execuções. Na métrica *F1-Score*, os valores variam de 0 a 1, sendo 0 o pior resultado, onde todas as amostras foram classificadas incorretamente e 1 o melhor, onde todas as amostras foram classificadas corretamente. As Tabelas 9, 10 e 11 apresentam os resultados obtidos. A última linha das tabela representa a média dos algoritmos, enquanto a última coluna, a média dos extratores.

Extrator \ Algoritmo	SVM	kNN	NB	LS	LP	PCC	
NASNetLarge	0,43±0,00	0,43±0,00	0,44±0,01	0,43±0,00	0,52±0,09	0,63±0,07	0,48±0,03
FCTH	0,43±0,00	0,43±0,00	0,48±0,18	0,43±0,00	0,52±0,11	0,57±0,07	0,48±0,06
Xception	0,43±0,00	0,43±0,00	0,44±0,01	0,43±0,00	0,65±0,12	0,67±0,08	0,51±0,03
ResNet50	0,43±0,00	0,43±0,00	0,44±0,01	0,43±0,00	0,63±0,13	0,66±0,08	0,50±0,04
SCD	0,58±0,16	0,43±0,00	0,77±0,04	0,63±0,16	0,68±0,12	0,70±0,07	**0,63±0,09**
CEDD	0,43±0,00	0,43±0,00	0,60±0,08	0,46±0,10	0,50±0,11	0,52±0,08	0,49±0,06
VGG16	0,43±0,00	0,43±0,00	0,44±0,02	0,43±0,00	0,57±0,14	0,69±0,09	0,50±0,04
EHD	0,43±0,01	0,43±0,00	0,48±0,04	0,43±0,00	0,43±0,00	0,53±0,06	0,46±0,02
	0,45±0,02	0,43±0,00	0,51±0,05	0,46±0,03	0,56±0,10	**0,62±0,08**	

Tabela 9: Resultado experimental com 5% da base de dados rotulada

Observando a Tabela 9, verifica-se que o melhor resultado obtido parte da combinação do algoritmo *Naive Bayes* com o extrator de características do *Scalable Color Descriptor* (SCD), onde a média *F1-Score* foi $0,77$ com desvio padrão de $0,04$. No entanto, o desempenho do mesmo algoritmo não é refletido para outros extratores, sendo, de modo geral, o modelo de partículas o algoritmo com a maior média *F1-Score*. Por outro lado,

os resultados obtidos com o uso do SCD foram muito superiores à outros extratores, dado que o maior valor *F1-Score* de cada algoritmo ocorreu na combinação com este extrator.

Ao aumentar a quantidade de dados rotulados, os métodos supervisionados obtiveram um ganho médio considerável quando combinados com o extrator SCD (Tabela 10), que novamente trouxe o melhor resultado. Para os algoritmos SVM e kNN, essa melhora foi de 13,79% e 55,81% respectivamente. O melhor resultado global se manteve na combinação de *Naive Bayes* e SCD. Análises mais detalhadas sobre o ganho em função da quantidade de dados rotulados serão abordados mais abaixo.

Extrator \ Algoritmo	SVM	kNN	NB	LS	LP	PCC	
NASNetLarge	0,43±0,00	0,46±0,04	0,53±0,05	0,43±0,00	0,51±0,08	0,65±0,05	0,50±0,04
FCTH	0,43±0,00	0,45±0,04	0,44±0,17	0,43±0,00	0,50±0,10	0,58±0,06	0,47±0,06
Xception	0,43±0,00	0,57±0,09	0,49±0,03	0,49±0,10	0,66±0,10	0,72±0,05	0,56±0,06
ResNet50	0,43±0,00	0,53±0,08	0,50±0,04	0,43±0,00	0,64±0,11	0,72±0,04	0,54±0,04
SCD	0,66±0,14	0,67±0,14	0,78±0,03	0,68±0,11	0,72±0,09	0,71±0,07	**0,70±0,10**
CEDD	0,43±0,00	0,46±0,07	0,49±0,12	0,48±0,09	0,49±0,10	0,51±0,08	0,47±0,08
VGG16	0,43±0,00	0,51±0,07	0,51±0,05	0,43±0,00	0,56±0,12	0,75±0,03	0,53±0,04
EHD	0,44±0,01	0,44±0,01	0,46±0,10	0,43±0,00	0,43±0,00	0,54±0,06	0,46±0,03
	0,46±0,02	0,51±0,07	0,52±0,07	0,48±0,04	0,56±0,09	**0,65±0,06**	

Tabela 10: Resultado experimental com 10% da base de dados rotulada

O extrator SCD não apenas se manteve com a maior média quando comparado a outros extratores, como também obteve um ganho de 11% quando comparado à base com 5% das amostras rotuladas, tendo uma média 25% superior ao extrator Xception, com média $0,56$. No entanto, o melhor resultado de cada algoritmo não mais originou-se da combinação com este extrator, pois o melhor resultado para o modelo de partículas foi o extrator VGG16, 7,14% superior ao melhor resultado obtido anteriormente. Para os

demais algoritmos semi-supervisionados, o ganho médio foi insignificante, no entanto, o melhor resultado dos algoritmos *Label Spreading* e *Label Propagation* foi, respectivamente, 7,93% e 5,88% superior. De modo geral, o aumento da quantidade de dados rotulados se mostrou interessante neste cenário, haja visto os ganhos obtidos.

Para o cenário onde se tem 15% da base de dados rotulada, o ganho médio dos algoritmos foi baixo na maioria dos casos, sendo relevante apenas para o algoritmo kNN. Ainda assim, o melhor resultado global teve um pequeno aumento, chegando a $0,79$ na combinação de *Naive Bayes* e SCD.

Extrator \ Algoritmo	SVM	kNN	NB	LS	LP	PCC	
NASNetLarge	0,43±0,00	0,54±0,07	0,60±0,04	0,43±0,00	0,50±0,07	0,67±0,05	0,53±0,04
FCTH	0,43±0,00	0,47±0,06	0,38±0,15	0,43±0,00	0,50±0,10	0,60±0,06	0,47±0,06
Xception	0,43±0,00	0,62±0,10	0,53±0,04	0,53±0,12	0,65±0,12	0,73±0,02	0,58±0,07
ResNet50	0,43±0,00	0,63±0,08	0,54±0,04	0,44±0,02	0,63±0,09	0,71±0,04	0,56±0,05
SCD	0,69±0,13	0,68±0,10	0,79±0,01	0,69±0,12	0,71±0,08	0,72±0,04	**0,71±0,08**
CEDD	0,43±0,00	0,46±0,07	0,41±0,11	0,47±0,08	0,46±0,07	0,48±0,06	0,45±0,07
VGG16	0,43±0,00	0,59±0,09	0,58±0,04	0,43±0,00	0,56±0,11	0,76±0,03	0,56±0,05
EHD	0,45±0,03	0,44±0,02	0,43±0,12	0,43±0,00	0,43±0,00	0,54±0,06	0,45±0,04
	0,47±0,02	0,55±0,07	0,53±0,07	0,48±0,04	0,56±0,08	**0,65±0,05**	

Tabela 11: Resultado experimental com 15% da base de dados rotulada

Com exceção do algoritmo *Label Propagation*, os algoritmos semi-supervisionados obtiveram seus melhores resultados quando comparados às bases com 5 e 10% dos dados, no entanto, o ganho que concedeu tal feito foi de apenas 1,47% e 1,33% para LS e PCC respectivamente. De modo geral, os extratores dos descritores de propósito geral não obtiveram bons resultados, sendo exceção apenas o extrator SCD que, em contrapartida, se mostrou superior ao restante em todos os cenários. Para uma melhor visualização, serão

apresentados a seguir os ganhos obtidos em função da quantidade de dados rotulados, sendo adicionados os cenários com 20% e 25% de dados rotulados.

A fim de avaliar o ganho dos algoritmos em função do aumento na quantidade de dados rotulados, serão exibidos na sequência os resultados referentes a cada extrator. Ao analisar os extratores separadamente, é possível mensurar o quão bem representadas estão as amostras, eliminando a ideia de que um baixo *F1-Score* possa ser consequência da falta de dados rotulados. Ainda que o valor *F1-Score* possa variar de 0 a 1, foram definidos como limite superior e inferior os valores $0,35$ e $0,9$, aprimorando a legibilidade do gráfico.

Analisando inicialmente os extratores que utilizam informações de borda, a Figura 32 traz no gráfico os eixos Cenário e *F1-Score*, sendo Cenário o eixo que representa a porcentagem de dados rotulados e, *F1-Score*, o eixo que traz o resultado da classificação para um dado algoritmo. O gráfico representa o desempenho do extrator CEDD.

