UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Diagnóstico de doenças mentais baseado em mineração de dados e redes complexas

Caroline Lourenço Alves

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Caroline Lourenço Alves

Diagnóstico de doenças mentais baseado em mineração de dados e redes complexas

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestra em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Francisco Aparecido Rodrigues

USP – São Carlos
Fevereiro de 2019

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

A474d Alves, Caroline Lourenço
 Diagnóstico de doenças mentais baseado em mineração
de dados e redes complexas / Caroline Lourenço
Alves; orientador Francisco Aparecido Rodrigues. --
São Carlos, 2019.
 156 p.

 Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2019.

 1. Mineração de dados. 2. Redes Complexas. 3.
Doenças neurológicas. I. Rodrigues, Francisco
Aparecido, orient. II. Título.

Caroline Lourenço Alves

**Diagnosis of mental disorders based on data mining and
complex networks**

Master dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP,
in partial fulfillment of the requirements for the
degree of the Master Program in Computer Science
and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Prof. Dr. Francisco Aparecido Rodrigues

USP – São Carlos
February 2019

*Este trabalho é dedicado à minha família,
e a todos os cientistas, que contra todas as adversidades impostas pelo país, arduamente lutam
pela Ciência brasileira.
Em especial, aos pesquisadores do Instituto de Ciências Matemáticas e de Computação (ICMC).*

AGRADECIMENTOS

Ao meu orientador, professor Francisco Aparecido Rodrigues, pela oportunidade e privilégio de estar fazendo Mestrado sob sua orientação e também por todo apoio que tem me dado.

Aos professores: Luciano da Fontoura da Costa e Paulino Ribeiro Villas Boas por todo apoio que me deram no início do Mestrado.

À Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo fornecimento da bolsa de estudo.

Ao Instituto de Ciências Matemáticas e de de Computação (ICMC) por me acolher e fornecer estrutura para realizar minha pesquisa.

Ao meu colega Guilherme Ferraz Arruda por toda contribuição no meu trabalho.

Aos meus pais, Itamar e Eunice, meus heróis, por todo apoio e inspiração que me deram até hoje e por sempre acreditarem em mim.

Ao meu amigo, Lucas, pela apoio, ajuda, sugestões e contribuição nesta pesquisa.

À minha irmã, Carine, por todo apoio e inspiração que sempre me deu, mesmo estando em outro país.

*“O herói é o homem da submissão autoconquistada.”
(Joseph Campbell)*

RESUMO

ALVES, C. L. **Diagnóstico de doenças mentais baseado em mineração de dados e redes complexas**. 2019. 156 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

O uso de técnicas de mineração de dados tem produzido resultados importantes em diversas áreas, tais como bioinformática, atividades de transações bancárias, auditorias de computadores relacionados à segurança, tráfego de redes, análise de textos, imagens e avaliação da qualidade em processos de fabricação. Em medicina, métodos de mineração de dados têm se revelado muito eficazes na realização de diagnósticos automáticos, ajudando na tomada de decisões por equipes médicas. Além do uso de mineração de dados, dados médicos podem ser representados por redes complexas, de modo a incluir conexões entre seus elementos. Por exemplo, no caso do cérebro, regiões corticais podem representar vértices em um grafo e as conexões podem ser definidas através das atividades corticais. Com isso, pode-se comparar a estrutura do cérebro de sujeitos sadios com a de pacientes que apresentam doenças mentais de modo a definir métodos para diagnóstico e obter conhecimento sobre como a estrutura do cérebro está relacionada com alterações comportamentais e neurológicas. Nesse trabalho, estamos interessados em usar métodos de mineração de dados e redes complexas para classificar pacientes portadores de quatro diferentes tipos de doenças mentais, isto é, esquizofrenia, autismo, déficit de atenção/desordem de hiperatividade e paralisia progressiva nuclear.

Palavras-chave: Inteligência Artificial, Mineração de dados, Redes complexas, Doenças neurológicas.

ABSTRACT

ALVES, C. L. **Diagnosis of mental disorders based on data mining and complex networks.** 2019. 156 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

A data mining and knowledge discovery is in a field of research, with applications in different areas such as bioinformatics, customer transaction activity, security related computer audits, network traffic, text analysis and quality evaluation in manufacturing. In medicine, data mining methods have proven very effective in performing automatic diagnostics, helping in making decisions by medical teams. In addition to the use of data mining, medical data can be represented by complex networks in order to include connections between its elements. For example, in the case of the brain, cortical regions can represent vertices in a graph and the connections can be defined through cortical activities. Thus, we can compare the brain structure of healthy patients with those of patients with mental disorder in order to define methods for diagnosis and to obtain knowledge about how the structure of the brain is related to behavioral and neurological changes. Here, we are interested in using data mining methods and complex networks to classify patients with four different types of mental disorders, that is, schizophrenia, autism, attention deficit / hyperactivity disorder, and progressive supranuclear paralysis.

Keywords: Artificial Intelligence, Data Mining, Complex Networks, Neurological Diseases.

LISTA DE ILUSTRAÇÕES

- Figura 1 – Esquematização da metodologia utilizada no presente trabalho. Foram utilizadas quatro bases, representadas em (A) por quatro retângulos: roxo, verde, azul e laranja, correspondentes, respectivamente, às doenças: esquizofrenia, autismo, deficit de atenção/desordem de hiperatividade (ADHD) e paralisia progressiva nuclear (PSP). Cada uma das bases contém matrizes de conectividades, representadas em B com retângulos, sendo que azul são aquelas referentes à pacientes sem a doença, e vermelho, àquelas referentes a pacientes com a doença. Cada uma dessas matrizes corresponde a uma rede; e, para cada rede extraiu-se uma série de medidas de redes que serviram como instâncias para a classificação, como pode ser visto em (C). 39
- Figura 2 – Figura modificada de (MAIMON; ROKACH, 2009) contendo o processo KDD. Em azul observam-se as principais etapas do processo: seleção, pré-processamento, transformação e mineração de dados. É importante ressaltar que previamente deve ser estabelecido o objetivo de aplicação do processo para que ao final dele seja dada a correta interpretação. 42
- Figura 3 – Figura extraída e modificada (FACELI *et al.*, 2011) contendo o esquema para melhor visualização da hierarquização dos tipos de algoritmos de aprendizado de Máquina. 43

- Figura 4 – Figura extraída e modificada de (LIU; MOTODA, 2012; GUYON; ELISSE-EFF, 2003), contendo o esquema dos tipos de seleção segundo os métodos de avaliação. Em (A), pode ser visto um esquema de como funciona uma seleção, em que um conjunto de atributos é submetido a um método de seleção (I), gerando um subconjunto que será avaliado segundo um critério (II) que será submetido a um critério de parada (III), se aprovado tem-se o subconjunto ideal, caso contrário outro subconjunto deve ser gerado pelo método de seleção. Em (B), há um esquema de método de seleção utilizando a abordagem tipo filtro, em que é gerado um subconjunto (fase 1) que será utilizado para classificação por algoritmo de máquina (V), onde é gerado um modelo a ser avaliado; assim a geração de um subconjunto é independente da etapa de classificação. Um esquema da abordagem wrapper pode ser vista em (C), em que um algoritmo de aprendizado de máquina (II) é utilizado para avaliar um subconjunto de atributos obtido pelo método de seleção (I); se a avaliação do modelo (III) for boa, é selecionado um subconjunto ideal que será utilizado por outra etapa de classificação (fase 2), se for ruim outro subconjunto é considerado. 46
- Figura 5 – Figura extraída e modificada de ((FACELI *et al.*, 2011)) contendo um modelo preditivo de um conjunto bidimensional com duas classes (1 e 2) e o objeto em verde é a instância a ser classificada. Os círculos representam a superfície de decisão encontrada pelo classificador em questão. O círculo 1 representa a superfície do classificador 1-NN e como o objeto mais próximo do verde possui rótulo vermelho ele será classificado como vermelho. De maneira análoga considerando 3-NN (círculo 2) como a maioria dos objetos próximos ao verde são azuis ele será classificado como azul. Seguindo esse raciocínio para o 5-NN o objeto verde será classificado como vermelho. Percebe-se, então, que o parâmetro k influencia na tomada de decisão do classificador k -NN. 52
- Figura 6 – Figura extraída e modificada de (FACELI *et al.*, 2011), contendo árvore de decisão obtida a partir de um espaço bidimensional (x_1, x_2) e cada nó corresponde a uma região desse espaço. 54
- Figura 7 – Figura contendo os esquemas de um neurônio biológico e de um artificial. Do lado esquerdo, um esquema de um neurônio artificial (extraído e adaptado de (HAYKIN, 1994)), em que os sinais de entrada $x_1, x_2, \dots, x_k$ são ponderados pelos pesos sinápticos $w_{k1}, w_{k2}, \dots, w_{km}$ através de uma combinação linear, passando pelos bias e função de ativação até obter o sinal de saída, y_k . Do lado direito, um neurônio biológico, extraído de (BORGES *et al.*, 2015), que possui os dendritos como entrada do pulso elétrico e os axônios a saída deles. 56

Figura 8 – Esquema de uma rede neural. A primeira camada contém os atributos de entrada conectada às camadas ocultas e à camada de saída.	57
Figura 9 – Figura extraída e modificada de (BERNARDINI, 2002), contendo um esquema da técnica de <i>Stacking</i> . No primeiro nível um conjunto de treinamento S , com N instâncias $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ é submetido a um conjunto de algoritmos $A_1, A_2, \dots, A_L$ induzindo um conjunto de hipóteses $h_1, h_2, \dots, h_L$. No nível-2, as hipóteses são novamente aplicada ao conjunto de treinamento S (formando um novo conjunto de treinamento) e é submetido a um meta classificador para então classificá-los.	59
Figura 10 – Figura contendo um esquema do processo de Validação Cruzada para o caso de $k=4$. Primeiramente o conjunto de dados é dividido em quatro partes (nomeados na figura de A, B, C e D) e com isso processo de validação foi realizado quatro vezes (1-4). Em 1 a parte foi retirada para a fase de teste e o resto do conjunto foi utilizado para induzir o modelo preditivo. Em 2 a parte C foi selecionada para a fase de avaliação e o restante para a classificação. E desse mesmo modo se deu para 3 e 4 em que as fases de teste selecionadas foram, respectivamente, D e B. Cada processo de avaliação obteve uma performance (Performance(1), Performance(2), Performance(3) e Performance(4)).	64
Figura 11 – Figura extraída de (RODRIGUES, 2007) que contém a representação matricial de dois grafos. Em (a) a matriz é simétrica representando uma rede não dirigida, o que não é observado em (b) caracterizando uma rede dirigida. Os valores das matrizes iguais a um representam que há conexão entre os vértices i e j ; enquanto que os valores zero representam a ausência de conexão.	65
Figura 12 – Figura extraída e modificada de (ALBERT; BARABÁSI, 2002) que contém a distribuição de grau de um grafo. Em (A) o grafo com 4 nós; os nós 4 e 3 (em verde) possuem grau igual a dois, o nó 2 (em azul) possui grau igual a 3 e o nó 1 (em vermelho) possui grau igual a 1. Em (B), contém a distribuição de grau, (P_k) , do grafo em (A); em vermelho com um quantidade de 0.25, representa a quantidade de nós (no caso somente o nó 1) que apresenta grau igual a 1; em verde a quantidade de nós (no caso os nós 3 e 4, representando a metade de nós presente na rede) que possuem grau igual a 2; e, por fim, em azul a quantidade de nós (no caso somente o nó 2) com grau igual a 3.	67
Figura 13 – Figura extraída de (RODRIGUES, 2007) , contendo três redes com diferentes coeficientes de aglomeração. Em (a) todos os vértices estão conectados representando o valor máximo da medida, $C=1$. Em (b) o coeficiente tem valor de $C = \frac{3}{10}$ e por fim, em (c), como não há a presença de nenhum triângulo $C=0$	69

Figura 14 – Figura extraída e modificada de (SPORNS; TONONI; KÖTTER, 2005), contendo os três principais níveis de organização do cérebro (microescala, meso escala e macro escala). Em (A), o nível macroscópico em que o cérebro pode ser dividido através de regiões anatômicas como os lobos corticais. Em (B), a meso escala, em que os neurônios se agregam em colunas, camadas e agrupamentos celulares. Em (C), a escala microscópica, onde a imagem extraída através de um microscópio de varredura mostra estruturas celulares como a vesícula sináptica. 75

Figura 15 – Figura adaptada de (SILVA, 2007; GAMEIRO, 2004) contendo um esquema do processo de RMN. Em (A), contém os spins (setas azuis) no estado degenerado. Com incidência do campo H_0 (em roxo), ocorre a quebra da degenerescência, gerando dois níveis energéticos (níveis de Zeeman), α e β , representado na figura em (B); o alinhamento dos spins é denominado magnetização (em verde na figura). Em (C), há a representação dos dois níveis energéticos, sendo a diferença entre eles, ΔE , proporcional ao campo e a frequência de precessão. Ao incidir um pulso de radiofrequência, (D), os spins de nível mais baixo são excitados ao nível de maior energia (na figura spin verde); enquanto os spins de nível mais alto são relaxados ao nível de energia mais baixo (na figura spin laranja). Como há maior quantidade de spins de energia mais baixa ocorre absorção de energia no próton. Após cessar o pulso, os spins voltam a sua organização anterior e assim é liberada a mesma quantidade de energia absorvida, sendo esta detectada por bobinas. 78

Figura 16 – Figura adaptada de (PAMPLONA, 2014; HEUVEL; POL, 2010) contendo um esquema do método da semente. Primeiramente identifica-se uma região específica de interesse, por exemplo as conexões do córtex motor primário esquerdo (A). Em seguida, os pacientes analisados são colocados no scanner em estado de repouso e é pedido para fechar os olhos e não pensar em nada, sem adormecer; com isso obtém-se as séries temporais do estado de repouso, (B). É pedido então que os voluntários movam a mão direita selecionando os voxels mais ativados do córtex motor, esse será o voxel semente; obtém-se, então, comparando com as séries temporais em repouso, a série temporal de repouso do voxel semente em repouso, (C). Assim com intuito de examinar conectividade funcional entre o voxel semente selecionado a uma segunda região cerebral j (por exemplo, uma região no córtex motor contralateral), é medido, por meio de medidas de correlação as séries temporais no estado de repouso dessa duas regiões, (D). Uma alta correlação indica uma alta conectividade entre o voxel semente e o voxel j . Por fim, (E), é feita a correlação do voxel semente com todos os outros voxels obtendo-se o mapa de conectividade cerebral. 80

Figura 17 – Figura extraída e modifica de (BECKMANN, 2012; BIJSTERBOSCH; SMITH; BECKMANN, 2017). Em (A) os dados de fMRI contêm os dados BOLD sendo que cada linha representa um volume tridimensional em um certo tempo e cada coluna representa todos os tempos de um voxel. Em (B), o conjunto de dados é então decomposto em duas matrizes pelo método de ICA, sendo que a matriz (I) contém a série temporal (coluna) de cada componente obtida e a matriz (II) contém o mapa espacial (linha) de cada componente. Para cada série temporal existe uma mapa espacial correspondente, assim o número de linhas do mapa espacial é o mesmo que o número de colunas da série temporal e representa o número de componentes (dimensionalidade do ICA). 81

Figura 18 – Figura extraída e modificada de (BULLMORE; SPORNS, 2009; LYNALL *et al.*, 2010) contendo o esquema da obtenção de um grafo a partir de dados de fMRI. Em (A) foram determinadas as regiões de interesses que serão os nós da rede. Por meio do scanner de fMRI são determinadas as séries temporais da região de interesse. Nessas séries são aplicados métodos de pré-processamento (por exemplo filtragem de séries temporais) com objetivo de eliminar ruídos de dados de fMRI. Com essas séries pré-processadas, é determinada a correlação entre elas que é feita utilizando algum dos métodos descritos nesta secção. Com a correlação das séries temporais obtém-se a matriz de conectividade em (C). A partir da matriz de conectividade é construído o grafo (o grafo apresentado na figura foi obtido da matriz de conectividade de um dos pacientes com esquizofrenia dos dados do capítulo 6 82

Figura 19 – Figura extraída e modificada de (RUBINOV; SPORNS, 2010; PARK; FRISTON, 2013; FIGUEIRA, 2013) contendo os tipos de arestas possíveis. Em (A) há dois tipos de conexão unindo três regiões do cérebro (1, 2, 3, nas cores azul, laranja e vermelho, respectivamente). As conexões em roxo pertencem a conectividade funcional e não são direcionadas, sendo a conexão (I) com o peso maior (linha mais grossa), indicando uma rede com peso. As conexões em azul são direcionadas (indicadas por flechas) pertencem a conectividade efetiva; como a aresta (IV) tem mais peso que as demais (mais grossa), trata-se de uma rede com peso. Em (B), há outros dois exemplos de rede direcionada e não direcionada, percebe-se que a matriz de conectividade desta é simétrica. As duas redes possuem pesos indicado pela escala acinzentada, sendo a preta a de maior conexão (valor igual a 1) e a branca a de menor (valor zero). 83

Figura 20 – Figura extraída e modificada de (FORNITO; ZALESKY; BULLMORE, 2016). Em (A) é mostrado uma matriz de conectividade obtida a partir da aplicação de fRMI em humanos; trata-se de uma matriz com peso pela presença de pixels com diferentes cores. Essa matriz é submetida ao método de <i>thresholding</i> resultando em uma matriz em (B), sendo que só foi mantido os pesos 20% maiores. A matriz em (B) é então submetida a binarização resultando na matriz em (C), uma matriz sem peso, de 0 e 1 (branco e preto, respectivamente).	84
Figura 21 – Figura extraída e modificada de (SPORNS, 2013). Em (A) se encontra destacado em laranja as comunidades densamente conectadas entre si e pouco conectadas com outras comunidades e essa conexão é feita por hubs (em azul). Em (B), se encontra destacado a segregação funcional. Em (C), a integração funcional, onde é destacado em azul os hubs da redes altamente conectados entre si, permitindo o fluxo de informação por toda rede.	85
Figura 22 – Figura contendo o algoritmo adotado pela a função train cujo intuito é determinar por intermédio da "força bruta" os valores de um conjunto de parâmetros do modelo que maximizam seu desempenho.	91
Figura 23 – Figura contendo o esquema da metodologia adotada. Primeiramente, em (A), as medidas de redes consideradas foram extraídas da base de dados (na imagem é mostrada a rede do primeiro paciente sem a esquizofrenia obtida por meio do pacote <i>igraph</i>). Essas medidas são arquivadas em um formato <i>csv</i> e com esse arquivo é realizado a classificação, (B), onde os algoritmos de aprendizado que obtiverem os melhores desempenhos são selecionados. Em seguida, em (C), aplicou-se os quatro métodos de seleção e, então realiza-se a predição com os melhores classificadores de (B), os métodos mais relevantes (aqueles que aumentaram a performance dos classificadores em questão) são selecionados. Por fim, em (D), aplica-se os três tipos de normalização, aos atributos selecionados pelos métodos mais relevante de (C), obtendo o desempenho final dos classificadores de (B).	93
Figura 24 – Matriz de correlação de Pearson obtida, percebe-se que as medidas altamente correlacionadas (superiores a 0.70) são: segundo momento da distribuição de grau e complexidade; segundo momento da distribuição de grau e knn; complexidade e knn; diâmetro e betweeness centrality; assortatividade e transitividade; menor caminho e diâmetro. Essas medidas foram eliminadas.	96
Figura 25 – Figura contendo o gráfico obtido ao se aplicar o processo de seleção de eliminação recursiva de características; o gráfico corresponde à raiz do erro quadrático médio pela quantidade de variáveis do grupo. Percebe-se que o grupo com menor erro é aquele que possui oito variáveis.	98

Figura 26 – Figura contendo as componentes geradas pelo método de PCA para cada medida; aquelas que tiveram menor contribuição foram: menor caminho, grau médio, pageRank e eficiência.	99
---	----

LISTA DE TABELAS

Tabela 1	– Tabela contendo a matriz confusão em problemas de duas classes, positivas e negativas. Em azul, trata-se dos rótulos obtidos durante a previsão; e, em amarelo, o verdadeiro rótulo das classes.	62
Tabela 2	– Tabela adaptada de (LANDIS; KOCH, 1977), fornece uma referência útil para descrever a força relativa associada ao kappa, ainda que as divisões em questão foram arbitrariamente criadas pelo autor.	63
Tabela 3	– Tabela adaptada de (HOSMER; LEMESHOW; STURDIVANT, 2013) em que os autores fornecem uma regra geral de utilização da área abaixo da curva ROC.	63
Tabela 4	– Tabela contendo os resultados obtidos após submeter o arquivo <i>csv</i> , que contém as medidas de redes como atributos e um atributo classe (com instâncias contendo os seguintes rótulos: COS e normal), aos algoritmos preditivos. As medidas utilizadas para verificar o desempenho de cada classificador foram: acurácia, kappa associado a acurácia, especificidade, sensibilidade e AUC. Quanto aos classificadores foram empregados: k-NN, Naive Bayes (NB), árvores de decisão (AD), redes neurais (RN). Foi feita, ainda, a combinação deles dois a dois, três a três e com os quatro, por meio da técnica de <i>stacking</i>	94
Tabela 5	– Tabela contendo os atributos correspondentes às medidas de rede com seu respectivo valor (organizada de forma decrescente) da medida de importância obtida pela função <i>varImp</i> . Foram destacados, na coloração avermelhada, aqueles atributos com menor medida encontrada.	97
Tabela 6	– Tabela contendo o resultado obtido em cada processo de seleção, sendo que cada um deles foi representado utilizando uma cor, para o classificador redes neurais. Percebe-se que os métodos de seleção não aumentaram o desempenho (em termos da acurácia e AUC) desse classificador.	99
Tabela 7	– Tabela contendo o desempenho do classificador árvore de decisão ao se aplicar os diferentes métodos de seleção (destacados com diferentes cores na tabela). Em negrito, encontra-se aqueles que mais aumentaram a performance (em termos da acurácia e AUC) do preditor em questão.	100
Tabela 8	– Tabela contendo o desempenho do classificador k-NN e naive bayes após submetido aos métodos de seleção. Percebe-se que os métodos de seleção (em termos da acurácia e AUC) pioraram (em termos da acurácia e AUC) a performance desse algoritmo preditivo.	100

Tabela 9 – Tabela contendo o desempenho obtido pelo classificador k-NN e redes neurais ao ser submetido aos diferentes métodos de seleção. Aquele que resultou na maior performance (em termos da acurácia e AUC) foi a seleção com PCA e a seleção por ordem de importância sem três medidas de rede (destacados em negrito na tabela).	101
Tabela 10 – Tabela contendo a performance obtida para o classificador oriundo da combinação entre naive bayes e redes neurais. Em negrito, destacou-se o processo de seleção que mais aumentou o desempenho (em termos da acurácia e AUC) desse classificador.	101
Tabela 11 – Tabela contendo a performance obtida com relação ao classificador: k-NN, árvore de decisão e naive bayes, ao empregar-se os diferentes métodos de seleção. O método que mais aumentou o desempenho (em termos da acurácia e AUC) desse classificador foi a remoção de características redundantes, ainda que os métodos RFE e seleção com PCA também obtiveram uma performance próxima ao maior valor.	102
Tabela 12 – Tabela contendo as performances do classificador k-NN, redes neurais e Naive Bayes, ao ser submetido aos diferentes tipos de seleção. Em negrito o método que proporcionou maior desempenho (em termos da acurácia e AUC) do algoritmo preditivo em questão.	102
Tabela 13 – Tabela contendo o resumo dos resultados obtidos ao se aplicar os métodos de seleção aos classificadores em questão. Assim em uma coluna se encontra o classificador e na coluna ao lado o processo de seleção que aumentou o seu desempenho. Aqueles em que nenhum método aumentou sua performance, a coluna de seleção se encontra como sem seleção.	103
Tabela 14 – Tabela contendo os resultados obtidos para o algoritmo de predição redes neurais, após realiza os três processos de normalização. Para esse classificador não foi utilizado nenhum processo de seleção, pois esses diminuíram o desempenho (em termos da acurácia e AUC) preditivo.	104
Tabela 15 – Tabela contendo os resultados obtidos para o classificador árvore de decisão após realizar os processos de normalização. Para este classificador foi realizado o processo de seleção RFE. Em negrito se encontra destacado o melhor desempenho (em termos da acurácia e AUC) obtido.	104
Tabela 16 – Tabela contendo os resultados, após aplicar os três processos de normalização, para o classificador formado entre a combinação do k-NN e redes neurais; utilizou-se, também o método de seleção por ordem de importância sem todas as medidas. Está destacado, em negrito, o processo de normalização que mais aumentou a performance (em termos da acurácia e AUC) do classificador em questão.	104

Tabela 17 – Desempenho obtido pelo classificador resultante da combinação entre o algoritmo naive bayes e k-NN após realizar os três processos de normalização. Pelo fato de nenhum dos processos de seleção ter contribuído por um aumento na performance desse algoritmo preditivo, eles não foram usados. Em negrito se encontra a melhor performance (em termos da acurácia e AUC) obtida.	105
Tabela 18 – Tabela contendo o desempenho obtido pelo classificador resultante da combinação de redes neurais e naive bayes, quando submetido aos três tipos de normalização. Em negrito é destacado a melhor performance (em termos da acurácia e AUC) encontrada.	105
Tabela 19 – Tabela contendo a performance do classificador obtido da combinação do k-NN, redes neurais e naive bayes, após submetido aos três processos de normalização. O classificador foi aplicado aos atributos obtidos se seleção por ordem de importância, excluindo-se a medida eficiência. Em negrito está destacado o melhor desempenho (em termos da acurácia e AUC) obtido.	105
Tabela 20 – Tabela contendo os resultados obtidos para o classificador resultante da combinação k-NN, árvore de decisão e naive bayes após a aplicação dos métodos de normalização. Os atributos utilizados foram aqueles obtidos pela remoção de características redundantes. Em negrito está destacado o melhor desempenho (em termos da acurácia e AUC) obtido.	106
Tabela 21 – Tabela contendo o resumo dos resultados obtidos, com todos os classificadores e os respectivos métodos de seleção e normalização que mais aumentaram seu desempenho. Nela também se encontra o valor desse desempenho.	107
Tabela 24 – Tabela contendo as medidas e sua respectiva importância. Em rosa destacou-se aquelas que obtiveram a menor medida medida de importância.	112
Tabela 22 – Tabela contendo os desempenhos de cada cada classificador ao se variar o threshold.	113
Tabela 23 – Tabela contendo as melhores performances para cada classificador. Em vermelho está destacado aqueles que possuem acurácia e AUC superior a 0.6; e em negrito o threshold que resultou na maioria do aumento de performance	114
Tabela 25 – Tabela contendo os resultados obtidos após o processo de seleção por importância extraíndo-se primeiramente a assortatividade, seguida pela medida pageRank. Em negrito, destacou-se os processos de seleção que mais aumentaram o desempenho de cada um dos classificadores.	114
Tabela 26 – Tabela contendo os resultados obtidos após submeter os classificadores ao processo de remoção de características com redundância superior a 70%. Percebe-se que esse processo de seleção não aumentou o desempenho de nenhum dos classificadores.	114
Tabela 27 – Tabela contendo os resultados obtidos após o processo de seleção RFE.	115

Tabela 28 – Tabela contendo os resultados após realizar os processos de normalização. Em negrito, destacou-se as melhores performance obtidas, e com as cores laranja e rosa, os resultados obtidos, respectivamente, para as normalizações softmax scaling e z-score.	115
Tabela 29 – Tabela contendo o resumo dos desempenhos obtidos pelos classificadores em questão e os processos que contribuíram para isso.	116
Tabela 30 – Tabela contendo o resultado obtido ao se variar o threshold de 0.3, 0.4, 0.5 e 0.6, que foram destacados, respectivamente, pelas cores: rosa, laranja, verde e azul. Em negrito destacou-se o threshold que resultou na maior performance (em termos da acurácia e AUC).	121
Tabela 31 – Tabela contendo o os thresholds que mais contribuíram para o aumento dos classificadores em questão. Em rosa destacou-se os desempenho de acurácia maiores que 0.6 e AUC maiores que 0.5.	122
Tabela 32 – Tabela contendo os processos de seleção por ordem de importância e RFE, destacados em azul e laranja, respectivamente e o desempenho obtido ao se utilizar esses processos. Em negrito foi destacado as melhores performances obtidas.	123
Tabela 33 – Resultados obtidos ao se aplicar os métodos de normalização softmax scaling e z-score, destacados, respectivamente, em laranja e rosa, aos conjuntos de atributos que contribuíram para aumentar o desempenho dos classificadores escolhidos. Em negrito, destacou-se a normalização que mais aumentou a performance de cada classificador.	124
Tabela 34 – Tabela contendo o resumo dos resultados obtidos nesta seção.	124
Tabela 35 – Tabela contendo os resultados ao se aplicar os algoritmos preditivos ao conjunto de atributos correspondentes as medidas de redes extraídas das matrizes de conectividade binárias de pacientes com e sem a doença PSP. Em vermelho encontram-se destacados aqueles cuja acurácia e AUC foram superiores a 0.6.	127
Tabela 36 – Tabela contendo os resultados obtidos após utilizar os métodos de seleção. Em azul, destacou-se aqueles obtidos pelo método de seleção por ordem de importância; e, em laranja, aqueles obtidos pelo método RFE. Em negrito, destacou-se os melhores desempenhos (em termos da acurácia e AUC) para cada classificador em questão.	128
Tabela 37 – Tabela contendo os resultados obtidos após aplicar os dois métodos de normalização em cada subconjunto resultante na seção anterior, e classificá-los com os algoritmos preditivos.	129
Tabela 38 – Tabela contendo o resumo dos resultados obtidos neste capítulo.	129

Tabela 39 – Tabela contendo os classificadores com melhores desempenhos para cada base. Destacou-se em rosa, laranja, amarelo e azul, respectivamente, as doenças: Esquizofrenia, Autismo, ADHD e PSP. Em negrito destacou-se a árvore decisão, algoritmo comum a todas as bases.	135
Tabela 40 – Tabela contendo o classificador de melhor desempenho para cada base e o subconjunto de atributos que contribuiu para isso. Além disso foi colocado o valor de binarização e thresholding utilizado para cada base, bem como o método de normalização utilizado.	136

LISTA DE ABREVIATURAS E SIGLAS

AD	Árvores de decisão
ADHD	Deficit/Hyperactivity Disorder
AFNI	Analysis of Functional NeuroImages
AM	Aprendizado de Máquina
ASD	autism spectrum disorders
AUC	Area Under Curve ROC
BOLD	Blood oxygenation level–dependent
CART	Classification and Regression Tree
COS	Childhood-onset schizophrenia
CSF	Cerebrospinal fluid
Esp.	Especificidade
fMRI	Functional Magnetic Resonance Imaging
FN	Falso Negativo
FP	Falso Positivo
FSL	FMRIB's Software Library
ICA	Independent component analysis
k-NN	k- Nearest Neighbors
KDD	Knowledge Discovery in Databases
LQV	Linear Quantization Vector
LSTUR	Linear Scaling to Unit Range
LTU	Logic Threshold Unit
MAP	Maximun a Posteriori
MD	Mineração de Dados
NB	Naive Bayes
NIH	National Institutes of Health
NN	Nearest Neighbors
PCA	Principal Component Analysis
PSP	Paralisia Supranuclear Progressiva
RFE	Recursive Feature Elimination
RMN	Ressonância magnética
RMSE	Root Mean Square Error

RNAs	Redes Neurais
ROC	Receiver Operating Characteristic
ROI	Region of interest
RP	Reconhecimento de Padrões
SCA	Seed-based Correlation Analysis
Sens.	Sensitividade
SS	Softmax Scaling
SVD	Singular value decomposition
TD	Pacientes sem a doença do grupo controle
UCLA	University of California-Los Angeles
UMCD	USC Multimodal Connectivity Database
varImp	função variable importance score do pacote caret do R
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
ZS	Z-score

LISTA DE SÍMBOLOS

ρ — Coeficiente de Pearson

μ — Média de determinado atributo

σ — Variância de determinado atributo

S — Conjunto de treinamento

$I(S)$ — Classificador gerado por um indutor I para um conjunto de treinamento S

$I(S)(x)$ — Classificação produzida por $I(S)$ aplicado ao padrão x

$\varepsilon(I(S)(x), D)$ — Erro de generalização de um classificador $I(S)$ em uma distribuição de probabilidade D sobre um conjunto de rótulos U

$Sens(f)$ — Sensitividade de um classificador f

$Esp(f)$ — Especificidade de um classificador f

$Acuracia(f)$ — Acurácia de um classificador f

κ — Kappa, coeficiente de concordância

G — Grafo

$A(G)$ — Matriz de adjacência de um grafo G

A_{ij} — Elementos da matriz de adjacência

K_i — Grau ou conectividade

$\langle K \rangle$ — Conectividade média da rede

P_k — Probabilidade de um vértice escolhido aleatoriamente ter grau k

H — Entropia da distribuição de grau

$knn(i)$ — Conectividade média entre os vizinhos para um nó i

D — Matriz de distância

$C(i)$ — O coeficiente de aglomeração local de um vértice i

$3N_{\Delta(i)}$ — Quantidade de triângulos formados envolvendo o vértice i

$3N_3(i)$ — Número de trios conectados que possuem o nó i como vértice central

d_{ij} — Elementos da matriz de distância

L — Caminho característico da rede

E — Eficiência da rede

b_i — *betweenness centrality*

$g_i^{(st)}$ — Número de caminhos geodésicos do vértice s ao vértice t passando por i

n_{st} — Número total de caminhos geodésicos de s a t

CDP — dominância do ponto central

Cl — Closeness centrality

x_i — Centralidade de autovetor de um vértice i

$P_R(i)$ — Medida pageRank de um vértice i

H_0 — Campo magnético

ω_0 — Frequência de precessão de um spin

SUMÁRIO

1	INTRODUÇÃO	37
1.1	Objetivos	38
1.2	Descrição dos capítulos	39
2	CONCEITOS BÁSICOS DE MINERAÇÃO DE DADOS	41
2.1	Mineração de dados e os tipos de algoritmos de aprendizado de Máquina	42
2.2	Métodos de seleção de atributos	43
2.2.1	<i>Métodos de seleção utilizados no presente trabalho</i>	45
2.3	Métodos de normalização	48
2.4	Métodos de classificação	49
2.4.1	<i>k-NN</i>	50
2.4.2	<i>Naive Bayes</i>	51
2.4.3	<i>Árvores de decisão</i>	53
2.4.4	<i>Redes neurais</i>	55
2.4.5	<i>Combinação de classificadores</i>	57
3	AVALIAÇÃO DE MODELOS	61
4	CARACTERIZAÇÃO DE REDES COMPLEXAS	65
4.0.1	<i>Medidas relacionadas à conectividade</i>	66
4.0.2	<i>Medidas relacionadas a ciclos</i>	68
4.0.3	<i>Medidas relacionadas à distância</i>	69
4.0.4	<i>Medidas de centralidade</i>	70
4.0.5	<i>K-core</i>	71
4.0.6	<i>Tipos de redes e outros conceitos</i>	71
5	CONECTIVIDADE CEREBRAL E A TÉCNICA DE RESSONÂN- CIA MAGNÉTICA FUNCIONAL	73
5.1	Introdução	73
5.2	Diferentes níveis de organização do cérebro	74
5.3	Redes funcionais	76
5.4	Técnica de imagem por ressonância magnética funcional	76
5.5	Aplicação de <i>threshold</i> e binarização em matrizes de conectividade	81

5.6	O conceito de integração e segregação e a organização complexa do cérebro	83
6	REDES CEREBRAIS DE PACIENTES COM ESQUIZOFRENIA . .	87
6.1	Introdução	87
6.2	Base de dados	88
6.3	Metodologia	90
6.4	Resultados	93
6.4.1	<i>Resultados após a aplicação dos processos de seleção</i>	96
6.4.2	<i>Resultados após à aplicação das normalizações</i>	103
6.5	Conclusões	106
7	REDES CEREBRAIS DE PACIENTES COM DESORDEM DO ESPECTRO AUTISTA	109
7.1	Introdução	109
7.2	Base de dados	110
7.3	Metodologia	111
7.4	Resultados	112
7.4.1	<i>Resultados após variar o threshold</i>	112
7.4.2	<i>Resultados após seleção</i>	112
7.5	Normalização	115
7.6	Conclusão	116
8	REDES CEREBRAIS DE PACIENTES COM DEFICIT DE ATENÇÃO/DESORDEM DE HIPERATIVIDADE	117
8.1	Introdução	117
8.2	Base de dados	118
8.3	Metodologia	119
8.4	Resultados	119
8.4.1	<i>Resultados dos classificadores variando o threshold da rede</i>	119
8.4.2	<i>Resultado após seleção</i>	120
8.4.3	<i>Resultados após normalização</i>	120
8.5	Conclusão	120
9	REDES CEREBRAIS DE PACIENTES COM PARALISIA SUPRANUCLEAR PROGRESSIVA	125
9.1	Introdução	125
9.2	Base de dados	125
9.3	Metodologia	126
9.4	Resultados	126
9.5	Resultados após seleção	127

9.6	Resultados após normalização	128
9.7	Conclusão	128
10	TRABALHOS FUTUROS	131
11	CONCLUSÃO	133
	REFERÊNCIAS	137

INTRODUÇÃO

Técnicas de mineração de dados vêm sendo amplamente utilizadas na área médica (ILYASOVA *et al.*, 2018; SHIRWALKAR *et al.*, 2018; BARKAN; KISILEV; WALACH, 2018) de forma a oferecer diagnósticos mais precisos e automáticos de diversas doenças (BELLAZZI; ZUPAN, 2008). Dentre as principais aplicações, podemos citar o diagnóstico por imagem (HU *et al.*, 2018; SAMANT; AGARWAL, 2018; MORRIS *et al.*, 2018), como na área oncológica (principalmente câncer de mama) (RHODES *et al.*, 2004; DELEN; WALKER; KADAM, 2005; TIBSHIRANI *et al.*, 2002; POLAKA *et al.*, 2017; ABDAR *et al.*, 2018; CHAURASIA; PAL; TIWARI, 2018; YUE *et al.*, 2018); na área neurofisiológica (RAMSAY; GIOVANNI, 2017; MAROCO *et al.*, 2011; GIOIA *et al.*, 2000; SHREE *et al.*, 2018; BERNSTEIN *et al.*, 2018); e doenças cardíacas (SONI *et al.*, 2011; PALANIAPPAN; AWANG, 2008; KUMAR; KOUSHIK; DEEPAK, 2018; SHIRWALKAR *et al.*, 2018).

Além do uso de mineração de dados, a representação por redes complexas vem sendo usada com sucesso de modo a caracterizar a estrutura de diversos sistemas biológicos (COSTA *et al.*, 2011; GREEN *et al.*, 2018; GOSAK *et al.*, 2018), principalmente o cérebro (BULLMORE; SPORNS, 2009; AVENA-KOENIGSBERGER; MISIC; SPORNS, 2018; MONACO *et al.*, 2018). A área de redes complexas consiste na utilização de grafos para a representação da estrutura de sistemas complexos (ARRUDA, 2013). No final da década de 1990, com intuito de representar e analisar a dinâmica de diferentes sistemas complexos, foi introduzida a teoria das redes complexas (WATTS; STROGATZ, 1998; BARABÁSI; ALBERT, 1999), ganhando importância a partir de 1999, quando a topologia das redes internet (FALOUTSOS; FALOUTSOS; FALOUTSOS, 1999) e *World Wide Web* (BARABÁSI; ALBERT, 1999) foram mapeadas (ARRUDA, 2013).

Em neurociência, o uso de redes complexas tem permitido um melhor entendimento sobre a organização do cérebro (SPORNS, 2011; AVENA-KOENIGSBERGER; MISIC; SPORNS, 2018; SHINE; POLDRACK, 2018). Várias doenças neurológicas têm sido estudadas sob a ótica de redes e diversos estudos foram realizados na tentativa de compreender como alterações

comportamentais estão relacionadas com a organização cerebral (FRISTON; FRITH, 1995; HALLQUIST; HILLARY, 2018; XU *et al.*, 2018). No entanto, várias doenças mentais ainda não foram estudadas em termos de redes e muitas delas são difíceis de serem diagnosticadas, sendo muitas vezes confundidas com outras doenças. Isso acarreta grande prejuízo social e mental ao paciente que, muitas vezes, permanece um longo tempo sendo tratado e diagnosticado de maneira equivocada.