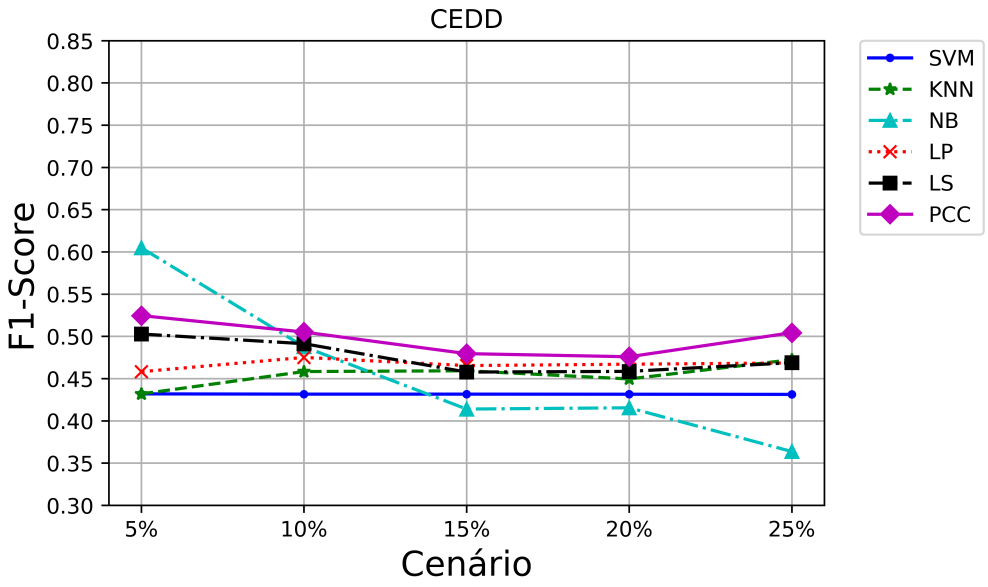


Figura 32: *F1-Score* × porcentagem de dados rotulados pra o extrator CEDD.

Fonte: Elaborada pelo autor.

É possível observar que ao aumentar a quantidade de dados rotulados de 5% para 10%, os resultados dos algoritmos LS e PCC decaem e permanecem estáveis até o cenário onde se tem 20% da base rotulada, havendo um pequeno aumento no cenário com 25%. Também há queda no resultado do algoritmo NB, que demonstra um comportamento inversamente proporcional ao crescimento de rótulo na base. Para os algoritmos LP e

kNN, há uma diferença positiva na mudança de 5% para 10% de rótulos, que são mantidos no restante dos cenários. O algoritmo SVM se manteve inalterado em todos os cenários. Baseado nisso, é possível concluir que, de modo geral, o aumento da quantidade de dados rotulados não trouxe melhorias no resultado dos algoritmos, ocasionando, inclusive, uma baixa no resultado de alguns. Desta forma, entende-se que o extrator não traz uma boa representação das imagens, uma vez que mais amostras rotuladas não agregaram à uma melhor identificação do padrão de cada classe.

Para o extrator EHD (Figura 33), houve um aumento crescente do valor *F1-Score* para os algoritmos PCC, SVM e kNN até o cenário de 20%. Na transição dos cenários de 20% e 25%, SVM e kNN se mantiveram crescentes, enquanto PCC decaiu levemente. O resultado do algoritmo NB se mostrou inversamente proporcional até o cenário de 20%, obtendo um ganho no cenário de 25%.

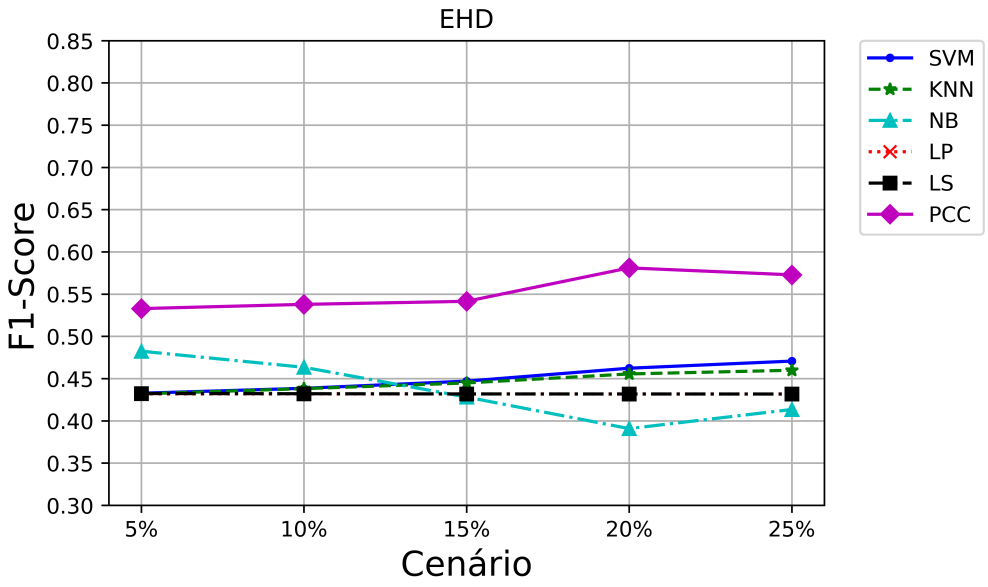


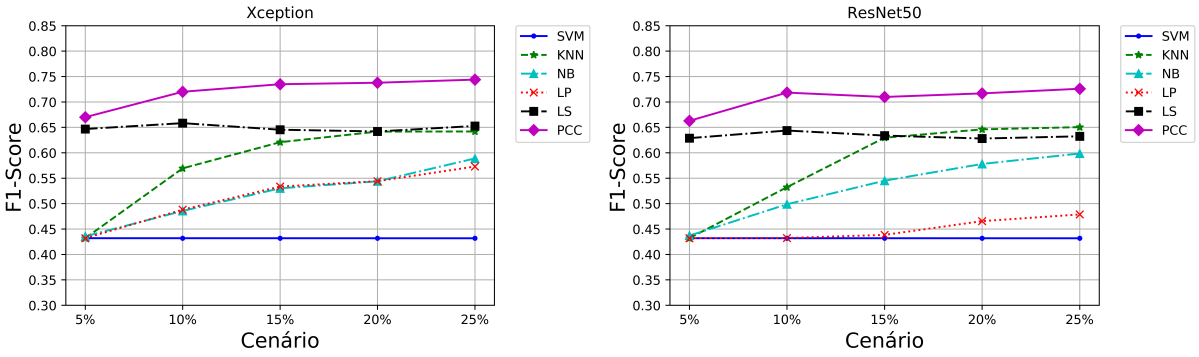
Figura 33: *F1-Score* × porcentagem de dados rotulados pra o extrator EHD.

Fonte: Elaborada pelo autor.

Ainda que, de modo geral, a resposta ao aumento de dados rotulados tenha sido crescente ou estável (salvo exceção do algoritmo NB), os resultados se mantiveram semelhantes aos apresentados pelo extrator CEDD, não trazendo uma boa representação.

Partindo para os extratores de redes profundas, a rede Xception demonstra um ganho considerável na transição do cenário de 5% para o cenário de 10% nos algoritmos kNN,

NB, LP e PCC (Figura 34A.). Para os mesmos algoritmos, o valor *F1-Score* se comportou de maneira diretamente proporcional no aumento de dados rotulados na base.

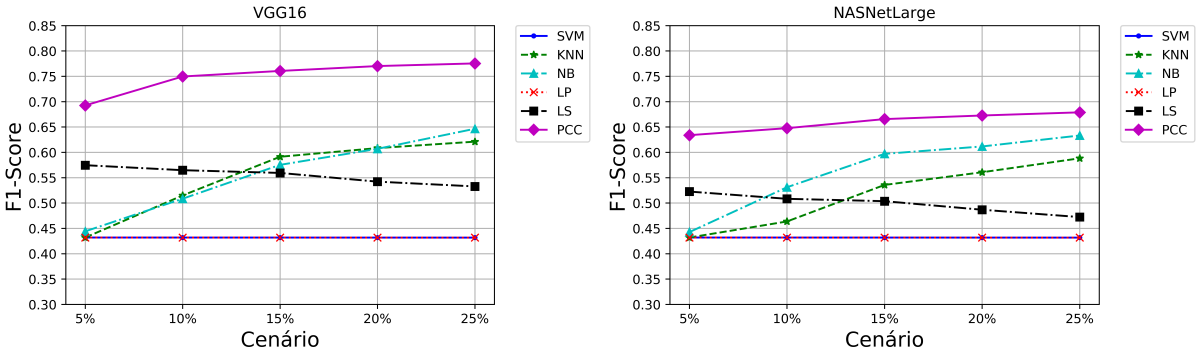


(A.) (B.)

Figura 34: *F1-Score* × porcentagem de dados rotulados para o extrator Xception (A) e ResNet50 (B)

Fonte: Elaborada pelo autor.

LS e SVM não trouxeram resultados superiores com o aumento da base de dados rotulada, no entanto, quando comparado com os extratores CEDD e EHD, o algoritmo LS alcançou melhores resultados, assim como ocorreu com o restante dos algoritmos a partir do cenário de 10%. O extrator ResNet50 (Figura 34B.) obteve resultados muito similares aos da rede Xception, sendo exceção o algoritmo LP, com desempenho inferior.



(A.) (B.)

Figura 35: *F1-Score* × porcentagem de dados rotulados para o extrator VGG16 (A) e NasNetLarge (B)

Fonte: Elaborada pelo autor.

Assim como Xception e ResNet50, os extratores das redes VGG16 e NasNetLarge (Figura 35B. e 35A. respectivamente) se comportaram de forma semelhante. Em ambos, o algoritmo LS demonstrou um comportamento inversamente proporcional ao aumento de

dados rotulados. Por outro lado, kNN, NB e PCC se comportaram de maneira contrária, tendo resultados crescentes à medida que a porcentagem de dados rotulados aumentava. SVM e LP não responderam às mudanças de cenário.

De modo geral, os extratores de redes profundas demonstraram resultados superiores em termos de *F1-Score* quando comparados aos extratores CEDD e EHD. Além disso, observado o ganho médio no aumento da quantidade dados rotulados e baseando-se na hipótese citada ao início desta seção, avalia-se que Xception, ResNet50, VGG16 e NasNetLarge trazem uma boa descrição dos dados.

Ainda serão abordados os extratores FCTH e SCD, que embora tratem características semelhantes, trouxeram resultados distintos. Analisando a Figura 36, é possível perceber que apenas o algoritmo kNN se mantém crescente ao decorrer de todos os cenários. Em contrapartida, o algoritmo NB decaiu até o cenário de 15%, onde tem um pequeno ganho na transição para o cenário de 20% e volta a cair novamente na sequência. Os algoritmos SVM e LP mais uma vez não responderam à troca de cenários.

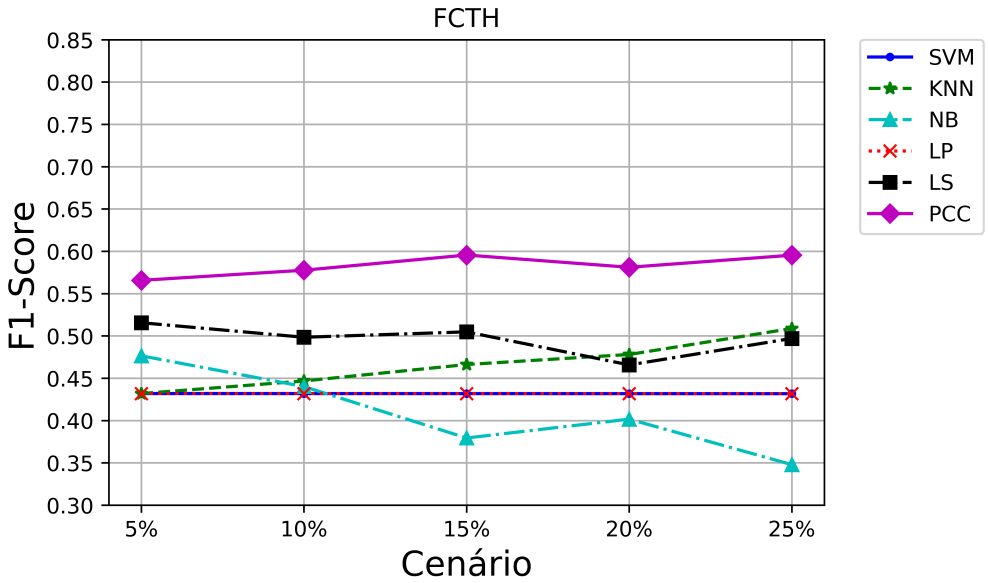


Figura 36: *F1-Score* × porcentagem de dados rotulados pra o extrator FCTH.

Fonte: Elaborada pelo autor.

Para o modelo de partículas, houve um pequeno ganho até o cenário de 15%, que não se repetiu para o restante dos cenários. O algoritmo LS se manteve estável, possuindo uma queda apenas no cenário com 20% dos dados rotulados.

Por fim, ao analisar a Figura 37 é possível observar que o extrator SCD foi o responsável pela maior média *F1-Score* entre os algoritmos (assim como já demonstrado nas tabelas 9, 10 e 11). De modo geral, o aumento da quantidade de dados rotulados não ocasionou no ganho contínuo do valor *F1-Score*, sendo exceção a transição de 5% para 10%. À vista disso, os resultados abaixo da média obtidos pelos algoritmos SVM, kNN e LP no cenário de 5% podem ser justificados justamente pela baixa quantidade de dados rotulados, e não pela capacidade de extração do método.

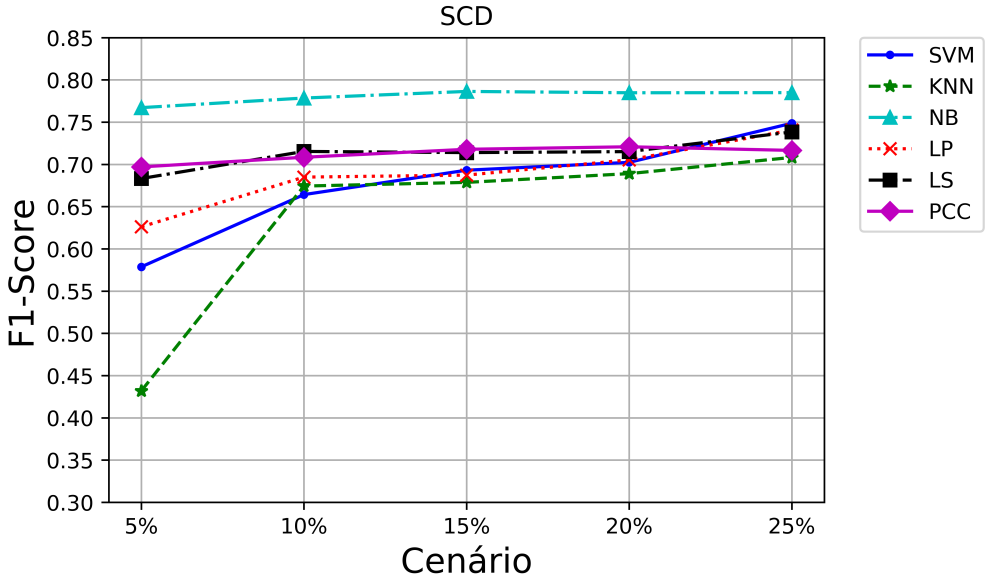


Figura 37: *F1-Score* × porcentagem de dados rotulados pra o extrator SCD.

Fonte: Elaborada pelo autor.

Desta forma, pode-se concluir que para o conjunto de imagens segmentadas da base OASIS-1, considerando parâmetros descritos em 9.1 e quantidades de dados rotulados estipuladas, o extrator SCD é o que melhor representa as diferenças entre o cérebro de pacientes saudáveis e com Alzheimer (dentre os extratores aqui representados).

A partir dos experimentos realizados, foi possível observar alguns fatores que serão descritos na sequência. De modo geral, extratores vindos de descritores de propósito geral não desempenharam uma boa extração de características, sendo exceção o extrator SCD. Para os extratores de redes profundas, ainda que tenham sido obtidos bons resultados, Xception e Resnet, VGG16 e NASNetLarge não demonstraram muita diversidade entre si, tendo comportamentos muito semelhantes. A falta de diversidade nos resultados

pode indicar que embora existam diferenças nas arquiteturas das redes, é possível que as informações extraídas sejam equivalentes. Ainda assim, Xception e VGG16 trouxeram melhores resultados entre os extratores de redes profundas. Ao analisar os resultados em função da quantidade de dados rotulados, identifica-se que na maior parte dos casos há um ganho significativo de *F1-Score* na transição de 5% para 10%. Também foram observados ganhos na transição de 10% para 15%, no entanto, uma vez que a proposta é fazer pouco uso deste tipo de dado, torna-se justificável definir o cenário de 10% como uma boa escolha.

Tendo em vista as conclusões acima, foram selecionados para os próximos experimentos os extratores Xception, VGG16 e SCD. Quanto a presença de dados rotulados, foi estipulada a quantia de 10%.

9.2.2 Impacto de Parâmetros no Modelo de Partículas

De modo a avaliar o impacto causado pela mudança de parâmetros no modelo de competição e cooperação entre partículas, serão apresentados os valores *F1-Score* em função da variação de k, P_{grd} e Δ_v. Além disso, pretende-se avaliar como a variação de parâmetros se comporta em combinação com os diferentes extratores de características. Uma vez que cada parâmetro será analisado singularmente, ficam mantidos os valores descritos em 9.1 para o restante dos parâmetros. Desta forma, na análise do valor k, fica definido $P_{grd} = 0,5$ e $\Delta_v = 0,35$. Para análise de P_{grd}, Δ_v se mantém $0,35$ e k possui valor 64. Por fim, na análise do parâmetro Δ_v, P_{grd} e k possuem valor $0,5$ e 64 respectivamente. Para todos os experimentos, a quantidade de iterações foi fixada em 1000. Ao fim, também será abordado o ganho em função da quantidade de iterações.

Inicialmente serão apresentados os impactos do parâmetro k. Em razão dos algoritmos semi-supervisionados abordados na etapa experimental serem baseados em grafos, foi possível compará-los ao modelo de partículas na variação deste parâmetro. Como descrito anteriormente, para esta análise serão abordados os extratores SCD, Xception e VGG16 (Figuras 38, 39,40). A fim de melhorar a visualização dos gráficos, o intervalo do eixo y representando o valor *F1-Score* foi definido de $0,6$ a $0,8$. Os valores de k iniciam em 16

e são acrescidos em 16 a cada experimento, até que se chegue em 112.

Iniciando pelo extrator SCD, o aumento do valor k traz um ganho em *F1-Score* para os algoritmos LP e LS até o valor 48. O algoritmo LS se mantém crescente até $k = 80$.