Nesse trabalho, estamos interessados no uso de métodos de mineração de dados e redes complexas a fim de classificar pacientes de acordo com quatro tipos de doenças mentais. Basicamente, vamos considerar dados obtidos por ressonância magnética funcional e representar a estrutura, em nível cortical, do cérebro de pacientes saudáveis e portadores de doenças mentais. A partir dessa representação, vamos extrair medidas das redes, que representam o conjunto de atributos de cada paciente, que constitui uma observação no conjunto de dados. Essa técnica será usada para alcançar os objetivos de nossa pesquisa, que são descritos a seguir.

1.1 Objetivos

O presente trabalho tem como principal objetivo gerar modelos preditivos capazes de discriminar duas classes (de pacientes com determinada doença mental e pessoas saudáveis) de forma a auxiliar no diagnóstico automático de doenças mentais.

Para alcançar tal objetivo foram analisadas redes corticais de quatro bases relacionadas às seguintes doenças neurológicas: (i) esquizofrenia, (ii) autismo, (iii) déficit de atenção/desordem de hiperatividade e (iv) paralisia progressiva nuclear. Cada base contém redes de pacientes que apresentam a doença e pessoas saudáveis. Para cada uma dessas doenças, extraiu-se algumas medidas de rede, que serviram como atributos para algoritmos preditivos discriminarem as duas classes (com ou sem a doença). Esses passos foram esquematizados na figura 1.

Para alcançar os objetivos propostos, as seguintes atividades também foram desenvolvidas:

- **Determinar os algoritmos preditivos que melhor discriminam as duas classes:** Para as quatro bases foram utilizados alguns algoritmos preditivos, com intuito de, primeiramente verificar qual deles gera um modelo preditivo de melhor desempenho; e, segundo, foram utilizados vários algoritmos comum às quatro bases, na tentativa de encontrar aqueles que geram um modelo de boa performance e, com isso, indicar aqueles que lidam melhor com esses tipos de atributos.
- **Determinar métodos de seleção e de normalização (etapas de pré-processamento) que são capazes de melhorar o desempenho dos algoritmos preditivos:** Assim no conjunto de medidas aplicou-se, primeiramente, métodos de seleção de atributos, com intuito

Figura 1 – Esquematização da metodologia utilizada no presente trabalho. Foram utilizadas quatro bases, representadas em (A) por quatro retângulos: roxo, verde, azul e laranja, correspondentes, respectivamente, às doenças: esquizofrenia, autismo, deficit de atenção/desordem de hiperatividade (ADHD) e paralisia progressiva nuclear (PSP). Cada uma das bases contém matrizes de conectividades, representadas em B com retângulos, sendo que azul são aquelas referentes à pacientes sem a doença, e vermelho, àquelas referentes a pacientes com a doença. Cada uma dessas matrizes corresponde a uma rede; e, para cada rede extraiu-se uma série de medidas de redes que serviram como instâncias para a classificação, como pode ser visto em (C).

de encontrar os subconjunto de medidas (atributos) mais discriminantes. Ademais, aplicou-se, também, alguns métodos de normalização para verificar se os métodos melhoram a performance dos classificadores.

1.2 Descrição dos capítulos

O Capítulo 2, contém os principais conceitos relacionados à mineração de dados que foram utilizados no presente trabalho. Ele foi dividido em:

- **Introdução:** em que é abordado as principais etapas do processo de extração de conhecimento.
- **Mineração de dados e os tipos de algoritmos de aprendizado de Máquina:** todos os algoritmos de aprendizado de máquina utilizados são descritos.
- **Métodos de seleção de atributos:** é definido o processo de seleção e, em seguida, são descritos os métodos utilizados.
- **Métodos de normalização:** são descritos os processos de normalização utilizados.

- **Métodos de classificação:** é definido o processo de classificação e são descritos os algoritmos de aprendizado de máquina utilizados na construção de modelos preditivos.

O Capítulo 3 contém as medidas de avaliação de modelos utilizados para analisar o modelo gerado, bem como os métodos de amostragem que aumentam a confiabilidade do modelo.

O Capítulo 4 introduz a descrição das medidas de redes utilizadas e os tipos de redes e conceitos relacionados ao presente trabalho. As medidas foram divididas em: (i) medidas relacionadas à conectividade; (ii) medidas relacionadas a ciclos; (iii) medidas relacionadas à distância; e (iv) medidas de centralidade.

O Capítulo 5 apresenta os conceitos de conectividade cerebral, a descrição da técnica de ressonância magnética; como, a partir dessa técnica, obtêm-se matrizes de adjacência; os tipos de matrizes de conectividade; e o conceito de integração e segregação e a organização do cérebro.

O Capítulo 6 descreve a análise de redes de pacientes com esquizofrenia, sendo, inicialmente feita a revisão bibliográfica do estudo dessa doença.

O Capítulo 7 mostra a análise de redes de pacientes com autismo.

O Capítulo 8 contém a análise de redes de pacientes com déficit de atenção/desordem de hiperatividade.

O Capítulo 9 apresenta a análise de redes de pacientes com paralisia progressiva supranuclear.

Nos capítulos 10 e 11, apresentamos as conclusões e discutimos algumas possibilidades de pesquisa futura.

CONCEITOS BÁSICOS DE MINERAÇÃO DE DADOS

O uso de ferramentas computacionais e estatísticas para análise de dados tem se tornado cada vez mais necessário (TAN; STEINBACH; KUMAR, 2005; WITTEN *et al.*, 2016). Técnicas de aprendizado de máquina (AM), mineração de dados (MD) e reconhecimento de padrões (RP) são importantes para obter conhecimento em grandes bases de dados, formadas por imagens, séries temporais, textos e som (TAN; STEINBACH; KUMAR, 2005). Particularmente, as técnicas de Mineração de Dados (MD) consistem na extração de padrões relevantes em dados. Segundo (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a), MD faz parte de um processo mais geral de descobrimento de conhecimento em bases de dados, processo denominado KDD (do inglês *Knowledge Discovery in Databases*). O processo KDD pode ser definido em (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996b) como: "Um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados, sendo o processo interativo por depender da tomada de decisão do usuário e iterativo pois pode conter *loops* entre quaisquer dois passos".

As várias etapas do processo de descoberta de conhecimento em dados podem ser resumidas em: a) **Seleção**, em que um conjunto de dados é selecionado para a descoberta a ser realizada; b) **Pré-processamento**, onde aumenta-se a confiabilidade dos dados por meio de processos que incluem a sua limpeza, manuseio de valores ausentes e a remoção de ruídos ou *outliers*; c) **Transformação de dados**, onde a geração de melhores dados para a mineração de dados é preparada e desenvolvida, incluindo-se processos de redução da dimensionalidade (como seleção e extração de atributos, bem como a sua amostragem) e transformação de atributos (tal como a discretização de atributos numéricos) (MAIMON; ROKACH, 2009); d) **Mineração de dados**, que é a etapa mais importante do processo KDD, que consiste na aplicação de análise de dados e algoritmos de aprendizado de máquina para reconhecimento de padrões. Por fim, esses

padrões devem ser interpretados de acordo com o objetivo estabelecido da aplicação do processo. A figura 2 contém um esquema para melhor entendimento do processo de KDD.

Figura 2 – Figura modificada de (MAIMON; ROKACH, 2009) contendo o processo KDD. Em azul observam-se as principais etapas do processo: seleção, pré-processamento, transformação e mineração de dados. É importante ressaltar que previamente deve ser estabelecido o objetivo de aplicação do processo para que ao final dele seja dada a correta interpretação.

As técnicas de mineração de dados vêm sendo usadas nas mais diversas áreas incluindo bioinformática (FRANK *et al.*, 2004), atividade de transações de clientes online e off-line, auditorias de computadores relacionados à segurança, tráfego de redes, texto e imagem e qualidade de fabricação (ADHIKARI; ADHIKARI, 2015). Na área médica, tem crescido o interesse da aplicação dessas técnicas para diagnóstico de doença como em: (SONI *et al.*, 2011; ALONSO *et al.*, 2002; MILJKOVIC *et al.*, 2016; KONONENKO, 2001; SRINIVAS; RANI; GOVRDHAN, 2010; YANG *et al.*, 2008). O presente trabalho, também utilizou essas técnicas na tentativa de diagnosticar quatro doenças neurais, que serão descritas posteriormente.

Nesta seção será introduzida uma série de conceitos referentes à mineração de dados e a alguns algoritmos de aprendizado de máquina que foram utilizados no presente trabalho.

2.1 Mineração de dados e os tipos de algoritmos de aprendizado de Máquina

Como já descrito anteriormente, a mineração de dados é a etapa mais importante do processo de KDD que utiliza algoritmos de aprendizado de máquina para extrair padrões de uma base de dados (ARRUDA, 2013).

Figura 3 – Figura extraída e modificada (FACELI *et al.*, 2011) contendo o esquema para melhor visualização da hierarquização dos tipos de algoritmos de aprendizado de Máquina.

Os algoritmos de aprendizado de máquina podem ser descritivos ou preditivos segundo a tarefa que eles realizam. Em tarefas de previsão, a meta é encontrar uma função que dado um conjunto de objetos rotulados constrói um estimador que permite encontrar um rótulo (dentro de um domínio conhecido) para novas instâncias do conjunto de treinamento. Por outro lado, em tarefas de descrição o objetivo é descrever ou explorar um conjunto de dados (FACELI *et al.*, 2011).

Os algoritmos preditivos seguem o paradigma do aprendizado supervisionado, em que há a necessidade de um supervisor externo que conhece o rótulo desejado para cada objeto podendo avaliar o modelo predito. As tarefas supervisionadas podem ser divididas ainda em classificação e regressão quanto ao domínio dos rótulos conhecidos. Em classificação, esse domínio é um conjunto de valores nominais (discretos) e a função estimadora é denominada de classificador. E em regressão, o domínio é um conjunto infinito e ordenado de valores (contínuos).

Os algoritmos descritivos seguem o paradigma do aprendizado não supervisionado uma vez que não utilizam os atributos de saída (os rótulos no caso de algoritmos preditivos). Esses algoritmos são divididos em: agrupamento (em inglês *Clustering*), em que dados semelhantes são agrupados entre si; sumarização, cuja meta é descrever um conjunto de dados de maneira sucinta; e associação, em que padrões de associações entre atributos de um conjunto de dados são encontrados. A figura 3 contém um esquema para melhor visualização dos tipos de algoritmos de aprendizado de máquina.

2.2 Métodos de seleção de atributos

A redução de atributos corresponde a uma das técnicas da etapa de transformação de dados no processo de KDD. Este processo é utilizado com várias finalidades, tal como a redução

da dimensionalidade, pois a alta dimensionalidade pode resultar em um aumento do número de parâmetros usados em alguns algoritmos (SAUNDERS; GAMMERMAN; VOVK, 1998) (com uma quantidade menor de atributos, haverá uma diminuição de parâmetros e um aumento da eficiência computacional) (LIU; MOTODA, 2012). As técnicas de redução da dimensionalidade podem ser divididas em dois tipos de abordagem: a agregação e seleção de subconjuntos (FACELI *et al.*, 2011).

A abordagem de agregação, também denominado de extração de características ou transformação, consiste na transformação ou combinação dos atributos (KANTARDZIC, 2011). A maior parte das técnicas realiza a redução dos atributos originais combinando-os por funções lineares ou não lineares (FACELI *et al.*, 2011). Uma dessas técnicas é a análise de componentes principais (em inglês *Principal Component Analysis*, PCA), que foi utilizada no presente trabalho e será explicada mais adiante na subsecção 2.2.1).

A abordagem para seleção de atributos é um processo de escolha de um subconjunto de atributos de acordo com determinados critérios (LIU; MOTODA, 2012). Em outras palavras, tem como objetivo encontrar um subconjunto de atributos do conjunto original, podendo ser manualmente ou por meio de processos automatizados. Esse processo pode ser, então, visto como um problema de busca (encontrar o subconjunto ideal) (FACELI *et al.*, 2011) e pode ser feito por meio de quatro passos, de acordo com (BLUM; LANGLEY, 1997; ISABELLE, 2006), que faz com que o processo tenha um caráter heurístico. Os passos são (ver figura 4):

1. **Determinação do ponto de partida que influencia a direção da busca:** Assim a seleção pode ser feita começando-se com todos os atributos, sendo um por vez eliminados até achar o subconjunto ideal (abordagem denominada de *backward generation*); ou ainda começando com nenhum atributo, sendo um por vez adicionado até encontrar o subconjunto ideal (abordagem denominada de *forward generation*). Contudo a busca não necessariamente precisa começar em um ponto específico (com todos ou nenhum atributo), sendo que os atributos podem ser adicionados ou removidos (abordagem denominado de *bidirection generation*); ou ainda, o ponto de partida ou o fato de atributos serem adicionados ou removidos pode ser feito de maneira estocástica (abordagem chamada *random generation*) (FACELI *et al.*, 2011).
2. **Determinação da estratégia de busca:** A busca pode ser feita de maneira exaustiva (buscar todos os possíveis conjuntos até encontrar o ideal), estratégia inviável se os dados possuírem alta dimensionalidade (BLUM; LANGLEY, 1997). Uma estratégia mais viável computacionalmente é busca heurística (que utilizam métodos de seleção), um exemplo é um método guloso que considera mudanças no conjunto de características (normalmente remoção ou adição de atributos da amostra), a cada uma dessas variações (correspondentes a interações) o algoritmo de busca seleciona o melhor subconjunto (HALL, 1999). Outros algoritmos de seleção utilizam técnicas de ordenação (em inglês, *ranking*) (GUYON;

ELISSEEFF, 2003) em que os atributos são ordenados de acordo com sua relevância.

3. **Escolha de critério para avaliar os subconjuntos gerados:** Os métodos podem ser divididos em filtros, *wrappers* e embutidos (KANTARDZIC, 2011). Os de wrappers, que normalmente utilizam técnicas exaustivas, utilizam algoritmos de aprendizado de máquina como uma “caixa preta” para selecionar os melhores subconjuntos de acordo com sua predição (GUYON; ELISSEEFF, 2003). Os filtros selecionam subconjuntos de atributos em uma etapa de pré-processamento, sem utilizar preditores; e os embutidos (ou integrados), diz em respeito àqueles contidos em alguns algoritmos de aprendizado de máquina (um exemplo é o algoritmo árvore de decisão que será descrito adiante) (FACELI *et al.*, 2011).
4. **Escolha de um critério de parada.** Um exemplo são as abordagens em wrappers e embutidas, em que o critério de parada é a obtenção do melhor valor preditivo encontrado.

2.2.1 Métodos de seleção utilizados no presente trabalho

Os métodos de seleção utilizados no presente trabalho foram: seleção por PCA, eliminação recursiva de atributos, ordenação de atributos por importância e remoção de características redundantes.

Como já mencionado na secção anterior, a análise de componentes principais (em inglês *Principal Component Analysis*, PCA) é um tipo de seleção que utiliza abordagem de agregação, sendo introduzida por Karl Pearson em (PEARSON, 1901) e desenvolvida por Hotelling em (HOTELLING, 1933; GOMES, 2013). A redução dos atributos é feita pela sua combinação linear (transformação linear ótima). Assim, havendo um conjunto de atributos com m dimensões, $X = (X_1, \dots, X_m)$, é possível combiná-los em um conjunto $Y = (Y_1, \dots, Y_p)$, de forma que a maior parte X seja representado por poucos elementos de Y (TEIXEIRA, 2013). Assim deve-se encontrar a combinação linear descrita em 2.1.

$$Y = AX, \quad (2.1)$$

onde A corresponde a uma matriz que maximiza a variância de Y ($\text{Var}(Y)$). Contudo na prática a matriz A não é calculada diretamente, calcula-se primeiramente a matriz de covariância S (primeiro passo da seleção) (KANTARDZIC, 2011) definida por 2.2.

$$S = 1/(n-1) \left[\sum_{j=1}^n (x_j - x')^T (x_j - x') \right], \quad (2.2)$$

sendo que $x' = [(1/n) \sum_{j=1}^n x_j]$ e S é uma matriz n -dimensional. Em seguida, são calculados os autovalores associados a matriz S e os respectivos autovetores. Cada componente principal é um autovetor associado a um dos autovalores, de forma que a primeira componente principal

Figura 4 – Figura extraída e modificada de (LIU; MOTODA, 2012; GUYON; ELISSEEFF, 2003), contendo o esquema dos tipos de seleção segundo os métodos de avaliação. Em (A), pode ser visto um esquema de como funciona uma seleção, em que um conjunto de atributos é submetido a um método de seleção (I), gerando um subconjunto que será avaliado segundo um critério (II) que será submetido a um critério de parada (III), se aprovado tem-se o subconjunto ideal, caso contrário outro subconjunto deve ser gerado pelo método de seleção. Em (B), há um esquema de método de seleção utilizando a abordagem tipo filtro, em que é gerado um subconjunto (fase 1) que será utilizado para classificação por algoritmo de máquina (V), onde é gerado um modelo a ser avaliado; assim a geração de um subconjunto é independente da etapa de classificação. Um esquema da abordagem wrapper pode ser vista em (C), em que um algoritmo de aprendizado de máquina (II) é utilizado para avaliar um subconjunto de atributos obtido pelo método de seleção (I); se a avaliação do modelo (III) for boa, é selecionado um subconjunto ideal que será utilizado por outra etapa de classificação (fase 2), se for ruim outro subconjunto é considerado.

corresponda ao maior autovalor, a segunda componente ao vetor próprio associado ao segundo maior autovalor, e assim por diante (GOMES, 2013). Matematicamente:

- Os autovalores de S são $\lambda_1, \lambda_2, \dots, \lambda_n$, onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.
- As principais componentes são autovetores $e_1, e_2, \dots, e_n$ associados a $\lambda_1, \lambda_2, \dots, \lambda_n$. Sendo $\text{Var}(Y_j) = \lambda_j$.

Para saber a quantidade de componentes principais capaz de representar os dados, deve-se calcular a taxa descrita pela equação 2.3.

$$R = (\sum_{i=1}^m \lambda_i) / (\sum_{i=1}^n \lambda_i). \quad (2.3)$$

Essa taxa representa a divisão dos m maiores autovalores de S pelo traço¹ da matriz S (KANTARDZIC, 2011). Quanto maior R melhor a representação dos dados pela componentes principais.

Outro método de seleção adotado aqui é a remoção de atributos redundantes, denominada de seleção de atributos baseada em correlações (em inglês *Correlation-based Feature Selection*, CFS), tratando-se de uma abordagem baseada em filtro (HALL; HOLMES, 2003).

Um atributo é redundante se outros atributos são altamente correlacionados a ele; assim um bom subconjunto de atributos é aquele altamente correlacionado com a classe e sem correlação entre seus elementos (HALL, 2000). Assim, este método de seleção utiliza medidas de correlação para quantificar essa correlação.

No presente trabalho, foi utilizado a correlação de Pearson (ρ dada pela equação 2.4). É medida, então, a correlação de todos os atributos entre si, que são armazenados em uma matriz. Aqueles altamente correlacionados (no caso considerou-se valores de correlação maiores que 0.7) são então removidos.

$$\rho = \frac{k\overline{r_{CA}}}{\sqrt{k + k(k-1)\overline{r_{AA}}}}. \quad (2.4)$$

Na equação 2.4, k , corresponde ao número de atributos; $\overline{r_{CA}}$ é a média de correlação entre a classe e os atributos; $\overline{r_{AA}}$ é a média de correlação entre atributos.

Outro método de seleção utilizado foi a ordenação de atributos por ordem de importância (em inglês, *rank features by importance*), em que considera-se como critério de busca, a ordenação. No caso foi utilizado a função *varimp* (a metodologia em si será mais detalhada adiante) presente no pacote *caret* do software *R*, que é capaz de utilizar ou não um modelo para sua avaliação (KUHN, 2012). Quando esta função utiliza um modelo preditivo para auxiliar

¹ O traço de uma matriz corresponde a soma dos elementos da diagonal principal.

na avaliação, trata-se de um critério de avaliação *wrapper* e quando não trata-se de um filtro; contudo é mais vantajoso utilizar a abordagem baseada em modelo já que é capaz de incorporar à avaliação dos preditores no cálculo de importância (KUHN, 2008b).

Dessa forma foi empregado o critério de avaliação *wrapper*, usando como modelo o algoritmo de aprendizado de quantização de vetor (em inglês, Linear Quantization Vector, LQV). Este método tem como proposta a divisão do espaço vetorial em regiões discretas (classes). Para cada uma dessas regiões é definido um vetor de reconstrução (KOHONEN, 1995); quando um novo vetor de entrada é apresentado ao quantizador, é determinado a região inicial em que ele se encontra e ela é então representada pelo vetor de reconstrução daquela região. A coleção de possíveis vetores de codificação é denominada livro de códigos do quantizador (em inglês, *code book*), e seus membros palavras códigos (em inglês, *code words*). A utilização desses vetores em substituição aos originais contribui para uma economia de armazenagem as custas de alguma distorção (HAYKIN, 2007).

A mínima distorção é obtida com um quantizador de Voronoi, que realiza a partição do espaço pela regra do vizinho mais próximo. Assim o algoritmo LQV pode ser resumido da seguinte maneira: um vetor de entrada x é escolhido aleatoriamente no espaço de entrada, se os rótulos de classe desse vetor concordarem com o vetor w de Voronoi, o vetor w é movido na direção de x ; se w e x discordarem, o vetor de Voronoi se afasta do vetor de entrada (MEYER-BAESE, 2004). Após vários passos (iterações) através dos dados de entrada, os vetores de Voronoi tipicamente convergem, e o treinamento está completo (HAYKIN, 1994).

Por fim, utilizou-se a eliminação recursiva de dados (em inglês, *recursive feature elimination*), também presente no pacote *caret*. Esse tipo de seleção utiliza para avaliação a abordagem tipo *wrapper* e para a direção de busca a abordagem *backward propagation* (KUHN, 2012). Utilizou-se como algoritmo de inteligência artificial, no modelo de avaliação, *random forest*.

O método *random forest* é um algoritmo resultado da combinação (do tipo *bagging*, que será explicado mais adiante) de árvores de decisão (algoritmo que também será explicado adiante) em um mesmo objeto. Cada árvore, seguindo sua predição, vota em uma classe dos atributos de entrada, a classe com um maior número de votos é o resultado da predição, sendo que o erro de generalização converge quanto maior o número de árvores (BREIMAN, 2001). Este algoritmo possui muitas vantagens em relação a outras técnicas de aprendizado de máquina em termos da capacidade de lidar com dados biológicos não-lineares, robustez a ruídos, sendo muito efetivos em processos de seleção de atributos (LEBEDEV *et al.*, 2014).

2.3 Métodos de normalização

A normalização dos dados é uma técnica da etapa de transformação de dados no processo de KDD. No caso trata-se de uma transformação de atributos numéricos empregada para evitar que um atributo predomine sobre o outro (em termos de escala) principalmente quando se deseja

utilizar classificadores baseados em distância .

No presente trabalho foram utilizadas três tipos de normalização: *Z-score* (ZS), *Linear Scaling to Unit Range* (LSTUR) e a *Softmax Scaling* (SS). A técnica *Linear Scaling to Unit Range* e *Softmax Scaling* consistem em normalizações por reescala onde para todos os atributos é definida uma nova escala de valores (com diferentes limites mínimo e máximo). Por outro lado, a técnica *Z-score* é uma normalização por padronização que define um valor central e um valor de espalhamento comum para todos os atributos (FACELI *et al.*, 2011).

A técnica *Linear Scaling to Unit Range* é uma transformação linear para normalizar os atributos na faixa de [0,1]. Dado o menor ($\min(x_i)$) e maior ($\max(x_i)$) valor de determinado atributo x_i , o valor dessa normalização ($\bar{x}_i$) é dado por 2.5 (SINGH; VERMA; THOKE, 2015).

$$\bar{x}_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}. \quad (2.5)$$

A normalização *Softmax Scaling* é uma técnica não-linear utilizada quando os dados não estão distribuídos uniformemente em torno da média. Os valores normalizados, $\bar{x}_i$, são dados por 2.6 e abrangem uma faixa de [0,1].

$$\bar{x}_i = \frac{1}{1 + e^{-y}}. \quad (2.6)$$

Onde $y = \frac{x_i - \mu}{r\sigma}$, μ é a média do atributo, σ é a variância e r é um parâmetro determinado numericamente pelo usuário (utilizou-se o valor de $r=1$). Pode-se notar que para valores de x_i próximos a μ , y se aproxima de uma função linear.

Por fim, para a normalização por padronização a cada valor do atributo é adicionada ou subtraída uma medida de localização e o valor resultante é multiplicado ou dividido por uma medida de escala. Deste modo, diferentes atributos possuirão as mesmas medidas de escala e espalhamento, ainda que não possuam os mesmos limites (FACELI *et al.*, 2011). Para o caso particular em que as medidas de localização e de escala forem respectivamente a média μ e a variância σ , a normalização é denominada *Z-score* e é dada pela expressão 2.7. Normalmente utiliza-se a normalização por padronização, pois esta é menos suscetível à presença de *outliers* que a normalização por reescala.

$$\bar{x}_i = \frac{x_i - \mu}{\sigma}. \quad (2.7)$$

2.4 Métodos de classificação

Em problemas de classificação procura-se encontrar uma função (classificador) que mapeie o conjunto de todos os exemplos (instâncias) possíveis dentro um conjunto pré-definido de rótulos (classes). Segundo (MAIMON; ROKACH, 2010), o objetivo de um classificador é

definido formalmente como: “Dado um conjunto de treinamento, S , com um conjunto de atributos de entrada $A=a_1, a_2, \dots, a_n$ e um atributo nominal y de uma distribuição fixada D desconhecida sob o espaço de instância rotulada, o objetivo é induzir um classificador ótimo com o mínimo erro de generalização.”

Mais formalmente², $dom(y) = c_1, c_2, \dots, c_k$ (domínio dos valores de y), constitui o conjunto dos rótulos (classes). O conjunto de todas as instâncias possíveis (espaço de instâncias), X , é definido como o produto cartesiano de todos os domínios dos atributos de entrada: $X = dom(a_1) \times dom(a_2) \times \dots \times dom(a_n)$. O espaço das instâncias rotuladas, U , é definido como o produto cartesiano de todos os domínios dos atributos de entrada pelo conjunto dos rótulos: $U = X \times dom(y)$. O conjunto de treinamento, $S(B)$, composto de m tuplas é dado pela equação 2.8.

$$S(B) = (< x_1, y_1 >, \dots, < x_m, y_m >), \quad (2.8)$$

onde $x_q \in X$ e $y_q \in dom(y)$, onde $q = 1, \dots, m$. Sendo que as tuplas são geradas de forma aleatórias e distribuídas de acordo com uma distribuição de probabilidade D sobre U . Formalmente, no caso de atributos nominais, essa taxa pode ser expressa pela expressão 2.9. O erro de generalização é definido como sendo a taxa de classificações que foram erradas sobre a distribuição D . Sendo I um indutor, tal que $I(S)$ é o classificador gerado por I para um conjunto de treinamento S e $I(S)(x)$, a classificação produzida por $I(S)$ aplicado ao padrão x . Dizemos por $I(S)$ o classificador que é gerado por I para o conjunto de treinamento S .

$$\varepsilon(I(S), D) = \sum_{(x,y) \in U} D(x,y) \cdot L(y, I(S)(x)), \quad (2.9)$$

onde $L(y, I(S)(x))$ é uma função de perda zero-um definida como zero se $y = I(S)(x)$ e um se $y \neq I(S)(x)$.

Existe uma grande quantidade de métodos preditivos de AM, contudo aqueles mais recorrentes em artigos (BALCÁZAR *et al.*, 2010) foram escolhidos para serem aplicados no presente trabalho. As subseções seguintes contêm uma breve descrição dos algoritmos utilizados.

2.4.1 k -NN

O algoritmo do vizinho mais próximo, do inglês *Nearest Neighbor (NN)*, proposto em (FIX; HODGES, 1951; FIX; HODGES, 1952), tem como objetivo caracterizar (rotular ou achar uma classe) uma amostra, a partir de uma ou mais medidas (normalmente utiliza-se medidas de dissimilaridade e de similaridade) de uma amostra, com base nas informações de um ou mais indivíduos previamente rotulados no espaço de busca.

² As definições mais formais, adiante, foram extraídas e adaptadas de (ROKACH, 2010).

O funcionamento deste algoritmo é muito simples e pode ser dividido na fase de treinamento e na fase teste. Na fase de treinamento ele memoriza todos os rótulos (classes) do conjunto de entrada. Para classificar um exemplo não rotulado, na fase de teste, é calculada a distância entre o vetor de valores de atributos e cada exemplo rotulado em memória. O rótulo da classe associada ao exemplo de treinamento mais próximo ao exemplo de teste é utilizado para classificar o novo exemplo.

Em (COVER; HART, 1967), foi demonstrado que a probabilidade de erro do método é menor que duas vezes a probabilidade de erro de Bayes (algoritmo que será explicado na subseção seguinte), sendo assim também inferior à probabilidade de erro das demais regras aplicáveis, sejam elas paramétricas ou não, para um conjunto de amostras infinitas.

Graficamente as superfícies de decisão utilizadas para resolver um problema de classificação NN são poliedros convexos com centros em cada objeto do conjunto de treinamento. Todos os pontos no interior deste poliedro pertencem a classe do objeto que se situa no centro dele. O conjunto desse poliedro é designado *diagrama de Voroni*.

Uma extensão do algoritmo, denominado de k-NN (do inglês, *k- nearest neighbor*), considera em vez de um vizinho mais próximo, os k vizinhos mais próximos do objeto de teste, sendo k um parâmetro desse algoritmo. Assim quando $k > 1$, para cada ponto de teste, são obtidos k vizinhos. Cada vizinho vota em uma classe, o objeto teste é classificado na classe mais votada, sendo que a função que minimiza a função de custo 0-1 (no caso o erro de generalização) é a moda (FACELI *et al.*, 2011).

Este algoritmo é denominado ainda de preguiçoso (em inglês *lazy*) porque adia o processo de aprendizagem (que consiste apenas em memorizar as instâncias) para até a fase de classificação. A figura 5 contém um exemplo ilustrativo do kNN.

Percebe-se da figura 5 que a classificação depende da escolha do valor do parâmetro k que é determinado pelo usuário. Normalmente utiliza-se valores pequenos e ímpares, evitando valores pares para que não ocorra empate na votação da classe.

Apesar de ser um algoritmo de fácil aplicação, a fase de treinamento possui um alto custo computacional. Outro aspecto negativo é o fato dele ser afetado por atributos redundantes e irrelevantes como todo método baseado em distância. Além disso, novamente por ser baseado em distância, ele é afetado pela dimensionalidade do problema que está relacionada à quantidade de atributos.

2.4.2 Naive Bayes

Para entendimento desse algoritmo é necessário primeiro introduzir o Teorema de Bayes. O Teorema de Bayes calcula a probabilidade *a posteriori* de um evento dado sua probabilidade *a priori* e a verossimilhança do novo dado (MITCHELL, 1997). Mais precisamente, supondo que $P(y_i|x)$ é a probabilidade de um evento x pertencer à classe y_i (probabilidade *a posteriori* de x).

Figura 5 – Figura extraída e modificada de ((FACELI *et al.*, 2011)) contendo um modelo preditivo de um conjunto bidimensional com duas classes (1 e 2) e o objeto em verde é a instância a ser classificada. Os círculos representam a superfície de decisão encontrada pelo classificador em questão. O círculo 1 representa a superfície do classificador 1-NN e como o objeto mais próximo do verde possui rótulo vermelho ele será classificado como vermelho. De maneira análoga considerando 3-NN (círculo 2) como a maioria dos objetos próximos ao verde são azuis ele será classificado como azul. Seguindo esse raciocínio para o 5-NN o objeto verde será classificado como vermelho. Percebe-se, então, que o parâmetro k influencia na tomada de decisão do classificador k -NN.

A função de custo zero-um é dada pelo custo de associar a classe incorreta, sendo minimizada quando x é associada à classe y_k quando $P(y_k|x)$ é máxima. Este método chamado de MAP (*Maximun A Posteriori*) é dado pela expressão 2.10.

$$y_{MAP} = \operatorname{argmax}_i P(y_i|x), \quad (2.10)$$

em que argmax retorna a classe y_i com a maior probabilidade de estar associada a x , em que $P(y_i|x)$ é a função discriminante, que separa exemplos de classes diferentes. O Teorema de Bayes provê um método para calcular esta função dada por 2.11 (FACELI *et al.*, 2011).

$$P(y_i|x) = \frac{P(y_i)P(x|y_i)}{P(x)}. \quad (2.11)$$

O Algoritmo Naive Bayes assume que os atributos de um exemplo são independentes da classe e por isso é denominado ingênuo (do inglês, *naive*). Assim $P(x|y_i)$ pode ser decomposto em: $P(x^1|y_i) \times \dots \times P(x^j|y_i)$ em que x^j é o j -ésimo atributo do exemplo x . Portanto, a probabilidade de um exemplo pertencer à classe y_i é proporcional a equação 2.12. O classificador obtido pelo

uso dessa função discriminante é denominado classificador Naive Bayes.

$$P(y_i|x) \propto P(y_i) \prod_{j=1}^j P(x^j|y_i). \quad (2.12)$$

Este algoritmo é muito utilizado na prática por resultar em classificação com elevada performance, sendo robusto a atributos irrelevantes (ADHIKARI; ADHIKARI, 2015).

A probabilidade de erro de classificação, ou erro de generalização, de um classificador h , $R(h)$ é definido pela equação 2.13 .

$$R(h) = \sum_{x \in U} P(h(x) \neq y_i) P(x), \quad (2.13)$$

em que $P(h(x) \neq y_i)$ é a probabilidade do rótulo obtido pelo classificador h ser diferente da classe y_i e U o espaço das instâncias (vide 2.9) e P_x é o erro de Bayes (RISH; HELLERSTEIN; THATHACHAR, 2001).

2.4.3 Árvores de decisão

As árvores de decisão são representações simples do conhecimento e têm sido amplamente aplicadas como, por exemplo, em diagnósticos médicos, análise de risco em créditos, entre outros exemplos (BARROS; CARVALHO; FREITAS, 2015).

Elas consistem em um método não paramétrico que pode ser utilizado tanto para problemas de classificação como de regressão (ROKACH; MAIMON, 2014). São estruturas hierárquicas do aprendizado supervisionado por onde o espaço de entrada é dividido em regiões locais de modo a prever a variável dependente.

Uma árvore de decisão pode ser vista como um grafo $G = (V, E)$ consistindo de um conjunto finito, não-vazio de nós (vértices, V) e um conjunto de arestas (ROKACH; MAIMON, 2014). Esses nós representam os atributos de um conjunto de dados e as arestas que os conectam recebem os valores possíveis para estes atributos. Existem, ainda, três nós possíveis: um nó raiz, que não possui nenhuma aresta de entrada e zero ou mais arestas de saída; nós internos, cada qual com exatamente uma aresta de entrada e duas ou mais arestas de saída; e nós folhas, cada qual com uma única aresta de entrada e nenhuma de saída, pois é o nó que determina a qual classe o exemplo pertence (FRIEDL *et al.*, 2010). A figura 6 contém uma árvore de decisão gerada a partir de um espaço bidimensional.

Figura 6 – Figura extraída e modificada de (FACELI *et al.*, 2011), contendo árvore de decisão obtida a partir de um espaço bidimensional (x_1, x_2) e cada nó corresponde a uma região desse espaço.

A construção de uma árvore de decisão utiliza a hipótese de indução (trata-se de um algoritmo de indução) (GUEIREZ, 2014). Esse processo possui um alto custo computacional porém depois da árvore ser gerada o seu uso é imediato e rápido. O algoritmo de indução escolhe o atributo preditivo que será utilizado em cada nó da árvore. Essa escolha é feita utilizando diversos critérios, sendo o principal as medidas de impureza. O algoritmo, então, tenta dividir os dados de um nó de forma a minimizar o grau de impureza dos nós filhos. Se um nó for totalmente puro (homogêneo), todos os exemplos que fazem parte desse nó pertencem à mesma classe.

O algoritmo utilizado no presente trabalho foi o *CART* (em inglês, *Classification and Regression Tree*) introduzido em (BREIMAN *et al.*, 1984), produz uma árvore binária (ROKACH; MAIMON, 2008) e a medida de impureza utilizada na divisão é o índice *Gini*, que para um nó t é dado pela equação 2.14.

$$Gini(t) = 1 - \sum_i p_i^2, \quad (2.14)$$

em que p_i é a probabilidade da classe i estar em t ; assim para cada divisão candidata, a impureza de todas as subdivisões é somada e aquela que mais reduzir a impureza é escolhida (APTÉ; WEISS, 1997). Para esta abordagem, ainda, após a árvore ser gerada é utilizado um mecanismo de poda por redução do fator custo complexidade cujo objetivo é a redução da taxa de erro, $R(T)$, levando em consideração a dimensão da árvore, $|T|$ (complexidade) (SHI, 2007). A redução do fator custo complexidade é dado por 2.15.

$$R_\alpha(T) = R(T) + \alpha|T|, \quad (2.15)$$

em que α é o parâmetro que pondera a importância relativa entre o tamanho da árvore e a taxa de erro. Para cada valor desse parâmetro, o objetivo é encontrar uma sub-árvore que minimize

$R_\alpha(T)$; é produzido, então, uma sequência de árvores com nós cada vez menores e aquela que minimiza $R_\alpha(T)$ é selecionada (GAMA, 1999).

2.4.4 Redes neurais

As redes neurais (RNAs) são algoritmos do aprendizado de máquina baseados nos neurônios cerebrais que são sistemas biológicos capazes de detectar padrões, fazer previsões e aprender (SUMATHI; SIVANANDAM, 2006). Elas são formadas por unidades interconectadas de neurônio capazes de receber sinais, processá-los e enviar um sinal como resposta (ZHANG, 2018).

O primeiro algoritmo deste tipo foi desenvolvido por McCulloch e Pitts (MCCULLOCH; PITTS, 1943), que propuseram um neurônio artificial, denominado de unidade lógica com limiar (LTU, em inglês *Logic Threshold Unit*), cujo o objetivo é executar funções lógicas simples (FACELI *et al.*, 2011). As redes neurais, então, são formadas por um conjunto destes neurônios artificiais que são unidades não-lineares (ANDERSON; ROSENFELD, 2000). Segundo (HAYKIN, 1994) há três elementos básicos de um modelo neuronal, sendo eles:

- Um conjunto de sinapses ou arestas conectivas, sendo que cada uma delas é caracterizada por um peso.
- Uma junção capaz de somar os sinais de entrada, ponderados pelas suas respectivas sinapses.
- Uma função de ativação, limitando o sinal de saída do neurônio, e os bias capazes de aumentar ou diminuir a entrada para função de ativação.

Com isso, para um neurônio k , onde $x_1, x_2, \dots, x_k$ são seus sinais de entrada ponderados, respectivamente, por seus pesos sinápticos $w_{k1}, w_{k2}, \dots, w_{km}$, a junção realizará uma combinação linear resultando em u_k , seguindo a equação 2.16; sendo os bias, b_k , e a função de ativação, ϕ , o sinal de saída, y_k , é dado matematicamente pela equação 2.17. A figura 7 contém um esquema de um neurônio artificial, bem como um neurônio biológico para comparação.

$$u_k = \sum_{j=1}^m w_{kj} x_j. \quad (2.16)$$

$$y_k = \phi(u_k + b_k). \quad (2.17)$$

Existem várias funções de ativação. No caso do modelo pioneiro proposto por McCulloch e Pitts a função de ativação era a limiar cujos valores de resultado possíveis são 0 e 1. Se o valor de ϕ ultrapassar o limiar estabelecido o processo seguirá para o neurônio seguinte (neurônio torna-se ativo). O que se assemelha aos neurônios biológicos após o impulso elétrico passar

Figura 7 – Figura contendo os esquemas de um neurônio biológico e de um artificial. Do lado esquerdo, um esquema de um neurônio artificial (extraído e adaptado de (HAYKIN, 1994)), em que os sinais de entrada $x_1, x_2, \dots, x_k$ são ponderados pelos pesos sinápticos $w_{k1}, w_{k2}, \dots, w_{km}$ através de uma combinação linear, passando pelos bias e função de ativação até obter o sinal de saída, y_k . Do lado direito, um neurônio biológico, extraído de (BORGES *et al.*, 2015), que possui os dendritos como entrada do pulso elétrico e os axônios a saída deles.

pelos dendritos se a intensidade dele ultrapassar o limiar de ativação o impulso se propaga para o neurônio seguinte.