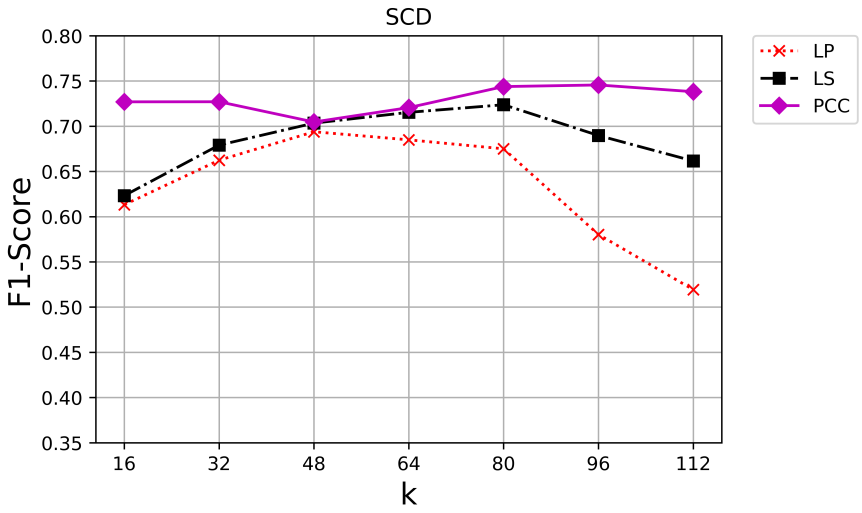


Figura 38: *F1-Score* × parâmetro k para o extrator SCD.

Fonte: Elaborada pelo autor.

Para o modelo de partículas, é possível observar que de modo geral há pouco impacto no valor *F1-Score* uma vez que os valores variam entre cerca de $0,7$ e $0,75$. O mesmo não ocorre para LS e LP, que possuem cerca do dobro da variação presente no resultado do PCC analisando os valores mínimos e máximos atingidos.

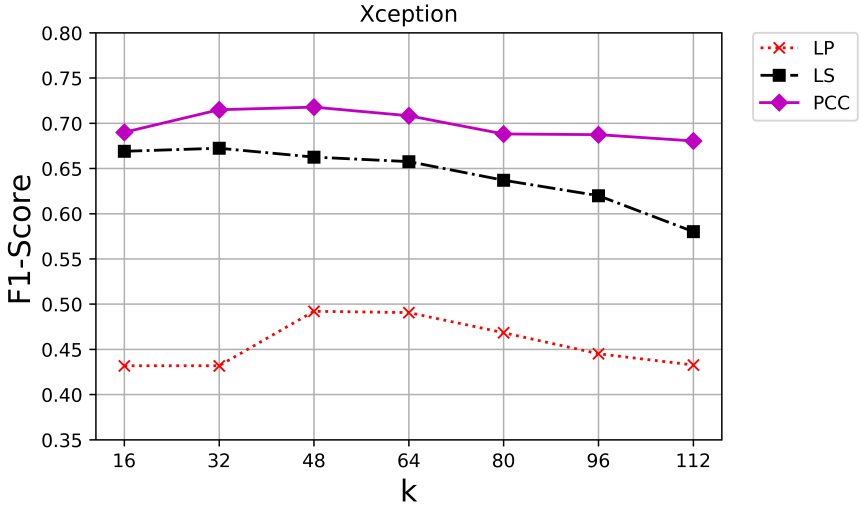


Figura 39: *F1-Score* × parâmetro k para o extrator Xception.

Fonte: Elaborada pelo autor.

Para o extrator Xception, os três algoritmos se comportaram de forma diferente. A variação do algoritmo de competição e cooperação entre partículas foi ainda menor quando comparada ao que foi obtido no extrator SCD. No entanto, os melhores resultados foram obtidos com $k < 64$, enquanto para o extrator SCD, ocorreu o inverso ($k > 64$). LS demonstrou um comportamento inversamente proporcional ao aumento do valor k a partir de $k = 32$. Para o algoritmo LP, embora de modo geral os resultados tenham sido inferiores a PCC e LS, os cenários com $k = 48$ e $k = 64$ trouxeram os melhores resultados.

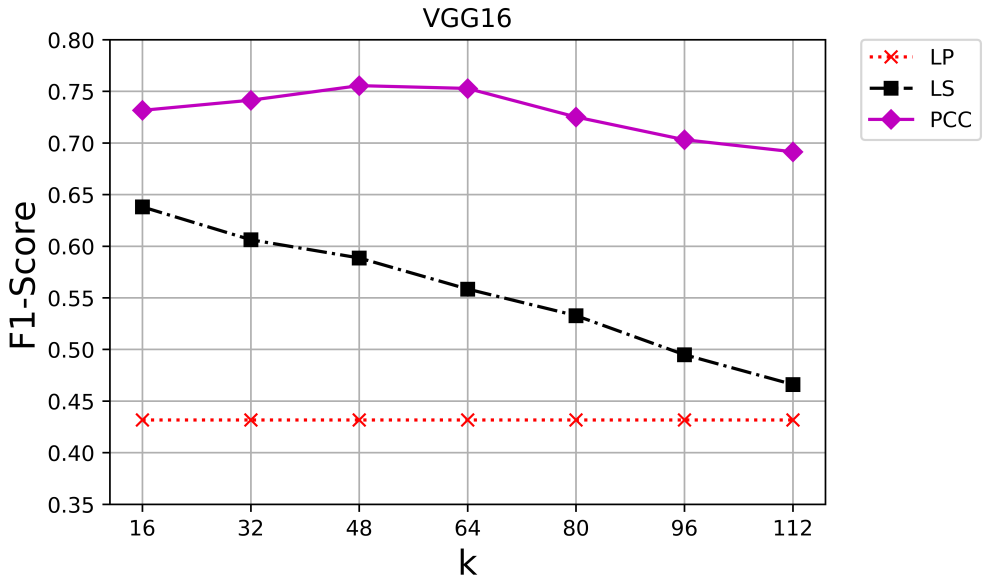


Figura 40: *F1-Score* × parâmetro k para o extrator VGG16.

Fonte: Elaborada pelo autor.

Por fim, considerando o extrator VGG16, foi possível perceber que os algoritmos PCC e LS se comportaram de forma semelhante à anterior. No etanto, a mudança no valor k não fez com que o algoritmo LP obtivesse uma melhora no resultado, se mantendo constante em todos os cenários. LS novamente obteve um comportamento inversamente proporcional em relação à k, variando ainda mais quando comparado ao extrator Xception. Assim como nos dois extratores anteriores, o modelo de partículas demonstrou pouca sensibilidade ao parâmetro k quando comparado aos outros algoritmos.

Desta forma, ao analisar os resultados do modelo de partículas, pode-se concluir que para os extratores SCD, Xception e VGG16, não houve nenhum cenário onde um valor k específico foi superior aos demais. Uma vez que o parâmetro k diz respeito ao relacio-

namento de amostras no grafo, é possível inferir algumas outras conclusões, descritas na sequência. O bom desempenho do algoritmo SCD com um valor alto de k pode, mais uma vez, representar o quão efetiva foi a descrição das imagens de pacientes saudáveis e pacientes com Alzheimer. Uma recuperação efetiva neste cenário faz com que haja maior distância entre vetores de características de classes distintas e menor distância entre amostras da mesma classe. No entanto, tal comportamento não foi refletido nos outros algoritmos. De modo geral, observa-se que a escolha de valores intermediários entre 16 e 112 se faz interessante. A seguir, serão apresentados os impactos de P_{grd} e Δ_v.

Uma vez que tais parâmetros estão presentes apenas no algoritmo de competição e cooperação entre partículas, serão inclusos no mesmo gráfico o desempenho dos três extratores. A fim de aprimorar a visualização do gráfico, fica definido o intervalo de $0,6$ a $0,8$ no eixo que representa o valor *F1-Score*. Os experimentos foram realizados variando os valores de P_{grd} e Δ_v em $0,1$, sendo cenário inicial $0,1$ e final $0,9$.

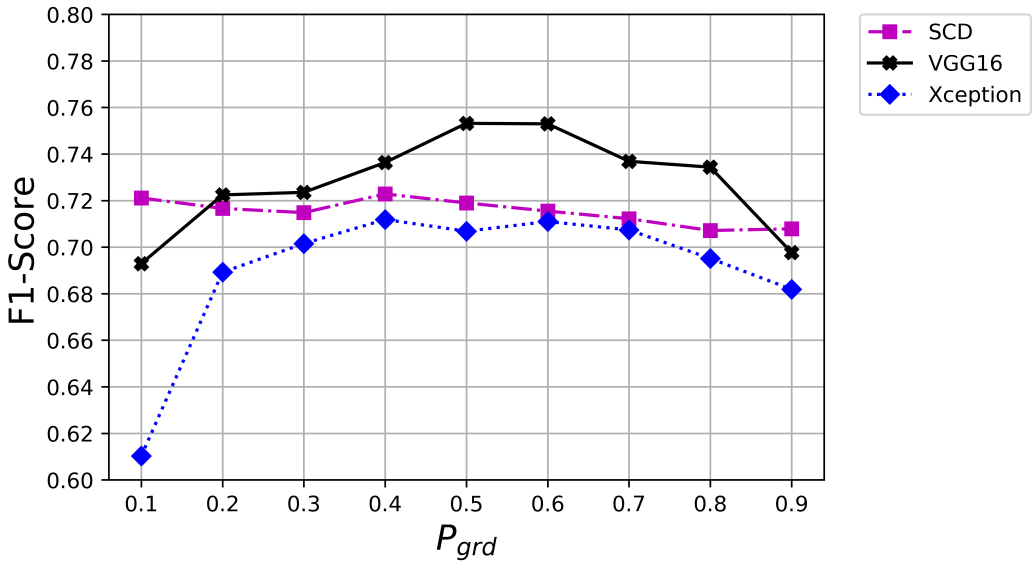


Figura 41: *F1-Score* $\times$ P_{grd} para o algoritmo de competição e cooperação entre partículas.

Fonte: Elaborada pelo autor.

Analisando inicialmente o impacto do parâmetro P_{grd} (Figura 41), ainda que o melhor resultado tenha sido obtido com $P_{grd} = 0,5$ e $P_{grd} = 0,6$ percebe-se que há pouca variação entre o intervalo $0,2$ e $0,8$. Assim como observado na análise do parâmetro k, o valor intermediário (escolhido no início dos experimentos) trouxe bons resultados para os três

extratores. O extrator VGG16 se manteve superior na maioria dos casos, sendo apenas inferior ao extrator SCD nos extremos do intervalo analisado ($0,1$ e $0,9$). Ainda que a diferença entre os cenários do valor P_{grd} tenha sido pequena, é possível avaliar que de modo geral, valores entre $0,4$ e $0,6$ ocasionam bons resultados. Valores muito baixos para o parâmetro P_{grd} fazem com que a busca por território no grafo seja, em sua maior parte, aleatória. Em contrapartida, valores muito altos fazem com que partículas sempre priorizem o movimento guloso, pouco explorando regiões visitadas por partículas inimigas, o que limita a busca por território.

A resposta em *F1-Score* para a variação do parâmetro Δ_v se mostrou ainda menor. O comportamento do modelo de partículas utilizando dados dos três extratores se manteve estável. Valores elevados para o parâmetro Δ_v resultam em trocas mais súbitas nos níveis de domínios dos nós. Valores mais baixos fazem com que os níveis de domínio mudem lentamente. A Figura 42 demonstra os resultados obtidos.

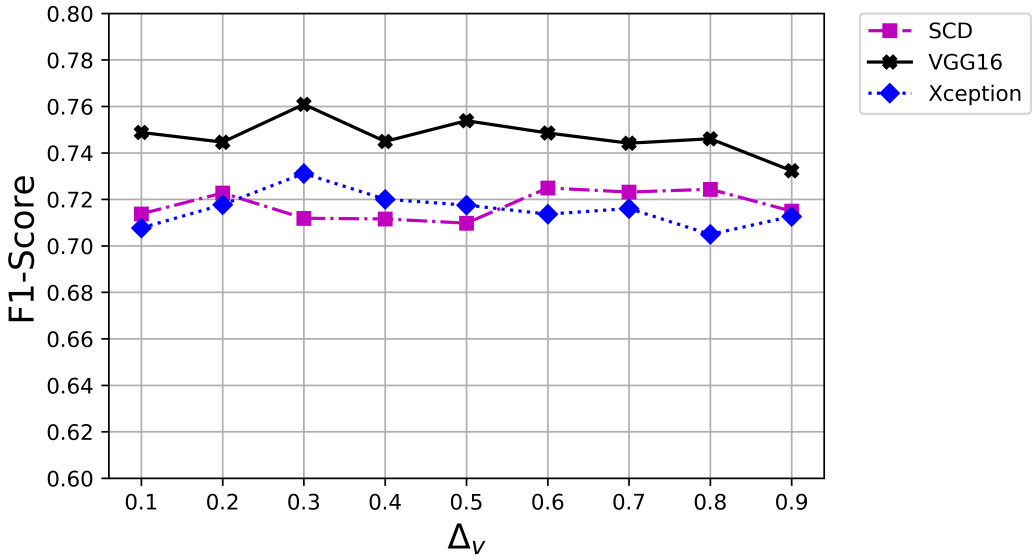


Figura 42: *F1-Score* $\times$ Δ_v para o algoritmo de competição e cooperação entre partículas.

Fonte: Elaborada pelo autor.

Embora os resultados tenham se mantido estáveis, foi possível perceber um ganho para o extrator VGG16 quando $\Delta_v = 0,3$, assim como ocorre com o extrator Xception. No entanto, como observado na análise dos parâmetros anteriores, não houve um valor de Δ_v que trouxesse o melhor resultado para todos os extratores. De qualquer forma, é

possível afirmar que o impacto do parâmetro Δ_v foi baixo quando analisados os valores *F1-Score* para os extratores SCD, VGG16 e Xception.

Por fim, a quantidade de iterações (ou épocas) representa quantos movimentos cada partícula irá fazer até que as classes do problema sejam definidas. Nos experimentos reportados na Figura 43, foram feitas execuções com 500, 1000, 1500 e 2000 iterações.

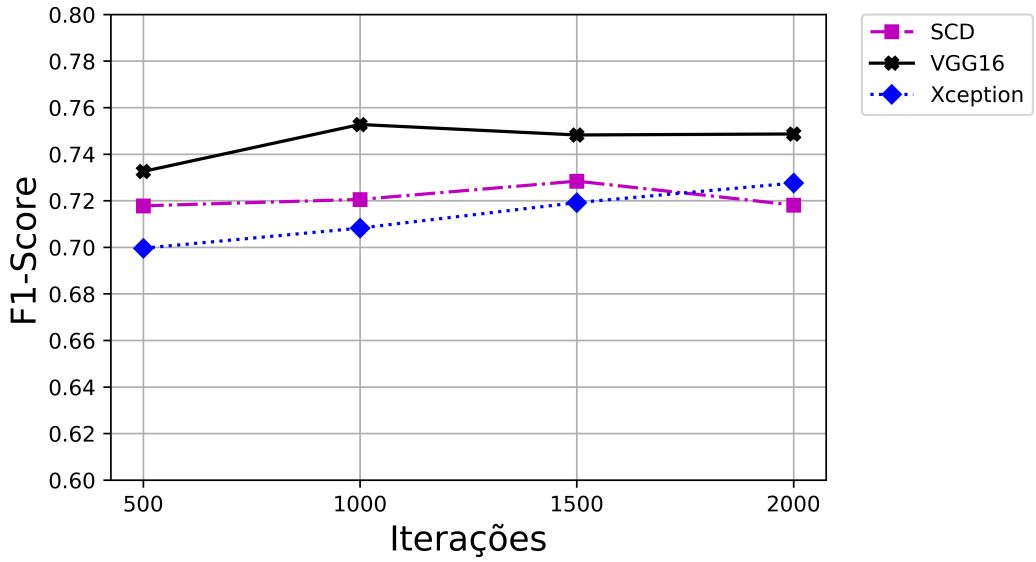


Figura 43: *F1-Score* × iterações para o algoritmo de competição e cooperação entre partículas.

Fonte: Elaborada pelo autor.

Observando o gráfico, percebe-se que na transição de 500 para 1000 iterações, há ganho em *F1-Score* para os três extratores. Este ganho se mantém no aumento para 1500 iterações nos extratores SCD e Xception. O extrator VGG16 teve um desempenho inferior para 1500 e 2000 quando comparado ao cenário com 1000 iterações.

Uma vez que foram explorados os impactos de cada parâmetro presente no algoritmo de competição e cooperação entre partículas, abaixo serão demonstrados experimentos relacionados ao tratamento dos dados, abordando o uso do algoritmo PCA e a segmentação das áreas do cérebro em imagens de ressonância magnética.

9.2.3 Impacto do Algoritmo *PCA*

Partindo do pressuposto de que a quantidade de características geradas por extratores de redes pré-treinadas é grande (Tabela 7), a representação de distância entre amostras se torna mais complexa. Uma vez que o modelo de partículas lida intrinsecamente com um grafo baseado em distância, obter melhores maneiras de representá-la se faz necessário. Como mencionado no Capítulo 6.3, a função do método PCA é reduzir a dimensionalidade dos dados fazendo com que variabilidade deste dado seja preservada. Desta forma, é possível reduzir um espaço de características contendo n componentes, para um outro espaço contendo $n/2$ componentes, por exemplo. Entende-se como componente um elemento do vetor de características. Espera-se que tal redução de componentes não afete a representação dos dados. Ou seja, as informações relevantes não deveriam ser perdidas durante o processo. Com o propósito de avaliar o impacto deste procedimento, foram realizados alguns experimentos. Além de comparar as diferenças para os resultados com e sem o uso deste algoritmo, também foram realizados experimentos com diferentes números de componentes. Para isso, a quantidade de componentes variou de 50 a 400, sendo acrescidos 50 componentes a cada experimento. Ao fim, é realizada a comparação entre o melhor resultado obtido com o uso do PCA e o resultado sem o uso de tal.