As redes neurais se organizam em multicamadas (vide figura 8) que consistem nas interconexões entre neurônios (nós da rede) organizados em três tipos de camadas: camada de entrada, camada oculta e camada de saída. Os nós fonte da camada de entrada correspondem a atributos de entrada e os neurônios da camada de saída correspondem a atributos de saída. Os neurônios das camadas ocultas são conectados tanto aos nós fonte como os de saída e induzem a classificação (SUMATHI; SIVANANDAM, 2006).

Figura 8 – Esquema de uma rede neural. A primeira camada contém os atributos de entrada conectada às camadas ocultas e à camada de saída.

2.4.5 Combinação de classificadores

A combinação de classificadores resultando em um único classificador (normalmente denominado de *ensemble*) (MACLIN; OPITZ, 1997) é uma técnica efetiva para aumentar a acurácia de alguns classificadores e reduzir significativamente o erro (FREUND; SCHAPIRE, 1996) quando usados individualmente. As técnicas mais utilizadas são: *Bagging*, *Boosting* e *Stacking*. As técnicas de *Bagging* e *Boosting* serão descritas brevemente a seguir, e o foco será a técnica de *Stacking* que foi utilizada no presente trabalho.

A técnica de *Bagging* (abreviatura de *bootstrap*³ *aggregating* (BREIMAN, 1996)) combina múltiplos algoritmos preditivos que foram aplicados a sub-amostras aleatórias de conjunto de dados e fusão deles em um único algoritmo é feita por meio de votação (TANWANI *et al.*, 2009). Assim é feito uma reamostragem aleatória de um conjunto de dados, em N conjuntos, e a cada uma dessas amostras é aplicada T classificadores, pela agregação de cada um deles é formado um único classificador. Assim, uma instância a ser classificada, cada um dos classificadores vota em uma classe, aquela mais votada é escolhida (QUINLAN, 1996a). O classificador *random forest* é formado pela combinação de árvores de decisão utilizando essa técnica.

³ Tipo de técnica de reamostragem que será descrita no capítulo de Avaliação de modelos.

A técnica *Boosting* (SCHAPIRE, 1990) funciona de forma semelhante a técnica *Bagging* (FREUND; SCHAPIRE, 1996) porém uma série de pesos são mantidos no conjunto de treinamento para cada instância (DIETTERICH, 2000), quanto maior o peso mais a instância influencia na predição (QUINLAN, 1996a).

A técnica de *Stacking* (WOLPERT, 1992) combina a saída de dois ou mais níveis de classificadores treinando-os por um meta-algoritmo (TANWANI *et al.*, 2009). Assim no primeiro nível, a um conjunto de treinamento S , com N instâncias $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, é aplicado a um conjunto de algoritmos $A_1, A_2, \dots, A_L$ induzindo a um conjunto de hipóteses $h_1, h_2, \dots, h_L$; essas hipóteses são, novamente, aplicadas ao conjunto de treinamento (nível 1) e o resultado é avaliado por um meta-algoritmo resultando em um meta classificador h (BERNARDINI, 2002). A figura 9, contém um esquema da técnica de *Stacking*.

No presente trabalho, foi utilizado como meta classificador a regressão logística. Modelos de regressão tem como intuito modelar relações entre variáveis, sendo uma ou mais variáveis independentes e uma variável resposta (dependente); um dos casos particulares dos modelos lineares generalizado são aqueles em que a variável resposta apresenta duas categorias, sendo um destes modelos mais conhecidos a regressão logística (CABRAL, 2013).

Matematicamente, segundo (MANLY, 2008), supondo m grupos a serem comparados, com um grupo i consistindo n_i itens, sendo λ_i exibem uma resposta positiva, e, $n_i - \lambda_i$, exibem uma resposta negativa; sendo, ainda, x_{ij} $j=1, \dots, p$, os valores da variável X_j , a probabilidade, π_i , de uma resposta positiva em um item no grupo i é dada pela equação 2.18. A probabilidade da resposta negativa é dada por $1 - \pi_i$.

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}. \quad (2.18)$$

A função que relaciona π_i às variáveis X_j é denominada regressão logística. Os parâmetros desconhecidos β_i são habitualmente estimados através de máxima verossimilhança.

Figura 9 – Figura extraída e modificada de (BERNARDINI, 2002), contendo um esquema da técnica de *Stacking*. No primeiro nível um conjunto de treinamento S , com N instâncias $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ é submetido a um conjunto de algoritmos $A_1, A_2, \dots, A_L$ induzindo um conjunto de hipóteses $h_1, h_2, \dots, h_n$. No nível-2, as hipóteses são novamente aplicadas ao conjunto de treinamento S (formando um novo conjunto de treinamento) e é submetido a um meta classificador para então classificá-los.

AVALIAÇÃO DE MODELOS

Estimar a performance de uma classificação induzida por algoritmos de aprendizado supervisionado é importante para avaliar a veracidade do modelo construído, ou seja, o quanto ele é capaz de prever o rótulo de determinado objeto corretamente (KOHAVI, 1995). Uma das medidas que mede o desempenho de um classificador é a taxa de erro, dada por 3.1.

$$err(\text{classificador}) = \frac{1}{n} \sum_{i=1}^n I(y_i \neq f(x_i)). \quad (3.1)$$

No caso n é o número de objetos que serão classificados, sendo que y_i é o rótulo e $f(x_i)$ é a classe do objeto x_i . No caso se $y_i = f(x_i)$, a função $I(y_i \neq f(x_i)) = 0$, caso contrário ela tem o valor igual a 1. Em que I é semelhante a função de custo 0-1 (FACELI *et al.*, 2011). A taxa de erro varia de zero a um sendo que quanto menor seu valor melhor a performance do classificador.

Outra medida muito utilizada é a acurácia dada pela equação 3.2, sendo a função complementar da taxa de erro. Logo quanto maior valor da acurácia melhor a performance de um classificador.

$$acur(\text{classificador}) = 1 - err(\text{classificador}). \quad (3.2)$$

Tratando-se de problemas de duas classes (negativas e positivas), uma alta acurácia pode não necessariamente ser aceitável, uma vez que classificador pode prever corretamente somente uma das classes; nestes casos, pode-se usar duas medidas denominadas: sensibilidade e especificidade (MAIMON; ROKACH, 2010).

Antes de explicá-las, vale a pena explicar a matriz de confusão, dada pela tabela 1. Em que, a taxa de verdadeiros positivos (VP) corresponde aos números de exemplos classificados corretamente; a taxa de falsos negativos (FN), que são as instâncias positivas preditas como negativas; a taxa de falsos positivos (FP), em que instâncias negativas foram classificadas como

Tabela 1 – Tabela contendo a matriz confusão em problemas de duas classes, positivas e negativas. Em azul, trata-se dos rótulos obtidos durante a previsão; e, em amarelo, o verdadeiro rótulo das classes.

		Classe predita	
		Positiva	Negativa
Classe verdadeira	Positiva	VP	FN
	Negativa	FP	VN

positivas; e a taxa de verdadeiros negativos (VN), correspondentes às instâncias negativas preditas corretamente.

Assim, a sensibilidade (*sens*) e a especificidade (*esp*) de um preditor f , segundo (FACELI *et al.*, 2011), são dadas pelas respectivas equações: 3.3 e 3.4 .

$$sens(f) = \frac{VP}{VP + FN}. \quad (3.3)$$

$$esp(f) = \frac{VN}{VN + FP}. \quad (3.4)$$

Assim a sensibilidade (também denominada de revocação) mede o quanto um classificador é capaz de prever os exemplos positivos (taxa de acerto na classe positiva). Enquanto a especificidade corresponde a taxa de acerto na classe negativa.

A acurácia, em problemas de duas classes, pode ser, ainda, descrita como na equação 3.5, em termo das taxas presentes na matriz de confusão.

$$acur(f) = \frac{VP + VN}{VP + VN + FN + FP}. \quad (3.5)$$

Associado a acurácia pode ser calculada a medida *kappa* (COHEN, 1960), κ , coeficiente de concordância que é capaz de avaliar a acurácia e é dada pela equação 3.6.

$$\kappa = \frac{P_o - P_e}{1 - P_e}, \quad (3.6)$$

onde, $P_o = VP + VN$ (soma dos elementos da diagonal principal da matriz confusão), e, $P_e = (VP + FN)(VP + FP) + (FN + VN)(FP + VN)$. A tabela 2 fornece uma regra geral para utilizar a medida em questão.

Tabela 2 – Tabela adaptada de (LANDIS; KOCH, 1977), fornece uma referência útil para descrever a força relativa associada ao kappa, ainda que as divisões em questão foram arbitrariamente criadas pelo autor.

Kappa estatístico	Força de concordância
<0.00	quase nenhuma
0.00-0.20	muito fraca
0.21-0.40	fraca
0.41-0.60	moderada
0.61-0.80	substancial
0.81-1.00	quase perfeita

Também tratando-se de problemas de duas classes é possível gerar um gráfico da taxa de verdadeiros positivos pela taxa de falsos positivos, a curva encontrada é denominada de *ROC* (em inglês, *Receiver Operating Characteristic*). A área abaixo desta curva, *AUC* (em inglês *Area Under Curve ROC*), tem sido muito empregada em problemas de classificação, como aqueles utilizados para diagnósticos médicos (HUANG; LING, 2005). A medida *AUC* varia de 0 a 1, e valores mais próximos a 1 são considerados melhores (FACELI *et al.*, 2011). A tabela 3 contém uma regra geral, segundo (HOSMER; LEMESHOW; STURDIVANT, 2013) para utilização da medida em questão.

Tabela 3 – Tabela adaptada de (HOSMER; LEMESHOW; STURDIVANT, 2013) em que os autores fornecem uma regra geral de utilização da área abaixo da curva ROC.

Valor de AUC	Regra geral
$AUC = 0.5$	sem discriminação
$0.7 < AUC < 0.8$	discriminação aceitável
$0.8 < AUC < 0.9$	discriminação excelente
$AUC \geq 0.9$	discriminação perfeita

Com o intuito de obter maior confiabilidade do modelo podem-se usar métodos de amostragem que dividem os objetos a serem classificados em um conjunto de teste e um de treinamento. O conjunto de teste será aquele em que o modelo será induzido e o teste aquele em que o modelo induzido será aplicado. Existem diversos métodos porém utilizou-se no presente trabalho uma das metodologia mais utilizadas: a validação cruzada (em inglês, *k-fold cross validation*) (ARRUDA, 2013).

Esse método proposto por (LEISCH; JAIN; HORNIK, 1998) consiste na divisão do conjunto em k subconjuntos. As $k-1$ instâncias são utilizadas pelo classificador na fase de treinamento para induzir o modelo de classificação que será testada na partição restante (fase de treinamento). Este processo é repetido k vezes sendo que cada vez é utilizado uma partição diferente. A figura 10 contém um esquema deste processo considerando $k = 4$ (THEODORIDIS; KOUTROUMBAS, 2001).

Figura 10 – Figura contendo um esquema do processo de Validação Cruzada para o caso de $k=4$. Primeiramente o conjunto de dados é dividido em quatro partes (nomeados na figura de A, B, C e D) e com isso processo de validação foi realizado quatro vezes (1-4). Em 1 a parte foi retirada para a fase de teste e o resto do conjunto foi utilizado para induzir o modelo preditivo. Em 2 a parte C foi selecionada para a fase de avaliação e o restante para a classificação. E desse mesmo modo se deu para 3 e 4 em que as fases de teste selecionadas foram, respectivamente, D e B. Cada processo de avaliação obteve uma performance (Performance(1), Performance(2), Performance(3) e Performance(4)).

CARACTERIZAÇÃO DE REDES COMPLEXAS

A Teoria das redes complexas é baseada na representação de sistemas complexos por meio de grafos formados por conjunto de vértices (nós) que são ligados por arestas (links) devido a algum tipo de interação.

Em termos computacionais, as redes podem ser armazenadas através de listas ou matrizes de adjacência. Assim, qualquer grafo G , com N nós, pode ser representado por uma matriz de adjacência, $A(G)$, com $n \times n$ elementos A_{ij} , cujos valores são 1 ($A_{ij} = A_{ji} = 1$), caso os nós i e j estiverem conectados, e 0 caso contrário (ALBERT; BARABÁSI, 2002). Em redes não dirigidas a matriz de adjacência é simétrica, sendo assim elas são iguais às suas transpostas; no caso de redes dirigidas isso não ocorre. A figura 11, a seguir, ilustra a representação de duas redes, uma dirigida e a outra não dirigida, na forma matricial.

Figura 11 – Figura extraída de (RODRIGUES, 2007) que contém a representação matricial de dois grafos. Em (a) a matriz é simétrica representando uma rede não dirigida, o que não é observado em (b) caracterizando uma rede dirigida. Os valores das matrizes iguais a um representam que há conexão entre os vértices i e j ; enquanto que os valores zero representam a ausência de conexão.

As medidas de rede são fundamentais para caracterização, classificação e modelagem de redes. Nas subseções seguintes serão descritas todas as medidas utilizadas neste trabalho.

4.0.1 Medidas relacionadas à conectividade

O grau, também conhecido como conectividade, K_i , corresponde ao número de conexões de um vértice i (a cardinalidade de sua vizinhança) cujo valor para redes não dirigidas e em termos da matriz de adjacência (BARABÁSI, 2016) é obtido pela equação 4.1.

$$K_i = \sum_{j=1}^N (A_{ij}). \quad (4.1)$$

Para redes dirigidas deve-se considerar o número de arestas que saem do nó K_i^{out} , e o número de arestas incidentes ao nó, K_i^{in} . Nesse caso o grau total, K_i^{in} , pode ser obtido pela equação 4.2.

$$k_i = k_i^{out} + k_i^{in}. \quad (4.2)$$

A partir de 1 é possível calcular a respectiva medida global, chamada conectividade média da rede, $\langle K \rangle$, cujo valor pode ser obtido pela equação 4.3 .

$$\langle K \rangle = \frac{1}{N} \sum_{i=1}^N (k_i). \quad (4.3)$$

Informações adicionais podem ser obtidas através da distribuição de graus, P_k , que representa a probabilidade de um vértice escolhido aleatoriamente ter grau k (ALBERT; BARABÁSI, 2002). Para uma rede com N nós, a distribuição de graus, P_k , é dado pela equação 4.4, onde N_k é um número de nós com grau k (BARABÁSI, 2016). A figura 12 contém a distribuição de grau de um grafo com 4 nós.

$$P_k = \frac{N_k}{N}. \quad (4.4)$$

Figura 12 – Figura extraída e modificada de (ALBERT; BARABÁSI, 2002) que contém a distribuição de grau de um grafo. Em (A) o grafo com 4 nós; os nós 4 e 3 (em verde) possuem grau igual a dois, o nó 2 (em azul) possui grau igual a 3 e o nó 1 (em vermelho) possui grau igual a 1. Em (B), contém a distribuição de grau, (P_k), do grafo em (A); em vermelho com um quantidade de 0.25, representa a quantidade de nós (no caso somente o nó 1) que apresenta grau igual a 1; em verde a quantidade de nós (no caso os nós 3 e 4, representando a metade de nós presente na rede) que possuem grau igual a 2; e, por fim, em azul a quantidade de nós (no caso somente o nó 2) com grau igual a 3.

Outra medida, relacionada ao grau é a complexidade, definida como a divisão do segundo momento da distribuição do grau pela média desta; essa medida mede o quão conectada é a rede. Além disso, pode-se calcular a entropia (medida de incerteza de uma variável aleatória) da distribuição do grau, H , dada pela equação 4.5.

$$H = -\sum_i (P_k) \log(P_k). \quad (4.5)$$

Uma das mais importantes propriedades de redes é a distribuição de graus seguir as leis da potência, e uma forma eficiente de quantificá-la é através da medida de "retidão", que pode ser determinada pelo cálculo do coeficiente de Pearson dos pontos da distribuição de graus na escala logarítmica. Para coordenadas x e y de um conjunto de N pontos, o coeficiente de Pearson, r_{xy} que pode ser calculado pela expressão 4.6.

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sum_{i=1}^N (x_i - \langle x \rangle)^2 \sum_{i=1}^N (y_i - \langle y \rangle)^2}, \quad (4.6)$$

onde x_i e y_i são valores do i -ésimo elemento dos vetores x e y , de tamanho n , e $\langle x \rangle$ e $\langle y \rangle$ representa a respectiva média de x e y . Para calcular, basta substituir x pelo logaritmo da conectividade k e y pelo logaritmo da probabilidade $P(k)$. O valor de r_{xy} pode variar entre -1 e 1. O caso -1 indica a presença de uma rede com distribuição livre de escala, pois o coeficiente

de Pearson tem valor mínimo quando duas variáveis são completamente anti-correlacionadas (BOAS, 2008).

Outra forma de utilização deste coeficiente é o cálculo do coeficiente de correlação da conectividade (graus) de ambos os lados da aresta (NEWMAN, 2003). Este cálculo recebe o nome de coeficiente de assortatividade, que varia de -1 a 1; para valores positivos a rede é dita assortativa, para valores negativos, dissortativa e para o valor igual a zero não há correlação (NEWMAN, 2002). Em redes assortativas, os *hubs* tendem a se conectar com outros *hubs* (resultando em um maior grau médio de seus vizinhos mais próximos); enquanto que em redes dissortativas, os *hubs* tendem a se ligar a nós com baixos graus (BARABÁSI, 2016).

Outra medida relacionada ao grau é a conectividade média entre os vizinhos, $k_{nn}(i)$, em que para um nó i (a medida varia para cada vértice presente na rede), ela pode ser calculada pela equação 4.7.

$$k_{nn}(i) = \frac{1}{k_i} \sum_j^N a_{ij} k_j. \quad (4.7)$$

4.0.2 Medidas relacionadas a ciclos

Muitas redes reais apresentam ciclos formados por três vértices cuja medição é feita pelo coeficiente de aglomeração. Essa medida foi introduzida em (WATTS; STROGATZ, 1998) no contexto de redes sociais (SCHANK; WAGNER, 2004) e pode ser interpretada como a probabilidade de dois vizinhos de um vértice estar conectados. Uma das maneiras de calcular essa medida é pela fórmula da transitividade (introduzida em (NEWMAN; WATTS; STROGATZ, 2002)) dada pela equação 4.8.

$$C = \frac{3N_{\Delta}}{3N_3}, \quad (4.8)$$

onde $3N_{\Delta}$ representa o número de triângulos presentes nas redes e N_3 é o número de triplas conectadas. Vale a pena observar que: $0 < C < 1$. A figura 13 contém três redes com diferentes coeficientes de aglomeração.

Além disso, existe o coeficiente de aglomeração local dado pela expressão 4.9, que calcula essa medida para cada vértice i .

$$C(i) = \frac{3N_{\Delta}(i)}{3N_3(i)}, \quad (4.9)$$

onde $3N_{\Delta}(i)$ é a quantidade de triângulos formados envolvendo o vértice i e $3N_3(i)$ é o número de trios conectados que possuem o nó i como vértice central. Por meio do coeficiente de aglomeração local pode-se calcular o coeficiente de aglomeração local médio, cuja medida de uma rede não dirigida e sem peso é dada por 4.10 .

$$C = \frac{1}{N} \sum_{i=1}^N C(i). \quad (4.10)$$

Figura 13 – Figura extraída de (RODRIGUES, 2007) , contendo três redes com diferentes coeficientes de aglomeração. Em (a) todos os vértices estão conectados representando o valor máximo da medida, $C=1$. Em (b) o coeficiente tem valor de $C = \frac{3}{10}$ e por fim, em (c), como não há a presença de nenhum triângulo $C=0$.

Outra medida, ainda, relacionada ao grau, é a complexidade, que consiste na divisão do seu segundo momento da distribuição de grau pela média da distribuição de grau.

4.0.3 Medidas relacionadas à distância

A maior distância entre qualquer par de vértices de uma rede é chamado diâmetro da rede. Por outro lado, o menor caminho (conhecido, também, como caminho geodésico) entre dois vértices i e j , d_{ij} , é aquele cujo comprimento é o menor de todos os caminhos possíveis entre esses vértices. Isso pode ser representado por uma matriz de distância D , cujos elementos, d_{ij} , expressam o valor do menor caminho entre esses vértices. O valor $d_{max_{ij}} = \max_{i,j}(d_{ij})$ é chamado diâmetro da rede. A média entre os valores dos elementos matriciais é denominada de *caminho característico* da rede, menor caminho médio da rede, e pode ser calculado pela equação 4.11.

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}. \quad (4.11)$$

Essa medida diverge quando há mais de um componente conexo. Assim como intuito de eliminar tal limitação da medida de L deve-se utilizar uma medida denominada de *eficiência da rede*, E , cujo valor pode ser obtido por 4.12.

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (4.12)$$

Vale observar que E difere de L , por considerar o inverso da média harmônica ao invés da média aritmética. Esta medida ainda é um indicador da capacidade de tráfego na rede (ARRUDA, 2013). Estas medidas estão diretamente relacionadas com transporte e comunicação na rede (COSTA *et al.*, 2007) (quanto menor o caminho, mais rápido é o transporte de informação pela rede, e, com isso, maior a medida de eficiência).

4.0.4 Medidas de centralidade

A partir das distâncias, ainda, é possível quantificar o quão central e importante é um vértice em uma rede (ARRUDA, 2013). Uma dessas medidas que indicam centralidade é o *betweenness centrality* (FREEMAN, 1977), ou, *centralidade de intermediação*; esta medida de um vértice i é definida como a fração de menores caminhos de pares de vértices que passam por ele (NEWMAN, 2005). Este mesmo autor, ainda, define matematicamente a medida *betweenness centrality*, b_i , como na equação 4.13.

$$b_i = \frac{\sum_{s < t} g_i^{(st)} / n_{st}}{(1/2)n(n-1)}, \quad (4.13)$$

onde $g_i^{(st)}$, é o número de caminhos geodésicos do vértice s ao vértice t passando por i ; n_{st} , o número total de caminhos geodésicos de s a t ; e n o número total de vértices da rede. Uma medida global relacionada ao grau de intermediação é a *dominância do ponto central*, CDP , definida pela equação 4.14, que varia de entre 0 (redes totalmente conectadas) e 1 (para redes em que existe um vértice central por onde todos os caminhos passam) (BOAS, 2008).

$$CDP = \frac{1}{N-1} \sum_i (B_{max} - B_i), \quad (4.14)$$

em que B_{max} é o maior valor do grau de intermediação da rede, e, B_i , é o *betweenness centrality* de determinado vértice i e n é o número total de caminhos geodésicos.

Outra medida, ainda de centralidade, é o *closeness centrality*, Cl , também denominada centralidade de proximidade, que em (FREEMAN, 1978) é descrito matematicamente como na equação .

$$Cl(i) = \frac{n-1}{\sum_{j=1}^n d_{ij}}, \quad (4.15)$$

onde d_{ij} é a distância de i a j , sendo a soma é baseada na distância de i a todos os $(n-1)$ pontos. A medida *closeness centrality*, $C(i)$, pode ser entendida como o inverso da média da distância de i a todos os outros pontos.

Outra medida de centralidade é o *eigenvector centrality* (BONACICH, 1987), ou centralidade de autovetor, que utiliza o conceito de que a importância de um vértice é aumentada por ter conexões com outros vértices; assim, ao invés de pontuar um ponto para cada vizinho, a centralidade de autovetor dá a cada vértice uma pontuação proporcional a soma das pontuações de seus vizinhos (NEWMAN, 2010). Matematicamente, a centralidade de autovetor de um vértice i , x_i , é proporcional a soma das centralidades de seus vizinhos, como na equação 4.16.

$$x_i = k_1^{-1} \sum_j A_{ij} x_j, \quad (4.16)$$

onde A_{ij} são os elementos da matriz de adjacência A e k_1 é o seu maior autovalor.

A medida *pageRank* foi proposta em (PAGE *et al.*, 1999), como medida de importância em páginas da internet, através do cálculo de um *ranking* para cada página baseada em um grafo da *world wide web*, sendo muito utilizado pela empresa *Google* (BRIN; PAGE, 2012). Matematicamente, o PageRank, $P_R(i)$, de um vértice i é dado pela equação 4.17.

$$P_R(i) = \frac{q}{N} + (1 - q) \sum_j A_{ij} \frac{P_R(j)}{k_{out,J}}, \quad (4.17)$$

onde A_{ij} são os elementos da matriz da adjacência; $k_{out,J}$, é o grau da saída do vértice J ; N , o número de páginas da internet; e, q , o *damping factor*, que indica a probabilidade do usuário parar a navegação seguindo os *links* da página que ele navegando e começar uma nova navegação em outra página inserindo seu endereço (CESTARI, 2017).

4.0.5 K-core

A estrutura de grafos muito grandes, muitas vezes difíceis de serem manipulados, são segmentados em partes menores para facilitar o estudo da rede (ARRUDA, 2013). Uma das medidas de segmentação hierárquica é o *k-core*.

O *k-core* é definido como o subconjunto máximo de vértice tais que cada um está conectado com ao menos k outros no subconjunto (NEWMAN, 2010). Proposto em (SEIDMAN, 1983), como: "Seja G um grafo. Se H é um subgrafo de G , e denotando $\delta(H)$ como o grau mínimo de H ; cada ponto de H é, então, adjacente no mínimo $\delta(H)$ outros pontos de H . Se H é o subgrafo de conexão máxima (induzida) de G , com $\delta(H) \geq k$, H é o *k-core* de G ".

Um algoritmo simples utilizado para obter essa medida foi proposto em (BATAGELJ; ZAVERSNIK, 2003), seguindo a seguinte propriedade: "Se dado um grafo $G(V, L)$, em que V é o número de vértices e L o número de arestas, recursivamente for deletado todos os vértices e arestas incidentes deles, de grau menores que k , então o grafo restante é o *k-core*".

4.0.6 Tipos de redes e outros conceitos

As redes complexas podem ser classificadas considerando suas propriedades estatísticas, principalmente distribuição de graus e o coeficiente de aglomeração, existindo vários modelos propostos na literatura (ARRUDA, 2013). Dentre estes, destacam-se: as redes regulares, redes aleatórias, redes livres de escala, redes de pequeno mundo, redes hierárquicas e modulares (AMANCIO; NUNES, 2009). A seguir, foram descritas brevemente todas as redes, focando-se nas redes de pequeno mundo, as quais foram utilizado no presente trabalho.

Em redes regulares, todos os vértices apresentam o mesmo grau. As redes aleatórias (propostas em (RÉNYI; ERDOS, 1959), por Erdős e Rényi (ER), sendo, assim conhecidas como grafos aleatórios de Erdős e Rényi), em que é gerado randomicamente grafos com N nós e k arestas (BOCCALETTI *et al.*, 2006), sendo que a probabilidade de que um nó se conectar a outro é mesma para todos vértices. A distribuição de nós encontrada para esta rede, P_k , é

aproximadamente uma distribuição de Poisson, dada pela equação 4.18, em que $\langle k \rangle$ é a média do grau fixada como $2m/n$, sendo m o número de nós e n o número de arestas.

$$P(K) = e^{-\langle k \rangle} \frac{\langle k \rangle^K}{K!}. \quad (4.18)$$

Na geração de uma rede livre de escala (um exemplo são as redes da *world wide web*), a distribuição de graus, P_k , segue a distribuição de grau dado pela equação 4.19, sendo que em sistemas biológicos, o λ varia de 2 a 3.

$$P_K \approx K^{-\lambda}. \quad (4.19)$$

A característica mais importante de uma rede hierárquica é a relação de lei de potência entre o coeficiente de aglomeração de um vértice e seu grau, sendo que implica em que vértices distantes são partes de áreas de alta aglomeração, denominadas de módulos, e a comunicação entre eles é feita por poucos vértices.

As redes denominadas pequeno mundo, em inglês *small-world*, foram propostas por Watts e Strogatz (muitas vezes conhecida como modelo de Watts e Strogatz, sendo sistemas com alto agrupamento, como redes regulares, e com pequenos caminhos característicos, como redes aleatórias (WATTS; STROGATZ, 1998). Segundos os autores a construção é feita iniciando com uma rede regular de N vértices e k arestas por vértice, sendo que cada novo vértice é inserido de maneira randômicas, com uma probabilidade p , tal que p se encontra entre 0 (rede totalmente regular) e 1 (rede totalmente aleatória).

Segundo (NEWMAN; BARABASI; WATTS, 2011), o modelo de pequeno mundo foi baseado na observação de muitas redes reais (como as redes neurais do verme *Caenorhabditis elegans*) que mostraram seguir duas propriedades:

- O efeito mundo-pequeno em que os vértices estão conectados por um curto caminho através das rede e a distância vértice-vértice aumenta logaritmicamente com o número total de vértices na rede.
- Alto agrupamento ou transitividade, ou seja, há alta probabilidade de que dois vértices que possuam vizinhos em comum estejam conectados.

Outro conceito, muito presente em redes cerebrais, são os *motifs*, que são definidos como padrões de interconexões em redes complexas em um número maior que aqueles em redes aleatórias (MILO *et al.*, 2002).

CONECTIVIDADE CEREBRAL E A TÉCNICA DE RESSONÂNCIA MAGNÉTICA FUNCIONAL

5.1 Introdução

A moderna ciência de redes complexas tem desempenhado um importante papel no entendimento de estruturas e funções de sistemas complexos (STAM, 2014). Por outro lado, sabe-se, desde o século XIX, que os elementos neuronais do cérebro formam uma rede estrutural muito complexa; e ainda, desde o século XX, que seu substrato anatômico suporta a dinâmica da atividade fisiológica (BULLMORE; SPORNS, 2009). Dessa forma é natural pensarmos na aplicação de redes complexas na neurociência (FORNITO; ZALESKY; BULLMORE, 2016).

A primeira aplicação de redes foi feita por Watts e Strogatz (WATTS; STROGATZ, 1998) para a análise do sistema nervoso do verme nematoide *C.elegans* (que até então era o único exemplo de uma rede neural mapeada), onde os nós eram representados pelos neurônios e a sinapse entre eles pela aresta. Os autores demonstraram que a rede em questão possuía um pequeno valor para a medida de menor caminho e um alto para coeficiente de aglomeração, sendo caracterizada topologicamente como rede “pequeno mundo”.

Este estudo foi o primeiro indício de que redes cerebrais compartilham algumas propriedades com outras redes complexas (como as redes sociais e as de recursos naturais) (FORNITO; ZALESKY; BULLMORE, 2016), levando a outras investigações mais sistemáticas da estrutura e conectividade do sistema nervoso deste mesmo nematoide (WHITE *et al.*, 1986), bem como a translação dos mesmos conceitos de redes a dados em larga escala¹ do córtex cerebral de gatos

¹ Conjuntos de dados de conectividade em grande escala correspondem a matrizes de conexão onde os grafos consistem de N áreas dos cérebros (vértices) ligado a K caminhos de conexão (arestas); isso será mais detalhadamente explicado na seções seguintes.

(SCANNELL *et al.*, 1999) e macacos (HILGETAG *et al.*, 2000), disponíveis até então. A análise desses dados demonstraram também se tratar de redes de pequeno mundo.

Redes deste tipo combinam características estruturais completamente aleatórias e topologias de conexões regulares. Preservam, ainda, um alto grau de conectividade entre vizinhanças locais, enquanto permitem que todos os seus nós estejam ligados a caminhos surpreendentemente curtos (SPORNS; ZWI, 2004).

A maioria destas informações sobre a conectividade anatômica foi obtida por técnicas invasivas em não-humanos e cérebros humanos *postmortem* (DIJK *et al.*, 2010). Métodos de imagens neurais (como a técnica de ressonância magnética funcional) têm oferecido rotinas experimentais mais promissoras para o entendimento das conexões cerebrais humanas (SPORNS; TONONI; KÖTTER, 2005) e, ainda, avançaram o campo da neurociência, evidenciando mudanças estruturais e funcionais no cérebro *in vivo* (CASEY *et al.*, 2005). No entanto, apesar desse impressionante aumento do conhecimento em neurociência, há poucas teorias disponíveis para explicar processos cerebrais (STAM; REIJNEVELD, 2007) e modelos estatísticos capazes de explicá-los de forma exata.

Esta seção tem como objetivo descrever a organização cerebral, a aplicação de redes complexas no cérebro e introduzir alguns conceitos relacionados. Isso será importante para o entendimento dos capítulos seguintes, onde serão analisados conjuntos de dados contendo a conectividade cerebral de pacientes que possuem determinados transtornos mentais.

5.2 Diferentes níveis de organização do cérebro

Como já visto no capítulo 4, redes são representações matemáticas de sistemas complexos reais e definidas por um conjunto de vértices cujas conexões são feitas por arestas. No caso do sistema nervoso é muito difícil determinar quais seriam suas representações em nós e arestas uma vez que o cérebro humano pode ser visto em diferentes escalas de organização (BULLMORE; SPORNS, 2009). Nos humanos, a anatomia neural é organizada em aproximadamente seis níveis de organização, indo do nível dos neurônios e sinapses ($<1\mu m$, microescala) à larga escala onde há distinção anatômica entre as regiões cerebrais e caminhos inter-regionais (≈ 10 cm, macro escala). Entre esses dois níveis, há ainda o nível de grupos neuronais ou grupos deles (≈ 1 cm, meso escala). A figura 14 ilustra os principais níveis de organização descritos. A dinâmica que ocorre nessas estruturas desenvolve-se em escalas temporais parecidas, abrangendo frequências entre 0.05 e 500 Hz.

Redes de larga escala (nível macroscópico) obtidas de imagens neurais *in vivo*, podem ter seus nós conectados por meio de três tipos de conexão: conectividade estrutural para ligações anatômicas, conectividade funcional e conectividade efetiva. Essas três categorias de conexões não se restringem apenas à escala macroscópica, mas também se aplicam a meso escala e a microscópica, apesar de ser mais comum análises de interações estruturais e funcionais entre

Figura 14 – Figura extraída e modificada de (SPORNS; TONONI; KÖTTER, 2005), contendo os três principais níveis de organização do cérebro (microescala, meso escala e macro escala). Em (A), o nível macroscópico em que o cérebro pode ser dividido através de regiões anatômicas como os lobos corticais. Em (B), a meso escala, em que os neurônios se agregam em colunas, camadas e agrupamentos celulares. Em (C), a escala microscópica, onde a imagem extraída através de um microscópio de varredura mostra estruturas celulares como a vesícula sináptica.

populações de neurônios em larga escala, como aquelas caracterizadas por ressonância magnética funcional (RMF, em inglês *Functional Magnetic Resonance Imaging*, fMRI) (PARK; FRISTON, 2013).

A conectividade funcional de um sistema neuronal corresponde à correlação temporal ou desvios de independência estatística dos padrões que existem entre unidades neurais. Tais correlações temporais são, normalmente, resultados de interações neurais dinâmicas ao longo de conexões estruturais ou anatômicas (SPORNS, 2002); além disso, essas dependências estatísticas não necessariamente estão no domínio do tempo, mas também podem estar no domínio de frequência e no domínio tempo-frequência (FORNITO; ZALESKY; BULLMORE, 2016). Por outro lado, os desvios de independência estatística entre elemento neuronais são frequentemente capturados pela matriz de covariância (ou a matriz de correlação) que representa a conectividade funcional do sistema. A completa descrição estatística desse sistema neural se dá pela junção da função de distribuição de probabilidade do sistema de variáveis, com efeitos de segunda ordem (interações) contidas na matriz de covariância (SPORNS; TONONI, 2001). Em termos de medição, esse tipo de conectividade pode ser inferido com base na ressonância magnética funcional dependente do nível de oxigenação do sangue (em inglês *blood oxygenation level-dependent*, BOLD) ou na coerência nos sinais de eletro ou magneto encefalograma (EEG / MEG) adquiridos durante a realização de uma tarefa de interesse ou em repouso (PARK; FRISTON, 2013).

Segundo (FRISTON, 1994): "A conectividade efetiva está mais próxima da noção intuitiva de uma conexão e pode ser definida como a influência que um sistema neural exerce sobre outro, seja em nível sináptico ou cortical". Dessa forma, a conectividade efetiva descreve

todos os efeitos causais entre elementos neurais e pode ser medida por meio de análise de séries temporais, modelagem estatística ou perturbação experimentais (SPORNS, 2010). Um exemplo de medida utilizada é a casualidade de Granger, baseada em medidas temporais que parte do pressuposto que para duas séries temporais medidas simultaneamente, uma delas pode ser a causa da outra se pode ser prevista utilizando-se um conhecimento prévio da primeira (CAMARGO, 2003). Assim como a conectividade funcional, a conectividade efetiva é dependente do tempo e é modulada por dois tipos de abordagens: as de estímulos externos ou mudanças do estado interno. Algumas dessas abordagens (como as medidas pela casualidade de Granger) derivam de interações da precedência temporal, não dependendo do modelo causal e assim são denominadas de "modelo livre" (do inglês, *model free*) (BULLMORE; SPORNS, 2009); já outras, dependem do modelo causal explícito, incluindo seus parâmetros estruturais (caminhos anatômicos).

A representação de todas as redes se dá por suas matrizes de conectividade (adjacência) que será explicada na subseção seguinte.

5.3 Redes funcionais

A construção de redes funcionais ou estruturais cerebrais deve ser feita por meio de alguns passos que basicamente podem ser resumidos em três. O primeiro é a definição dos nós da rede, bem como estimar a medida de associação entre pares deles. O segundo passo é a compilação dessas associações estimadas em uma matriz, denominada matriz de conectividade ou adjacência (o nome adjacente advém do fato de que em teoria de grafos, nós unidos por uma aresta são denominados adjacentes ou vizinhos). Caso a matriz em questão seja esparsa, ou seja, contenha vários valores iguais a zero fora da diagonal principal, podem-se remover as ligações mais fracas com intuito de se examinar as associações mais fortes. E o passo final é o cálculo das medidas do grafo (SPORNS, 2010). Vale lembrar que para cada um desses passos é necessário fazer escolhas dependendo do tipo de dado analisado.

No caso do presente trabalho, pelo fato dos dados terem sido extraídos pela técnica de ressonância magnética funcional, na seção seguinte, a seguir, descreveremos essa técnica e como é feita a construção de redes baseadas nela.

5.4 Técnica de imagem por ressonância magnética funcional

A ressonância magnética nuclear é observada em isótopos que não contenham spin (movimento de rotação de núcleos carregados em torno de seu eixo dando origem ao momento magnético nuclear (SLICHTER, 2013)) iguais a zero. Na ausência de campo magnético, os spins se encontram em estado degenerado (sem separação de nível de energia). Contudo havendo um campo magnético (H_0), os spins se alinham na mesma direção dele e o momento magnético nuclear adquire um movimento de precessão ao seu redor. Esse alinhamento é denominado

magnetização e a quebra da degenerescência resulta em dois níveis de energia (níveis de Zeeman, α e β) (GIL; GERALDES, 1987). A diferença de energia presente entre esses dois níveis (ΔE) é proporcional ao campo H_0 e à frequência de precessão, ω_0 . Quando é aplicado um pulso de radiofrequência (com direção perpendicular a H_0), os spins no estado inferior são excitados para o nível de energia superior, enquanto os spins com nível energético superior são estimulados a relaxar para o nível inferior libertando sua energia; como existem mais spins em um nível energético inferior ocorre absorção líquida de energia no sistema (BROWN; SEMELKA; NISHINO, 2004). Assim prótons absorvem parte da energia correspondente à frequência de precessão ω_0 (uma vez que a diferença de energia de seus níveis é proporcional a esta), energia denominada de absorção de ressonância. Com o fim do estímulo, os spins voltam ao seu estado e os prótons libertam a energia absorvida (fenômeno de relaxação) que será detectado por bobinas. Assim durante o exame de RMN, deve-se conhecer a frequência exata de absorção do tecido exposto. (GAMEIRO, 2004). A figura 15 contém um esquema do fenômeno de RMN.