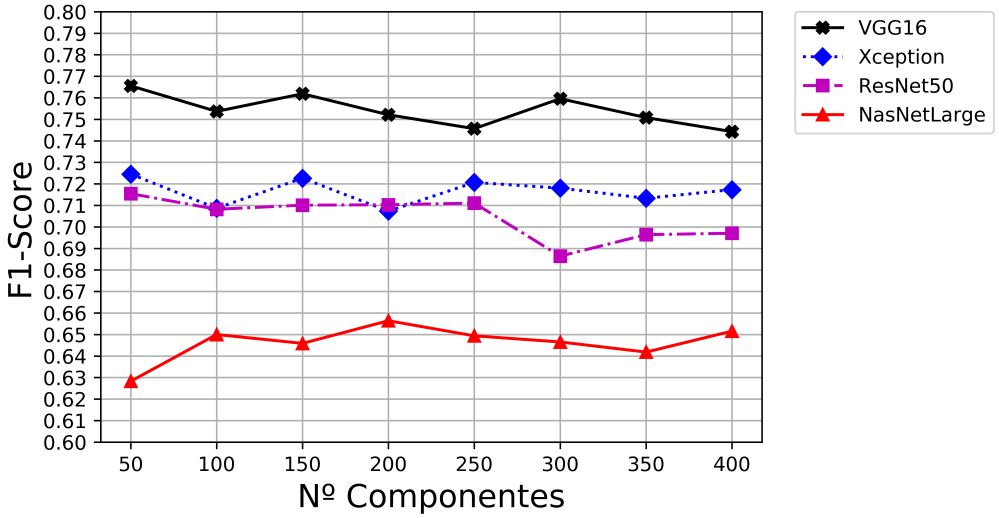


Figura 44: *F1-Score* × componentes do algoritmo PCA para os extratores de redes profundas utilizando o modelo de partículas.

Fonte: Elaborada pelo autor.

Analisando a Figura 44, percebe-se que bons resultados foram obtidos de modo geral. Dentre os experimentos realizados, os melhores valores *F1-Score* para VGG16, Xception e ResNet50 ocorreram no cenário com o menor número de componentes. Para o extrator NasNetLarge, o menor número de componentes representou o pior resultado. O cenário com 200 componentes foi o que trouxe o maior *F1-Score* para este extrator. No entanto, não houve nenhum comportamento direta ou inversamente proporcional à quantidade de componentes. Ao comparar o valor *F1-Score* de um experimento sem o uso de PCA com o melhor resultado utilizando PCA, pode-se perceber pequenas diferenças (Tabela 12). Para o extrator VGG16 e ResNet50, foi percebido um aumento do desvio padrão com o uso do algoritmo PCA. Os extratores NasNetLarge e Xception demonstraram o oposto, tendo maior desvio padrão sem o uso do PCA. Além disso, Xception obteve cerca de 5% de ganho no valor *F1-Score*.

Extrator	*F1-Score* (PCA)	*F1-Score*
VGG16	0,76±0,05	0,76±0,02
Xception	0,72±0,10	0,68±0,12
ResNet50	0,71±0,10	0,72±0,00
NasNetLarge	0,65±0,04	0,65±0,07

Tabela 12: Comparação de resultados do modelo de partículas utilizando o método PCA

Ainda que os resultados tenham sido semelhantes, pode-se justificar o uso do método PCA como uma melhor escolha pelo fato de ser possível obter uma representação mais compacta dos dados. Uma vez que o cálculo da distância Euclidiana envolve todas as características de cada amostra, reduzi-las pode influenciar no tempo de execução durante a montagem do grafo no modelo de partículas. Portanto, pode-se concluir que o uso do algoritmo PCA não influenciou a taxa de acerto do algoritmo de competição e cooperação entre partículas para os extratores VGG16, Xception, ResNet50 e NasNetLarge.

Por fim, mantendo o uso do algoritmo PCA, a próxima seção tratará o impacto da segmentação de regiõs cerebrais na base de dados OASIS.

9.2.4 Impacto da Segmentação de Imagens

Como mencionado no Capítulo 8.1.1, a base de dados OASIS fornece amostras segmentadas e não-segmentadas dos exames de ressonância magnética. O foco de tal segmentação é destacar as regiões de massa cinzenta, massa branca e líquido cefalorraquidiano. As imagens foram segmentadas com o uso do método *K-Means*. Uma vez que a abordagem adotada neste trabalho tem como foco a análise das áreas citadas anteriormente, foi utilizado o conjunto segmentado de imagens em toda etapa experimental. No entanto, a fim de verificar o impacto de tal segmentação, serão demonstrados os resultados atingidos com o uso de todos algoritmos e extratores utilizados no Capítulo 9.2.1 para imagens não segmentadas. Para este conjunto de experimentos, foram replicados os parâmetros usados em 9.2.1. A quantidade de amostras rotuladas foi de 10%. A tabela 13 apresenta os resultados.

Extrator \ Algoritmo	SVM	kNN	NB	LS	LP	PCC	
NasNetLarge	0,22±0,22	0,22±0,22	0,23±0,22	0,43±0,00	0,43±0,00	0,46±0,02	0,33±0,11
FCTH	0,22±0,22	0,22±0,22	0,30±0,12	0,44±0,02	0,44±0,02	0,47±0,03	0,34±0,10
Xception	0,22±0,22	0,22±0,22	0,24±0,23	0,43±0,00	0,43±0,01	0,48±0,02	0,34±0,12
ResNet50	0,22±0,22	0,22±0,22	0,23±0,22	0,43±0,00	0,43±0,01	0,46±0,02	0,33±0,11
SCD	0,22±0,22	0,22±0,22	0,36±0,10	0,43±0,01	0,45±0,02	0,48±0,03	**0,36±0,10**
CEDD	0,22±0,22	0,22±0,22	0,32±0,12	0,44±0,02	0,44±0,02	0,48±0,02	0,35±0,10
VGG16	0,22±0,22	0,22±0,22	0,24±0,23	0,43±0,00	0,43±0,00	0,47±0,02	0,33±0,11
EHD	0,22±0,22	0,22±0,22	0,30±0,10	0,43±0,00	0,43±0,00	0,46±0,02	0,34±0,09
	0,22±0,22	0,22±0,22	0,28±0,17	0,43±0,01	0,44±0,01	**0,47±0,02**	

Tabela 13: Resultados dos extratores e algoritmos considerando o conjunto de imagens não-segmentado.

Ao analisar os resultados, é possível perceber uma média *F1-Score* muito menor

quando comparado ao mesmo experimento utilizando imagens segmentadas. A Tabela 14 demonstra a média de cada algoritmo considerando todos os extratores.

Algoritmo / Conjunto	SVM	kNN	NB	LS	LP	PCC
Conjunto não-segmentado	0,22±0,22	0,22±0,22	0,28±0,17	0,43±0,01	0,44±0,01	0,47±0,02
Conjunto segmentado	0,46±0,02	0,51±0,07	0,52±0,07	0,48±0,04	0,56±0,09	**0,65±0,06**

Tabela 14: Média *F1-Score* de cada algoritmo para os conjuntos de dados segmentados e não-segmentados.

O mesmo comportamento ocorre ao analisar a média obtida por cada extrator. A Tabela 15 demonstra a comparação dos resultados.

Algoritmo / Conjunto	NasNetLarge	FCTH	Xception	ResNet50	SCD	CEDD	VGG16	EHD
Conjunto não-segmentado	0,33±0,11	0,34±0,10	0,34±0,12	0,33±0,11	0,36±0,10	0,35±0,10	0,33±0,11	0,34±0,09
Conjunto segmentado	0,50±0,04	0,47±0,06	0,56±0,06	0,54±0,04	**0,70±0,10**	0,47±0,08	0,53±0,04	0,46±0,03

Tabela 15: Média *F1-Score* de cada extrator para os conjuntos de dados segmentados e não-segmentados.

Considerando os resultados obtidos, pode-se concluir que a segmentação das regiões de massa cinzenta, massa branca e líquido cefalorraquidiano trazem um grande impacto em *F1-Score* para todos os algoritmos e extratores de modo geral. Sendo assim, a segmentação de imagens se torna uma etapa importante anterior à classificação.

10 Considerações Finais

Neste trabalho foram apresentados métodos que auxiliam a detecção da doença de Alzheimer a partir de imagens de ressonância magnética. Dentre os temas abordados, destacam-se as técnicas de extração de características de imagens e aprendizado de máquina.

Inicialmente, foram discutidos os tipos de abordagens no diagnóstico da doença de Alzheimer. Tal discussão teve como propósito avaliar a melhor abordagem no contexto de ressonância magnética. Para que esses fatores pudessem ser abordados, foi realizada a revisão da literatura em busca de explorar os melhores caminhos a serem tomados no decorrer da pesquisa.

Como principal objetivo, foi investigado o uso do algoritmo de competição e cooperação entre partículas para resolução do problema abordado. Este objetivo foi atingido a partir da comparação realizada com outros métodos de aprendizado supervisionados e semi-supervisionados. Consequentemente, também foi possível avaliar o impacto no uso de poucos dados rotulados, sendo esta a principal contribuição deste trabalho.

Além do estudo de diferentes métodos de aprendizado de máquina, foi abordado o conceito de extração de características de imagens. Para isso, duas formas de extração de características foram utilizadas, sendo a extração através de descritores de imagem, e a extração através da transferência de aprendizagem em redes profundas. De modo geral, foi possível qualificar os métodos de extração que obtiveram o melhor desempenho. Destacam-se as redes profundas pela boa aptidão em descrever as imagens utilizadas e o descritor *Scalable Color*, o qual obteve os melhores resultados.

Com o estudo da aplicação do método PCA para redução de dimensionalidade junto ao impacto da segmentação de imagens, informações relevantes sobre a maneira de pré-tratar os dados foram obtidas. Desta forma, avalia-se que, de modo geral, os resultados do presente trabalho indicaram uma abordagem promissora, com diversos espaços para possíveis melhorias. Na sequência serão levantados trabalhos futuros a serem desenvolvidos.

Devido ao impacto da segmentação das áreas cerebrais, avalia-se o uso de outros algoritmos para este propósito. Além disso, pretende-se englobar novas bases de dados para os próximos experimentos.

De modo geral, junto a imagens médicas, é possível obter exames clínicos com informações relevantes sobre pacientes. Como consequência, torna-se viável adicionar essas informações à etapa de classificação de modo a obter melhores resultados.

Por fim, em busca de aprimorar o desempenho do modelo de partículas, podem ser adicionados alguns passos anteriores à classificação. Para montagem do grafo de relacionamento, pode-se utilizar técnicas de ranqueamento de imagens em busca de uma melhor representação de similaridade. Ademais, ainda que de modo geral o impacto de parâmetros não tenha sido relevante, também é possível aplicar algoritmos de otimização em busca do melhor conjunto de parâmetros para cada extrator ou base de dados.

REFERÊNCIAS

ABADI, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>.

ABDI, H.; WILLIAMS, L. J. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, Wiley Online Library, v. 2, n. 4, p. 433–459, 2010.

AL-RFOU, R. et al. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, maio 2016. Disponível em: <http://arxiv.org/abs/1605.02688>.

BLUM, A. Machine learning theory. *Carnegie Melon Universit, School of Computer Science*, p. 26, 2007.

BLUM, A.; MITCHELL, T. Combining labeled and unlabeled data with co-training. In: ACM. *Proceedings of the eleventh annual conference on Computational learning theory*. [S.l.], 1998. p. 92–100.

BREVE, F. et al. Particle competition and cooperation in networks for semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 24, n. 9, p. 1686–1698, 2011.

BREVE, F. et al. Particle competition and cooperation in networks for semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 24, n. 9, p. 1686–1698, 2012.

BREVE, F. A. *Aprendizado de máquina em redes complexas*. Tese (Doutorado) — Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. Citado na, 2010.

BREVE, F. A.; ZHAO, L.; QUILES, M. G. Particle competition and cooperation for semi-supervised learning with label noise. *Neurocomputing*, Elsevier, v. 160, p. 63–72, 2015.

BUCK, I. Gpu computing: Programming a massively parallel processor. In: IEEE. *International Symposium on Code Generation and Optimization (CGO'07)*. [S.l.], 2007. p. 17–17.

CAICEDO, J. C.; GONZALEZ, F. A.; ROMERO, E. A semantic content-based retrieval method for histopathology images. In: SPRINGER. *Asia Information Retrieval Symposium*. [S.l.], 2008. p. 51–60.

CASTELLI, V.; COVER, T. M. On the exponential value of labeled samples. *Pattern Recognition Letters*, Elsevier, v. 16, n. 1, p. 105–111, 1995.

CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. *Semi-supervised learning, ser. Adaptive computation and machine learning*. [S.l.]: Cambridge, MA: The MIT Press, 2006.

CHAPELLE, O.; SCHOLKOPF, B.; ZIEN, A. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, IEEE, v. 20, n. 3, p. 542–542, 2009.

CHATZICHRISTOFIS, S. A.; BOUTALIS, Y. S. Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In: SPRINGER. *International Conference on Computer Vision Systems*. [S.l.], 2008. p. 312–322.

CHATZICHRISTOFIS, S. A.; BOUTALIS, Y. S. Fcth: Fuzzy color and texture histogram-a low level feature for accurate image retrieval. In: IEEE. *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*. [S.l.], 2008. p. 191–196.

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

CHRISTOFOLETTI, G. et al. Risco de quedas em idosos com doença de parkinson e demência de alzheimer: um estudo transversal. *Revista brasileira de fisioterapia*, SciELO Brasil, v. 10, n. 4, 2006.

COLAS, F. P. R. et al. *Data mining scenarios for the discovery of subtypes and the comparison of algorithms*. [S.l.]: Leiden Institute of Advanced Computer Science (LIACS), Faculty of Science . . . , 2009.

COSTA, G. B. P. da et al. An empirical study on the effects of different types of noise in image classification tasks. *arXiv preprint arXiv:1609.02781*, 2016.

CUINGNET, R. et al. Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the adni database. *neuroimage*, Elsevier, v. 56, n. 2, p. 766–781, 2011.

DAVATZIKOS, C. et al. Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiology of aging*, Elsevier, v. 29, n. 4, p. 514–523, 2008.

DENG, L.; YU, D. et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, Now Publishers, Inc., v. 7, n. 3–4, p. 197–387, 2014.

D'ERRICO, G. E.; DIVIETO, C. Diagnostic uncertainty in alzheimer's disease: A bio-metrology perspective addressed to microrna-based biomarkers. In: IEEE. *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. [S.l.], 2018. p. 1–5.

DESIKAN, R. S. et al. Automated mri measures identify individuals with mild cognitive impairment and alzheimer's disease. *Brain*, Oxford University Press, v. 132, n. 8, p. 2048–2057, 2009.

DESPOTOVIĆ, I.; GOOSSENS, B.; PHILIPS, W. Mri segmentation of the human brain: challenges, methods, and applications. *Computational and mathematical methods in medicine*, Hindawi, v. 2015, 2015.

DHARANI, T.; AROQUIARAJ, I. L. A survey on content based image retrieval. In: IEEE. *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on*. [S.l.], 2013. p. 485–490.

DIETTERICH, T. G. Machine learning for sequential data: A review. In: SPRINGER. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. [S.l.], 2002. p. 15–30.

DUCHENNE, O. et al. A tensor-based algorithm for high-order graph matching. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 33, n. 12, p. 2383–2395, 2011.

FUJINO, A.; ISOZAKI, H.; SUZUKI, J. Multi-label text categorization with model combination based on f1-score maximization. In: *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*. [S.l.: s.n.], 2008.

GIEDD, J. N. Structural magnetic resonance imaging of the adolescent brain. *Annals of the New York Academy of Sciences*, Wiley Online Library, v. 1021, n. 1, p. 77–85, 2004.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. [S.l.: s.n.], 2010. p. 249–256.

GOLDMAN, S.; ZHOU, Y. Enhancing supervised learning with unlabeled data. In: *ICML*. [S.l.: s.n.], 2000. p. 327–334.

GRIRA, N.; CRUCIANU, M.; BOUJEMAA, N. Unsupervised and semi-supervised clustering: a brief survey. *A review of machine learning techniques for processing multimedia content*, Report of the MUSCLE European Network of Excellence, v. 1, p. 9–16, 2004.

GUDIVADA, V. N.; RAGHAVAN, V. V. Content based image retrieval systems. *Computer*, IEEE, v. 28, n. 9, p. 18–22, 1995.

GUO, Y.; ZHAO, G.; PIETIKÄINEN, M. Texture classification using a linear configuration model based descriptor. In: CITESEER. *BMVC*. [S.l.], 2011. p. 1–10.

GUSTAFSON, D. E.; KESSEL, W. C. Fuzzy clustering with a fuzzy covariance matrix. In: IEEE. *Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on*. [S.l.], 1979. p. 761–766.

GUZIOR, N. et al. Recent development of multifunctional agents as potential drug candidates for the treatment of alzheimer's disease. *Current medicinal chemistry*, Bentham Science Publishers, v. 22, n. 3, p. 373–404, 2015.

HAFIANE, A. et al. Rotationally invariant hashing of median binary patterns for texture classification. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2008. p. 619–629.

HAYKIN, S. S. et al. *Neural networks and learning machines*. [S.l.]: Pearson Upper Saddle River, NJ, USA:, 2009. v. 3.

HEBERT, L. E. et al. Alzheimer disease in the us population: prevalence estimates using the 2000 census. *Archives of neurology*, American Medical Association, v. 60, n. 8, p. 1119–1122, 2003.

HINTON, G. E. et al. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

HONG, Z.-Q. Algebraic feature extraction of image for recognition. *Pattern recognition*, Elsevier, v. 24, n. 3, p. 211–219, 1991.

HOYER, P. O.; HYVÄRINEN, A. Independent component analysis applied to feature extraction from colour and stereo images. *Network: computation in neural systems*, Taylor & Francis, v. 11, n. 3, p. 191–210, 2000.

IYAPPAN, A. et al. Neuroimaging feature terminology: A controlled terminology for the annotation of brain imaging features. *Journal of Alzheimer's Disease*, IOS Press, v. 59, n. 4, p. 1153–1169, 2017.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. *ACM computing surveys (CSUR)*, Acm, v. 31, n. 3, p. 264–323, 1999.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, The Royal Society Publishing, v. 374, n. 2065, p. 20150202, 2016.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, American Association for the Advancement of Science, v. 349, n. 6245, p. 255–260, 2015.