A técnica de imagem por ressonância magnética funcional (fMRI, do inglês Functional Magnetic Resonance Imaging) utiliza a ressonância magnética nuclear para avaliar as mudanças de metabolismo ou de função cerebral em um certo intervalo de tempo. O cérebro não possui uma reserva energética alta, necessitando de uma constante suplementação sanguínea de glicose e oxigênio. A molécula de hemoglobina (responsável pelo transporte de oxigênio no sangue) ao se ligar com o oxigênio forma a oxihemoglobina e, ao perdê-la, forma a deoxihemoglobina, a qual possui propriedades paramagnéticas. Em outras palavras, quando a hemoglobina perde a molécula de oxigênio (se transforma em deoxihemoglobina) (FIGUEIRA, 2013), os seus spins em um campo magnético se tornam degenerados e assim pode-se utilizar a técnica de ressonância magnética descrita anteriormente e esquematizada na figura 15. Assim a diminuição da taxa de deoxihemoglobina pode ser detectada com o aumento do sinal de RMN, este efeito é denominado de BOLD (do inglês, blood-oxygen-level dependent). A evolução temporal da série de BOLD é denominada função hemodinâmica de resposta e pode ser determinada pela intensidade do pixel nas imagens de fMRI (STURZBECHER, 2006). Em (BISWAL *et al.*, 1995), a função de conectividade foi definida como: “correlação temporal entre eventos neurofisiológicos em diferentes áreas cerebrais”, sendo essa dependência medidas pelo sinal de BOLD.

O valor de cada série temporal registrada nos voxels nas imagens de ressonância magnética funcional é medido em escala de cinza, sendo que cada voxel mapeia anatomicamente uma posição no cérebro. Assim é obtido um conjunto de séries temporais a serem correlacionadas (LOPES, 2016). Essas séries temporais, antes de serem correlacionadas passam por etapas de pré-processamento de dados com intuito de remover os dados que não interessam ao estudo e prepara-los para a análise estatística; dessas etapas destaca-se a filtragem temporal e normalização (PAMPLONA, 2014). Na filtragem temporal é feita a remoção, pelo filtro passa-banda, de frequências não consistentes com redes funcionais (na faixa de 0.01-0.08), correspondentes na maioria das vezes a ruídos cardíaco ou respiratório (WHITFIELD-GABRIELI; NIETO-CASTANON, 2012). Para algumas bases, ainda, com intuito de reduzir a dimensionalidade do

Figura 15 – Figura adaptada de (SILVA, 2007; GAMEIRO, 2004) contendo um esquema do processo de RMN. Em (A), contém os spins (setas azuis) no estado degenerado. Com incidência do campo H_0 (em roxo), ocorre a quebra da degenerescência, gerando dois níveis energéticos (níveis de Zeeman), α e β , representado na figura em (B); o alinhamento dos spins é denominado magnetização (em verde na figura). Em (C), há a representação dos dois níveis energéticos, sendo a diferença entre eles, ΔE , proporcional ao campo e a frequência de precessão. Ao incidir um pulso de radiofrequência, (D), os spins de nível mais baixo são excitados ao nível de maior energia (na figura spin verde); enquanto os spins de nível mais alto são relaxados ao nível de energia mais baixo (na figura spin laranja). Como há maior quantidade de spins de energia mais baixa ocorre absorção de energia no próton. Após cessar o pulso, os spins voltam a sua organização anterior e assim é liberada a mesma quantidade de energia absorvida, sendo esta detectada por bobinas.

problema, projeções dos dados de fMRI são construídas em um espaço de menor dimensão utilizando a transformada de *wavelet* (isso ocorreu na base de dados relacionada a esquizofrenia).

O próximo passo é a análise estatística das séries temporais obtidas para determinar como elas estão relacionadas, determinando assim a conectividade cerebral. Essa conectividade pode ser de séries temporais provenientes de uma região de interesse (ROI, em inglês *region of interest*), de pelo menos dois voxels distintos, ou estrutura cerebral a serem analisadas; no caso trata-se dos nós presentes na rede. Existem umas séries de métodos para quantificar a correlação das séries temporais obtidas, que podem ser basicamente divididos em dois grupos: modelo dependente (em inglês, *model-dependent*) e modelo livre (em inglês, *model-free*). Os modelos dependentes são baseados no método da semente, e os métodos de modelo livre incluem: análise da componente principal (em inglês principal component analysis, PCA), decomposição do valor singular (em inglês singular value decomposition, SVD), análise de componentes independentes (em inglês independent component analysis, ICA) e algoritmos de agrupamento (em inglês clustering) (FIGUEIRA, 2013). Os métodos de modelo livres são denominados dessa forma por não dependerem da especificação do modelo temporal e possuem um custo computacional menor em relação aqueles modelos baseados em sementes.

Desses dois tipos de métodos, os mais utilizados na literatura (HEUVEL; POL, 2010; MAGALHÃES, 2013) é a análise da correlação baseada em sementes (em inglês, Seed-based Correlation Analysis, SCA), que é modelo dependente; e a análise da componente principal, modelo livre. No método da semente é primeiramente definido uma região de interesse (sendo essa ROI denominada semente, na figura 16 foi chamada de *voxel seed*) nas aquisições de fMRI com alguma tarefa de ativação envolvida (na figura 16, utilizou-se, por exemplo, a movimentação dos dedos). Assim para determinar se outra região *j* (na figura 16 *voxel j*) está correlacionada com a semente, utiliza-se medidas de correlação, sendo a mais usada a de Pearson. Se a medida de correlação obtida foi alta, significa que essas duas regiões possuem alta conectividade. Seguindo esse procedimento para outras regiões do cérebro resulta em um mapa de conectividade cerebral. Neste mapa é feito outros procedimentos estatísticos (como a transformada de fisher para que os coeficiente de correlação envolvidos possuam uma distribuição normal) que resultarão na matriz de conectividade. A figura 16 contém um esquema resumindo esse método da semente.

Enquanto o método baseado em semente utiliza séries temporais, a análise de componentes independente pode ser usada temporalmente e espacialmente (CALHOUN *et al.*, 2001). Ela é uma abordagem multivariada pois considera todos os dados provenientes de cada voxel para encontrar componentes descritos como um mapa espacial (reflete onde a porção espacial está sendo detectada) e série temporal (descrevendo como o sinal evoluiu ao longo do tempo) (BIJSTERBOSCH; SMITH; BECKMANN, 2017). O modelo de ICA é um modelo linear sendo que os dados originais podem ser representados como a soma de todas essas componentes, consideradas estatisticamente independentes. A figura 17 contém um esquema do processo em questão. São procurados sinais, então, que maximizam a independência estatística de cada um.

Figura 16 – Figura adaptada de (PAMPLONA, 2014; HEUVEL; POL, 2010) contendo um esquema do método da semente. Primeiramente identifica-se uma região específica de interesse, por exemplo as conexões do córtex motor primário esquerdo (A). Em seguida, os pacientes analisados são colocados no scanner em estado de repouso e é pedido para fechar os olhos e não pensar em nada, sem adormecer; com isso obtém-se as séries temporais do estado de repouso, (B). É pedido então que os voluntários movam a mão direita selecionando os voxels mais ativados do córtex motor, esse será o voxel semente; obtém-se, então, comparando com as séries temporais em repouso, a série temporal de repouso do voxel semente em repouso, (C). Assim com intuito de examinar conectividade funcional entre o voxel semente selecionado a uma segunda região cerebral j (por exemplo, uma região no córtex motor contralateral), é medido, por meio de medidas de correlação as séries temporais no estado de repouso dessa duas regiões, (D). Uma alta correlação indica uma alta conectividade entre o voxel semente e o voxel j . Por fim, (E), é feita a correlação do voxel semente com todos os outros voxels obtendo-se o mapa de conectividade cerebral.

Figura 17 – Figura extraída e modificada de (BECKMANN, 2012; BIJSTERBOSCH; SMITH; BECKMANN, 2017). Em (A) os dados de fMRI contêm os dados BOLD sendo que cada linha representa um volume tridimensional em um certo tempo e cada coluna representa todos os tempos de um voxel. Em (B), o conjunto de dados é então decomposto em duas matrizes pelo método de ICA, sendo que a matriz (I) contém a série temporal (coluna) de cada componente obtida e a matriz (II) contém o mapa espacial (linha) de cada componente. Para cada série temporal existe uma mapa espacial correspondente, assim o número de linhas do mapa espacial é o mesmo que o número de colunas da série temporal e representa o número de componentes (dimensionalidade do ICA).

Apesar do resultado ser mais difícil de interpretar que o mapa de conectividade, é a forma mais eficiente de procurar grupos (FIGUEIRA, 2013). Pode-se, ainda utilizar os dados obtidos no método de semente.

A figura 18 representa o resumo da construção de matrizes de conectividade a partir de dados de fMRI. Na sessão seguinte se encontra os tipos de arestas armazenadas na matriz de adjacência e o processo de *threshold* para torná-las binárias.

5.5 Aplicação de *threshold* e binarização em matrizes de conectividade

A conectividade entre os nós da seção anterior é representada por arestas em grafos. As arestas podem ser distinguidas com base em seu peso e direção (RUBINOV; SPORNS, 2010).

Arestas podem ser diferenciadas pela ausência ou presença de direção; grafos não direcionados possuem matrizes simétricas (SPORNS *et al.*, 2004). Conectividades funcionais dão origem a grafos não direcionados enquanto as efetivas dão origem a grafos direcionados. O peso em redes funcionais representa a magnitude das interações, ele pode variar de -1 (correlação negativa perfeita ou anti-correlação) a 1 (correlação perfeita) (FORNITO; ZALESKY; BULLMORE, 2016). Redes sem peso são denominadas de binárias, apresentando, assim, somente valores 0 e 1; sendo que 1 indica regiões conectadas e 0 sem conexões. A figura 19 contém dois

Figura 18 – Figura extraída e modificada de (BULLMORE; SPORNS, 2009; LYNALL *et al.*, 2010) contendo o esquema da obtenção de um grafo a partir de dados de fMRI. Em (A) foram determinadas as regiões de interesses que serão os nós da rede. Por meio do scanner de fMRI são determinadas as séries temporais da região de interesse. Nessas séries são aplicados métodos de pré-processamento (por exemplo filtragem de séries temporais) com objetivo de eliminar ruídos de dados de fMRI. Com essas séries pré-processadas, é determinada a correlação entre elas que é feita utilizando algum dos métodos descritos nesta secção. Com a correlação das séries temporais obtém-se a matriz de conectividade em (C). A partir da matriz de conectividade é construído o grafo (o grafo apresentado na figura foi obtido da matriz de conectividade de um dos pacientes com esquizofrenia dos dados do capítulo 6

exemplos demonstrando os grafos dos dois tipos de conectividade.

Muitos estudos têm descartado arestas com pesos, transformando-as em binárias, uma vez que redes sem peso são mais simples de caracterizar e há muito mais medidas de redes complexas para redes sem peso; a remoção de pesos ainda elimina arestas fracas, sem significância, que podem ser vistas como ruídos em redes funcionais e efetivas (RUBINOV; SPORNS, 2010). A transformação de redes com peso para redes binárias pode ser feito por dois processos mais comuns, o de *thresholding* e o de binarização. Quando estes dois processos são aplicados juntos a uma matriz com peso, o resultado é uma matriz binária, sem peso.

O processo de *thresholding* consiste na escolha de um valor de corte ou *threshold*, τ , tais que as conexões, C_{ij} são mantidas tais como a equação 5.1.

$$A_{ij} = \begin{cases} C_{ij} & \text{se } C_{ij} > \tau, \\ 0 & \text{caso contrário} \end{cases} \quad (5.1)$$

Em seguida, é feita a etapa de binarização em que os elementos mantidos se tornam binários por meio de 5.2.

$$A_{ij} = \begin{cases} 1 & \text{se } C_{ij} > \tau, \\ 0 & \text{caso contrário} \end{cases} \quad (5.2)$$

Figura 19 – Figura extraída e modificada de (RUBINOV; SPORNS, 2010; PARK; FRISTON, 2013; FIGUEIRA, 2013) contendo os tipos de arestas possíveis. Em (A) há dois tipos de conexão unindo três regiões do cérebro (1, 2, 3, nas cores azul, laranja e vermelho, respectivamente). As conexões em roxo pertencem a conectividade funcional e não são direcionadas, sendo a conexão (I) com o peso maior (linha mais grossa), indicando uma rede com peso. As conexões em azul são direcionadas (indicadas por flechas) pertencem a conectividade efetiva; como a aresta (IV) tem mais peso que as demais (mais grossa), trata-se de uma rede com peso. Em (B), há outros dois exemplos de rede direcionada e não direcionada, percebe-se que a matriz de conectividade desta é simétrica. As duas redes possuem pesos indicado pela escala acinzentada, sendo a preta a de maior conexão (valor igual a 1) e a branca a de menor (valor zero).

O processo de escolher o valor de τ para os dois métodos é tipicamente arbitrário e requer a exploração de vários valores e comparação entre eles (RUBINOV; SPORNS, 2011). A figura 20 contém a matriz de conectividade obtida a partir da fMRI em humanos, após ser submetida ao processo de *thresholding* e binarização.

5.6 O conceito de integração e segregação e a organização complexa do cérebro

No cérebro de invertebrados há a presença de dois conceitos contrastantes: o de segregação funcional (ou especialização) e a integração (ou processos distribuídos) (TONONI; SPORNS; EDELMAN, 1994).

A segregação anatômica e funcional refere-se a existência de neurônios especializados e as áreas cerebrais organizados em módulos (SPORNS, 2002). Esses módulos correspondem a comunidades em que seus membros possuem alta conectividade entre eles e poucas conexões com membros de outro módulos (SPORNS, 2013). Algumas medidas utilizadas para quantificar essa segregação é o coeficiente de agrupamento e a transitividade, que refletem, em média, na prevalência de grupos (em inglês, *cluster*) a nós individuais (RUBINOV; SPORNS, 2010).

Em oposição à segregação, unidades de neurônios não operam isoladamente (SPORNS, 2002), existindo regiões do cérebro (sistema distribuído do córtex cerebral) capazes de combinar

Figura 20 – Figura extraída e modificada de (FORNITO; ZALESKY; BULLMORE, 2016). Em (A) é mostrado uma matriz de conectividade obtida a partir da aplicação de fMRI em humanos; trata-se de uma matriz com peso pela presença de pixels com diferentes cores. Essa matriz é submetida ao método de *thresholding* resultando em uma matriz em (B), sendo que só foi mantido os pesos 20% maiores. A matriz em (B) é então submetida a binarização resultando na matriz em (C), uma matriz sem peso, de 0 e 1 (branco e preto, respectivamente).

informações especializadas, caracterizando o conceito de integração (RUBINOV; SPORNS, 2010). Essas regiões possuem uma função executora, beneficiando-se de uma alta eficiência global de transferência de informação ao longo de toda rede (BULLMORE; SPORNS, 2012). As medidas utilizadas para a medição da integração é a média dos menores caminhos e a eficiência global.

Assim as redes cerebrais podem ser estruturalmente descritas como módulos com nós regionais densamente interconectados, onde cada um deles compartilha entre si funções especializadas (BULLMORE; SPORNS, 2012). Essas estruturas moldam o fluxo de informação para esses módulos densamente conectados e fortemente acoplados e promovem a integração funcional através de *hubs* (nós densamente conectados) que permitem a comunicação global dessa comunidades, havendo, assim, um equilíbrio entre os dois conceitos mencionados. A figura 21 mostra a organização em comunidades.

Figura 21 – Figura extraída e modificada de (SPORNS, 2013). Em (A) se encontra destacado em laranja as comunidades densamente conectadas entre si e pouco conectadas com outras comunidades e essa conexão é feita por hubs (em azul). Em (B), se encontra destacado a segregação funcional. Em (C), a integração funcional, onde é destacado em azul os hubs da rede altamente conectados entre si, permitindo o fluxo de informação por toda rede.

REDES CEREBRAIS DE PACIENTES COM ESQUIZOFRENIA

6.1 Introdução

Pessoas ao longo de séculos têm sofrido com esquizofrenia. Arqueólogos encontraram escritos da antiga civilização egípcia descrevendo sintomas típicos dessa doença (VEAGUE; COLLINS, 2007). Contudo somente no século XIX ela foi caracterizada pelo psiquiatra Emil Kraepelin, como um demência precoce (ele acreditava que pacientes com essa doença sofriam uma contínua e irreversível deterioração mental iniciando-se na infância); e em 1908 o psiquiatra Eugen Bleuler, nomeou essa desordem psicológica como esquizofrenia (BLEULER; JUNG, 1908), que, em grego, significa "mente dividida", uma vez que um dos seus sintomas era a perda da unidade da mente e consciência (SPORNS, 2010).

Apesar de séculos de pesquisa, o que causa biologicamente a esquizofrenia ainda não é conhecido. Em (FRISTON; FRITH, 1995), foi proposta a desconexão da conectividade funcional e estrutural de redes cerebrais resultando em uma integração disfuncional delas, refletindo em vários sintomas de cognição e comportamentais da esquizofrenia (CALHOUN; EICHELE; PEARLSON, 2009). Essa desconexão em larga escala reflete na topologia estrutural e funcional de pacientes com a doença e assim tem-se aplicado medidas de rede nelas (SPORNS, 2010). Em (TAN *et al.*, 2006) e (LIU *et al.*, 2008) são sugeridos propriedades de pequeno mundo alterado nessas redes usando dados de fMRI e em (MICHELOYANNIS *et al.*, 2006), por meio de dados de eletroencefalograma é relatada a diminuição dessas propriedades. Em (RUBINOV *et al.*, 2009), utilizando, novamente, dados de eletroencefalograma e as medidas de rede: coeficiente de agrupamento (sabido da seção 5 ser uma medida de segregação) e a média dos menores caminhos (sabido da seção 5 ser uma medida de integração), constatou-se nas redes de pacientes com esquizofrenia a diminuição de agrupamentos e menores caminhos em relação a redes de pacientes saudáveis; apesar das redes ainda serem de pequeno mundo há uma sutil randomização

resultando em uma perturbação no balanço de integração e segregação do cérebro (SPORNS, 2010).

Essas medidas sugerem uma metodologia não invasiva de diagnóstico (ARRUDA, 2013) como proposta por (SUPEKAR *et al.*, 2008) em que o coeficiente de agrupamento conseguiu distinguir pacientes com Alzheimer (em cujas redes cerebrais é observada a mesma hipótese de desconexão que as de pacientes com esquizofrenia) de pacientes saudáveis com uma sensibilidade de 72% e especificidade de 78%. Esse fato é muito promissor, uma vez que, o diagnóstico da esquizofrenia até hoje é baseado em critérios subjetivos como a ocorrência de sintomas psicóticos (por exemplo, alucinação e delírios), de déficit (como perda de motivação e depressão) e cognitivos (por exemplo, perda de memória e déficit de atenção) (RUBINOV, 2013), sendo que muitas vezes é confundida com outras doenças como o transtorno bipolar.

Assim no presente trabalho, semelhante ao que foi feito em (ARRUDA, 2013; BABSHIRANI *et al.*, 2013) utilizou-se algumas medidas de redes extraídas de redes de dois grupos de voluntários: com e sem esquizofrenia, obtidas por meio da técnica de ressonância magnética funcional; e, em seguida, utilizou-se alguns algoritmos de aprendizado de máquina na tentativa de distinguir os pacientes normais dos que apresentaram esquizofrenia através da classificação das medidas de redes obtidas. A seção seguinte contém a descrição da base de dados utilizada.

6.2 Base de dados

A base de dados foi extraída de (VÉRTES *et al.*, 2012)¹. Os dados foram obtidos a partir da técnica de fMRI (vide descrição da técnica no capítulo 5) em dois grupos de voluntários: 20 participantes saudáveis e 19 pacientes com esquizofrenia infantil (em inglês, *childhood-onset schizophrenia*, COS, sendo a idade média deles de 18,7) recrutados do National Institutes of Health (NIH). A esquizofrenia infantil é uma forma rara e mais grave da doença, sendo crônica, com início da psicose antes dos 13 anos de idade (ORDÓÑEZ; LUSCHER; GOGTAY, 2016). Entre os dois grupos, não houve diferenças significativas entre sexo, idade ou deslocamento máximo devido a movimentação da cabeça (inclusive movimentação excessiva da cabeça foi um critério de exclusão) (VÉRTES *et al.*, 2012).

Os participantes foram então submetidos ao scanner de ressonância magnética operando a 1,5T General Electric Signa no centro clínico NIH em Bethesda, Maryland nos Estados Unidos. Primeiramente, foi adquirido o volume anatômico dos voluntários. Em seguida, foi pedido que eles ficassem quietos e com olhos fechados e utilizando um tempo de relaxação de 2.3s em um voxel de volume 3.75 X 3.75 X 5 mm (vide capítulo 5). Foram adquiridas duas varreduras sequenciais de imagem eco-planar (em inglês, Echo Planar Imaging). A principal característica

¹ Disponível em <https://www.nimh.nih.gov/labs-at-nimh/research-areas/clinics-and-labs/chp/research-articles.shtml>

das sequências EPI é a existência de um único pulso de excitação varrendo de forma rápida todo espaço, o intuito é minimizar o tempo de aquisição das imagens de fMRI (BUENO, 2004).

As imagens foram processadas por dois softwares: AFNI² (do inglês, Analysis of Functional NeuroImages) (COX, 1996) e FSL³ (do inglês, FMRIB's Software Library) (SMITH *et al.*, 2004). As imagens obtidas foram submetidas às seguintes etapas de pré-processamento: suavização temporal para limitar valores extremos (função *3DDespike* do AFNI) (FERREIRA, 2014) e para correção de movimento da cabeça (movimento causa mudança no sinal e assim alteram as correlações em imagens de ressonância magnética) utilizou-se a função *3DVOLREG* do AFNI (POWER *et al.*, 2014).

Em seguida é feita a normalização dessas imagens para o espaço estereotático (sistema de coordenadas padrão que permite que diferentes imagens de cérebro sejam comparadas independentemente de seu tamanho, posição ou sistema de orientação) do Montreal Neurological Institute (no software AFNI). Neste espaço, o líquido cefalorraquidiano (em inglês, Cerebrospinal fluid, CSF) e a matéria branca são segmentados, ou seja, ocorre subdivisão da imagem em objetos de interesse (ROCHA, 2015) utilizando threshold de 0.8. As séries temporais de cada voxel destas duas regiões, por meio de regressão são removidas pois são variáveis que posteriormente podem confundir a análise. As áreas de matéria cinza foram identificadas usando o atlas probabilístico presente no software FSL e foi feita a eliminação dos voxels utilizando um threshold de 25% (VÉRTES *et al.*, 2012). O resultado foi 300 regiões, contudo o foco era o hemisfério direito para facilitar a aproximação do comprimento da aresta pela distância euclidiana entre as regiões do cérebro (ARRUDA, 2013). Através da transformada de wavelet (utilizada para redução da dimensionalidade, como já falado no Capítulo 5) obteve-se 140 regiões de interesse. As séries temporais correspondentes a essas regiões passaram por um etapa de pré-processamento, antes de serem correlacionadas (também foi mencionado no capítulo 5), no caso foi utilizado o filtro passa banda no intervalo de frequências: 0.05–0.111 (neste tipo de filtro somente as frequência nesse intervalo serão retidas) com intuito de remover frequências que não serão interessantes para o estudo.

Para construção dos grafos binários utilizou-se a matriz de correlação do espaço wavelet estimada fazendo o thresholding para cada participante; no artigo (VÉRTES *et al.*, 2012) não é especificado o método de correlação e o valor do thresholding utilizado. Para garantir que não havia nós desconectados utilizou-se a árvore geradora mínima como suporte (é um algoritmo guloso que divide o conjunto de vértices em conjuntos disjuntos, o objetivo é determinar a árvore de custo mínimo capaz de cobrir exatamente um vértice de cada componente) (FERREIRA; OCHI; MACAMBIRA, 2007). Assim novas arestas são adicionadas de maneira gradual pela ordem decrescente de correlação e assim é obtido redes binárias cujos graus médios são controlados (ARRUDA, 2013). Dessa forma são obtidas redes binárias com 140 regiões e grau médio

² Disponível para aquisição em: <http://afni.nimh.nih.gov/>

³ Disponível para aquisição em: www.fmrib.ox.ac.uk

igual entre elas.

6.3 Metodologia

Nessa seção será formalizada a metodologia utilizada para a análise desta base. Primeiramente extraiu-se algumas medidas de rede por meio do pacote *igraph* versão 1.2.2 presente no Software *R* versão 3.0.2 (R core Team 2016). Além das medidas presentes em (VÉRTES *et al.*, 2012) (média da distribuição do grau, menor caminho, eficiência e coeficiente de aglomeração, que foi calculado pela fórmula da transitividade), e das presentes em (ARRUDA, 2013) (*betweenness centrality*, *closeness centrality*, *k-core*, *assortatividade*), foram extraídas, também, as medidas: *segundo momento da distribuição de grau*, *complexidade*, *eigenvector centrality*, *diâmetro*, *entropia da distribuição de grau*, *k-nearest neighbors (knn)* e *pagerank*, as quais encontram-se descritas no capítulo 4. A escolha dessas medidas se deu baseada na literatura (SPORNS; HONEY; KÖTTER, 2007; BASSETT *et al.*, 2008; ALEXANDER-BLOCH *et al.*, 2012; SKÅTUN *et al.*, 2016; ZUO *et al.*, 2011; LYNALL *et al.*, 2010; HEUVEL *et al.*, 2010; ALEXANDER-BLOCH *et al.*, 2010; MOTA *et al.*, 2014).

Essas medidas foram extraídas para cada uma das redes resultando em arquivo no formato *csv* com 39 linhas e 14 colunas (13 atributos correspondendo às medidas e a classe à que eles pertencem: normais ou com esquizofrenia, COS).

A classificação foi feita no arquivo obtido com os algoritmos descritos no capítulo 2, são eles: *Naive Bayes*, *KNN*, *árvores de decisão*, *redes neurais*. A utilização destes se deu baseada nos classificadores utilizados para classificar dados relacionados à esquizofrenia (MOTA *et al.*, 2014; ARBABSHIRANI *et al.*, 2013; AGUIAR-PULIDO *et al.*, 2010; RISH *et al.*, 2013; ARRUDA *et al.*, 2014), sendo alguns destes dados também são derivados da técnica de ressonância magnética funcional.

Com intuito de tornar o modelo preditivo mais confiável, utilizou-se o método de validação cruzada *k-fold cross-validation*, cuja descrição se encontra no capítulo 3. Empregou-se $k=10$, baseado em alguns artigos encontrados (QUINLAN, 1996b; PEREIRA; MITCHELL; BOTVINICK, 2009; POLAT; GÜNEŞ, 2007; LIU; LEE; LIN, 2010; LATIFOĞLU *et al.*, 2008; SHAFFER *et al.*, 2013) que empregaram esse valor no método em questão, tratando-se de modelos preditivos aplicados a dados médicos.

Para cada um desses classificadores foram utilizados alguns parâmetros presentes no pacote *caret* presente no *R*. No caso do classificador *k-NN*, utilizou-se o método '*knn*' (VENABLES; RIPLEY, 2013) e como métrica de distância a euclidiana (PREMRAJ; HERZIG, 2011). No caso do algoritmo da árvore de decisão, utilizou-se o método '*rpart*' (THERNEAU; ATKINSON; RIPLEY, 2010) que implementa a metodologia CART (em inglês, *Classification and Regression Trees*), cuja descrição se encontra no Capítulo 2; a escolha deste tipo de árvore de decisão se deu baseada em artigos que empregaram este tipo de algoritmo a dados médicos (STROBL; MAL-

LEY; TUTZ, 2009; LAVANYA; RANI, 2011; PODGORELEC *et al.*, 2002). Tratando-se das redes neurais, empregou-se o algoritmo *backpropagation* baseando-se na literatura (TU, 1996; ABBASS, 2002), com uma taxa de aprendizado de 0.25 (valor baseado em (YAO; LIU, 1995; KAYAER; YILDIRIM, 2003)); o método utilizado para as redes neurais foi '*nnet*' (VENABLES; RIPLEY, 2013) presente no pacote *caret* do R.

Para todos esses classificadores, utilizou-se a função *train* presente no pacote *caret*. Essa função pode ser usada, de acordo com (KUHN, 2015) para: "avaliar, usando amostragem, o efeito dos ajustes nos parâmetros modelo em seu desempenho; escolhendo um modelo "ótimo" através desses parâmetros". A figura 22 adaptada de (KUHN, 2015) contém o algoritmo utilizado por essa função. Essa função, então, consegue achar os valores dos parâmetros que resultam em um maior desempenho do modelo, utilizando para isso "força bruta" (ou seja, tentando todos os valores possíveis para o modelo); com isso para o classificador k-NN, a função determina o valor de K que resulta em um modelo de maior performance, bem como para o classificador redes neurais, são determinados os valores dos parâmetros tamanho (número de camadas) e decaimento (parâmetro relacionado ao decaimento do peso, utilizado no processo de otimização para evitar um super ajustamento (em inglês, *overfitting*)).

Figura 22 – Figura contendo o algoritmo adotado pela a função *train* cujo intuito é determinar por intermédio da "força bruta" os valores de um conjunto de parâmetros do modelo que maximizam seu desempenho.

Além destes classificadores também foi feita a combinação deles por meio da técnica de *stacking* utilizando um modelo linear generalizado (ajustado no R pela função "*glm*", presente no pacote *caret*) modelo de regressão logística (vide capítulo 2). Escolheu-se realizar este tipo de combinação, pois em (TANWANI *et al.*, 2009) houve um aumento do desempenho de alguns classificadores (incluindo Naive Bayes, árvores de decisão, redes neurais) quando combinados comparando com seus desempenhos individuais, aplicados a dados biomédicos. Dessa forma foram feitas as seguintes combinações: k-NN e Redes Neurais; k-NN e Árvore de decisão; k-NN

e Naive Bayes; Árvore de decisão e Redes Neurais; Árvore de decisão e Naive Bayes; Naive Bayes e Redes Neurais; k-NN, Redes Neurais e Árvore de decisão; k-NN, Redes Neurais e Naive Bayes; Naive Bayes, Redes Neurais e Árvore de decisão; k-NN, Naive Bayes e Árvore de decisão; k-NN, Naive Bayes, Árvore de decisão e Redes Neurais.

Após realizar as predições empregando os algoritmos de aprendizagem, aqueles que obtiveram um melhor desempenho foram selecionados; sendo que para a avaliação dos modelos utilizou-se as seguintes medidas: acurácia, sensibilidade, especificidade, área abaixo da curva ROC e kappa, cujas descrições se encontram no Capítulo 3.

Os classificadores selecionados foram, então, submetidos aos processos de seleção, com intuito de verificar se por meio deles há um aumento no desempenho dos modelos e, ainda, para encontrar quais atributos (no caso as medidas utilizadas) são mais relevantes. Os processos de seleção utilizados foram: seleção por PCA, remoção de características redundantes, eliminação recursiva de características e seleção pela ordem de importância; estes, também, descritos no capítulo 2. No caso para a remoção de características redundantes, foi feita a correlação de pares de atributos e aqueles com uma correlação maior que 50% foram removidos. Quanto a seleção pela ordem de importância, após encontrar a relevância de cada atributo, aqueles com menor relevância foram eliminados um por vez a cada iteração, enquanto houvesse aumento da acurácia do modelo preditivo. Os resultados encontram-se na seção 6.4.

Para os processos de seleção que de fato contribuíram para a melhora do modelo, foi aplicado a cada um deles às três normalizações descritas no capítulo 2 (*Linear Scaling to Unit Range*, *Softmax Scaling*, *Z-score*). Os resultados obtidos se encontram na subseção 6.4.2. A figura 23 contém um esquema para a metodologia adotada.

Figura 23 – Figura contendo o esquema da metodologia adotada. Primeiramente, em (A), as medidas de redes consideradas foram extraídas da base de dados (na imagem é mostrada a rede do primeiro paciente sem a esquizofrenia obtida por meio do pacote *igraph*). Essas medidas são arquivadas em um formato *csv* e com esse arquivo é realizada a classificação, (B), onde os algoritmos de aprendizado que obtiverem os melhores desempenhos são selecionados. Em seguida, em (C), aplicou-se os quatro métodos de seleção e, então realiza-se a predição com os melhores classificadores de (B), os métodos mais relevantes (aqueles que aumentaram a performance dos classificadores em questão) são selecionados. Por fim, em (D), aplica-se os três tipos de normalização, aos atributos selecionados pelos métodos mais relevante de (C), obtendo o desempenho final dos classificadores de (B).

6.4 Resultados

Seguindo a Metodologia descrita anteriormente, primeiro obteve-se a classificação sem aplicar nenhum método de pré-processamento. Os resultados obtidos se encontram na tabela 4.

Tabela 4 – Tabela contendo os resultados obtidos após submeter o arquivo *csv*, que contém as medidas de redes como atributos e um atributo classe (com instâncias contendo os seguintes rótulos: COS e normal), aos algoritmos preditivos. As medidas utilizadas para verificar o desempenho de cada classificador foram: acurácia, kappa associado a acurácia, especificidade, sensibilidade e AUC. Quanto aos classificadores foram empregados: k-NN, Naive Bayes (NB), árvores de decisão (AD), redes neurais (RN). Foi feita, ainda, a combinação deles dois a dois, três a três e com os quatro, por meio da técnica de *stacking*.

Classificadores	Acurácia	Kappa(Acurácia)	Especificidade	Sensibilidade	AUC
k-NN	0.480	-0.25	0.70	-0.043	0.46
Naive Bayes	0.66	0.30	0.75	0.50	0.67
Redes Neurais	0.72	0.44	0.85	0.70	0.8250
Árvore de decisão	0.70	0.40	0.85	0.45	0.65
k-NN e RN	0.69	0.35	0.70	0.65	0.72
k-NN e AD	0.69	0.35	0.80	0.60	0.80
k-NN e NB	0.77	0.50	0.85	0.65	0.75
AD e RN	0.68	0.35	0.75	0.55	0.73
AD e NB	0.63	0.25	0.70	0.60	0.68
NB e RD	0.75	0.50	0.80	0.65	0.75
k-NN, RN e AD	0.68	0.35	0.80	0.60	0.7875
k-NN, RN e NB	0.72	0.44	0.70	0.70	0.65
NB, RN e AD	0.65	0.30	0.70	0.65	0.73
k-NN, NB e AD	0.69	0.35	0.75	0.75	0.78
k-NN, NB, AD e RN	0.65	0.30	0.75	0.55	0.825

Da tabela 4, percebe-se que dos classificadores individuais (ou seja, aqueles que não foram combinados com outros pela técnica *stacking*) aquele que obteve melhor desempenho pelas medidas de acurácia, especificidade e AUC, foram as redes neurais, cujos valores são, respectivamente 0.7167, 0.85 e 0.8250; sendo o kappa relativo a acurácia com o valor de 0.44, que garante uma confiabilidade regular no modelo segundo (GUYATT *et al.*, 1995). Constatase, ainda, que esse classificador discrimina melhor a classe negativa (referente aos pacientes saudáveis) que a classe positiva (pacientes com esquizofrenia), uma vez que, a especificidade obtida (taxa de acertos na classe negativa) é maior que a sensibilidade (taxa de acertos na classe positiva).

Analisando os classificadores combinados dois a dois pelo método *stacking*, aquele que obteve melhor desempenho pela acurácia e especificidade (cujos valores foram, respectivamente, 0.767 e 0.85) foi a combinação do Naive Bayes e k-NN; inclusive foi o que alcançou o maior valor de kappa (0.5), o que faz com que o modelo gerado seja o de maior confiabilidade. Essa combinação resultou na melhora de todas as medidas em relação aquelas obtidas individualmente (quando o k-NN e o Naive Bayes foram empregados separadamente).

Quando os classificadores foram combinados três a três, a combinação que obteve melhor acurácia e valor de kappa associado a ela, foi: k-NN, RN e NB; contudo a combinação entre k-NN e NB resultou em uma performance melhor e quando comparada com os classificadores

individuais, resultou em um desempenho igual a RN (mesmo valor de acurácia e kappa associado a ela), porém as outras medições foram inferiores, em outras palavras, a especificidade, sensibilidade e AUC, resultaram em um valor maior quando se utilizou somente a RN.

Por fim, a combinação dos quatro, melhorou a acurácia do classificador k-NN mais piorou a dos outros. Analisando, ainda, todas as classificações, a especificidade sempre resultou em valor maior que a sensibilidade. Isso pode ser explicado pelo fato do dado ser desbalanceado, uma vez que a quantidade de pacientes com esquizofrenia (classe positiva, medida pela sensibilidade) é menor que a de pacientes normais (classe negativa, medida pela especificidade). Assim é provável que alguns classificadores acabem favorecendo a classe majoritária.

Como critério de seleção dos classificadores que obtiveram maior desempenho, para a próxima etapa (utilização dos quatro tipos de seleção), analisou-se, primeiro o valor de acurácia e kappa associado a ela. Foram considerados os classificadores que obtiveram um valor de acurácia superior a 0.6, que apesar de não diferir significativamente de uma tentativa randômica (KEOGH; KASETTY, 2003), é considerada aceitável (MAROCO *et al.*, 2011) e um valor de kappa superior a 0.4 (baseado em (CARNEC; CALLET; BARBA, 2003; ALAGIAKRISHNAN *et al.*, 2013)) e também no fato de que a faixa de valor: 0.4-0.6, faz com que a confiabilidade no modelo seja moderado (GUYATT *et al.*, 1995). O intuito desse primeiro critério é excluir aqueles classificadores com desempenho, em termos da acurácia, inferior ao randômico e que a confiabilidade do modelo seja inferior ao moderado (MUNOZ; BANGDIWALA, 1997), levando-se em consideração a tabela 2 descrita no Capítulo 3. Com isso, os classificadores que satisfazem os dois requisitos são: redes neurais, árvore de decisão, combinação entre k-NN e Naive Bayes, combinação entre Naive Bayes e redes neurais, e combinação entre K-NN, redes neurais e Naive Bayes.

O segundo critério utilizado leva em consideração os valores das medidas: AUC, especificidade e sensibilidade. O valor de AUC escolhido foi de 0.7, pois de acordo com a tabela 3 do Capítulo 3 representa uma discriminação aceitável. Para a sensibilidade e especificidade o valor adotado foi acima de 0.6, de forma que a taxas de acerto referentes à classe positiva e negativa, sejam superiores a uma discriminação aleatória (0.5). Assim, com esses três valores, selecionou-se os seguintes classificadores: redes neurais, combinação entre k-NN e RN, combinação entre NB e RN, combinação entre k-NN e NB, combinação entre k-NN, NB e AD.

Unindo os dois conjuntos de classificadores obtidos segundo os dois critérios adotados obtém-se a seleção dos seguintes classificadores: **combinação entre k-NN e redes neurais, combinação entre naïve Bayes e redes neurais, combinação entre k-NN, Naive Bayes e árvore de decisão, combinação entre k-NN, redes neurais e Naive Bayes, combinação entre k-NN e Naive Bayes, redes neurais e árvore de decisão.**

6.4.1 Resultados após a aplicação dos processos de seleção

Como descrito anteriormente (seção 6.3), após a escolha dos melhores classificadores segundo o desempenho deles (vide seção 6.3 anteriormente), os atributos (medidas extraídas das redes de pacientes normais e com esquizofrenia) foram submetidos a quatro tipos de seleção, com intuito de verificar se essas reduções de dimensionalidade aumenta a performance deles.

Os tipos de seleção empregados, como já mencionados anteriormente, foram: seleção por PCA, remoção de características redundantes, eliminação recursiva de características e seleção pela ordem de importância, cujas descrições se encontram no capítulo 2.

A primeira seleção utilizada foi a remoção de características redundantes (cuja descrição detalhada se encontra no capítulo 2). Para isso, foi calculado a matriz de correlação de Pearson dos atributos, sendo que foi utilizado um *threshold* de 0.7 (valor de correlação considerado alto por (HINKLE; WIERMA; JURS, 2003)), ou seja, os atributos com uma correlação maior que 70% foram eliminados. A figura 24 contém a matriz de correlação obtida.

Figura 24 – Matriz de correlação de Pearson obtida, percebe-se que as medidas altamente correlacionadas (superiores a 0.70) são: segundo momento da distribuição de grau e complexidade; segundo momento da distribuição de grau e knn; complexidade e knn; diâmetro e betweenness centrality; assortatividade e transitividade; menor caminho e diâmetro. Essas medidas foram eliminadas.

Em seguida, aplicou-se no arquivo csv, contendo todas as medidas de rede, seleção pela

ordem de importância (em inglês, *Rank features by importance*). Esse tipo de seleção usa a abordagem *wrapper* (vide capítulo 2) que se caracteriza por ser sempre realizada atrelada a um algoritmo de aprendizado de máquina e onde se busca um subconjunto de atributos que consiga minimizar o erro de predição (BOCCA, 2014). No caso, o algoritmo escolhido para construção do modelo foi o de quantização vetorial por aprendizagem (em inglês, *learning vector quantization*), lqv, cuja descrição se encontra na subsecção 2.2.1, por meio do pacote *class* presente no R. Com intuito de medir a importância de cada atributo no modelo de predição construído, empregou-se a função *varImp* (BREIMAN, 2001) (em inglês, *variable importance score* (WIESE, 2016)) do pacote *caret* do R, descrito por (KUHN, 2012). Para problemas de duas classes, a função realiza um série de cortes nos atributos preditivos. Em cada um desses cortes é gerado um modelo de classificação (no caso utiliza-se como classificador o lqv) e para cada um desses modelos é computado a especificidade e sensibilidade, e com isso sua curva ROC, bem como a área abaixo da curva ROC. O valor dessa área é usada para medir a variável de importância (KUHN, 2008a). A tabela 5 contém os valores encontrados pela função *varImp* para cada uma das medidas de rede utilizadas como atributos.

Tabela 5 – Tabela contendo os atributos correspondentes às medidas de rede com seu respectivo valor (organizada de forma decrescente) da medida de importância obtida pela função *varImp*. Foram destacados, na coloração avermelhada, aqueles atributos com menor medida encontrada.

Medidas	Importância
k-core	0.79
média da distribuição do grau	0.70
knn	0.68
complexidade	0.67
segundo momento da distribuição do grau	0.66
closeness centrality	0.60
transitividade	0.58
eigenvector centrality	0.57
menor caminho	0.57
betweenness centrality	0.57
diâmetro	0.57
entropia da distribuição de grau	0.55
assortatividade	0.54
grau médio	0.50
PageRank	0.50
Eficiência	0.50

Da tabela 5 percebe-se que as piores medidas obtidas foram: *Eficiência*, *PageRank* e *grau médio*. Assim, excluiu-se, primeiramente a *Eficiência*, seguida pelo *PageRank* e depois excluindo as três medidas juntas.

Utilizou-se, então, um método de seleção automático presente no pacote *caret* denominado de eliminação recursiva de características (em inglês, *Recursive Feature Elimination* ou

RFE), vide sua descrição no capítulo 2. Utilizou-se para avaliação do modelo gerado o algoritmo *Random Forest*, cuja descrição também se encontra no capítulo 2 e como medida a raiz do erro quadrático médio (*Root Mean Square Error*, ou RMSE). Esse tipo de seleção é configurado para explorar todos os grupos de atributos possíveis; a figura 25 contém o gráfico da raiz do erro quadrático médio e seu respectivo grupo obtido pelo processo de seleção, com uma determinada quantidade de atributos. Percebe-se que o grupo que obteve o menor erro foi o que continha oito variáveis, sendo elas: *closeness centrality*, *betweenness centrality*, *eigenvector centrality*, assortatividade, menor caminho, k-core, complexidade e *diâmetro*. Dessa forma, utilizaram-se essas oito medidas para realizar a predição com os classificadores anteriores.

Figura 25 – Figura contendo o gráfico obtido ao se aplicar o processo de seleção de eliminação recursiva de características; o gráfico corresponde à raiz do erro quadrático médio pela quantidade de variáveis do grupo. Percebe-se que o grupo com menor erro é aquele que possui oito variáveis.

Por fim, foi empregado o método de seleção utilizando PCA (vide capítulo 2). A figura 26 contém as componentes obtidas do PCA, percebe-se que as medidas que tiveram menor contribuição foram: menor caminho, grau médio, pageRank e eficiência. Dessa forma, elas foram excluídas e novamente as classificações foram realizadas.

Os resultados encontram-se nas tabelas 6, 7, 8, 9, 10, 11 e 12, referentes, respectivamente, aos classificadores redes neurais; árvore de decisão; k-NN e naive bayes; k-NN e redes neurais; naive bayes e redes neurais; k-NN, naive bayes e árvore de decisão; k-NN, rede neurais e naive bayes. As cores foram usadas para destacar os diferentes tipos de seleção; em rosa, a seleção por remoção de características redundantes; em azul a seleção por ordem de importância; em laranja, a eliminação recursiva de característica; em verde, a seleção utilizando o método PCA.

Figura 26 – Figura contendo as componentes geradas pelo método de PCA para cada medida; aquelas que tiveram menor contribuição foram: menor caminho, grau médio, PageRank e eficiência.

Tabela 6 – Tabela contendo o resultado obtido em cada processo de seleção, sendo que cada um deles foi representado utilizando uma cor, para o classificador redes neurais. Percebe-se que os métodos de seleção não aumentaram o desempenho (em termos da acurácia e AUC) desse classificador.

Modelo de seleção	Acurácia	kappa	sensitividade	especificidade	AUC
sem seleção	0.72	0.44	0.70	0.85	0.83
remoção de características redundantes	0.67	0.35	0.70	0.75	0.8000
seleção por ordem de importância sem eficiência	0.69	0.39	0.65	0.70	0.67
seleção por ordem de importância sem PageRank	0.77	0.55	0.70	0.75	0.75
seleção por ordem de importância sem grau médio	0.74	0.49	0.65	0.80	0.77
seleção por ordem de importância sem eficiência, PageRank, grau médio	0.74	0.45	0.75	0.85	0.72
RFE	0.51	0.00	0.10	0.90	0.54
seleção com PCA	0.68	0.35	0.65	0.85	0.70

Tabela 7 – Tabela contendo o desempenho do classificador árvore de decisão ao se aplicar os diferentes métodos de seleção (destacados com diferentes cores na tabela). Em negrito, encontra-se aqueles que mais aumentaram a performance (em termos da acurácia e AUC) do preditor em questão.

Modelo de seleção	Acurácia	kappa	sensitividade	especificidade	AUC
sem seleção	0.70	0.40	0.45	0.85	0.65
remoção de características redundantes	0.72	0.44	0.55	0.80	0.67
seleção por ordem de importância sem eficiência	0.67	0.34	0.50	0.80	0.65
seleção por ordem de importância sem PageRank	0.69	0.39	0.60	0.70	0.66
seleção por ordem de importância sem grau médio	0.68	0.35	0.65	0.75	0.71
seleção por ordem de importância sem eficiência, PageRank, grau médio	0.64	0.29	0.60	0.60	0.64
RFE	0.73	0.45	0.65	0.65	0.68
seleção com PCA	0.68	0.35	0.55	0.75	0.65

Tabela 8 – Tabela contendo o desempenho do classificador k-NN e naive bayes após submetido aos métodos de seleção. Percebe-se que os métodos de seleção (em termos da acurácia e AUC) pioraram (em termos da acurácia e AUC) a performance desse algoritmo preditivo.

Modelo de seleção	Acurácia	kappa	sensitividade	especificidade	AUC
sem seleção	0.77	0.50	0.65	0.85	0.75
remoção de características redundantes	0.74	0.48	0.75	0.80	0.78
seleção por ordem de importância sem eficiência	0.74	0.49	0.70	0.60	0.75
seleção por ordem de importância sem PageRank	0.66	0.30	0.55	0.70	0.80
seleção por ordem de importância sem grau médio	0.72	0.44	0.7	0.85	0.85
seleção por ordem de importância sem eficiência, PageRank, grau médio	0.66	0.35	0.65	0.80	0.75
RFE	0.74	0.48	0.60	0.65	0.75
seleção com PCA	0.76	0.51	0.65	0.80	0.78

Tabela 9 – Tabela contendo o desempenho obtido pelo classificador k-NN e redes neurais ao ser submetido aos diferentes métodos de seleção. Aquele que resultou na maior performance (em termos da acurácia e AUC) foi a seleção com PCA e a seleção por ordem de importância sem três medidas de rede (destacados em negrito na tabela).

Modelo de seleção	Acurácia	kappa	sensitividade	especificidade	AUC
sem seleção	0.69	0.35	0.65	0.70	0.73
remoção de características redundantes	0.68	0.35	0.60	0.80	0.75
seleção por ordem de importância sem eficiência	0.6667	0.34	0.60	0.75	0.8500
seleção por ordem de importância sem PageRank	0.66	0.35	0.65	0.80	0.85
seleção por ordem de importância sem grau médio	0.73	0.45	0.55	0.70	0.78
seleção por ordem de importância sem eficiência, PageRank, grau médio	0.75	0.50	0.55	0.65	0.70
RFE	0.63	0.25	0.50	0.55	0.58
seleção com PCA	0.71	0.39	0.65	0.80	0.73

Tabela 10 – Tabela contendo a performance obtida para o classificador oriundo da combinação entre naive bayes e redes neurais. Em negrito, destacou-se o processo de seleção que mais aumentou o desempenho (em termos da acurácia e AUC) desse classificador.

Modelo de seleção	Acurácia	kappa	sensitividade	especificidade	AUC
sem seleção	0.75	0.50	0.65	0.80	0.75
remoção de características redundantes	0.70	0.38	0.75	0.65	0.65
seleção por ordem de importância sem eficiência	0.75	0.50	0.65	0.80	0.7250
seleção por ordem de importância sem PageRank	0.69	0.39	0.60	0.75	0.60
seleção por ordem de importância sem grau médio	0.74	0.49	0.70	0.80	0.68
seleção por ordem de importância sem eficiência, PageRank, grau médio	0.75	0.35	0.55	0.70	0.70
RFE	0.75	0.50	0.60	0.75	0.73
seleção do PCA	0.67	0.36	0.70	0.70	0.75

Tabela 11 – Tabela contendo a performance obtida com relação ao classificador: k-NN, árvore de decisão e naive bayes, ao empregar-se os diferentes métodos de seleção. O método que mais aumentou o desempenho (em termos da acurácia e AUC) desse classificador foi a remoção de características redundantes, ainda que os métodos RFE e seleção com PCA também obtiveram uma performance próxima ao maior valor.

Modelo de seleção	Acurácia	kappa	sensitividade	especificidade	AUC
sem seleção	0.69	0.35	0.75	0.75	0.78
remoção de características redundantes	0.82	0.65	0.60	0.70	0.67
seleção por ordem de importância sem eficiência	0.69	0.39	0.70	0.75	0.80
seleção por ordem de importância sem PageRank	0.69	0.35	0.50	0.70	0.78
seleção por ordem de importância sem grau médio	0.60	0.22	0.65	0.70	0.68
seleção por ordem de importância sem eficiência, PageRank, grau médio	0.64	0.25	0.65	0.70	0.82
RFE	0.7167	0.44	0.65	0.75	0.7250
seleção com PCA	0.75	0.51	0.75	0.80	0.83

Tabela 12 – Tabela contendo as performances do classificador k-NN, redes neurais e Naive Bayes, ao ser submetido aos diferentes tipos de seleção. Em negrito o método que proporcionou maior desempenho (em termos da acurácia e AUC) do algoritmo preditivo em questão.

Modelo de seleção	Acurácia	kappa	sensitividade	especificidade	AUC
sem seleção	0.72	0.44	0.70	0.70	0.65
remoção de características redundantes	0.64	0.28	0.60	0.80	0.75
seleção por ordem de importância sem eficiência	0.7750	0.55	0.65	0.80	0.83
seleção por ordem de importância sem PageRank	0.71	0.40	0.65	0.75	0.8500
seleção por ordem de importância sem grau médio	0.68	0.37	0.55	0.70	0.70
seleção por ordem de importância sem eficiência, PageRank, grau médio	0.75	0.50	0.65	0.80	0.65
RFE	0.73	0.45	0.65	0.75	0.75
seleção com PCA	0.68	0.33	0.65	0.80	0.73

Das tabelas 6, 8, 10, percebe-se que, respectivamente, para os classificadores redes neurais, o classificador resultante da combinação entre k-NN e naive bayes, e classificador formado pelo naive bayes e redes neurais, nenhum processo de seleção aumentaram seus desempenhos. Para o classificador k-NN era esperado que a redução da dimensionalidade resultasse no aumento em sua performance (uma vez que tanto a alta dimensionalidade quanto a presença de atributos irrelevantes e redundantes interferem de forma negativa no desempenho desse classificador), contudo, como isso não foi observado conclui-se que a combinação com o algoritmo naive

bayes (robusto à presença de atributos irrelevantes e ruídos) resultou em um classificador menos susceptível a esse tipo de atributos.

Para os classificador árvores de decisão (tabela 7), a seleção que aumentou a performance foi a eliminação recursiva de atributos. Houve um baixo aumento na acurácia e na medida AUC (para ambas 0.025), o que era esperado uma vez que este algoritmo é robusto a presença de atributos irrelevantes e redundantes. Isso ocorre, pois durante a construção de uma árvore de decisão há a seleção dos atributos a ser usados no modelo, o que produz um modelo robusto a esses tipos de atributos.

Para os classificadores oriundos da combinação dos classificadores k-NN e redes neurais (vide tabela 9); e da combinação de k-NN, redes neurais e naive bayes (tabela 12); o método de seleção que aumentou seus desempenhos foi seleção por ordem de importância, sendo que para o primeiro sem as medidas: grau médio, eficiência e PageRank, e para o segundo sem a medida de eficiência.

Por fim, para o classificador obtido da combinação k-NN, redes neurais e naive bayes, o método de seleção que aumentou seu desempenho foi a remoção de características redundantes. A tabela 13 contém o resumo do que foi descrito anteriormente, com o classificador e o respectivo método de seleção que aumentou sua performance.

Tabela 13 – Tabela contendo o resumo dos resultados obtidos ao se aplicar os métodos de seleção aos classificadores em questão. Assim em uma coluna se encontra o classificador e na coluna ao lado o processo de seleção que aumentou o seu desempenho. Aqueles em que nenhum método aumentou sua performance, a coluna de seleção se encontra como sem seleção.

Classificador	Processo de seleção
Redes neurais	sem seleção
Árvore de decisão	RFE
k-NN e redes neurais	seleção por ordem de importância sem grau médio, eficiência e PageRank
k-NN e naive bayes	sem seleção
naive bayes e redes neurais	sem seleção
k-NN, árvore de decisão e naive bayes	remoção de características redundantes
k-NN, redes neurais e naive bayes	seleção por ordem de importância sem eficiência

6.4.2 Resultados após à aplicação das normalizações

Nessa seção, serão aplicados os métodos de normalização descritos anteriormente, com intuito de verificar se eles aumentam o desempenho dos classificadores considerados.

Essas normalizações serão aplicadas ao conjunto de atributos que de fato contribuíram para cada modelo preditivo, ou seja, serão levados em consideração os métodos de seleção que se encontram na tabela 13. Os resultados se encontram nas tabelas 14, 15, 16, 17, 18, 19 e 20.

Tabela 14 – Tabela contendo os resultados obtidos para o algoritmo de predição redes neurais, após realiza os três processos de normalização. Para esse classificador não foi utilizado nenhum processo de seleção, pois esses diminuam o desempenho (em termos da acurácia e AUC) preditivo.

Tipo de normalização	Acurácia	Kappa	Sensitividade	Especificidade	AUC
Sem normalização	0.72	0.44	0.70	0.85	0.83
Z-score	0.69	0.39	0.55	0.60	0.78
Linear Scaling to Unit Range	0.72	0.40	0.60	0.80	0.80
Softmax Scaling	0.63	0.25	0.50	0.80	0.78

Tabela 15 – Tabela contendo os resultados obtidos para o classificador árvore de decisão após realizar os processos de normalização. Para este classificador foi realizado o processo de seleção RFE. Em negrito se encontra destacado o melhor desempenho (em termos da acurácia e AUC) obtido.

Tipo de normalização	Acurácia	Kappa	Sensitividade	Especificidade	AUC
Sem normalização	0.73	0.45	0.65	0.65	0.68
Z-score	0.68	0.35	0.75	0.65	0.73
Linear Scaling to Unit Range	0.67	0.34	0.65	0.75	0.70
Softmax Scaling	0.73	0.45	0.65	0.75	0.70

Tabela 16 – Tabela contendo os resultados, após aplicar os três processos de normalização, para o classificador formado entre a combinação do k-NN e redes neurais; utilizou-se, também o método de seleção por ordem de importância sem todas as medidas. Está destacado, em negrito, o processo de normalização que mais aumentou a performance (em termos da acurácia e AUC) do classificador em questão.

Tipo de normalização	Acurácia	Kappa	Sensitividade	Especificidade	AUC
Sem normalização	0.69	0.35	0.65	0.70	0.73
Z-score	0.72	0.43	0.70	0.75	0.83
Linear Scaling to Unit Range	0.76	0.50	0.60	0.70	0.75
Softmax Scaling	0.73	0.47	0.60	0.70	0.70

Nas tabelas destacou-se cada normalização com uma cor: rosa, azul e laranja, correspondente, respectivamente, as normalizações z-score, linear scaling to unit range e softmax scaling.

Da tabela 14, percebe-se que os processos de normalização não aumentaram o desempenho do classificador rede neurais.

Da tabela 15, constata-se que o melhor desempenho obtido para o classificador árvore de decisão foi utilizando a normalização softmax scaling. Apesar de não haver aumento na acurácia, as medidas especificidade e AUC foram aumentadas. Além disso houve um aumento na especificidade em relação a sensibilidade, ou seja, houve um aumento da distinção entre os indivíduos verdadeiramente negativos.

O maior desempenho obtido para o classificador obtido da combinação k-NN e redes neurais (sendo que foi aplicado a seleção por ordem de importância, obtida na secção anterior), após utilizar os processos de normalização, foi ao se utilizar o Z-score (vide tabela 16), pois

Tabela 17 – Desempenho obtido pelo classificador resultante da combinação entre o algoritmo naive bayes e k-NN após realizar os três processos de normalização. Pelo fato de nenhum dos processos de seleção ter contribuído por um aumento na performance desse algoritmo preditivo, eles não foram usados. Em negrito se encontra a melhor performance (em termos da acurácia e AUC) obtida.

Tipo de normalização	Acurácia	Kappa	Sensitividade	Especificidade	AUC
Sem normalização	0.7167	0.50	0.65	0.85	0.75
Z-score	0.67	0.33	0.65	0.80	0.83
Linear Scaling to Unit Range	0.73	0.45	0.65	0.80	0.70
Softmax Scaling	0.64	0.27	0.50	0.80	0.60

Tabela 18 – Tabela contendo o desempenho obtido pelo classificador resultante da combinação de redes neurais e naive bayes, quando submetido aos três tipos de normalização. Em negrito é destacado a melhor performance (em termos da acurácia e AUC) encontrada.

Tipo de normalização	Acurácia	Kappa	Sensitividade	Especificidade	AUC
Sem normalização	0.7500	0.50	0.65	0.80	0.75
Z-score	0.69	0.38	0.60	0.75	0.73
Linear Scaling to Unit Range	0.70	0.39	0.60	0.75	0.68
Softmax Scaling	0.69	0.39	0.65	0.80	0.75

Tabela 19 – Tabela contendo a performance do classificador obtido da combinação do k-NN, redes neurais e naive bayes, após submetido aos três processos de normalização. O classificador foi aplicado aos atributos obtidos se seleção por ordem de importância, excluindo-se a medida eficiência. Em negrito está destacado o melhor desempenho (em termos da acurácia e AUC) obtido.

Tipo de normalização	Acurácia	Kappa	Sensitividade	Especificidade	AUC
Sem normalização	0.78	0.55	0.65	0.80	0.83
Z-score	0.74	0.48	0.65	0.75	0.68
Linear Scaling to Unit Range	0.66	0.32	0.75	0.80	0.83
Softmax Scaling	0.62	0.23	0.55	0.80	0.73

obteve-se um valor muito alto da medida AUC, além de aumentar todas as outras medidas (foi a única normalização que aumentou a sensibilidade e especificidade). Novamente a taxa de verdadeiros negativos (especificidade) foi superior a taxa de verdadeiros positivos.

Percebe-se que o maior desempenho obtido para o classificador obtido da combinação do naive bayes e k-NN, foi sem a normalização. Apesar da normalização linear scaling to unit range ter aumentado a acurácia do classificador, todas as outras medidas diminuíram (inclusive o kappa associado a ela indicando a diminuição da confiabilidade da medida). Da mesma forma, a normalização z-score aumentou a medida AUC, contudo houve diminuição da acurácia, kappa associada a ela e a especificidade.

Percebe-se, da tabela 18, que nenhum processo de normalização aumentou o desempenho do classificador obtido da combinação da redes neurais com naive bayes.

Percebe-se da tabela 19 que nenhum processo de normalização aumentou o desempenho do classificador obtido da combinação k-NN, redes neurais e naive bayes.

Tabela 20 – Tabela contendo os resultados obtidos para o classificador resultante da combinação k-NN, árvore de decisão e naive bayes após a aplicação dos métodos de normalização. Os atributos utilizados foram aqueles obtidos pela remoção de características redundantes. Em negrito está destacado o melhor desempenho (em termos da acurácia e AUC) obtido.

Tipo de normalização	Acurácia	Kappa	Sensitividade	Especificidade	AUC
Sem normalização	0.82	0.65	0.60	0.70	0.68
Z-score	0.66	0.31	0.60	0.70	0.70
Linear Scaling to Unit Range	0.68	0.33	0.75	0.75	0.78
Softmax Scaling	0.64	0.28	0.65	0.70	0.75

Para o classificador obtido da combinação do k-NN, árvore de decisão e naive bayes (vide tabela 20) não houve processo de normalização que aumentou seu desempenho (em termos da acurácia e AUC).

A tabela 21, a seguir, contém o resumo de todos os resultados obtidos nessa seção para facilitar a análise global.

Por fim, da tabela 21, percebe-se que os métodos de normalização aumentaram o desempenho de poucos classificadores (somente a árvore de decisão e o classificador obtido da combinação k-NN e redes neurais).

6.5 Conclusões

Até hoje o diagnóstico da esquizofrenia é baseado em critérios subjetivos e muitas vezes é confundida com outras doenças (como o transtorno bipolar). Com advento de técnicas não invasivas, como a ressonância magnética funcional, e a possibilidade de construção de redes complexas baseadas nelas, possibilitam novas alternativas para prever se determinada pessoa possui essa doença.

Nesse trabalho (semelhante ao que foi feito em (ARRUDA, 2013) e (ARBABSHIRANI *et al.*, 2013)), utilizando a base de dados de (VÉRTES *et al.*, 2012), contendo matrizes de conectividade (obtidas por meio da técnica de ressonância magnética funcional) de dois grupos de voluntários, com e sem a doença; extraiu-se algumas medidas de redes com a intenção de classificá-las, utilizando algoritmos de aprendizado de máquina, com a intenção de distinguir os dois grupos e assim auxiliar no diagnóstico da doença. Para medir o desempenho dos classificadores foram utilizadas as medidas: Acurácia e o kappa associado a ela, sensibilidade, especificidade e área sob a curva ROC. As medidas extraídas de rede utilizadas foram: média da distribuição do grau, menor caminho, coeficiente de aglomeração (calculado pela fórmula da transitividade), betweenness centrality, closeness centrality, k-core, assortatividade, segundo momento da distribuição de grau, complexidade, eigenvector centrality, diâmetro, entropia da distribuição de grau, k-nearest neighbors (knn).

No início, uma série de classificadores foram utilizados mas seguindo alguns critérios

Tabela 21 – Tabela contendo o resumo dos resultados obtidos, com todos os classificadores e os respectivos métodos de seleção e normalização que mais aumentaram seu desempenho. Nela também se encontra o valor desse desempenho.

Classificadores	Método de seleção	Método de normalização	Acurácia	Kappa (Ac.)	Sens.	Esp.	AUC
RN	sem seleção	sem normalização	0.72	0.44	0.70	0.85	0.83
AD	RFE	softmax scaling	0.73	0.45	0.65	0.75	0.70
k-NN e RN	seleção por ordem de importância sem grau médio, eficiência e pageRank	z-score	0.72	0.43	0.70	0.75	0.83
k-NN e NB	sem seleção	sem normalização	0.72	0.50	0.65	0.85	0.75
NB e RN	sem seleção	sem normalização	0.75	0.50	0.65	0.8	0.75
k-NN, NB e AD	remoção de características redundantes	sem normalização	0.78	0.55	0.65	0.80	0.83
k-NN, NB e RN	seleção por ordem de importância sem eficiência	sem normalização	0.82	0.65	0.60	0.70	0.68

(possuir acurácia superior a 0.6 e kappa maior que 0.4; AUC maior que 0.7; e sensibilidade e especificidade superior a 0.6) foram selecionados os classificadores presentes na tabela 21. O interessante foi que combinação de classificadores, por meio da técnica de stacking, resultou para alguns classificadores um aumento em sua performance (por exemplo, o k-NN que individualmente não obteve uma boa performance mais quando combinado com os outros classificadores obteve melhores medidas).

Processos de seleção foram aplicados às medidas de redes para reduzir a sua dimensionalidade e para alguns classificadores isso resultou em um aumento de desempenho (vide na tabela 21). Já os processos de normalização aplicados, em seguidas, contribuíram para o aumento da performance somente de dois dos algoritmos de AM.

Em termos de acurácia, o classificador que obteve melhor performance (cujo valor foi 82%) foi aquele resultante da combinação dos classificadores k-NN, naive bayes e redes neurais, valor superior ao de (ARRUDA, 2013), cuja acurácia para redes bayesianas foi de 79,95%. Para esse classificador utilizou-se a seleção por ordem de importância, sendo que foram usadas todas as medidas de rede menos a eficiência.

Em termos de área sob a curva ROC, o maior valor obtido, de 0.825, foi utilizando-se os classificadores redes neurais; combinação entre k-NN e naive bayes (onde se utilizou a seleção por ordem de importância excluindo o grau médio, eficiência e pageRank; e a normalização z-score); combinação entre k-NN, naive bayes e árvore de decisão (em que se utilizou a remoção de características redundantes).

Sendo que em todos os classificadores, a especificidade (a maior obtida foi de 85% pelos classificadores: redes neurais e a combinação do k-NN e naive bayes) e a sensibilidade (a maior obtida foi de 70% pelos classificadores: redes neurais e a combinação do k-NN e redes neurais) apresentaram-se com taxas bem diferentes para cada dos algoritmos. Esse fato é positivo, uma vez que a intenção era conseguir a separação entre as duas classes. Contudo a especificidade que media a taxa de verdadeiros negativos (classe de pacientes normais) foi sempre superior a sensibilidade.

REDES CEREBRAIS DE PACIENTES COM DESORDEM DO ESPECTRO AUTISTA

7.1 Introdução

O autismo ou desordem do espectro autista (em inglês, *autism spectrum disorders, ASD*) é uma doença relativamente comum, sendo que suas causas ainda não são inteiramente conhecidas (SPORNS, 2010). Esta desordem normalmente se manifesta desde a infância (ao menos nos três primeiros anos de vida), sendo definida pelos déficits de comunicação social, falta de empatia, e comportamentos não usuais repetidos e restritos (LORD *et al.*, 2000) (como a resistência a mudança). Uma das teoria que tenta explicar esses sintomas é a baixa coerência central, baseada nas diferenças no sistema de informação em crianças com autismo que está relacionada a falta da tendência em juntar partes da informação pra formar um "todo" com significado (coerência central) (FRITH, 1989), refletindo uma dominância de estratégia cognitiva que enfatiza um processamento focado e uma perda do poder integrativo (SPORNS, 2010).

Em adição ao paradigma da coerência central descrito, outras teorias foram propostas (como deficit da função cognitiva (OZONOFF; PENNINGTON; ROGERS, 1991), teoria da mente (BARON-COHEN; LESLIE; FRITH, 1985), processamento de informação complexa (MINSHEW; GOLDSTEIN; SIEGEL, 1997), entre outras), contudo estas teoria ainda possuem várias questões em aberto, assim, tem-se focado na conectividade neural anormal (BELMONTE *et al.*, 2004). Em (BELMONTE; YURGELUN-TODD, 2003), através da técnica de fMRI, foi demonstrado que regiões do cérebro de pacientes com autismo com função de integração possuem suas entradas cortadas, manifestando reduções na ativação e correlações funcionais com regiões sensoriais. Em (BELMONTE; BARON-COHEN, 2004), também utilizando a técnica de fMRI em crianças com autismo e sem a doença, sugere uma forte ativação do córtex parietal¹

¹ Responsável pela recepção de sensações como tato e dor.

durante a supressão de distrações, ao mesmo tempo que regiões do córtex pré-frontal médio². Em (KENNEDY; REDCAY; COURCHESNE, 2006), ainda, além da região do córtex pré-frontal médio, outras regiões como o córtex cingulado anterior rostral³ e córtex cingulado posterior⁴, que possuem atividades metabólicas durante o repouso e são desativadas durante tarefas cognitivas exigentes, indicando interrupção da atividade mental que persiste durante o repouso; os autores observaram que em pacientes com autismo (em imagem de fMRI) essa desativação não ocorre.

Em termos estruturais, em (KELLER; KANA; JUST, 2007) é sugerido que o cérebro de crianças possuem um desenvolvimento atípico como um crescimento excessivo precoce da substância branca, seguido pela sua redução na adolescência e vida adulta; e sugerem, ainda, pelo resultado de difusão de fMRI, a desorganização de caminhos de matéria branca.

O diagnóstico do autismo é feito em termos da observação prática de seus sintomas, critério subjetivo, sendo muitas vezes descoberto tardiamente em períodos críticos de desenvolvimento (DANIELS; MANDELL, 2014). Assim, por intermédio de uma base de dados de fMRI de pacientes normais e com autismo, algumas medidas de redes complexas foram extraídas e por meio de um alguns classificadores de AM, foi realizada a predição de pacientes com a desordem em questão, com intuito de auxiliar em seu diagnóstico.

7.2 Base de dados

A base de dados utilizada nesta secção (assim como as outras bases dos dois capítulos seguintes) foram extraídas do repositório UMCD⁵ (em inglês, *USC Multimodal Connectivity Database*), fazendo parte do projeto *Human project Connectome*, que possui vários estudos, cujo o objetivo é desvendar a conexão do cérebro humano utilizando neuroimagens (BROWN; HORN, 2016). Esses estudos possuem em comum o formato de seus dados, a matriz de conectividade com as regiões de interesse (ROIs) ao longo das linhas e colunas e a força da conectividade entre um dado par de ROIs armazenados na célula onde essas duas regiões se cruzam.

A base de dados em questão, de 60 crianças e adolescentes com ASD (52 homens e 8 mulheres) e 45 sem a doença (38 homens e 7 mulheres) recrutados pelo Centro de Pesquisa e Tratamento para Autismo (em inglês, *Center for Autism Research and Treatment*, CART) da Universidade da Califórnia em Los Angeles (em inglês, *University of California-Los Angeles*, UCLA). As imagens de fMRI foram adquiridas por meio de um scanner Simiens na UCLA, por meio de uma varredura eco-planar de alta resolução, sendo que em uma única sessão, os sujeitos

² Possui função executiva definida como um conjunto de habilidades que possibilitam o indivíduo a tomada de decisão, atenção, diferenciação de pensamentos conflitantes entre outros (EUSTON; GRUBER; MCNAUGHTON, 2012)

³ Região cuja função engloba evocação de memória e aprendizado.

⁴ Função cognitiva, emocional e de aprendizado.

⁵ Extraída de <http://umcd.humanconnectomeproject.org>

foram convidados a relaxarem e manterem os olhos abertos enquanto uma cruz de fixação foi exibida em um fundo branco por 6 min (RUDIE *et al.*, 2013). As imagens adquiridas foram pré-processadas por meio dos softwares *FSL* e *AFNI*. Os objetos com uma alta movimentação foram excluídos, restando ao fim 79 amostras (42 com ASD e 37 normais). Utilizou-se um filtro passa banda ($0.1 \text{ Hz} > t > 0.01 \text{ Hz}$) para minimizar as flutuações cardíacas e respiratórias.

O esquema de parcelamento do cérebro utilizado foi o mesmo que em (POWER *et al.*, 2011), em que um conjunto de 264 regiões funcionais foram mostradas com mais precisão, portanto, essas regiões foram escolhidas e registradas no espaço funcional. As séries temporais dessas regiões foram correlacionadas pelo coeficiente de z-transformação, gerando uma matriz de conectividade 264×264 .

7.3 Metodologia

Nessa secção será formalizada a metodologia utilizada para a análise da presente base de dados. Seguiu-se uma metodologia semelhante a do capítulo 6, divergindo em alguns pontos, uma vez que as matrizes de conectividade disponíveis para esta base eram com peso. Como foi feito em (RUDIE *et al.*, 2013) (que utilizou a mesma base) a matriz foi binarizada, utilizando para isso um *threshold* variando de 0.15 (que segundo o artigo, era correlação mínima necessária para ser estatisticamente significante) a 0.34. Assim, com intuito de verificar se os thresholds interferiam na predição da desordem, foram escolhidos os seguintes valores: 0.3, 0.4, 0.5 e 0.6, assim gerando grafos binários, sendo 79 grafos diferentes para cada um dos valores.

Extraiu-se as medidas de redes, por meio do pacote *igraph*, semelhante ao que foi feito no capítulo 6. As medidas escolhidas foram média da distribuição de grau, segundo momento da distribuição de grau, coeficiente médio de aglomeração, transitividade, assortatividade, média dos menores caminhos, complexidade, betweenness centrality, eigenvector centrality, closeness centrality, pageRank, diâmetro e dominância central, baseadas no capítulo anterior e nos artigos em questão e em (RUDIE *et al.*, 2013; ZHUKOV; PETROV; DODONOVA, 2015). Assim 4 arquivos no formato csv (cada um correspondendo a um threshold), sendo suas colunas as medidas de redes mais a classe (ASD ou TD) e o número de linhas iguais a quantidade de amostras (79 objetos). Os arquivos foram submetidos aos mesmos classificadores do capítulo anterior.

Nesta primeira etapa o threshold que melhor aumentou a performance dos classificadores foi selecionado. E os classificadores que obtiveram as melhores performances são selecionados para próxima etapa que consiste na etapa de seleção e normalização. As seleções e normalizações utilizadas foram aquelas que aumentaram a performance dos classificadores selecionados do capítulo anterior, ou seja, as seleções: por ordem de importância, remoção de características redundantes e RFE; e as normalizações: softmax scaling e z-score. Os resultados obtidos se encontram na seção seguinte 7.4.

7.4 Resultados

7.4.1 Resultados após variar o threshold

Os resultados obtidos após variar o threshold se encontram na tabela 22. Da tabela, percebe-se que os resultados da predição foram inferiores aos obtidos pela esquizofrenia, assim como critério para a próxima etapa, utilizou-se um valor de acurácia superior a 0.6 e de AUC superior a 0.6. A tabela 23 contém os melhores desempenhos obtidos para cada classificador e o threshold correspondente a esse desempenho. Assim os classificadores considerados serão: árvore de decisão; combinação k-NN e árvore de decisão; combinação entre k-NN, redes neurais e árvore de decisão; combinação entre k-NN, naive bayes e redes neurais. E o threshold escolhido foi o de 0.6, uma vez que foi esse que mais aumentou o desempenho dos classificadores.

7.4.2 Resultados após seleção

Assim aos classificadores selecionados anteriormente foram submetidos aos processos de seleção: por ordem de importância, remoção de características redundantes e RFE.

Primeiramente utilizou-se o processo de remoção de características por importância, a tabela 24 contém as medidas e sua respectiva importância obtida, percebe-se que menores foram o pageRank e a assortatividade; assim primeiramente foi excluída a assortatividade, seguida pela medida pageRank. Os resultados após esse processo de seleção se encontram na tabela 25.

Tabela 24 – Tabela contendo as medidas e sua respectiva importância. Em rosa destacou-se aquelas que obtiveram a menor medida medida de importância.

Medidas	Importâncias
Complexidade	0.60
segundo momento da distribuição de grau	0.60
grau médio	0.59
média dos menores caminhos	0.56
diâmetro	0.55
transitividade	0.55
eigenvector centrality	0.55
closeness centrality	0.54
betweenness centrality	0.53
coeficiente de aglomeração médio	0.52
dominancia.central	0.51
pageRank	0.51
assortatividade	0.51

Da tabela 25, percebe-se que a seleção só aumentou a performance do classificador combinação de classificadores k-NN, redes neurais e árvore de decisão; sendo que para o restante dos classificadores esse processo de seleção contribuiu para a piora de seus desempenhos.

Tabela 22 – Tabela contendo os desempenhos de cada cada classificador ao se variar o threshold.

	Threshold	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
k-NN	0.3	0.48	-0.08	0.61	0.23	0.47
	0.4	0.55	0.047	0.69	0.28	0.44
	0.5	0.47	-0.07	0.71	0.31	0.49
	0.6	0.47	-0.07	0.61	0.17	0.39
Naive Bayes	0.3	0.58	0.13	0.74	0.43	0.52
	0.4	0.54	0	1	0	0.50
	0.5	0.55	0.03	0.98	0.03	0.49
	0.6	0.5	0.09	1	0.09	0.55
Árvore de decisão	0.3	0.64	0.25	0.90	0.43	0.66
	0.4	0.58	0.12	0.82	0.44	0.62
	0.5	0.63	0.21	0.96	0.23	0.5925
	0.6	0.66	0.29	0.88	0.38	0.66
Redes neurais	0.3	0.56	0.06	0.73	0.36	0.55
	0.4	0.54	0.03	0.69	0.41	0.54
	0.5	0.63	0.23	0.63	0.38	0.52
	0.6	0.59	0.14	0.77	0.25	0.57
k-NN e RN	0.3	0.55	0.1	0.86	0.22	0.53
	0.4	0.56	0.11	0.71	0.44	0.66
	0.5	0.56	0.11	0.71	0.44	0.66
	0.6	0.61	0.20	0.86	0.33	0.54
k-NN e AD	0.3	0.66	0.29	0.65	0.19	0.52
	0.4	0.59	0.15	0.78	0.35	0.63
	0.5	0.50	-0.08	0.69	0.28	0.54
	0.6	0.64	0.23	0.90	0.35	0.69
k-NN e NB	0.3	0.54	0.05	0.81	0	0.25
	0.4	0.65	0.28	0.76	0.31	0.57
	0.5	0.59	0.12	0.81	0.23	0.59
	0.6	0.54	0.05	0.92	0.13	0.51
AD e RN	0.3	0.65	0.28	0.83	0.04	0.33
	0.4	0.58	0.14	0.75	0.49	0.57
	0.5	0.53	-0.01	0.77	0.46	0.60
	0.6	0.63	0.22	0.74	0.39	0.52
AD e NB	0.3	0.61	0.20	0.74	0.47	0.60
	0.4	0.50	-0.06	0.72	0.55	0.54
	0.5	0.52	-0.02	0.90	0.12	0.47
	0.6	0.65	0.25	0.92	0.32	0.56
NB e RN	0.3	0.60	0.17	0.74	0.48	0.60
	0.4	0.56	0.06	0.92	0	0.46
	0.5	0.53	-0.01	0.86	0.16	0.46
	0.6	0.57	0.09	0.88	0.26	0.62
k-NN, RN e NB	0.3	0.53	0.02	0.71	0.41	0.61
	0.4	0.61	0.19	0.78	0.22	0.46
	0.5	0.59	0.18	0.80	0.34	0.66
	0.6	0.57	0.08	0.8	0.35	0.64
k-NN,RN e AD	0.3	0.55	0.08	0.76	0.49	0.60
	0.4	0.60	0.16	0.69	0.22	0.51
	0.5	0.54	0.05	0.77	0.42	0.67
	0.6	0.62	0.20	0.93	0.34	0.63
NB , RN e AD	0.3	0.59	0.16	0.75	0.45	0.65
	0.4	0.60	0.18	0.76	0.45	0.69
	0.5	0.59	0.14	0.73	0.48	0.60
	0.6	0.59	0.15	0.75	0.41	0.64
k-NN, NB e AD	0.3	0.65	0.28	0.75	0.55	0.69
	0.4	0.60	0.17	0.74	0.43	0.67
	0.5	0.60	0.17	0.80	0.48	0.63
	0.6	0.59	0.13	0.74	0.45	0.63
k-NN, RN, AD e NB	0.3	0.67	0.31	0.65	0.43	0.54
	0.4	0.63	0.24	0.70	0.32	0.46
	0.5	0.59	0.13	0.81	0.26	0.58
	0.6	0.58	0.13	0.76	0.46	0.63

Tabela 23 – Tabela contendo as melhores performances para cada classificador. Em vermelho está destacado aqueles que possuem acurácia e AUC superior a 0.6; e em negrito o threshold que resultou na maioria do aumento de performance .