KARAS, G. et al. A comprehensive study of gray matter loss in patients with alzheimer?s disease using optimized voxel-based morphometry. *Neuroimage*, Elsevier, v. 18, n. 4, p. 895–907, 2003.

KLÖPPEL, S. et al. Accuracy of dementia diagnosis?a direct comparison between radiologists and a computerized method. *Brain*, Oxford University Press, v. 131, n. 11, p. 2969–2974, 2008.

KLÖPPEL, S. et al. Automatic classification of mr scans in alzheimer's disease. *Brain*, Oxford University Press, v. 131, n. 3, p. 681–689, 2008.

KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. *Ijcai*. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.

KOKETSU, K.; FUJIWARA, H.; IKEGAMI, Y. Finite-element simulation of seismic ground motion with a voxel mesh. *Pure and Applied Geophysics*, Springer, v. 161, n. 11-12, p. 2183–2198, 2004.

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.

KUNZE, L. F. et al. Classification analysis of ndvi time series in metric spaces for sugarcane identification. In: *ICEIS (1)*. [S.l.: s.n.], 2018. p. 162–169.

KWONG, K. K. et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 89, n. 12, p. 5675–5679, 1992.

LATECKI, L. J.; LAKAMPER, R.; ECKHARDT, T. Shape descriptors for non-rigid shapes with a single closed contour. In: IEEE. *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. [S.l.], 2000. v. 1, p. 424–429.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015.

LONG, M. et al. Transfer feature learning with joint distribution adaptation. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2013. p. 2200–2207.

LUX, M.; CHATZICHRISTOFIS, S. A. Lire: lucene image retrieval: an extensible java cbir library. In: ACM. *Proceedings of the 16th ACM international conference on Multimedia*. [S.l.], 2008. p. 1085–1088.

MARCUS, D. S. et al. Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *J. Cognitive Neuroscience*, MIT Press, Cambridge, MA, USA, v. 19, n. 9, p. 1498–1507, set. 2007. ISSN 0898-929X. Disponível em: <http://dx.doi.org/10.1162/jocn.2007.19.9.1498>.

MAZZOLA, A. A. Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional. *Revista Brasileira de Física Médica*, v. 3, n. 1, p. 117–129, 2009.

MELLO, R. F. de; PONTI, M. A. *Machine Learning: A Practical Approach on the Statistical Learning Theory*. [S.l.]: Springer, 2018.

MURTY, M. N.; JAIN, A.; FLYNN, P. Data clustering: a review acm compt. surv. *ACM Computing Surveys*, Citeseer, v. 31, n. 3, 1999.

NAGAHARA, A. H. et al. Neuroprotective effects of brain-derived neurotrophic factor in rodent and primate models of alzheimer's disease. *Nature medicine*, Nature Publishing Group, v. 15, n. 3, p. 331–337, 2009.

NETO, J. G.; TAMELINI, M. G.; FORLENZA, O. V. Diagnóstico diferencial das demências. *Rev Psiq Clín*, SciELO Brasil, v. 32, n. 3, p. 119–30, 2005.

NIGAM, K. et al. Text classification from labeled and unlabeled documents using em. *Machine learning*, Springer, v. 39, n. 2, p. 103–134, 2000.

OLIVEIRA, J. E. D. et al. Mammosys: A content-based image retrieval system using breast density patterns. *Computer methods and programs in biomedicine*, Elsevier, v. 99, n. 3, p. 289–297, 2010.

OLSON, D. L.; DELEN, D. *Advanced data mining techniques*. [S.l.]: Springer Science & Business Media, 2008.

OQUAB, M. et al. Learning and transferring mid-level image representations using convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 1717–1724.

PADOVESE, B. T. Suporte ao diagnóstico da doença de alzheimer a partir de imagens de ressonância magnética. Universidade Estadual Paulista (UNESP), 2017.

PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 22, n. 10, p. 1345–1359, 2009.

PARK, C. et al. Integrative gene network construction to analyze cancer recurrence using semi-supervised learning. *PloS one*, Public Library of Science, v. 9, n. 1, p. e86309, 2014.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

PEDRONETTE, D. C. G.; TORRES, R. da S. Shape retrieval using contour features and distance optimization. In: *VISAPP (2)*. [S.l.: s.n.], 2010. p. 197–202.

PEIKARI, M. et al. A cluster-then-label semi-supervised learning approach for pathology image classification. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 7193, 2018.

PEREIRA, C. R. et al. A new computer vision-based approach to aid the diagnosis of parkinson's disease. *Computer methods and programs in biomedicine*, Elsevier, v. 136, p. 79–88, 2016.

PETERSEN, R. C. et al. Mild cognitive impairment: a concept in evolution. *Journal of internal medicine*, Wiley Online Library, v. 275, n. 3, p. 214–228, 2014.

PONTI, M.; NAZARÉ, T. S.; THUMÉ, G. S. Image quantization as a dimensionality reduction procedure in color and texture feature extraction. *Neurocomputing*, Elsevier, v. 173, p. 385–396, 2016.

PRAJAPATI, N.; NANDANWAR, A. K.; PRAJAPATI, G. Edge histogram descriptor, geometric moment and sobel edge detector combined features based object recognition and retrieval system. *Int. J. Comput. Sci. Inf. Technol.*, v. 7, n. 1, p. 407–412, 2016.

QUATTONI, A.; COLLINS, M.; DARRELL, T. Transfer learning for image classification with sparse prototype representations. In: IEEE. *2008 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.], 2008. p. 1–8.

QUILES, M. G. et al. Particle competition for complex network community detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, AIP, v. 18, n. 3, p. 033107, 2008.

RIEDMILLER, M. Advanced supervised learning in multi-layer perceptrons?from backpropagation to adaptive learning algorithms. *Computer Standards & Interfaces*, Elsevier, v. 16, n. 3, p. 265–278, 1994.

ROLI, F.; MARCIALIS, G. L. Semi-supervised pca-based face recognition using self-training. In: SPRINGER. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. [S.l.], 2006. p. 560–568.

RUI, Y.; HUANG, T. S.; CHANG, S.-F. Image retrieval: Current techniques, promising directions, and open issues. *Journal of visual communication and image representation*, Elsevier, v. 10, n. 1, p. 39–62, 1999.

SANCHES, M. K. *Aprendizado de máquina semi-supervisionado: proposta de um algoritmo para rotular exemplos a partir de poucos exemplos rotulados*. Tese (Doutorado) — Universidade de São Paulo, 2003.

SEIDE, F.; AGARWAL, A. Cntk: Microsoft's open-source deep-learning toolkit. In: ACM. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.], 2016. p. 2135–2135.

SERMANET, P.; CHINTALA, S.; LECUN, Y. Convolutional neural networks applied to house numbers digit classification. *arXiv preprint arXiv:1204.3968*, 2012.

SILVA, B. R. d.; BREVE, F. A. Segmentação de imagens utilizando competição e cooperação entre partículas. *Interciência & Sociedade*, p. 75–85, 2015.

STANISZ, G. J. et al. T1, t2 relaxation and magnetization transfer in tissue at 3t. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, Wiley Online Library, v. 54, n. 3, p. 507–512, 2005.

TAN, S. Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, Elsevier, v. 28, n. 4, p. 667–671, 2005.

TEIXEIRA, J. de F. *Inteligência artificial*. [S.l.]: Pia Sociedade de São Paulo-Editora Paulus, 2014.

TORRES, R. d. S.; FALCÃO, A. X. Recuperação de imagens baseada em conteúdo. In: *Workshop de Visão Computacional*. [S.l.: s.n.], 2008. v. 4.

TUFAIL, A. B. et al. Automatic classification of initial categories of alzheimer's disease from structural mri phase images: a comparison of psvm, knn and ann methods. In: WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY (WASET). *Proceedings of World Academy of Science, Engineering and Technology*. [S.l.], 2012. p. 1731.

VEMURI, P. et al. Alzheimer's disease diagnosis in individual subjects using structural mr images: validation studies. *Neuroimage*, Elsevier, v. 39, n. 3, p. 1186–1197, 2008.

WELLER, J.; BUDSON, A. Current understanding of alzheimer's disease diagnosis and treatment. *F1000Research*, Faculty of 1000 Ltd, v. 7, 2018.

WON, C. S.; PARK, D. K.; PARK, S.-J. Efficient use of mpeg-7 edge histogram descriptor. *ETRI journal*, Wiley Online Library, v. 24, n. 1, p. 23–30, 2002.

YANG, S.-T. et al. Discrimination between alzheimer's disease and mild cognitive impairment using som and pso-svm. *Computational and mathematical methods in medicine*, Hindawi Publishing Corporation, v. 2013, 2013.

ZHANG, D. et al. Multimodal classification of alzheimer's disease and mild cognitive impairment. *Neuroimage*, Elsevier, v. 55, n. 3, p. 856–867, 2011.

ZHOU, S.-M.; GAN, J. Q.; SEPULVEDA, F. Classifying mental tasks based on features of higher-order statistics from eeg signals in brain–computer interface. *Information Sciences*, Elsevier, v. 178, n. 6, p. 1629–1640, 2008.

ZHU, X. Semi-supervised learning literature survey. Citeseer, 2005.

ZHU, X. Semi-supervised learning. In: *Encyclopedia of machine learning*. [S.l.]: Springer, 2011. p. 892–897.