	Melhor threshold	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
k-NN	0.4	0.57	0.05	0.69	0.28	0.44
Naive Bayes	0.3	0.58	0.13	0.74	0.43	0.54
Árvore de decisão	0.6	0.67	0.30	0.98	0.26	0.62
Redes neurais	0.5	0.63	0.23	0.63	0.38	0.52
k-NN e RN	0.6	0.62	0.20	0.90	0.33	0.54
k-NN e AD	0.6	0.64	0.23	0.90	0.35	0.69
k-NN e NB	0.4	0.65	0.29	0.76	0.31	0.57
AD e RN	0.6	0.63	0.22	0.74	0.39	0.52
AD e NB	0.6	0.65	0.25	0.92	0.32	0.56
NB e RN	0.6	0.57	0.09	0.88	0.26	0.63
k-NN,RN e NB	0.5	0.60	0.18	0.80	0.34	0.66
k-NN,RN e AD	0.6	0.62	0.20	0.92	0.34	0.63
NB , RN e AD	0.3	0.59	0.16	0.75	0.45	0.65
k-NN,NB e AD	0.3	0.65	0.28	0.75	0.55	0.69
k-NN,RN, AD e NB	0.3	0.67	0.31	0.65	0.43	0.54

Tabela 25 – Tabela contendo os resultados obtidos após o processo de seleção por importância extraindo-se primeiramente a assortatividade, seguida pela medida pageRank. Em negrito, destacou-se os processos de seleção que mais aumentaram o desempenho de cada um dos classificadores.

		Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	com todas as medidas	0.67	0.30	0.98	0.26	0.62
	sem assortatividade	0.68	0.23	0.96	0.30	0.63
	sem assortatividade e pageRank	0.67	0.28	0.87	0.34	0.63
k-NN e AD	com todas as medidas	0.64	0.23	0.90	0.35	0.69
	sem assortatividade	0.67	0.28	0.95	0.32	0.58
	sem assortatividade e pageRank	0.67	0.29	0.94	0.33	0.61
k-NN, RN e AD	com todas as medidas	0.62	0.20	0.93	0.34	0.63
	sem assortatividade	0.61	0.18	0.90	0.37	0.71
	sem assortatividade e pageRank	0.61	0.17	0.88	0.38	0.71
k-NN, NB e RN	com todas as medidas	0.6522222	0.28	0.75	0.55	0.69
	sem assortatividade	0.62	0.19	0.82	0.31	0.57
	sem assortatividade e pageRank	0.59	0.14	0.84	0.35	0.55

Tabela 26 – Tabela contendo os resultados obtidos após submeter os classificadores ao processo de remoção de características com redundância superior a 70%. Percebe-se que esse processo de seleção não aumentou o desempenho de nenhum dos classificadores.

		Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	com todas as medidas	0.67	0.30	0.98	0.26	0.62
	remoção de características redundantes	0.65	0.25	0.89	0.32	0.61
k-NN e AD	com todas as medidas	0.63	0.23	0.90	0.35	0.69
	remoção de características redundantes	0.64	0.23	0.73	0.48	0.63
k-NN, RN e AD	com todas as medidas	0.62	0.20	0.93	0.34	0.63
	remoção de características redundantes	0.57	0.11	0.90	0.28	0.46
k-NN, NB e RN	com todas as medidas	0.65	0.28	0.75	0.56	0.69
	remoção de características redundantes	0.61	0.19	0.73	0.12	0.38

Em seguida, realizou-se a remoção de atributos redundantes, eliminando-se aqueles com correlação superior a 70%, os resultados se encontram na tabela 26. Percebe-se desta tabela que esse processo de seleção não contribuiu para aumentar a performance de nenhum dos classificadores.

Por fim, realizou-se o processo de seleção remoção recursiva de atributos. Como foi visto

anteriormente, esse tipo de seleção é configurado para explorar todos os grupos de atributos possíveis. O grupo que obteve o menor erro foi o que continha nove variáveis, sendo elas: *segundo momento da distribuição de grau, complexidade, grau.médio, assortatividade, diâmetro, coeficiente de aglomeração médio, média dos menores caminhos, closeness centrality, pageRank*. Dessa forma, utilizaram-se essas nove medidas para realizar a predição com os classificadores anteriores. Os resultados se encontram na tabela 27.

Tabela 27 – Tabela contendo os resultados obtidos após o processo de seleção RFE.

		Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	com todas as medidas	0.67	0.30	0.98	0.27	0.63
	RFE	0.68	0.33	0.95	0.33	0.64
k-NN e AD	com todas as medidas	0.64	0.23	0.90	0.35	0.69
	RFE	0.65	0.25	0.94	0.31	0.56
k-NN, RN e AD	com todas as medidas	0.62	0.20	0.93	0.34	0.63
	RFE	0.65	0.25	0.79	0.30	0.59
k-NN, NB e RN	com todas as medidas	0.65	0.28	0.75	0.56	0.69
	RFE	0.61	0.16	0.90	0.32	0.45

Da tabela 27, percebe-se que o processo de seleção RFE aumentou a performance somente do algoritmo árvore de decisão. Os melhores desempenhos serão utilizados na próxima etapa de normalização da seção seguinte.

7.5 Normalização

Após o processo de seleção foram realizadas as duas normalizações: *z-score* e *softmax scaling* (elas foram escolhidas por terem sido as únicas que contribuíram para o aumento do desempenho dos classificadores na base relacionada a esquizofrenia), cujos resultados se encontram na tabela 28. Percebe-se que somente o *z-score* aumentou a performance do classificador árvore de decisão, sendo que para todos os outros classificadores o desempenho piorou com o uso das normalizações. A tabela 29 contém o resumo dos melhores desempenhos obtidos e os processos que contribuíram para isso.

Tabela 28 – Tabela contendo os resultados após realizar os processos de normalização. Em negrito, destacou-se as melhores performance obtidas, e com as cores laranja e rosa, os resultados obtidos, respectivamente, para as normalizações softmax scaling e z-score.

	Processo de seleção	Normalização	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	RFE	sem normalização	0.68	0.33	0.94	0.33	0.64
		softmax scaling	0.67	0.29	0.94	0.39	0.66
		z-score	0.68	0.34	0.96	0.37	0.67
k-NN e AD	sem seleção	sem normalização	0.64	0.23	0.90	0.35	0.69
		softmax scaling	0.58	0.12	0.86	0.29	0.61
		z-score	0.63	0.22	0.79	0.37	0.68
k-NN, RN e AD	seleção por ordem de importância sem assortatividade	sem normalização	0.61	0.81	0.90	0.37	0.70
		softmax scaling	0.59	0.16	0.66	0.39	0.57
		z-score	0.59	0.14	0.64	0.53	0.65
k-NN, NB e RN	sem seleção	sem normalização	0.65	0.28	0.75	0.56	0.69
		softmax scaling	0.61	0.21	0.69	0.37	0.51
		z-score	0.62	0.21	0.64	0.45	0.56

Tabela 29 – Tabela contendo o resumo dos desempenhos obtidos pelos classificadores em questão e os processos que contribuíram para isso.

	Processo de Seleção	Processo de normalização	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	RFE	z-score	0.68	0.34	0.96	0.38	0.67
k-NN e AD	sem seleção	sem normalização	0.64	0.23	0.90	0.35	0.69
k-NN, RN e AD	seleção por ordem de importância sem assortatividade	sem normalização	0.61	0.81	0.90	0.37	0.70
k-NN, NB e AD	sem seleção	sem normalização	0.65	0.28	0.75	0.56	0.69

7.6 Conclusão

A tabela 29 contém o resumo dos resultados obtidos nesta seção, em que pode ser observado que o maior desempenho obtido foi para o classificador obtido da combinação k-NN, RN e AD, que apesar de ter resultado na menor acurácia, obteve as outras maiores medidas. Quanto à normalização, somente a z-score contribui para o aumento da acurácia para o classificador árvore de decisão; e, ao método de seleção, somente o RFE e seleção por ordem de importância que contribuíram, respectivamente, para o aumento da performance dos classificadores: árvore de decisão e aquele obtido da combinação dos classificadores k-NN, RN e AD.

A sensibilidade que mede a taxa de verdadeiro positivo que se refere às amostras da classe ASD, ou seja, aquelas que apresentam a doença, foi superior à taxa especificidade. Sendo a maior sensibilidade obtida foi de 0.96 pela árvore de decisão.

Em relação à literatura, em (PETROV *et al.*, 2016), utilizando a mesma base e modelo de validação (validação cruzada) do presente trabalho, porém considerando diferentes medidas de rede (Coeficiente de aglomeração com peso, Coeficiente de aglomeração médio normalizado, comprimento de caminho característico, comprimento de característica normalizado, modularidade, e coeficiente de mundo pequeno) e classificadores (regressão logística, Máquina de vetores de suporte com kernel linear, *Random forest*, e árvore de decisão com a técnica *boosting*), a maior AUC obtida foi de 0.77 (pelo classificador Máquina de vetores de suporte), e a pior 0.57 (pela árvore de decisão com a técnica *boosting*). Comparando, apesar dos classificadores serem diferentes, a maior AUC obtida (0.71) é bem próxima à do artigo e pelo fato de ter sido utilizado outras medidas, torna o modelo obtido mais confiável. Em outro artigo (ZHUKOV; PETROV; DODONOVA, 2015), também foi utilizada a mesma base e várias medidas de rede (grau médio, knn médio, closennes centrality, betweenness centrality, eigenvector centrality, coeficiente de aglomeração com peso, número de triângulo ao redor do nó, densidade de rede, coeficiente de assortatividade, tamanho máximo de clique, transitividade e diâmetro) classificadas pela máquina de vetores de suporte, e foram obtidos os seguintes desempenhos: 0.64 de acurácia, 0.61 de precisão e 0.88 de revocação (também denominada de sensibilidade). Apesar dos classificadores serem diferentes, a árvore de decisão obteve uma acurácia e sensibilidade superior ao do artigo em questão.

REDES CEREBRAIS DE PACIENTES COM DEFICIT DE ATENÇÃO/DESORDEM DE HIPERATIVIDADE

8.1 Introdução

Déficit de atenção/desordem de hiperatividade (em inglês, *Attention Deficit/Hyperactivity Disorder, ADHD*) é um transtorno muito prevalente no mundo, sendo na maioria das vezes observado na infância e persistindo na fase adulta ([SATO *et al.*, 2012](#)). Os sintomas clínicos dessa desordem incluem desatenção, hiperatividade e impulsividade ([POLANCZYK *et al.*, 2007](#)), afetando o processo de cognição, emocional e motor ([ANDERSON *et al.*, 2014](#)).

Trata-se, também, de outra desordem associada ao interrompimento da conectividade funcional ([SPORNS, 2010](#)). Assim como em outros distúrbios neuropsiquiátricos (como no autismo e esquizofrenia), os modelos emergentes mudaram o foco da suposição de anomalias patológicas em regiões cerebrais para disfunção na distribuição da organização da rede ([KONRAD; EICKHOFF, 2010](#)). Em termos de conectividade funcional, em ([CASTELLANOS *et al.*, 2008](#)) foi utilizada uma base de dados proveniente da ressonância magnética funcional de 20 adultos com ADHD e 20 voluntários saudáveis, foi revelado que no grupo controle há presença de uma relação negativa do córtex anterior com outras componentes da rede; no grupo com ADHD houve diminuição da conectividade funcional entre o córtex anterior e posterior, e, também, da conexão do córtex com outras componentes da rede. Assim em termos comportamentais, esse efeito reflete em déficit de desempenho periódicos e transitórios ([KONRAD; EICKHOFF, 2010](#)).

Em termos de topologia de rede, em ([WANG *et al.*, 2009](#)), com uma base de dados também proveniente de ressonância magnética funcional de 19 crianças com ADHD e 20 saudáveis, demonstraram que, apesar de em ambos os grupos apresentarem uma topologia de pequeno mundo, foram encontradas algumas alterações no cérebro dos pacientes com a doença,

sendo a mais notável, o aumento da eficiência local e diminuição da eficiência global.

No entanto, apesar do grande aumento do número de estudos (principalmente relacionados a conectividade funcional e estrutural do cérebro de pacientes com o distúrbio), em termos clínicos, o diagnóstico do ADHD é baseado em critérios subjetivos (ANDERSON *et al.*, 2014) (presença de pelo menos dois sintomas observados (ASSOCIATION, 2002)) e não confiáveis (SATO *et al.*, 2012), sendo muitas vezes confundidos com outras doenças.

No presente trabalho empregou-se medidas de redes e algoritmos de aprendizado de máquina em uma base de dados relacionada a disfunção em questão (descrita na seção 8.2, a seguir) com intuito de discriminar pacientes com ADHD de pacientes saudáveis e com isso auxiliar no diagnóstico da doença de forma não-invasiva.

8.2 Base de dados

A base de dados¹ utilizada consiste em 520 matrizes de conectividade obtidas de dois tipos de grupos: aqueles que possuíam déficit de atenção/hiperatividade (DAHD) de desordem (190 pacientes) e o controle saudável (330 pessoas). Os pacientes com doença, ainda, eram subdivididos em três classes: DAHD-desatento (em inglês, *ADHD-Inattentive*, com 74 pacientes), em que os pacientes possuem sintomas como desatenção e dificuldade em se concentrar em atividades que exijam atenção; DAHD-Hiperativo/impulsivo (em inglês, *ADHD-Hyperactive/Impulsive*, 7 pacientes), em que os sintomas são caracterizados como inquietação, impaciência e instabilidade de humor; e, DAHD-combinado que apresenta simultaneamente os sintomas dos dois tipos anteriores (em inglês, *ADHD-Combined*, 109 pacientes). Contudo, os três tipos de classes foram aglutinados em uma única classe denominada de *ADHD*, uma vez que o objetivo principal era discriminar entre pacientes com a doença e aqueles que estão desenvolvendo.

A base originalmente era do estudo *ADHD200 Consortium*² em que foram obtidas 759 imagens de Ressonância magnética funcional de crianças e adolescentes tipicamente em desenvolvimento e com o distúrbio, provenientes de oito sites diferentes (Univerdidade de Peking, Hospital da universidade Brown, Instituto Kennedy Krieger, Centro de estudo infantil da Unviversidade de Nova York, Univerdidade de saúde e ciência de Oregon, Universidade de Pittsburgh e Universidade de Washington) (SATO *et al.*, 2012). O pré-processamento das imagens foi novamente feito com os software *AFN* e *FSL* e pode ser resumida nos passos: remoção dos quatro primeiros volumes de EPI, correção de fatia temporal, realinhamento, correção de movimento, co-registro da imagem de EPI média para a imagem anatomia correspondente e normalização para o espaço temporário (LIANG *et al.*, 2012). Em seguida, é aplicado um filtro passa banda (cujas frequências se encontram no intervalo $0.009 < f < 0.08$ Hz), com intuito de remover aquelas que não são usadas na análise da conectividade funcional em estado de repouso.

¹ Disponível em: <http://umcd.humanconnectomeproject.org/umcd/default/update/520>

² Disponível em <http://fcon1000.projects.nitrc.org/indi/adhd200/>.

As regiões de interesses são obtidas baseada na análise de correlação ou análise das componentes independentes, utilizando uma série de seleção tipo filtro (CRADDOCK *et al.*, 2009), é obtido, ao final, 190 regiões de interesse.

Para obter a correlação entre as áreas de interesse foi utilizado o método da semente usando a correlação de *Pearson* (LIANG *et al.*, 2012).

8.3 Metodologia

Foi utilizada a mesma metodologia da base relacionada ao autismo (capítulo anterior), uma vez que também se trata de uma base de dados extraída do repositório de base de dados *USC Multimodal Connectivity Database*, sendo as matrizes de conectividade com peso. Assim, com o programa *python*, as matrizes de conectividade com peso foram transformadas em binárias como em outros estudos utilizando a base de dados em questão (CAO *et al.*, 2014; GUO *et al.*, 2014; DEY; RAO; SHAH, 2012; BOHLAND *et al.*, 2012; CHENG *et al.*, 2012), para isso foi realizado a normalização *z-score* e utilizado o processo de *thresholding* e binarização com os valores de 0.3, 0.4, 0.5 e 0.6 (em (CAO *et al.*, 2014) utilizou esse valor); como mencionado em (CAO *et al.*, 2014), os valores de escolha para transformar os grafos em binários é muito subjetivo, assim essa faixa de valores foi escolhida para avaliar quais deles contribuem para melhor a performance desses classificadores.

Dessa forma a cada um desses valores, é gerada uma matriz binária diferente e em cada uma delas são extraídas as mesmas medidas de rede utilizadas na base relacionada ao autismo (média da distribuição de grau, segundo momento da distribuição de grau, coeficiente médio de aglomeração, transitividade, assortatividade, média dos menores caminhos, complexidade, *betweenness centrality*, *eigenvector centrality*, *closeness centrality*, *pageRank*, diâmetro e dominância central); estas medidas são classificadas pelos mesmos algoritmos de aprendizado de máquina. O *threshold* e os melhores classificadores são selecionados para a próxima etapa de seleção e normalização (sendo os métodos utilizados o mesmo da base relacionada ao autismo).

8.4 Resultados

8.4.1 Resultados dos classificadores variando o *threshold* da rede

Assim, seguindo a Metodologia descrita, primeiramente variou-se o *threshold* para verificar se ele interfere na performance dos classificadores em questão, sendo que o resultado obtido se encontra na tabela 22.

A tabela 31 contém o *threshold* que mais aumentou a performance de cada classificador. Percebe-se que o *threshold* 0.6 foi o que aumentou o desempenho da maioria dos classificadores, dessa forma foi o escolhido. E, ainda, foram selecionados os classificadores para a próxima

etapa, segundo o critério de acurácia superior a 0.6 (mesmo valor de acurácia utilizado para a base relacionada ao autismo) e o de AUC superior a 0.5, pois obteve-se menores valores para esta medida em relação aos obtidos na base relacionada ao autismo. Com esse critério, foram escolhidos os classificadores destacados em rosa na tabela 31.

8.4.2 Resultado após seleção

Os classificadores obtidos anteriormente, foram, então, submetidos aos dois processos de seleção: seleção por ordem de importância e RFE (os dois processos que contribuíram para aumentar o desempenho da base relacionada ao autismo). Na seleção por ordem de importância, a medida de menor importância foi o pageRank seguida pelo diâmetro; assim primeiramente a medida pageRank foi excluída e posteriormente o diâmetro, e assim verificou-se se essa redução de dimensionalidade contribuiu para um aumento da performance dos algoritmos preditivos.

Quanto ao processo RFE, o conjunto escolhido por esse método de seleção foram: eigenvector centrality, closeness centrality, assortatividade, média dos menores caminhos e coeficiente de aglomeração médio.

Os resultados ao se utilizar os dois métodos de seleção se encontra na tabela 32. Para cada classificador, foi considerado o conjunto de atributos que mais aumentou seu desempenho para a etapa de normalização.

8.4.3 Resultados após normalização

Assim aplicou-se a normalização aos conjuntos de atributos que mais aumentou a performance de cada classificador. As duas normalizações utilizadas foram o z-score e softmax scaling, que foram os dois métodos que mais contribuíram para aumentar o desempenho dos classificadores na base relacionada à esquizofrenia. Os resultados se encontram na tabela 33. Percebe-se que o desempenho de todos os classificadores aumentou com a utilização das normalizações, sendo que o método softmax scaling foi o que contribuiu na maioria deles.

8.5 Conclusão

A tabela 34 contém o resumo dos resultados obtidos neste capítulo, nela pode ser observado que o classificador obtido da combinação árvore de decisão e naive bayes, apesar de vários outros classificadores terem obtido uma performance parecida.

A especificidade, que mede a taxa de verdadeiros negativos (neste caso trata-se da classe dos portadores de ADHD) é superior à taxa de verdadeiros positivos (sensitividade).

Ainda pode-se verificar que para todos os classificadores os métodos de seleção contribuíram para o aumento de seus desempenhos, principalmente o RFE. Quanto a normalização,

Tabela 30 – Tabela contendo o resultado obtido ao se variar o threshold de 0.3, 0.4, 0.5 e 0.6, que foram destacados, respectivamente, pelas cores: rosa, laranja, verde e azul. Em negrito destacou-se o threshold que resultou na maior performance (em termos da acurácia e AUC).

	Threshold	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
k-NN	0.3	0.58	0.01	0.30	0.74	0.53
	0.4	0.58	-0.01	0.21	0.75	0.47
	0.5	0.56	-0.05	0.18	0.76	0.44
	0.6	0.58	-0.02	0.16	0.80	0.47
Naive Bayes	0.3	0.63	0.00	0.00	1.00	0.50
	0.4	0.60	0.02	0.14	0.85	0.50
	0.5	0.38	-0.02	0.92	0.09	0.52
	0.6	0.36	-0.01	0.98	0.02	0.53
Árvore de decisão	0.3	0.63	-0.01	0.43	0.60	0.51
	0.4	0.62	0.02	0.20	0.86	0.55
	0.5	0.63	0.01	0.14	0.85	0.51
	0.6	0.61	-0.03	0.18	0.83	0.55
Redes neurais	0.3	0.64	0.01	0.00	1.00	0.50
	0.4	0.63	0.00	0.01	0.99	0.51
	0.5	0.63	0.00	0.00	1.00	0.50
	0.6	0.65	0.051	0.03	0.98	0.50
k-NN e RN	0.3	0.63	0.00	0.00	1.00	0.54
	0.4	0.63	0.01	0.00	1.00	0.52
	0.5	0.63	0.02	0.00	1.00	0.51
	0.6	0.63	-0.01	0.04	0.99	0.54
k-NN e AD	0.3	0.63	0.00	0.00	1.00	0.53
	0.4	0.63	0.00	0.00	1.00	0.50
	0.5	0.62	0.01	0.00	0.99	0.52
	0.6	0.63	0.00	0.00	0.99	0.55
k-NN e NB	0.3	0.63	0.00	0.00	1.00	0.48
	0.4	0.63	0.00	0.00	0.99	0.52
	0.5	0.64	0.04	0.00	1.00	0.52
	0.6	0.63	0.00	0.00	1.00	0.51
AD e RN	0.3	0.63	0.00	0.00	1.00	0.47
	0.4	0.63	-0.01	0.01	0.98	0.49
	0.5	0.63	-0.01	0.00	0.99	0.48
	0.6	0.64	0.021	0.05	0.99	0.53
AD e NB	0.3	0.63	0.00	0.00	1.00	0.49
	0.4	0.63	0.00	0.00	1.00	0.54
	0.5	0.63	0.00	0.00	1.00	0.50
	0.6	0.63	0.00	0.016	0.99	0.53
NB e RN	0.3	0.63	0.00	0.02	0.98	0.52
	0.4	0.63	0.00	0.00	1.00	0.48
	0.5	0.63	0.00	0.00	1.00	0.40
	0.6	0.63	0.00	0.00	1.00	0.47
k-NN,RN e NB	0.3	0.63	-0.01	0.00	1.00	0.49
	0.4	0.63	-0.01	0.03	0.99	0.48
	0.5	0.62	-0.01	0.00	0.99	0.51
	0.6	0.63	-0.01	0.00	1.00	0.47
k-NN,RN e AD	0.3	0.64	0.01	0.00	1.00	0.52
	0.4	0.63	0.00	0.01	0.99	0.54
	0.5	0.63	0.01	0.00	0.99	0.53
	0.6	0.64	0.01	0.02	0.99	0.42
NB , RN e AD	0.3	0.63	0.00	0.02	1.00	0.54
	0.4	0.64	0.00	0.01	0.99	0.54
	0.5	0.63	0.01	0.00	0.99	0.53
	0.6	0.64	0.01	0.02	0.99	0.42
k-NN,NB e AD	0.3	0.63	-0.01	0.00	1.00	0.47
	0.4	0.63	-0.01	0.06	0.97	0.54
	0.5	0.63	0.01	0.00	1.00	0.46
	0.6	0.63	0.00	0.02	0.99	0.46
k-NN,RN, AD e NB	0.3	0.65	0.08	0.01	0.99	0.51
	0.4	0.63	0.01	0.00	1.00	0.49
	0.5	0.63	-0.01	0.04	0.98	0.54
	0.6	0.63	-0.01	0.03	0.99	0.48

Tabela 31 – Tabela contendo o os thresholds que mais contribuíram para o aumento dos classificadores em questão. Em rosa destacou-se os desempenho de acurácia maiores que 0.6 e AUC maiores que 0.5.

	Melhor threshold	Acurácia	Kappa (Ac.)	Sensitividade	Especificidade	AUC
k-NN	0.3	0.58	0.01	0.30	0.74	0.53
Naive Bayes	0.3	0.63	0.00	0.00	1.00	0.50
Árvore de decisão	0.5	0.63	0.01	0.14	0.85	0.51
Redes neurais	0.6	0.65	0.05	0.03	0.98	0.50
k-NN e RN	0.4	0.63	0.01	0.00	1.00	0.52
k-NN e AD	0.6	0.63	0.00	0.00	0.99	0.55
k-NN e NB	0.5	0.64	0.04	0.00	1.00	0.52
AD e RN	0.6	0.64	0.02	0.05	0.99	0.53
AD e NB	0.6	0.63	0.00	0.02	0.99	0.53
NB e RN	0.3	0.63	0.00	0.02	0.98	0.52
k-NN, RN e NB	0.5	0.62	-0.01	0.00	0.99	0.51
k-NN, NB e AD	0.6	0.62	0.01	0.02	0.99	0.42
NB, RN e AD	0.5	0.63	0.01	0.00	0.99	0.53
k-NN, NB e AD	0.6	0.63	0.00	0.02	0.99	0.46
k-NN, RN, AD e NB	0.3	0.65	0.08	0.01	0.99	0.52

principalmente o método softmax scaling, contribuíram para aumento da performance de todos os classificadores.

Comparando-se com a literatura, em (SIQUEIRA *et al.*, 2014) utilizando as medidas: grau, Closeness, Betweenness, Eigenvector e *Burt's constraint*, e o classificador SVM obteve-se uma acurácia de 61% (sendo que o valor de threshold utilizado foi de 0.25, e foi utilizado redes sem pesos); no caso do presente trabalho a performance obtida foi superior a este estudo.

Em (DEY; RAO; SHAH, 2014), foram utilizadas as medidas: densidade, eficiência global, *Rich club coefficient* e *High power node fraction*, e o classificador SVM, contudo considerou-se a base de cada região (Kennedy Krieger Institute, Neuro Image Sample, Oregon Health, Science University, Peking University (Peking)) de maneira separada, calculando-se depois a média das performances de cada região. Considerando o gênero feminino e masculino juntos, obteve-se uma média de: 64.48% de acurácia, 84.71% de especificidade e 30.66% de sensibilidade; considerando-se os gêneros de maneira separada obteve-se uma média de: 70.49% de acurácia, 84.53% de especificidade e 46.72% de especificidade. Assim a acurácia de alguns classificadores no presente trabalho foi superior a obtida considerando os dois gêneros e inferior àquela obtida ao se separar os gêneros, com isso a separação dos gêneros em trabalho futuro deve ser considerado. A especificidade, ainda, do presente trabalho foi superior à do artigo em questão.

Em (GUO *et al.*, 2014), utilizando segundo o artigo as medidas de redes tradicionais e características novas: *assortative mixing* e sincronização, obteve-se uma acurácia média de 63.75%, valor inferior a acurácia obtidas por alguns classificadores no presente trabalho. Por fim, ainda, o valor de acurácia máxima alcançada na competição *ADHD-200* foi de 61.04% cujo valor também é inferior aos valores obtidos no presente trabalho.

Tabela 32 – Tabela contendo os processos de seleção por ordem de importância e RFE, destacados em azul e laranja, respectivamente e o desempenho obtido ao se utilizar esses processos. Em negrito foi destacado as melhores performances obtidas.

Classificadores	Processo de seleção	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	sem seleção	0.63	0.01	0.14	0.85	0.52
	com todas as medidas	0.62	-0.02	0.43	0.67	0.56
	sem pageRank	0.62	-0.03	0.24	0.78	0.58
	sem pageRank e diâmetro	0.63	-0.01	0.18	0.83	0.53
Redes neurais	sem seleção	0.65	0.05	0.03	0.98	0.50
	com todas as medidas	0.65	0.04	0.03	0.98	0.56
	sem pageRank	0.65	0.05	0.02	0.99	0.56
	sem pageRank e diâmetro	0.64	0.01	0.1	0.96	0.57
k-NN e RN	sem seleção	0.63	0.01	0.04	0.99	0.54
	com todas as medidas	0.63	-0.01	0.00	0.99	0.55
	sem pageRank	0.64	0.01	0.03	0.99	0.54
	sem pageRank e diâmetro	0.64	0.05	0.03	0.99	0.57
k-NN e AD	sem seleção	0.63	-0.01	0.04	0.99	0.54
	com todas as medidas	0.63	0.00	0.02	1.00	0.54
	sem pageRank	0.63	0.00	0.00	1.00	0.49
	sem pageRank e diâmetro	0.63	-0.01	0.02	0.99	0.55
k-NN e NB	sem seleção	0.63	0.00	0.00	1.00	0.52
	com todas as medidas	0.64	0.01	0.00	0.99	0.53
	sem pageRank	0.64	0.04	0.00	0.99	0.53
	sem pageRank e diâmetro	0.62	-0.01	0.06	0.96	0.57
AD e NB	sem seleção	0.63	0.00	0.02	0.99	0.53
	com todas as medidas	0.64	0.035	0.037	0.99	0.50
	sem pageRank	0.62	-0.03	0.04	0.97	0.50
	sem pageRank e diâmetro	0.64	0.03	0.06	0.95	0.56
AD e RN	sem seleção	0.64	0.02	0.05	0.99	0.53
	com todas as medidas	0.64	0.02	0.01	1.00	0.40
	sem pageRank	0.65	0.05	0.01	0.99	0.54
	sem pageRank e diâmetro	0.64	0.03	0.04	0.98	0.56
NB e RN	sem seleção	0.63	0.00	0.00	1.00	0.47
	com todas as medidas	0.63	0.00	0.02	0.99	0.53
	sem pageRank	0.64	0.04	0.01	0.98	0.53
	sem pageRank e diâmetro	0.64	0.03	0.05	0.96	0.56
k-NN, RN e NB	sem seleção	0.63	-0.01	0.00	1.00	0.49
	com todas as medidas	0.64	0.03	0.03	0.99	0.52
	sem pageRank	0.65	0.04	0.01	1.00	0.52
	sem pageRank e diâmetro	0.63	0.01	0.07	0.96	0.57
NB, RN e AD	sem seleção	0.64	0.01	0.02	0.99	0.42
	com todas as medidas	0.63	-0.01	0.00	0.99	0.47
	sem pageRank	0.64	0.04	0.06	0.97	0.55
	sem pageRank e diâmetro	0.63	0.03	0.06	0.98	0.57
k-NN, RN, AD e NB	sem seleção	0.63	-0.01	0.01	0.99	0.48
	com todas as medidas	0.66	0.01	0.02	0.98	0.52
	sem pageRank	0.63	0.01	0.03	0.97	0.51
	sem pageRank e diâmetro	0.63	0.03	0.02	0.98	0.53

Tabela 33 – Resultados obtidos ao se aplicar os métodos de normalização softmax scaling e z-score, destacados, respectivamente, em laranja e rosa, aos conjuntos de atributos que contribuíram para aumentar o desempenho dos classificadores escolhidos. Em negrito, destacou-se a normalização que mais aumentou a performance de cada classificador.

Classificadores	Processo de seleção	Processo de normalização	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	sem seleção com todas as medidas	sem normalização	0.63	0.01	0.14	0.85	0.51
	sem pageRank	softmax scaling	0.62	-0.02	0.07	0.93	0.54
	sem pageRank e diâmetro	z-score	0.63	-0.00	0.10	0.90	0.52
Redes neurais	sem seleção com todas as medidas	sem normalização	0.65	0.05	0.03	0.98	0.50
	sem pageRank	softmax scaling	0.65	0.05	0.15	0.90	0.54
	sem pageRank e diâmetro	z-score	0.63	0.04	0.09	0.95	0.56
k-NN e RN	sem seleção com todas as medidas	sem normalização	0.63	0.01	0.04	0.99	0.54
	sem pageRank	softmax scaling	0.63	0.01	0.06	0.95	0.60
	sem pageRank e diâmetro	z-score	0.63	-0.01	0.03	0.98	0.56
k-NN e AD	sem seleção com todas as medidas	sem normalização	0.63	-0.01	0.04	0.99	0.54
	sem pageRank	softmax scaling	0.63	0.01	0.00	0.99	0.54
	sem pageRank e diâmetro	z-score	0.63	0.00	0.00	0.98	0.56
k-NN e NB	sem seleção com todas as medidas	sem normalização	0.63	0.00	0.00	1.00	0.51
	sem pageRank	softmax scaling	0.63	0.02	0.00	0.99	0.57
	sem pageRank e diâmetro	z-score	0.64	0.03	0.04	0.98	0.55
AD e NB	sem seleção com todas as medidas	sem normalização	0.63	0.00	0.02	0.99	0.53
	sem pageRank	softmax scaling	0.64	0.05	0.09	0.96	0.57
	sem pageRank e diâmetro	z-score	0.63	0.03	0.01	0.99	0.51
AD e RN	sem seleção com todas as medidas	sem normalização	0.64	0.02	0.05	0.99	0.53
	sem pageRank	softmax scaling	0.63	0.01	0.03	0.98	0.58
	sem pageRank e diâmetro	z-score	0.62	-0.05	0.07	0.96	0.57
NB e RN	sem seleção com todas as medidas	sem normalização	0.63	0.00	0.00	1.00	0.47
	sem pageRank	softmax scaling	0.65	0.08	0.06	0.96	0.58
	sem pageRank e diâmetro	z-score	0.63	-0.01	0.00	0.99	0.51
k-NN, RN e NB	sem seleção com todas as medidas	sem normalização	0.63	-0.01	0.00	1.00	0.49
	sem pageRank	softmax scaling	0.62	0.01	0.06	0.95	0.53
	sem pageRank e diâmetro	z-score	0.63	-0.01	0.01	0.98	0.51
NB, RN e AD	sem seleção com todas as medidas	sem normalização	0.64	0.01	0.02	0.99	0.42
	sem pageRank	softmax scaling	0.65	0.1	0.03	0.95	0.57
	sem pageRank e diâmetro	z-score	0.63	0.04	0.05	0.97	0.53
k-NN, RN, AD e NB	sem seleção com todas as medidas	sem normalização	0.63	-0.01	0.03	0.99	0.48
	sem pageRank	softmax scaling	0.64	0.04	0.14	0.93	0.61
	sem pageRank e diâmetro	z-score	0.63	0.02	0.15	0.93	0.60

Tabela 34 – Tabela contendo o resumo dos resultados obtidos nesta seção.

Classificadores	Processo de seleção	Processo de normalização	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore Decisão	seleção por ordem de importância sem pageRank	softmax scaling	0.62	-0.02	0.07	0.93	0.54
Redes Neurais	seleção por ordem sem pageRank e diâmetro	z-score	0.63	0.04	0.09	0.95	0.56
k-NN e RN	seleção por ordem sem pageRank	softmax scaling	0.63	0.01	0.06	0.95	0.60
k-NN e AD	seleção por ordem sem pageRank e diâmetro	z-score	0.63	0.00	0.00	0.98	0.56
k-NN e NB	sem pageRank	softmax scaling	0.63	0.02	0.00	0.99	0.57
AD e NB	seleção por ordem sem pageRank	softmax scaling	0.64	0.05	0.09	0.96	0.57
AD e RN	sem pageRank	softmax scaling	0.63	0.01	0.03	0.98	0.58
NB e RN	sem pageRank	softmax scaling	0.65	0.08	0.06	0.96	0.58
k-NN, RN e NB	sem pageRank	softmax scaling	0.62	0.01	0.06	0.95	0.53
NB, RN e AD	sem pageRank	softmax scaling	0.65	0.1	0.03	0.95	0.57
k-NN, RN, AD e NB	sem pageRank	softmax scaling	0.64	0.04	0.14	0.93	0.61

REDES CEREBRAIS DE PACIENTES COM PARALISIA SUPRANUCLEAR PROGRESSIVA

9.1 Introdução

A Paralisia Supranuclear Progressiva (PSP) foi descrita pela primeira vez em (STEELE; RICHARDSON; OLSZEWSKI, 1964), também conhecida como síndrome de Richardson, sendo uma doença rara que produz deterioração cognitiva progressiva, apraxia da pálpebra, perda neuronal, entre outros ¹ (OLIVEIRA; MUNARI; PELZER, 2010).

Em termos de rede, assim como as outras doenças no presente trabalho, é caracterizada por uma interrupção da rede em larga escala, ainda que o alvo específico da rede não tenha sido totalmente caracterizado (GARDNER *et al.*, 2013).

O diagnóstico clínico dessa doença é muito difícil (RAJPUT; RAJPUT, 2001), sendo ela muitas vezes confundida com a doença de Parkinson. O objetivo do presente capítulo, similar aos anteriores, foi propor uma forma de diagnóstico não invasiva baseada na classificação de medidas de redes complexas.

9.2 Base de dados

A base de dados ² continha matrizes de conectividade para cada uma correspondente a um dos pacientes que foram avaliados na Universidade da Califórnia de São Francisco (em inglês, *University of California*, UCSF), no centro de envelhecimento e Memória; sendo 20 deles diagnosticados com PSP segundo o critério determinado em (LITVAN *et al.*, 1996) e 12

¹ A Paralisia Supranuclear Progressiva possui essa denominação devido à degeneração progressiva das estruturas cerebrais localizadas na região superior dos núcleos oculomotores, e, assim, causando paralisia eventual dos movimentos oculares. (OLIVEIRA; MUNARI; PELZER, 2010)

² Disponível em: http://umcd.humanconnectomeproject.org/umcd/default/browse_studies

saudáveis, sendo que foi assegurado que a movimentação da cabeça fosse menor que 3mm da translação máxima, 3 graus de rotação relativa máxima, com níveis de movimentação da cabeça aceitáveis (BROWN *et al.*, 2017).

Todos os objetos foram scaneados no centro de imageamento e neurociência da UCSF no scanner Siemens Trio 3T, sendo o tempo de aquisição foi de 8 min 06s e a resolução do voxel: 2.5 x 2.5 x 3.0, resultando em um total de 240 volumes. Para cada scans de fMRI, o primeiro de cinco volumes foram descartados; foi utilizado para pré-processamento das imagens o software FSL. Utilizou-se, também, o método baseado na semente e a correlação de Spearman para determinar a conectividade funcional, e, ao final do processo, 27 regiões de interesses foram conectadas, ou seja, cada paciente possuía 2 matrizes de conectividade de dimensões 27 x 27 em dois momentos diferentes (T1 e T2, no presente trabalho foi utilizado somente as matrizes T2), todas com peso.

9.3 Metodologia

Esta seção contém a metodologia utilizada na base de dados descrita na seção anterior. Utilizou-se uma Metodologia similar à utilizada nos dois capítulos anteriores, uma vez que as matrizes de conectividade em questão possuíam peso. Contudo utilizou-se para threshold e binarização um valor de 0.3 baseado em (GARDNER *et al.*, 2013), após realizar a normalização z-score em cada matriz de conectividade (que também foi feito no artigo citado anteriormente), resultando em matrizes binárias.

Extraíu-se as mesmas medidas de redes anteriores (média da distribuição de grau, segundo momento da distribuição de grau, coeficiente médio de aglomeração, transitividade, assortatividade, média dos menores caminhos, complexidade, betweenness centrality, eigenvector centrality, closeness centrality, pageRank, diâmetro e dominância central), que serão os atributos a serem classificados pelos mesmos algoritmos anteriores, por intermédio do pacote *caret* do R. Por meio de alguns critérios são selecionados alguns classificadores que passarão para etapa de seleção, em que dois métodos de seleção (seleção por ordem de importância e RFE) são utilizados, com intuito de verificar se eles aumentam o desempenho dos algoritmos preditivos.

E por fim, são aplicados dois métodos de normalização (softmax scaling e z-score) para verificar se eles aumentam a performance dos classificadores em questão. A seção seguinte contém os resultados obtidos anteriormente.

9.4 Resultados

Esta seção contém os resultados obtidos neste capítulo após ter empregado a metodologia descrita na seção anterior.

Tabela 35 – Tabela contendo os resultados ao se aplicar os algoritmos preditivos ao conjunto de atributos correspondentes as medidas de redes extraídas das matrizes de conectividade binárias de pacientes com e sem a doença PSP. Em vermelho encontram-se destacados aqueles cuja acurácia e AUC foram superiores a 0.6.

	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
k-NN	0.45	-0.11	0.69	0.19	0.43
Naive Bayes	0.58	0.08	0.98	0.10	0.57
Árvore de decisão	0.65	0.26	1.00	0.26	0.63
Redes neurais	0.54	0.00	0.72	0.40	0.58
k-NN e RN	0.65	0.28	0.90	0.17	0.52
k-NN e AD	0.65	0.27	0.84	0.29	0.61
k-NN e NB	0.51	-0.04	0.72	0.53	0.65
AD e RN	0.64	0.23	0.72	0.45	0.58
AD e NB	0.64	0.22	0.90	0.27	0.60
NB e RN	0.54	0.02	0.84	0.32	0.66
k-NN,RN e NB	0.61	0.19	0.72	0.35	0.58
k-NN,RN e AD	0.62	0.20	0.67	0.44	0.61
NB , RN e AD	0.67	0.29	0.90	0.36	0.60
k-NN,NB e AD	0.63	0.25	0.88	0.36	0.67
k-NN,RN, AD e NB	0.67	0.32	0.92	0.36	0.57

Primeiramente, então, aplicou-se os algoritmos preditivos em questão cujos resultados se encontram na tabela 35. Aqueles que obtiveram uma acurácia e AUC superiores a 0.6 foram selecionados para a etapa de seleção cujos resultados se encontram na subsecção seguinte.

9.5 Resultados após seleção

Os dois processos de seleção foram utilizados no conjunto de atributos e, então, os classificadores selecionados da secção seguinte foram utilizados para discriminação das duas classes cujos resultados se encontram na tabela 36. As duas medidas com menor importância são a dominância central, seguida pelo diâmetro; assim, primeiramente retirou-se o atributo dominância central, realizou-se a classificação, e depois retirou-se o atributo diâmetro e, novamente, foi realizado a classificação. Ao se aplicar o método RFE obteve-se como subconjunto de atributos o seguinte: coeficiente de aglomeração médio, transitividade, complexidade, segundo momento da distribuição de grau e eigenvector centrality e nesse subconjunto realizou-se a classificação.

Da tabela 36 percebe-se que o RFE aumentou o desempenho dos classificadores árvore de decisão e o classificador obtido da combinação de classificadores k-NN, naive bayes e árvore de decisão. A seleção por ordem de importância sem a medida dominância central aumentou o desempenho do classificador formado da combinação entre os classificadores: naive bayes, redes neurais e árvore de decisão. Para os outros classificadores nenhum outro subconjunto de atributos contribuiu para aumentar sua performance.

Tabela 36 – Tabela contendo os resultados obtidos após utilizar os métodos de seleção. Em azul, destacou-se aqueles obtidos pelo método de seleção por ordem de importância; e, em laranja, aqueles obtidos pelo método RFE. Em negrito, destacou-se os melhores desempenhos (em termos da acurácia e AUC) para cada classificador em questão.

	Processo de Seleção	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	sem seleção	0.65	0.26	1.00	0.26	0.63
	seleção por ordem de importância	0.60	0.142	0.58	0.50	0.57
	sem dominância central					
	seleção por ordem de importância	0.67	0.19	0.55	0.68	0.64
	sem dominância central e diâmetro					
	RFE	0.66	0.19	0.68	0.48	0.68
k-NN e AD	sem seleção	0.651111	0.2732581	0.29	0.84	0.61
	seleção por ordem de importância	0.64	0.13	0.95	0.08	0.57
	sem dominância central					
	seleção por ordem de importância	0.61	-0.03	0.88	0.17	0.57
	sem dominância central e diâmetro					
	RFE	0.52	-0.19	1.00	0.00	0.38
k-NN, RN e AD	sem seleção	0.62	0.20	0.67	0.44	0.61
	seleção por ordem de importância	0.57	-0.10	0.95	0.03	0.44
	sem dominância central					
	seleção por ordem de importância	0.65	0.21	0.93	0.00	0.36
	sem dominância central e diâmetro					
	RFE	0.53	-0.13	0.80	0.17	0.60
NB, RN e AD	sem seleção	0.67	0.29	0.90	0.36	0.60
	seleção por ordem de importância	0.71	0.37	0.80	0.23	0.66
	sem dominância central					
	seleção por ordem de importância	0.67	0.27	0.75	0.28	0.64
	sem dominância central e diâmetro					
	RFE	0.58	0.08	0.78	0.40	0.62
k-NN, NB e AD	sem seleção	0.63	0.25	0.88	0.36	0.67
	seleção por ordem de importância	0.60	0.10	0.78	0.42	0.65
	sem dominância central					
	seleção por ordem de importância	0.58	-0.10	0.875	0.23	0.56
	sem dominância central e diâmetro					
	RFE	0.71	0.39	0.82	0.38	0.65

9.6 Resultados após normalização

Para finalizar, aplicou-se os dois métodos de normalização resultando na performance dos classificadores na tabela 37.

Na tabela 37, verifica-se que na maioria dos classificadores a normalização não contribui para aumentar sua performance; a softmax scaling aumentou o desempenho do classificador árvore de decisão e a z-score aumentou o desempenho do classificador obtido pela combinação dos classificadores: k-NN, RN e AD.

9.7 Conclusão

Da tabela 38, percebe-se que os processos de seleção para a maioria dos classificadores contribuiu para aumentar suas performances (em termos da acurácia e AUC). Enquanto os métodos de normalização somente para dois classificadores (árvore de decisão e classificador obtido da combinação de k-NN, RN e AD) contribui para aumentar sua performances.

Tabela 37 – Tabela contendo os resultados obtidos após aplicar os dois métodos de normalização em cada subconjunto resultante na secção anterior, e classificá-los com os algoritmos preditivos.

	Processo de Seleção	Processo de normalização	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	RFE	sem normalização	0.66	0.19	0.68	0.48	0.68
		softmax scaling	0.65	0.17	0.90	0.40	0.64
		z-score	0.64	0.18	0.80	0.43	0.59
k-NN e AD	sem seleção	sem normalização	0.65	0.27	0.84	0.29	0.61
		softmax scaling	0.60	0.08	0.80	0.28	0.60
		z-score	0.61	0.13	0.80	0.23	0.63
k-NN, RN e AD	sem seleção	sem normalização	0.62	0.20	0.67	0.44	0.61
		softmax scaling	0.66	0.24	0.75	0.13	0.61
		z-score	0.66	0.24	0.90	0.22	0.54
NB, RN e AD	seleção por ordem de importância sem dominância	sem normalização	0.71	0.39	0.80	0.23	0.66
		softmax scaling	0.64	0.18	0.88	0.28	0.53
		z-score	0.60	0.10	0.73	0.20	0.59
k-NN, NB e AD	RFE	sem normalização	0.71	0.39	0.82	0.38	0.65
		softmax scaling	0.58	0.01	0.78	0.35	0.64
		z-score	0.69	0.29	0.75	0.40	0.67

Tabela 38 – Tabela contendo o resumo dos resultados obtidos neste capítulo.

	Processo de Seleção	Processo de normalização	Acurácia	Kappa(Ac.)	Sensitividade	Especificidade	AUC
Árvore de decisão	RFE	softmax scaling	0.65	0.17	0.90	0.40	0.63
k-NN e AD	sem seleção	sem normalização	0.65	0.27	0.84	0.29	0.61
k-NN, RN e AD	sem seleção	z-score	0.66	0.24	0.90	0.22	0.54
NB, RN e AD	seleção por ordem de importância sem dominância central	sem normalização	0.71	0.37	0.80	0.23	0.65
k-NN, AD e NB	RFE	sem normalização	0.71	0.39	0.83	0.38	0.65

A taxa de sensibilidade (nesse caso estava relacionada à classe PSP) foi muito superior a especificidade (relacionada à classe de pacientes saudáveis). Dessa forma, a sensibilidade seria um forte indicativo para discriminação das duas classes.

Quanto à literatura, os artigos encontrados relacionados a essa base, como em: (BROWN *et al.*, 2017; GARDNER *et al.*, 2013; MANDELLI *et al.*, 2016), focaram-se na análise estrutural da rede dos pacientes com PSP, tentando identificar as regiões que mais diferem de uma rede de paciente sem a doença. No presente trabalho, focou-se em algo novo, no sentido de classificar medidas de rede na tentativa de distinguir as duas classes (de pacientes com PSP e pacientes saudáveis) e conseguir prever pacientes com PSP.

A tabela 38 contém o resumo de todos os resultados obtidos neste capítulo.

TRABALHOS FUTUROS

Uma possível complementação do trabalho desenvolvido seria a análise de processos dinâmicos das redes consideradas, considerando-se os *motifs*, para verificar se há nelas os processos de integração e segregação relatados na literatura.

Além disso, ainda poderia ser considerado o classificador muito utilizado na literatura a Máquina de Vetores de Suporte.

Considerando as bases relacionadas ao autismo e ao ADHD, pode-se, ainda, realizar uma variação maior do valor de binarização e thresholding, para verificar se esses outros valores melhoram as performances.

O mesmo pode ser feito para a base relacionada à PSP, cujo único valor utilizado foi de 0.3.

CONCLUSÃO

No presente trabalho foram analisadas redes de quatro bases relacionadas às doenças neurológicas: esquizofrenia, autismo, déficit de atenção/desordem de hiperatividade e paralisia progressiva nuclear, contendo redes de pacientes com a doença e saudáveis. Para cada uma dessas doenças, extraiu-se algumas medidas de rede, que serviram como atributos para algoritmos preditivos discriminar as duas classes (com ou sem a doença).

Assim, primeiramente construiu-se uma tabela contendo o desempenho dos melhores classificadores obtidos por cada base. O objetivo é verificar se há algoritmos preditivos em comum. Na tabela 39, percebe-se que o classificador árvore de decisão foi comum às quatro bases, indicando ser, com isso, um bom algoritmo para esse tipo de classificação.

Com intuito de verificar qual o melhor subconjunto de atributos, selecionou-se os melhores algoritmos preditivos (analisando as medidas de desempenho obtidas). A tabela 40 contém os subconjuntos de atributos para os classificadores de maior desempenho para cada base. Nela pode-se verificar que as medidas de cada subconjunto são variadas, sendo a única medida comum a todas as bases a eigenvector centrality. Percebe-se que para as bases Autismo e ADHD em que houve variação dos valores de binarização e thresholding (0,3-0.4-0.5-0.6), em ambas o valor que mais aumentou o desempenho dos classificadores foi 0.6.

Em relação à base Esquizofrenia, em termos de acurácia o classificador que obteve melhor performance (cujo valor foi 82%) foi aquele resultante da combinação dos classificadores k-NN, naive bayes e redes neurais, valor superior ao de (ARRUDA, 2013), cuja acurácia para redes bayesianas foi de 79%.

Quanto a base relacionada ao Autismo, em (PETROV *et al.*, 2016), utilizando a mesma base, a maior AUC obtida foi de 0.77 (pelo classificador Máquina de vetores de suporte), e a pior 0.57 (pela árvore de decisão com a técnica boosting). Comparando, apesar dos classificadores serem diferentes, a maior AUC obtida (0.71) é próximo à do artigo e pelo fato de ter sido utilizado outras medidas, torna o modelo obtido mais confiável. Em outro artigo (ZHUKOV; PETROV;

[DODONOVA, 2015](#)), também foi utilizado a mesma base, o classificador Máquina de vetores de suporte, obteve 0.64 de acurácia, 0.61 de precisão e 0.88 de revocação (também denominada de sensibilidade). Apesar dos classificadores serem diferentes, a árvore de decisão obteve uma acurácia e sensibilidade superior ao do artigo em questão.

Em relação à base ADHD, em ([SIQUEIRA et al., 2014](#)) utilizando a mesma base, obteve-se, por meio do classificador SVM, a acurácia de 61% (sendo que o valor de threshold utilizado foi de 0.25, e foi utilizado redes sem pesos); no caso do presente trabalho a performance obtida foi superior a este estudo. Em ([DEY; RAO; SHAH, 2014](#)), o classificador SVM teve como desempenho: 64.48% de acurácia, 84.71% de especificidade e 30.66% de sensibilidade; assim a acurácia e especificidade de alguns classificadores no presente trabalho foi superior a obtida neste artigo. Em ([GUO et al., 2014](#)), obteve-se uma acurácia média de 63.75%, valor inferior a acurácia obtidas por alguns classificadores no presente trabalho. Por fim, ainda, o valor de acurácia máxima alcançada na competição *ADHD-200* foi de 61.04% cujo valor também é inferior aos valores obtidos aqui.

Para a base relacionada à PSP, quanto à literatura, os artigos encontrados relacionados a ela, focaram-se na análise estrutural da rede dos pacientes com PSP, tentando identificar as regiões que mais diferem de uma rede de paciente sem a doença. No presente trabalho, focou-se em algo novo, no sentido de classificar medidas de rede na tentativa de distinguir as duas classes (de pacientes com PSP e pacientes saudáveis) e conseguir prever pacientes com PSP. Sendo que houve uma boa discriminação entre as classes, principalmente utilizando a medida sensibilidade.

Tabela 39 – Tabela contendo os classificadores com melhores desempenhos para cada base. Destacou-se em rosa, laranja, amarelo e azul, respectivamente, as doenças: Esquizofrenia, Autismo, ADHD e PSP. Em negrito destacou-se a árvore decisão, algoritmo comum a todas as bases.

Base	Classificadores	Ac.	K(Ac.)	Sens.	Esp.	AUC
Esquizofrenia	RN	0.72	0.44	0.70	0.85	0.83
	AD	0.72	0.45	0.65	0.75	0.70
	k-NN e RN	0.72	0.43	0.70	0.75	0.82
	k-NN e NB	0.72	0.50	0.65	0.85	0.75
	NB e RN	0.75	0.50	0.65	0.80	0.75
	k-NN, NB e AD	0.78	0.55	0.55	0.80	0.82
	k-NN, NB e RN	0.82	0.65	0.60	0.70	0.67
Autismo	AD	0.68	0.34	0.96	0.37	0.67
	k-NN e AD	0.63	0.23	0.90	0.35	0.69
	k-NN, RN e AD	0.61	0.81	0.90	0.37	0.70
	k-NN, NB e AD	0.65	0.28	0.75	0.55	0.69
ADHD	AD	0.62	0.02	0.07	0.93	0.54
	RN	0.631	0.04	0.09	0.95	0.56
	k-NN e RN	0.63	0.01	0.06	0.95	0.60
	k-NN e AD	0.63	0	0	0.98	0.56
	k-NN e NB	0.63	0.02	0	0.99	0.57
	AD e NB	0.64	0.05	0.09	0.96	0.57
	AD e RN	0.63	0.01	0.02	0.98	0.57
	NB e RN	0.65	0.1	0.06	0.96	0.56
	k-NN, RN e NB	0.62	0.01	0.06	0.95	0.53
	NB, RN e AD	0.65	0.07	0.03	0.95	0.57
	k-NN, RN, AD e NB	0.64	0.04	0.14	0.93	0.60
PSP	AD	0.65	0.17	0.90	0.40	0.63
	k-NN e AD	0.65	0.27	0.84	0.29	0.61
	k-NN, RN e AD	0.66	0.24	0.90	0.22	0.54
	NB, RN e AD	0.71	0.37	0.80	0.23	0.65
	k-NN, AD e NB	0.71	0.39	0.82	0.38	0.65

Tabela 40 – Tabela contendo o classificador de melhor desempenho para cada base e o subconjunto de atributos que contribuiu para isso. Além disso foi colocado o valor de binarização e thresholding utilizado para cada base, bem como o método de normalização utilizado.

Base	Subconjuntos de atributos	Normalização	Threshold	Classificadores
Esquizofrenia	-média da distribuição do grau -segundo momento da distribuição de grau -média dos menores caminhos -transitividade -betweenness centrality -closeness centrality -eigenvector centrality -k-core -assortatividade -complexidade	-	-	RN
Autismo	-média da distribuição de grau -segundo momento da distribuição de grau -coeficiente médio de aglomeração médio -transitividade -assortatividade -média dos menores caminhos -complexidade -betweenness centrality -eigenvector centrality -closeness centrality -pageRank -diâmetro -dominância central	-	0.6	k-NN, RN e AD
ADHD	-média dos menores caminhos -eigenvector centrality -closeness centrality -assortatividade -coeficiente de aglomeração médio	softmax scaling	0.6	k-NN e NB
PSP	-coeficiente de aglomeração médio -transitividade -complexidade -segundo momento da distribuição de grau -eigenvector centrality	-	0.3	k-NN, AD e NB

REFERÊNCIAS

- ABBASS, H. A. An evolutionary artificial neural networks approach for breast cancer diagnosis. **Artificial intelligence in Medicine**, Elsevier, v. 25, n. 3, p. 265–281, 2002. Citado na página 91.
- ABDAR, M.; ZOMORODI-MOGHADAM, M.; ZHOU, X.; GURURAJAN, R.; TAO, X.; BARUA, P. D.; GURURAJAN, R. A new nested ensemble technique for automated diagnosis of breast cancer. **Pattern Recognition Letters**, Elsevier, 2018. Citado na página 37.
- ADHIKARI, A.; ADHIKARI, J. **Advances in knowledge discovery in databases**. [S.l.]: Springer, 2015. Citado nas páginas 42 e 53.
- AGUIAR-PULIDO, V.; SEOANE, J. A.; RABUÑAL, J. R.; DORADO, J.; PAZOS, A.; MUNTEANU, C. R. Machine learning techniques for single nucleotide polymorphism—disease classification models in schizophrenia. **Molecules**, Molecular Diversity Preservation International, v. 15, n. 7, p. 4875–4889, 2010. Citado na página 90.
- ALAGIAKRISHNAN, K.; ZHAO, N.; MEREU, L.; SENIOR, P.; SENTHILSELVAN, A. Montreal cognitive assessment is superior to standardized mini-mental status exam in detecting mild cognitive impairment in the middle-aged and elderly patients with type 2 diabetes mellitus. **BioMed research international**, Hindawi Publishing Corporation, v. 2013, 2013. Citado na página 95.
- ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of modern physics**, APS, v. 74, n. 1, p. 47, 2002. Citado nas páginas 17, 65, 66 e 67.
- ALEXANDER-BLOCH, A. F.; GOGTAY, N.; MEUNIER, D.; BIRN, R.; CLASEN, L.; LALONDE, F.; LENROOT, R.; GIEDD, J.; BULLMORE, E. T. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. **Frontiers in systems neuroscience**, Frontiers Media SA, v. 4, 2010. Citado na página 90.
- ALEXANDER-BLOCH, A. F.; VÉRTES, P. E.; STIDD, R.; LALONDE, F.; CLASEN, L.; RAPOPORT, J.; GIEDD, J.; BULLMORE, E. T.; GOGTAY, N. The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. **Cerebral cortex**, Oxford University Press, v. 23, n. 1, p. 127–138, 2012. Citado na página 90.
- ALONSO, F.; CARAÇA-VALENTE, J. P.; GONZÁLEZ, A. L.; MONTES, C. Combining expert knowledge and data mining in a medical diagnosis domain. **Expert Systems with Applications**, Elsevier, v. 23, n. 4, p. 367–375, 2002. Citado na página 42.
- AMANCIO, D. R.; NUNES, M. d. G. V. Avaliando tradução automática e simplificação textual com redes complexas. 2009. Citado na página 71.
- ANDERSON, A.; DOUGLAS, P. K.; KERR, W. T.; HAYNES, V. S.; YUILLE, A. L.; XIE, J.; WU, Y. N.; BROWN, J. A.; COHEN, M. S. Non-negative matrix factorization of multimodal mri, fmri and phenotypic data reveals differential changes in default mode subnetworks in adhd. **NeuroImage**, Elsevier, v. 102, p. 207–219, 2014. Citado nas páginas 117 e 118.

- ANDERSON, J. A.; ROSENFELD, E. **Talking nets: An oral history of neural networks**. [S.l.]: MiT Press, 2000. Citado na página 55.
- APTÉ, C.; WEISS, S. Data mining with decision trees and decision rules. **Future generation computer systems**, Elsevier, v. 13, n. 2-3, p. 197–210, 1997. Citado na página 54.
- ARBABSHIRANI, M. R.; KIEHL, K. A.; PEARLSON, G. D.; CALHOUN, V. D. Classification of schizophrenia patients based on resting-state functional network connectivity. **Frontiers in neuroscience**, Frontiers Media SA, v. 7, 2013. Citado nas páginas 88, 90 e 106.
- ARRUDA, G. F. d. **Mineração de dados em redes complexas: estrutura e dinâmica**. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado nas páginas 37, 42, 63, 69, 70, 71, 88, 89, 90, 106, 108 e 133.
- ARRUDA, G. F. de; COSTA, L. da F.; SCHUBERT, D.; RODRIGUES, F. A. Structure and dynamics of functional networks in child-onset schizophrenia. **Clinical Neurophysiology**, Elsevier, v. 125, n. 8, p. 1589–1595, 2014. Citado na página 90.
- ASSOCIATION, A. P. **Manual diagnóstico e estatístico de transtornos mentais: texto revisado (DSM-IV-TR)**. [S.l.]: Artmed, 2002. Citado na página 118.
- AVENA-KOENIGSBERGER, A.; MISIC, B.; SPORNS, O. Communication dynamics in complex brain networks. **Nature Reviews Neuroscience**, Nature Publishing Group, v. 19, n. 1, p. 17, 2018. Citado na página 37.
- BALCÁZAR, J. L.; BONCHI, F.; GIONIS, A.; SEBAG, M. Machine learning and knowledge discovery in databases. **Lecture Notes in Computer Science**, v. 6323, 2010. Citado na página 50.
- BARABÁSI, A.-L. **Network science**. [S.l.]: Cambridge university press, 2016. Citado nas páginas 66 e 68.
- BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **Science**, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999. Citado na página 37.
- BARKAN, E.; KISILEV, P.; WALACH, E. **Method for automatic visual annotation of radiological images from patient clinical data**. [S.l.]: Google Patents, 2018. US Patent App. 15/249,415. Citado na página 37.
- BARON-COHEN, S.; LESLIE, A. M.; FRITH, U. Does the autistic child have a “theory of mind”? **Cognition**, Elsevier, v. 21, n. 1, p. 37–46, 1985. Citado na página 109.
- BARROS, R. C.; CARVALHO, A. C. de; FREITAS, A. A. **Automatic design of decision-tree induction algorithms**. [S.l.]: Springer, 2015. Citado na página 53.
- BASSETT, D. S.; BULLMORE, E.; VERCHINSKI, B. A.; MATTAY, V. S.; WEINBERGER, D. R.; MEYER-LINDENBERG, A. Hierarchical organization of human cortical networks in health and schizophrenia. **Journal of Neuroscience**, Soc Neuroscience, v. 28, n. 37, p. 9239–9248, 2008. Citado na página 90.
- BATAGELJ, V.; ZAVERSNIK, M. An o (m) algorithm for cores decomposition of networks. **arXiv preprint cs/0310049**, 2003. Citado na página 71.

- BECKMANN, C. F. Modelling with independent components. **Neuroimage**, Elsevier, v. 62, n. 2, p. 891–901, 2012. Citado nas páginas 19 e 81.
- BELLAZZI, R.; ZUPAN, B. Predictive data mining in clinical medicine: current issues and guidelines. **International journal of medical informatics**, Elsevier, v. 77, n. 2, p. 81–97, 2008. Citado na página 37.
- BELMONTE, M.; BARON-COHEN, S. Normal sibs of children with autism share negative frontal but not positive sensory abnormalities: preliminary evidence from fmri during processing of visual distractors. In: **Society for Neuroscience Abstract**. [S.l.: s.n.], 2004. v. 30, n. 582.10. Citado na página 109.
- BELMONTE, M. K.; ALLEN, G.; BECKEL-MITCHENER, A.; BOULANGER, L. M.; CARPER, R. A.; WEBB, S. J. Autism and abnormal development of brain connectivity. **Journal of Neuroscience**, Soc Neuroscience, v. 24, n. 42, p. 9228–9231, 2004. Citado na página 109.
- BELMONTE, M. K.; YURGELUN-TODD, D. A. Functional anatomy of impaired selective attention and compensatory processing in autism. **Cognitive brain research**, Elsevier, v. 17, n. 3, p. 651–664, 2003. Citado na página 109.
- BERNARDINI, F. C. **Combinação de classificadores simbólicos para melhorar o poder preditivo e descritivo de ensembles**. Tese (Doutorado) — Universidade de São Paulo, 2002. Citado nas páginas 17, 58 e 59.
- BERNSTEIN, A.; BURNAEV, E.; KONDRATYEVA, E.; SUSHCHINSKAYA, S.; SHARAEV, M.; ANDREEV, A.; ARTEMOV, A.; AKZHIGITOV, R. Machine learning pipeline for discovering neuroimaging-based biomarkers in neurology and psychiatry. **arXiv preprint arXiv:1804.10163**, 2018. Citado na página 37.
- BIJSTERBOSCH, J.; SMITH, S. M.; BECKMANN, C. F. **An Introduction to Resting State fMRI Functional Connectivity**. [S.l.]: Oxford University Press, 2017. Citado nas páginas 19, 79 e 81.
- BISWAL, B.; YETKIN, F. Z.; HAUGHTON, V. M.; HYDE, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. **Magnetic resonance in medicine**, Wiley Online Library, v. 34, n. 4, p. 537–541, 1995. Citado na página 77.
- BLEULER, E.; JUNG, C. G. Komplexe und krankheitsursachen bei dementia praecox. **Zentralblatt fur Nervenheilkunde und Psychiatrie**, n. XIX, p. 220–227, 1908. Citado na página 87.
- BLUM, A. L.; LANGLEY, P. Selection of relevant features and examples in machine learning. **Artificial intelligence**, Elsevier, v. 97, n. 1, p. 245–271, 1997. Citado na página 44.
- BOAS, P. R. V. **Efeito da amostragem nas propriedades topológicas de redes complexas**. Tese (Doutorado) — Universidade de São Paulo, 2008. Citado nas páginas 68 e 70.
- BOCCA, F. F. Produtividade de cana-de-açúcar: caracterização dos contextos de decisão e utilização de técnicas de mineração de dados para modelagem. [sn], 2014. Citado na página 97.
- BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D.-U. Complex networks: Structure and dynamics. **Physics reports**, Elsevier, v. 424, n. 4, p. 175–308, 2006. Citado na página 71.

- BOHLAND, J. W.; SAPERSTEIN, S.; PEREIRA, F.; RAPIN, J.; GRADY, L. Network, anatomical, and non-imaging measures for the prediction of adhd diagnosis in individual subjects. **Frontiers in systems neuroscience**, Frontiers Media SA, v. 6, 2012. Citado na página 119.
- BONACICH, P. Power and centrality: A family of measures. **American journal of sociology**, University of Chicago Press, v. 92, n. 5, p. 1170–1182, 1987. Citado na página 70.
- BORGES, R. R.; IAROSZ, K. C.; BATISTA, A. M.; CALDAS, I. L.; BORGES, F. S.; LA-MEU, E. L. Sincronização de disparos em redes neuronais com plasticidade sináptica. **Caderno Brasileiro de Ensino de Física**, v. 37, n. 2, 2015. Citado nas páginas 16 e 56.
- BREIMAN, L. Bagging predictors. **Machine learning**, Springer, v. 24, n. 2, p. 123–140, 1996. Citado na página 57.
- _____. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001. Citado nas páginas 48 e 97.
- BREIMAN, L.; FRIEDMAN, J.; STONE, C. J.; OLSHEN, R. A. **Classification and regression trees**. [S.l.]: CRC press, 1984. Citado na página 54.
- BRIN, S.; PAGE, L. Reprint of: The anatomy of a large-scale hypertextual web search engine. **Computer networks**, Elsevier, v. 56, n. 18, p. 3825–3833, 2012. Citado na página 71.
- BROWN, J. A.; HORN, J. D. V. Connected brains and minds—the umcd repository for brain connectivity matrices. **Neuroimage**, Elsevier, v. 124, p. 1238–1241, 2016. Citado na página 110.
- BROWN, J. A.; HUA, A. Y.; TRUJILLO, A.; ATTYGALLE, S.; BINNEY, R. J.; SPINA, S.; LEE, S. E.; KRAMER, J. H.; MILLER, B. L.; ROSEN, H. J. Advancing functional dysconnectivity and atrophy in progressive supranuclear palsy. **NeuroImage: Clinical**, Elsevier, v. 16, p. 564–574, 2017. Citado nas páginas 126 e 129.
- BROWN, M.; SEMELKA, R.; NISHINO, T. K. Mri: basic principles and applications. **Medical Physics**, Wiley Online Library, v. 31, n. 1, p. 170–170, 2004. Citado na página 77.
- BUENO, L. S. **Aquisição rápida de imagens com técnicas tipo Echo Planar Imaging-Implementação das sequências EPI e SEPI**. Tese (Doutorado) — Universidade de São Paulo, 2004. Citado na página 89.
- BULLMORE, E.; SPORNS, O. Complex brain networks: graph theoretical analysis of structural and functional systems. **Nature Reviews Neuroscience**, v. 10, n. 3, 2009. Citado nas páginas 19, 37, 73, 74, 76 e 82.
- _____. The economy of brain network organization. **Nature Reviews Neuroscience**, Nature Publishing Group, v. 13, n. 5, p. 336–349, 2012. Citado na página 84.
- CABRAL, C. I. S. **Aplicação do modelo de regressão logística num estudo de mercado**. Tese (Doutorado), 2013. Citado na página 58.
- CALHOUN, V. D.; ADALI, T.; PEARLSON, G. D.; PEKAR, J. A method for making group inferences from functional mri data using independent component analysis. **Human brain mapping**, Wiley Online Library, v. 14, n. 3, p. 140–151, 2001. Citado na página 79.

CALHOUN, V. D.; EICHELE, T.; PEARLSON, G. Functional brain networks in schizophrenia: a review. **Frontiers in human neuroscience**, Frontiers Research Foundation, 2009. Citado na página 87.

CAMARGO, R. Y. d. **Inferência de conexões sinápticas em redes neurais biologicamente plausíveis**. Tese (Doutorado) — Universidade de São Paulo (USP). Instituto de Física, 2003. Citado na página 76.

CAO, M.; SHU, N.; CAO, Q.; WANG, Y.; HE, Y. Imaging functional and structural brain connectomics in attention-deficit/hyperactivity disorder. **Molecular neurobiology**, Springer, v. 50, n. 3, p. 1111–1123, 2014. Citado na página 119.

CARNEC, M.; CALLET, P. L.; BARBA, D. Full reference and reduced reference metrics for image quality assessment. In: IEEE. **Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on**. [S.l.], 2003. v. 1, p. 477–480. Citado na página 95.

CASEY, B.; TOTTENHAM, N.; LISTON, C.; DURSTON, S. Imaging the developing brain: what have we learned about cognitive development? **Trends in cognitive sciences**, Elsevier, v. 9, n. 3, p. 104–110, 2005. Citado na página 74.

CASTELLANOS, F. X.; MARGULIES, D. S.; KELLY, C.; UDDIN, L. Q.; GHAFARI, M.; KIRSCH, A.; SHAW, D.; SHEHZAD, Z.; MARTINO, A. D.; BISWAL, B. Cingulate-precuneus interactions: a new locus of dysfunction in adult attention-deficit/hyperactivity disorder. **Biological psychiatry**, Elsevier, v. 63, n. 3, p. 332–337, 2008. Citado na página 117.

CESTARI, D. M. **Classificação de sinais de epilepsia utilizando redes complexas**. Tese (Doutorado) — Universidade de São Paulo, 2017. Citado na página 71.

CHAURASIA, V.; PAL, S.; TIWARI, B. Prediction of benign and malignant breast cancer using data mining techniques. **Journal of Algorithms & Computational Technology**, SAGE Publications Sage UK: London, England, v. 12, n. 2, p. 119–126, 2018. Citado na página 37.

CHENG, W.; JI, X.; ZHANG, J.; FENG, J. Individual classification of adhd patients by integrating multiscale neuroimaging markers and advanced pattern recognition techniques. **Frontiers in systems neuroscience**, Frontiers Media SA, v. 6, 2012. Citado na página 119.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and psychological measurement**, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 37–46, 1960. Citado na página 62.

COSTA, L. d. F.; JR, O. N. O.; TRAVIESO, G.; RODRIGUES, F. A.; BOAS, P. R. V.; ANTI-QUEIRA, L.; VIANA, M. P.; ROCHA, L. E. C. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. **Advances in Physics**, Taylor & Francis, v. 60, n. 3, p. 329–412, 2011. Citado na página 37.

COSTA, L. d. F.; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: A survey of measurements. **Advances in physics**, Taylor & Francis, v. 56, n. 1, p. 167–242, 2007. Citado na página 69.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE transactions on information theory**, IEEE, v. 13, n. 1, p. 21–27, 1967. Citado na página 51.

COX, R. W. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. **Computers and Biomedical research**, Elsevier, v. 29, n. 3, p. 162–173, 1996. Citado na página 89.

CRADDOCK, R. C.; HOLTZHEIMER, P. E.; HU, X. P.; MAYBERG, H. S. Disease state prediction from resting state functional connectivity. **Magnetic resonance in Medicine**, Wiley Online Library, v. 62, n. 6, p. 1619–1628, 2009. Citado na página 119.

DANIELS, A. M.; MANDELL, D. S. Explaining differences in age at autism spectrum disorder diagnosis: A critical review. **Autism**, Sage Publications Sage UK: London, England, v. 18, n. 5, p. 583–597, 2014. Citado na página 110.

DELEN, D.; WALKER, G.; KADAM, A. Predicting breast cancer survivability: a comparison of three data mining methods. **Artificial intelligence in medicine**, Elsevier, v. 34, n. 2, p. 113–127, 2005. Citado na página 37.

DEY, S.; RAO, A. R.; SHAH, M. Exploiting the brain's network structure in identifying ADHD subjects. **Frontiers in systems neuroscience**, Frontiers Media SA, v. 6, 2012. Citado na página 119.

_____. Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects. **Frontiers in neural circuits**, Frontiers, v. 8, p. 64, 2014. Citado nas páginas 122 e 134.

DIETTERICH, T. G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. **Machine learning**, Springer, v. 40, n. 2, p. 139–157, 2000. Citado na página 58.

DIJK, K. R. V.; HEDDEN, T.; VENKATARAMAN, A.; EVANS, K. C.; LAZAR, S. W.; BUCKNER, R. L. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. **Journal of neurophysiology**, Am Physiological Soc, v. 103, n. 1, p. 297–321, 2010. Citado na página 74.

EUSTON, D. R.; GRUBER, A. J.; MCNAUGHTON, B. L. The role of medial prefrontal cortex in memory and decision making. **Neuron**, Elsevier, v. 76, n. 6, p. 1057–1070, 2012. Citado na página 110.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. Inteligência artificial: Uma abordagem de aprendizado de máquina. **Rio de Janeiro: LTC**, v. 2, p. 192, 2011. Citado nas páginas 15, 16, 43, 44, 45, 49, 51, 52, 54, 55, 61, 62 e 63.

FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the internet topology. In: ACM. **ACM SIGCOMM computer communication review**. [S.l.], 1999. v. 29, n. 4, p. 251–262. Citado na página 37.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996. Citado na página 41.

FAYYAD, U. M.; PIATETSKY-SHAPIO, G.; SMYTH, P. Knowledge discovery and data mining: Towards a unifying framework. In: **KDD**. [S.l.: s.n.], 1996. v. 96, p. 82–88. Citado na página 41.

FERREIRA, C. M. S.; OCHI, L. S.; MACAMBIRA, E. M. **Desenvolvimento e Análise Experimental de Heurísticas GRASP para uma Generalização do Problema da Árvore Geradora Mínima**. [S.l.]: XXXIX SBPO-Simpósio Brasileiro de Pesquisa Operacional, Fortaleza-CE, 2007. Citado na página 89.

FERREIRA, L. R. K. **Funcionamento cerebral de repouso em idosos e recuperação de memória autobiográfica: um estudo de ressonância magnética funcional**. Tese (Doutorado) — Universidade de São Paulo, 2014. Citado na página 89.

FIGUEIRA, A. F. C. **Human brain networks: an investigation on cortical thickness resting-state fMRI and structural connectivity as assessed by tractography**. Tese (Doutorado) — Universidade de Lisboa, 2013. Citado nas páginas 19, 77, 79, 81 e 83.

FIX, E.; HODGES, J. L. **Discriminatory analysis-nonparametric discrimination: consistency properties**. [S.l.], 1951. Citado na página 50.

_____. **Discriminatory analysis-nonparametric discrimination: Small sample performance**. [S.l.], 1952. Citado na página 50.

FORNITO, A.; ZALESKY, A.; BULLMORE, E. **Fundamentals of brain network analysis**. [S.l.]: Academic Press, 2016. Citado nas páginas 20, 73, 75, 81 e 84.

FRANK, E.; HALL, M.; TRIGG, L.; HOLMES, G.; WITTEN, I. H. Data mining in bioinformatics using weka. **Bioinformatics**, Oxford University Press, v. 20, n. 15, p. 2479–2481, 2004. Citado na página 42.

FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, JSTOR, p. 35–41, 1977. Citado na página 70.

_____. Centrality in social networks conceptual clarification. **Social networks**, Elsevier, v. 1, n. 3, p. 215–239, 1978. Citado na página 70.

FREUND, Y.; SCHAPIRE, R. E. Experiments with a new boosting algorithm. In: **Icml**. [S.l.: s.n.], 1996. v. 96, p. 148–156. Citado nas páginas 57 e 58.

FRIEDL, M. A.; SULLA-MENASHE, D.; TAN, B.; SCHNEIDER, A.; RAMANKUTTY, N.; SIBLEY, A.; HUANG, X. Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets. **Remote sensing of Environment**, Elsevier, v. 114, n. 1, p. 168–182, 2010. Citado na página 53.

FRISTON, K. J. Functional and effective connectivity in neuroimaging: a synthesis. **Human brain mapping**, Wiley Online Library, v. 2, n. 1-2, p. 56–78, 1994. Citado na página 75.

FRISTON, K. J.; FRITH, C. D. Schizophrenia: a disconnection syndrome. **Clinical Neuroscience**, v. 3, n. 2, p. 89–97, 1995. Citado nas páginas 38 e 87.

FRITH, U. **Autism: Explaining the enigma**. [S.l.]: Wiley Online Library, 1989. v. 1989. Citado na página 109.

GAMA, J. M. P. d. Combining classification algorithms. Universidade do Porto. Reitoria, 1999. Citado na página 55.

- GAMEIRO, S. R. **Nanopartículas poliméricas coloidais como meios de contraste em imagiologia por ressonância magnética: síntese de quelatos de Gd (III), preparação, caracterização físico-química e eficácia in vitro**. Dissertação (Mestrado) — Universidade de Coimbra, 2004. Citado nas páginas 18, 77 e 78.
- GARDNER, R. C.; BOXER, A. L.; TRUJILLO, A.; MIRSKY, J. B.; GUO, C. C.; GENNATAS, E. D.; HEUER, H. W.; FINE, E.; ZHOU, J.; KRAMER, J. H. Intrinsic connectivity network disruption in progressive supranuclear palsy. **Annals of neurology**, Wiley Online Library, v. 73, n. 5, p. 603–616, 2013. Citado nas páginas 125, 126 e 129.
- GIL, V. M.; GERALDES, C. F. d. G. C. **Ressonância magnética nuclear: fundamentos, métodos e aplicações**. [S.l.: s.n.], 1987. Citado na página 77.
- GIOIA, G. A.; ISQUITH, P. K.; GUY, S. C.; KENWORTHY, L. Test review behavior rating inventory of executive function. **Child Neuropsychology**, Taylor & Francis, v. 6, n. 3, p. 235–238, 2000. Citado na página 37.
- GOMES, V. S. P. S. **Análise estatística multivariada aplicada a dados hidrogeológicos**. Tese (Doutorado), 2013. Citado nas páginas 45 e 47.
- GOSAK, M.; MARKOVIČ, R.; DOLENŠEK, J.; RUPNIK, M. S.; MARHL, M.; STOŽER, A.; PERC, M. Network science of biological systems at different scales: a review. **Physics of life reviews**, Elsevier, v. 24, p. 118–135, 2018. Citado na página 37.
- GREEN, S.; ȘERBAN, M.; SCHOLL, R.; JONES, N.; BRIGANDT, I.; BECHTEL, W. Network analyses in systems biology: new strategies for dealing with biological complexity. **Synthese**, Springer, v. 195, n. 4, p. 1751–1777, 2018. Citado na página 37.
- GUEIREZ, J. E. de J. Árvores de decisão desequilibradas para detecção de erros em transações de comércio externo usando técnicas de data mining. 2014. Citado na página 54.
- GUO, X.; AN, X.; KUANG, D.; ZHAO, Y.; HE, L. Adhd-200 classification based on social network method. In: SPRINGER. **International Conference on Intelligent Computing**. [S.l.], 2014. p. 233–240. Citado nas páginas 119, 122 e 134.
- GUYATT, G. H.; LEFCOE, M.; WALTER, S.; COOK, D.; TROYAN, S.; GRIFFITH, L.; KING, D.; ZYLAK, C.; HICKEY, N.; CARRIER, G. Interobserver variation in the computed tomographic evaluation of mediastinal lymph node size in patients with potentially resectable lung cancer. **Chest**, Elsevier, v. 107, n. 1, p. 116–119, 1995. Citado nas páginas 94 e 95.
- GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **Journal of machine learning research**, v. 3, n. Mar, p. 1157–1182, 2003. Citado nas páginas 16, 45 e 46.
- HALL, M. A. Correlation-based feature selection for machine learning. University of Waikato Hamilton, 1999. Citado na página 44.
- _____. Correlation-based feature selection of discrete and numeric class machine learning. University of Waikato, Department of Computer Science, 2000. Citado na página 47.
- HALL, M. A.; HOLMES, G. Benchmarking attribute selection techniques for discrete class data mining. **IEEE Transactions on Knowledge and Data engineering**, IEEE, v. 15, n. 6, p. 1437–1447, 2003. Citado na página 47.

HALLQUIST, M. N.; HILLARY, F. G. Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world. **Network Neuroscience**, MIT Press, n. Just Accepted, p. 1–58, 2018. Citado na página 38.

HAYKIN, S. **Neural networks: a comprehensive foundation**. [S.l.]: Prentice Hall PTR, 1994. Citado nas páginas 16, 48, 55 e 56.

_____. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2007. Citado na página 48.

HEUVEL, M. P. V. D.; POL, H. E. H. Exploring the brain network: a review on resting-state fmri functional connectivity. **European neuropsychopharmacology**, Elsevier, v. 20, n. 8, p. 519–534, 2010. Citado nas páginas 18, 79 e 80.

HEUVEL, M. P. van den; MANDL, R. C.; STAM, C. J.; KAHN, R. S.; POL, H. E. H. Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. **Journal of Neuroscience**, Soc Neuroscience, v. 30, n. 47, p. 15915–15926, 2010. Citado na página 90.

HILGETAG, C.-C.; BURNS, G. A.; O'NEILL, M. A.; SCANNELL, J. W.; YOUNG, M. P. Anatomical connectivity defines the organization of clusters of cortical areas in the macaque and the cat. **Philosophical Transactions of the Royal Society of London B: Biological Sciences**, The Royal Society, v. 355, n. 1393, p. 91–110, 2000. Citado na página 74.

HINKLE, D. E.; WIERSMA, W.; JURIS, S. G. Applied statistics for the behavioral sciences. JSTOR, 2003. Citado na página 96.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. [S.l.]: John Wiley & Sons, 2013. v. 398. Citado nas páginas 23 e 63.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of educational psychology**, Warwick & York, v. 24, n. 6, p. 417, 1933. Citado na página 45.

HU, Y.; DUAN, K.; ZHANG, Y.; HOSSAIN, M. S.; RAHMAN, S. M. M.; ALELAIWI, A. Simultaneously aided diagnosis model for outpatient departments via healthcare big data analytics. **Multimedia Tools and Applications**, Springer, v. 77, n. 3, p. 3729–3743, 2018. Citado na página 37.

HUANG, J.; LING, C. X. Using auc and accuracy in evaluating learning algorithms. **IEEE Transactions on knowledge and Data Engineering**, IEEE, v. 17, n. 3, p. 299–310, 2005. Citado na página 63.

ILYASOVA, N.; KUPRIYANOV, A.; PARINGER, R.; KIRSH, D. Particular use of big data in medical diagnostic tasks. **Pattern Recognition and Image Analysis**, Springer, v. 28, n. 1, p. 114–121, 2018. Citado na página 37.

ISABELLE, G. Feature extraction foundations and applications. pattern recognition. Springer, 2006. Citado na página 44.

KANTARDZIC, M. **Data mining: concepts, models, methods, and algorithms**. [S.l.]: John Wiley & Sons, 2011. Citado nas páginas 44, 45 e 47.

- KAYAER, K.; YILDIRIM, T. Medical diagnosis on pima indian diabetes using general regression neural networks. In: **Proceedings of the international conference on artificial neural networks and neural information processing (ICANN/ICONIP)**. [S.l.: s.n.], 2003. p. 181–184. Citado na página 91.
- KELLER, T. A.; KANA, R. K.; JUST, M. A. A developmental study of the structural integrity of white matter in autism. **Neuroreport**, LWW, v. 18, n. 1, p. 23–27, 2007. Citado na página 110.
- KENNEDY, D. P.; REDCAY, E.; COURCHESNE, E. Failing to deactivate: resting functional abnormalities in autism. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 103, n. 21, p. 8275–8280, 2006. Citado na página 110.
- KEOGH, E.; KASETTY, S. On the need for time series data mining benchmarks: a survey and empirical demonstration. **Data Mining and knowledge discovery**, Springer, v. 7, n. 4, p. 349–371, 2003. Citado na página 95.
- KOHAVI, R. 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. Citado na página 61.
- KOHONEN, T. Learning vector quantization. In: **Self-Organizing Maps**. [S.l.]: Springer, 1995. p. 175–189. Citado na página 48.
- KONONENKO, I. Machine learning for medical diagnosis: history, state of the art and perspective. **Artificial Intelligence in medicine**, Elsevier, v. 23, n. 1, p. 89–109, 2001. Citado na página 42.
- KONRAD, K.; EICKHOFF, S. B. Is the adhd brain wired differently? a review on structural and functional connectivity in attention deficit hyperactivity disorder. **Human brain mapping**, Wiley Online Library, v. 31, n. 6, p. 904–916, 2010. Citado na página 117.
- KUHN, M. Building predictive models in r using the caret package. **Journal of Statistical Software**, v. 28, n. 5, p. 1–26, 2008. Citado na página 97.
- _____. Caret package. **Journal of Statistical Software**, v. 28, n. 5, p. 1–26, 2008. Citado na página 48.
- _____. Variable importance using the caret package. 2012. Citado nas páginas 47, 48 e 97.
- _____. A short introduction to the caret package. **R Found Stat Comput**, p. 1–10, 2015. Citado na página 91.
- KUMAR, M. N.; KOUSHIK, K.; DEEPAK, K. Prediction of heart diseases using data mining and machine learning algorithms and tools. 2018. Citado na página 37.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, JSTOR, p. 159–174, 1977. Citado nas páginas 23 e 63.
- LATIFOĞLU, F.; POLAT, K.; KARA, S.; GÜNEŞ, S. Medical diagnosis of atherosclerosis from carotid artery doppler signals using principal component analysis (pca), k-nn based weighting pre-processing and artificial immune recognition system (airs). **Journal of Biomedical Informatics**, Elsevier, v. 41, n. 1, p. 15–23, 2008. Citado na página 90.

LAVANYA, D.; RANI, K. U. Performance evaluation of decision tree classifiers on medical datasets. **International Journal of Computer Applications**, International Journal of Computer Applications, 244 5 th Avenue, # 1526, New York, NY 10001, USA India, v. 26, n. 4, 2011. Citado na página [91](#).

LEBEDEV, A.; WESTMAN, E.; WESTEN, G. V.; KRAMBERGER, M.; LUNDERVOLD, A.; AARSLAND, D.; SOININEN, H.; KŁOSZEWSKA, I.; MECOCCHI, P.; TSOLAKI, M. Random forest ensembles for detection and prediction of alzheimer's disease with a good between-cohort robustness. **NeuroImage: Clinical**, Elsevier, v. 6, p. 115–125, 2014. Citado na página [48](#).

LEISCH, F.; JAIN, L. C.; HORNIK, K. Cross-validation with active pattern selection for neural-network classifiers. **IEEE Transactions on Neural Networks**, IEEE, v. 9, n. 1, p. 35–41, 1998. Citado na página [63](#).

LIANG, S.-F.; HSIEH, T.-H.; CHEN, P.-T.; WU, M.-L.; KUNG, C.-C.; LIN, C.-Y.; SHAW, F.-Z. Differentiation between resting-state fmri data from adhd and normal subjects: based on functional connectivity and machine learning. In: IEEE. **Fuzzy Theory and it's Applications (iFUZZY), 2012 International Conference on**. [S.l.], 2012. p. 294–298. Citado nas páginas [118](#) e [119](#).

LITVAN, I.; AGID, Y.; CALNE, D.; CAMPBELL, G.; DUBOIS, B.; DUVOISIN, R.; GOETZ, C.; GOLBE, L. I.; GRAFMAN, J.; GROWDON, J. Clinical research criteria for the diagnosis of progressive supranuclear palsy (steele-richardson-olszewski syndrome) report of the ninds-spsp international workshop. **Neurology**, AAN Enterprises, v. 47, n. 1, p. 1–9, 1996. Citado na página [125](#).

LIU, C.-L.; LEE, C.-H.; LIN, P.-M. A fall detection system using k-nearest neighbor classifier. **Expert systems with applications**, Elsevier, v. 37, n. 10, p. 7174–7181, 2010. Citado na página [90](#).

LIU, H.; MOTODA, H. **Feature selection for knowledge discovery and data mining**. [S.l.]: Springer Science & Business Media, 2012. v. 454. Citado nas páginas [16](#), [44](#) e [46](#).

LIU, Y.; LIANG, M.; ZHOU, Y.; HE, Y.; HAO, Y.; SONG, M.; YU, C.; LIU, H.; LIU, Z.; JIANG, T. Disrupted small-world networks in schizophrenia. **Brain**, Oxford University Press, v. 131, n. 4, p. 945–961, 2008. Citado na página [87](#).

LOPES, E. d. O. Análise de medidas em grafos para conectividade funcional em redes de modo padrão na demência da doença de alzheimer leve utilizando técnicas de aprendizado de máquina. [sn], 2016. Citado na página [77](#).

LORD, C.; COOK, E. H.; LEVENTHAL, B. L.; AMARAL, D. G. Autism spectrum disorders. **Neuron**, Elsevier, v. 28, n. 2, p. 355–363, 2000. Citado na página [109](#).

LYNALL, M.-E.; BASSETT, D. S.; KERWIN, R.; MCKENNA, P. J.; KITZBICHLER, M.; MULLER, U.; BULLMORE, E. Functional connectivity and brain networks in schizophrenia. **Journal of Neuroscience**, Soc Neuroscience, v. 30, n. 28, p. 9477–9487, 2010. Citado nas páginas [19](#), [82](#) e [90](#).

MACLIN, R.; OPITZ, D. An empirical evaluation of bagging and boosting. **AAAI/IAAI**, v. 1997, p. 546–551, 1997. Citado na página [57](#).

MAGALHÃES, R. J. d. S. **Modelação de padrões de conectividade cerebral funcional**. Tese (Doutorado) — Universidade de Minho, 2013. Citado na página 79.

MAIMON, O.; ROKACH, L. Introduction to knowledge discovery and data mining. In: **Data Mining and Knowledge Discovery Handbook**. [S.l.]: Springer, 2009. p. 1–15. Citado nas páginas 15, 41 e 42.

_____. **Data mining and knowledge discovery handbook, 2nd edn**. [S.l.]: Springer, 2010. Citado nas páginas 49 e 61.

MANDELLI, M. L.; VILAPLANA, E.; BROWN, J. A.; HUBBARD, H. I.; BINNEY, R. J.; ATTYGALLE, S.; SANTOS-SANTOS, M. A.; MILLER, Z. A.; PAKVASA, M.; HENRY, M. L. Healthy brain connectivity predicts atrophy progression in non-fluent variant of primary progressive aphasia. **Brain**, Oxford University Press, v. 139, n. 10, p. 2778–2791, 2016. Citado na página 129.

MANLY, B. F. M. **Métodos estatísticos multivariados: uma introdução**. [S.l.]: Bookman, 2008. Citado na página 58.

MAROCO, J.; SILVA, D.; RODRIGUES, A.; GUERREIRO, M.; SANTANA, I.; MENDONÇA, A. de. Data mining methods in the prediction of dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. **BMC research notes**, BioMed Central, v. 4, n. 1, p. 299, 2011. Citado nas páginas 37 e 95.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 55.

MEYER-BAESE, A. **Pattern recognition for medical imaging**. [S.l.]: Academic Press, 2004. Citado na página 48.

MICHELOYANNIS, S.; PACHOU, E.; STAM, C. J.; VOURKAS, M.; ERIMAKI, S.; TSIRKA, V. Using graph theoretical analysis of multi channel eeg to evaluate the neural efficiency hypothesis. **Neuroscience letters**, Elsevier, v. 402, n. 3, p. 273–277, 2006. Citado na página 87.

MILJKOVIC, D.; ALEKSOVSKI, D.; PODPEČAN, V.; LAVRAČ, N.; MALLE, B.; HOLZINGER, A. Machine learning and data mining methods for managing parkinson's disease. In: **Machine Learning for Health Informatics**. [S.l.]: Springer, 2016. p. 209–220. Citado na página 42.

MILO, R.; SHEN-ORR, S.; ITZKOVITZ, S.; KASHTAN, N.; CHKLOVSKII, D.; ALON, U. Network motifs: simple building blocks of complex networks. **Science**, American Association for the Advancement of Science, v. 298, n. 5594, p. 824–827, 2002. Citado na página 72.

MINSHEW, N. J.; GOLDSTEIN, G.; SIEGEL, D. J. Neuropsychologic functioning in autism: Profile of a complex information processing disorder. **Journal of the International Neuropsychological Society**, Cambridge University Press, v. 3, n. 4, p. 303–316, 1997. Citado na página 109.

MITCHELL, T. M. Machine learning. 1997. **Burr Ridge, IL: McGraw Hill**, v. 45, n. 37, p. 870–877, 1997. Citado na página 51.

MONACO, A.; MONDA, A.; AMOROSO, N.; BERTOLINO, A.; BLASI, G.; CARLO, P. D.; PAPALINO, M.; PERGOLA, G.; TANGARO, S.; BELLOTTI, R. A complex network approach reveals a pivotal substructure of genes linked to schizophrenia. **PloS one**, Public Library of Science, v. 13, n. 1, p. e0190110, 2018. Citado na página 37.

MORRIS, M. A.; SABOURY, B.; BURKETT, B.; GAO, J.; SIEGEL, E. L. Reinventing radiology: big data and the future of medical imaging. **Journal of thoracic imaging**, Wolters Kluwer, v. 33, n. 1, p. 4–16, 2018. Citado na página 37.

MOTA, N. B.; FURTADO, R.; MAIA, P. P.; COPELLI, M.; RIBEIRO, S. Graph analysis of dream reports is especially informative about psychosis. **Scientific reports**, Nature Publishing Group, v. 4, 2014. Citado na página 90.

MUNOZ, S. R.; BANGDIWALA, S. I. Interpretation of kappa and b statistics measures of agreement. **Journal of Applied Statistics**, Taylor & Francis, v. 24, n. 1, p. 105–112, 1997. Citado na página 95.

NEWMAN, M. **Networks: an introduction**. [S.l.]: Oxford university press, 2010. Citado nas páginas 70 e 71.

NEWMAN, M.; BARABASI, A.-L.; WATTS, D. J. **The structure and dynamics of networks**. [S.l.]: Princeton University Press, 2011. Citado na página 72.

NEWMAN, M. E. Assortative mixing in networks. **Physical review letters**, APS, v. 89, n. 20, p. 208701, 2002. Citado na página 68.

_____. The structure and function of complex networks. **SIAM review**, SIAM, v. 45, n. 2, p. 167–256, 2003. Citado na página 68.

_____. A measure of betweenness centrality based on random walks. **Social networks**, Elsevier, v. 27, n. 1, p. 39–54, 2005. Citado na página 70.

NEWMAN, M. E.; WATTS, D. J.; STROGATZ, S. H. Random graph models of social networks. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 99, n. suppl 1, p. 2566–2572, 2002. Citado na página 68.

OLIVEIRA, F. V. B.; MUNARI, D. B.; PELZER, M. T. Bases para o cuidado de idosos portadores de paralisia supra-nuclear progressiva. 2010. Citado na página 125.

ORDÓÑEZ, A. E.; LUSCHER, Z. I.; GOGTAY, N. Neuroimaging findings from childhood onset schizophrenia patients and their non-psychotic siblings. **Schizophrenia research**, Elsevier, v. 173, n. 3, p. 124–131, 2016. Citado na página 88.

OZONOFF, S.; PENNINGTON, B. F.; ROGERS, S. J. Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. **Journal of child Psychology and Psychiatry**, Wiley Online Library, v. 32, n. 7, p. 1081–1105, 1991. Citado na página 109.

PAGE, L.; BRIN, S.; MOTWANI, R.; WINOGRAD, T. **The PageRank citation ranking: Bringing order to the web**. [S.l.], 1999. Citado na página 71.

PALANIAPPAN, S.; AWANG, R. Intelligent heart disease prediction system using data mining techniques. In: IEEE. **Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on**. [S.l.], 2008. p. 108–115. Citado na página 37.

PAMPLONA, G. S. P. **Conectividade funcional no cérebro: uma análise das associações com desempenho intelectual e atenção sustentada usando imagens por ressonância magnética**. Tese (Doutorado) — Universidade de São Paulo, 2014. Citado nas páginas 18, 77 e 80.

PARK, H.-J.; FRISTON, K. Structural and functional brain networks: from connections to cognition. **Science**, American Association for the Advancement of Science, v. 342, n. 6158, p. 1238411, 2013. Citado nas páginas 19, 75 e 83.

PEARSON, K. Principal components analysis. **The London, Edinburgh and Dublin Philosophical Magazine and Journal**, v. 6, n. 2, p. 566, 1901. Citado na página 45.

PEREIRA, F.; MITCHELL, T.; BOTVINICK, M. Machine learning classifiers and fmri: a tutorial overview. **Neuroimage**, Elsevier, v. 45, n. 1, p. S199–S209, 2009. Citado na página 90.

PETROV, D.; DODONOVA, Y.; ZHUKOV, L.; BELYAEV, M. Boosting connectome classification via combination of geometric and topological normalizations. In: IEEE. **Pattern Recognition in Neuroimaging (PRNI), 2016 International Workshop on**. [S.l.], 2016. p. 1–4. Citado nas páginas 116 e 133.

PODGORELEC, V.; KOKOL, P.; STIGLIC, B.; ROZMAN, I. Decision trees: an overview and their use in medicine. **Journal of medical systems**, Springer, v. 26, n. 5, p. 445–463, 2002. Citado na página 91.

POLAKA, I.; GAŠENKO, E.; BARASH, O.; HAICK, H.; LEJA, M. Constructing interpretable classifiers to diagnose gastric cancer based on breath tests. **Procedia Computer Science**, Elsevier, v. 104, p. 279–285, 2017. Citado na página 37.

POLANCZYK, G.; LIMA, M. S. de; HORTA, B. L.; BIEDERMAN, J.; ROHDE, L. A. The worldwide prevalence of adhd: a systematic review and metaregression analysis. **American journal of psychiatry**, Am Psychiatric Assoc, v. 164, n. 6, p. 942–948, 2007. Citado na página 117.

POLAT, K.; GÜNEŞ, S. Breast cancer diagnosis using least square support vector machine. **Digital Signal Processing**, Elsevier, v. 17, n. 4, p. 694–701, 2007. Citado na página 90.

POWER, J. D.; COHEN, A. L.; NELSON, S. M.; WIG, G. S.; BARNES, K. A.; CHURCH, J. A.; VOGEL, A. C.; LAUMANN, T. O.; MIEZIN, F. M.; SCHLAGGAR, B. L. *et al.* Functional network organization of the human brain. **Neuron**, Elsevier, v. 72, n. 4, p. 665–678, 2011. Citado na página 111.

POWER, J. D.; MITRA, A.; LAUMANN, T. O.; SNYDER, A. Z.; SCHLAGGAR, B. L.; PETERSEN, S. E. Methods to detect, characterize, and remove motion artifact in resting state fmri. **Neuroimage**, Elsevier, v. 84, p. 320–341, 2014. Citado na página 89.

PREMRAJ, R.; HERZIG, K. Network versus code metrics to predict defects: A replication study. In: IEEE. **Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on**. [S.l.], 2011. p. 215–224. Citado na página 90.

QUINLAN, J. R. Bagging, boosting, and c4. 5. In: **AAAI/IAAI, Vol. 1**. [S.l.: s.n.], 1996. p. 725–730. Citado nas páginas 57 e 58.

_____. Improved use of continuous attributes in c4. 5. **Journal of artificial intelligence research**, v. 4, p. 77–90, 1996. Citado na página 90.

RAJPUT, A.; RAJPUT, A. H. Progressive supranuclear palsy. **Drugs & aging**, Springer, v. 18, n. 12, p. 913–925, 2001. Citado na página 125.

RAMSAY, R. R.; GIOVANNI, G. D. Structure-based drug design for diagnosis and treatment of neurological diseases. **Frontiers in pharmacology**, Frontiers, v. 8, p. 13, 2017. Citado na página 37.

RÉNYI, A.; ERDOS, P. On random graphs. **Publicationes Mathematicae**, v. 6, n. 290-297, p. 5, 1959. Citado na página 71.

RHODES, D. R.; YU, J.; SHANKER, K.; DESHPANDE, N.; VARAMBALLY, R.; GHOSH, D.; BARRETTE, T.; PANDER, A.; CHINNAIYAN, A. M. Oncomine: a cancer microarray database and integrated data-mining platform. **Neoplasia**, Elsevier, v. 6, n. 1, p. 1–6, 2004. Citado na página 37.

RISH, I.; CECCHI, G.; THYREAU, B.; THIRION, B.; PLAZE, M.; PAILLIERE-MARTINOT, M. L.; MARTELLI, C.; MARTINOT, J.-L.; POLINE, J.-B. Schizophrenia as a network disease: disruption of emergent brain function in patients with auditory hallucinations. **PloS one**, Public Library of Science, v. 8, n. 1, p. e50625, 2013. Citado na página 90.

RISH, I.; HELLERSTEIN, J.; THATHACHAR, J. An analysis of data characteristics that affect naive bayes performance. **IBM TJ Watson Research Center**, v. 30, 2001. Citado na página 53.

ROCHA, H. S. C. **Estudo da técnica de VBM em imagens de ressonância magnética de cérebros com lesões**. Tese (Doutorado) — Universidade Federal de Lavras, 2015. Citado na página 89.

RODRIGUES, F. A. Caracterização, classificação e análise de redes complexas. **Instituto de Física de Sao Carlos, Universidade de Sao Paulo, Sao Carlos**, 2007. Citado nas páginas 17, 65 e 69.

ROKACH, L. **Pattern classification using ensemble methods**. [S.l.]: World Scientific, 2010. v. 75. Citado na página 50.

ROKACH, L.; MAIMON, O. **Data mining with decision trees: theory and applications**. [S.l.]: World Scientific, 2008. Citado na página 54.

_____. **Data mining with decision trees: theory and applications**. [S.l.]: World scientific, 2014. Citado na página 53.

RUBINOV, M. Schizophrenia and abnormal brain network hubs. **Dialogues in clinical neuroscience**, Les Laboratoires Servier, v. 15, n. 3, p. 339, 2013. Citado na página 88.

RUBINOV, M.; KNOCK, S. A.; STAM, C. J.; MICHELOYANNIS, S.; HARRIS, A. W.; WILLIAMS, L. M.; BREAKSPEAR, M. Small-world properties of nonlinear brain activity in schizophrenia. **Human brain mapping**, Wiley Online Library, v. 30, n. 2, p. 403–416, 2009. Citado na página 87.

RUBINOV, M.; SPORNS, O. Complex network measures of brain connectivity: uses and interpretations. **Neuroimage**, Elsevier, v. 52, n. 3, p. 1059–1069, 2010. Citado nas páginas 19, 81, 82, 83 e 84.

_____. Weight-conserving characterization of complex functional brain networks. **Neuroimage**, Elsevier, v. 56, n. 4, p. 2068–2079, 2011. Citado na página 83.

RUDIE, J. D.; BROWN, J.; BECK-PANCER, D.; HERNANDEZ, L.; DENNIS, E.; THOMPSON, P.; BOOKHEIMER, S.; DAPRETTO, M. Altered functional and structural brain network organization in autism. **NeuroImage: clinical**, Elsevier, v. 2, p. 79–94, 2013. Citado na página 111.

SAMANT, P.; AGARWAL, R. Machine learning techniques for medical diagnosis of diabetes using iris images. **Computer methods and programs in biomedicine**, Elsevier, v. 157, p. 121–128, 2018. Citado na página 37.

SATO, J. R.; HOEXTER, M. Q.; FUJITA, A.; ROHDE, L. A. Evaluation of pattern recognition and feature extraction methods in adhd prediction. **Frontiers in systems neuroscience**, Frontiers Media SA, v. 6, 2012. Citado nas páginas 117 e 118.

SAUNDERS, C.; GAMMERMAN, A.; VOVK, V. Ridge regression learning algorithm in dual variables. 1998. Citado na página 44.

SCANNELL, J.; BURNS, G.; HILGETAG, C.; O'NEIL, M.; YOUNG, M. P. The connectional organization of the cortico-thalamic system of the cat. **Cerebral Cortex**, Oxford University Press, v. 9, n. 3, p. 277–299, 1999. Citado na página 74.

SCHANK, T.; WAGNER, D. **Approximating clustering-coefficient and transitivity**. [S.l.]: Universität Karlsruhe, Fakultät für Informatik, 2004. Citado na página 68.

SCHAPIRE, R. E. The strength of weak learnability. **Machine learning**, Springer, v. 5, n. 2, p. 197–227, 1990. Citado na página 58.

SEIDMAN, S. B. Network structure and minimum degree. **Social networks**, Elsevier, v. 5, n. 3, p. 269–287, 1983. Citado na página 71.

SHAFFER, J. L.; PETRELLA, J. R.; SHELDON, F. C.; CHOUDHURY, K. R.; CALHOUN, V. D.; COLEMAN, R. E.; DORAISWAMY, P. M.; INITIATIVE, A. D. N. Predicting cognitive decline in subjects at risk for alzheimer disease by using combined cerebrospinal fluid, mr imaging, and pet biomarkers. **Radiology**, Radiological Society of North America, Inc., v. 266, n. 2, p. 583–591, 2013. Citado na página 90.

SHI, H. **Best-first decision tree learning**. Tese (Doutorado) — The University of Waikato, 2007. Citado na página 54.

SHINE, J. M.; POLDRACK, R. A. Principles of dynamic network reconfiguration across diverse brain states. **NeuroImage**, Elsevier, v. 180, p. 396–405, 2018. Citado na página 37.

SHIRWALKAR, N.; GURSALKAR, S.; TAK, T.; KALSHETTI, A. Human heart disease prediction system using data mining techniques. 2018. Citado na página 37.

SHREE, S. B.; SHESHADRI, H.; NAGARAJ, M. K.; PRINCE, M.; FALL, C. H.; KRISHNA, M. Application of machine learning methods for diagnosis of dementia based on the 10/66 battery of cognitive function tests in south india. **Social psychiatry and psychiatric epidemiology**, Europe PMC Funders, v. 53, n. 1, p. 77, 2018. Citado na página 37.

SILVA, C. B. d. S. **Processamento de sinais de ressonância magnética nuclear usando classificador neural para reconhecimento de carne bovina**. Tese (Doutorado) — Universidade de São Paulo, 2007. Citado nas páginas 18 e 78.

SINGH, B. K.; VERMA, K.; THOKE, A. Investigations on impact of feature normalization techniques on classifier's performance in breast tumor classification. **International Journal of Computer Applications**, Foundation of Computer Science, v. 116, n. 19, 2015. Citado na página 49.

SIQUEIRA, A. dos S.; JUNIOR, B.; EDUARDO, C.; COMFORT, W. E.; ROHDE, L. A.; SATO, J. R. Abnormal functional resting-state networks in adhd: graph theory and pattern recognition analysis of fmri data. **BioMed research international**, Hindawi, v. 2014, 2014. Citado nas páginas 122 e 134.

SKÅTUN, K. C.; KAUFMANN, T.; TØNNESEN, S.; BIELE, G.; MELLE, I.; AGARTZ, I.; ALNÆS, D.; ANDREASSEN, O. A.; WESTLYE, L. T. Global brain connectivity alterations in patients with schizophrenia and bipolar spectrum disorders. **Journal of psychiatry & neuroscience: JPN**, Canadian Medical Association, v. 41, n. 5, p. 331, 2016. Citado na página 90.

SLICHTER, C. P. **Principles of magnetic resonance**. [S.l.]: Springer Science & Business Media, 2013. v. 1. Citado na página 76.

SMITH, S. M.; JENKINSON, M.; WOOLRICH, M. W.; BECKMANN, C. F.; BEHRENS, T. E.; JOHANSEN-BERG, H.; BANNISTER, P. R.; LUCA, M. D.; DROBNJAK, I.; FLITNEY, D. E. Advances in functional and structural mr image analysis and implementation as fsl. **Neuroimage**, Elsevier, v. 23, p. S208–S219, 2004. Citado na página 89.

SONI, J.; ANSARI, U.; SHARMA, D.; SONI, S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. **International Journal of Computer Applications**, v. 17, n. 8, p. 43–48, 2011. Citado nas páginas 37 e 42.

SPORNS, O. Network analysis, complexity, and brain function. **Complexity**, Wiley Online Library, v. 8, n. 1, p. 56–60, 2002. Citado nas páginas 75 e 83.

_____. **Networks of the Brain**. [S.l.]: MIT press, 2010. Citado nas páginas 76, 87, 88, 109 e 117.

_____. The human connectome: a complex network. **Annals of the New York Academy of Sciences**, Wiley Online Library, v. 1224, n. 1, p. 109–125, 2011. Citado na página 37.

_____. Network attributes for segregation and integration in the human brain. **Current opinion in neurobiology**, Elsevier, v. 23, n. 2, p. 162–171, 2013. Citado nas páginas 20, 83 e 85.

SPORNS, O.; CHIALVO, D. R.; KAISER, M.; HILGETAG, C. C. Organization, development and function of complex brain networks. **Trends in cognitive sciences**, Elsevier, v. 8, n. 9, p. 418–425, 2004. Citado na página 81.

SPORNS, O.; HONEY, C. J.; KÖTTER, R. Identification and classification of hubs in brain networks. **PloS one**, Public Library of Science, v. 2, n. 10, p. e1049, 2007. Citado na página 90.

SPORNS, O.; TONONI, G. Classes of network connectivity and dynamics. **Complexity**, Wiley Online Library, v. 7, n. 1, p. 28–38, 2001. Citado na página 75.

- SPORNS, O.; TONONI, G.; KÖTTER, R. The human connectome: a structural description of the human brain. **PLoS computational biology**, Public Library of Science, v. 1, n. 4, p. e42, 2005. Citado nas páginas 18, 74 e 75.
- SPORNS, O.; ZWI, J. D. The small world of the cerebral cortex. **Neuroinformatics**, Springer, v. 2, n. 2, p. 145–162, 2004. Citado na página 74.
- SRINIVAS, K.; RANI, B. K.; GOVRDHAN, A. Applications of data mining techniques in healthcare and prediction of heart attacks. **International Journal on Computer Science and Engineering (IJCSE)**, v. 2, n. 02, p. 250–255, 2010. Citado na página 42.
- STAM, C. J. Modern network science of neurological disorders. **Nature reviews. Neuroscience**, v. 15 10, p. 683–95, 2014. Citado na página 73.
- STAM, C. J.; REIJNEVELD, J. C. Graph theoretical analysis of complex networks in the brain. **Nonlinear biomedical physics**, BioMed Central, v. 1, n. 1, p. 3, 2007. Citado na página 74.
- STEELE, J. C.; RICHARDSON, J. C.; OLSZEWSKI, J. Progressive supranuclear palsy: a heterogeneous degeneration involving the brain stem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. **Archives of neurology**, American Medical Association, v. 10, n. 4, p. 333–359, 1964. Citado na página 125.
- STROBL, C.; MALLEY, J.; TUTZ, G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. **Psychological methods**, American Psychological Association, v. 14, n. 4, p. 323, 2009. Citado na página 91.
- STURZBECHER, M. J. **Detecção e caracterização da resposta hemodinâmica pelo desenvolvimento de novos métodos de processamento de imagens funcionais por ressonância magnética**. Tese (Doutorado) — Universidade de São Paulo, 2006. Citado na página 77.
- SUMATHI, S.; SIVANANDAM, S. **Introduction to data mining and its applications**. [S.l.]: Springer, 2006. v. 29. Citado nas páginas 55 e 57.
- SUPEKAR, K.; MENON, V.; RUBIN, D.; MUSEN, M.; GREICIUS, M. D. Network analysis of intrinsic functional brain connectivity in alzheimer’s disease. **PLoS computational biology**, Public Library of Science, v. 4, n. 6, p. e1000100, 2008. Citado na página 88.
- TAN, H.-Y.; SUST, S.; BUCKHOLTZ, J. W.; MATTAY, V. S.; MEYER-LINDENBERG, A.; EGAN, M. F.; WEINBERGER, D. R.; CALLICOTT, J. H. Dysfunctional prefrontal regional specialization and compensation in schizophrenia. **American Journal of Psychiatry**, Am Psychiatric Assoc, v. 163, n. 11, p. 1969–1977, 2006. Citado na página 87.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. [S.l.]: Boston: Pearson Addison Wesley., 2005. Citado na página 41.
- TANWANI, A. K.; AFRIDI, J.; SHAFIQ, M. Z.; FAROOQ, M. Guidelines to select machine learning scheme for classification of biomedical datasets. In: SPRINGER. **European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics**. [S.l.], 2009. p. 128–139. Citado nas páginas 57, 58 e 91.
- TEIXEIRA, S. C. **Utilização de análise de componentes principais em séries temporais**. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 45.

- THEODORIDIS, S.; KOUTROUMBAS, K. Pattern recognition and neural networks. In: **Machine Learning and Its Applications**. [S.l.]: Springer, 2001. p. 169–195. Citado na página [63](#).
- THERNEAU, T. M.; ATKINSON, B.; RIPLEY, M. B. **The rpart package**. 2010. Citado na página [90](#).
- TIBSHIRANI, R.; HASTIE, T.; NARASIMHAN, B.; CHU, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 99, n. 10, p. 6567–6572, 2002. Citado na página [37](#).
- TONONI, G.; SPORNS, O.; EDELMAN, G. M. A measure for brain complexity: relating functional segregation and integration in the nervous system. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 91, n. 11, p. 5033–5037, 1994. Citado na página [83](#).
- TU, J. V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. **Journal of clinical epidemiology**, Elsevier, v. 49, n. 11, p. 1225–1231, 1996. Citado na página [91](#).
- VEAGUE, H. B.; COLLINS, C. E. **Personality disorders**. [S.l.]: Infobase Publishing, 2007. Citado na página [87](#).
- VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S-PLUS**. [S.l.]: Springer Science & Business Media, 2013. Citado nas páginas [90](#) e [91](#).
- VÉRTES, P. E.; ALEXANDER-BLOCH, A. F.; GOGTAY, N.; GIEDD, J. N.; RAPOPORT, J. L.; BULLMORE, E. T. Simple models of human brain functional networks. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 109, n. 15, p. 5868–5873, 2012. Citado nas páginas [88](#), [89](#), [90](#) e [106](#).
- WANG, L.; ZHU, C.; HE, Y.; ZANG, Y.; CAO, Q.; ZHANG, H.; ZHONG, Q.; WANG, Y. Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. **Human brain mapping**, Wiley Online Library, v. 30, n. 2, p. 638–649, 2009. Citado na página [117](#).
- WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **Nature**, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998. Citado nas páginas [37](#), [68](#), [72](#) e [73](#).
- WHITE, J. G.; SOUTHGATE, E.; THOMSON, J. N.; BRENNER, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*: the mind of a worm. **Philosophical Transactions of the Royal Society of London**, v. 314, p. 1–340, 1986. Citado na página [73](#).
- WHITFIELD-GABRIELI, S.; NIETO-CASTANON, A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. **Brain connectivity**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 2, n. 3, p. 125–141, 2012. Citado na página [77](#).
- WIESE, I. S. **Predição de mudanças conjuntas de artefatos de software com base em informações contextuais**. Tese (Doutorado) — Universidade de São Paulo, 2016. Citado na página [97](#).
- WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J. **Data Mining: Practical machine learning tools and techniques**. [S.l.]: Morgan Kaufmann, 2016. Citado na página [41](#).

- WOLPERT, D. H. Stacked generalization. **Neural networks**, Elsevier, v. 5, n. 2, p. 241–259, 1992. Citado na página [58](#).
- XU, X.; GUAN, X.; GUO, T.; ZENG, Q.; YE, R.; WANG, J.; ZHONG, J.; XUAN, M.; GU, Q.; HUANG, P. *et al.* Brain atrophy and reorganization of structural network in parkinson's disease with hemiparkinsonism. **Frontiers in human neuroscience**, Frontiers, v. 12, p. 117, 2018. Citado na página [38](#).
- YANG, Y.; IYER, L. K.; ADELSTEIN, S. J.; KASSIS, A. I. Integrative genomic data mining for discovery of potential blood-borne biomarkers for early diagnosis of cancer. **PloS one**, Public Library of Science, v. 3, n. 11, p. e3661, 2008. Citado na página [42](#).
- YAO, X.; LIU, Y. Evolving artificial neural networks for medical applications. In: **Proc. of**. [S.l.: s.n.], 1995. p. 1–16. Citado na página [91](#).
- YUE, W.; WANG, Z.; CHEN, H.; PAYNE, A.; LIU, X. Machine learning with applications in breast cancer diagnosis and prognosis. **Designs**, Multidisciplinary Digital Publishing Institute, v. 2, n. 2, p. 13, 2018. Citado na página [37](#).
- ZHANG, Z. Artificial neural network. In: **Multivariate Time Series Analysis in Climate and Environmental Research**. [S.l.]: Springer, 2018. p. 1–35. Citado na página [55](#).
- ZHUKOV, L. E.; PETROV, D.; DODONOVA, Y. Differences in structural connectomes between typically developing and autism groups. In: "**Information Technologies and Systems 2015**". [S.l.]: Institute for Information Transmission Problems. AA Kharkevich RAS, 2015. p. 1–15. Citado nas páginas [111](#), [116](#) e [134](#).
- ZUO, X.-N.; EHMKE, R.; MENNES, M.; IMPERATI, D.; CASTELLANOS, F. X.; SPORNS, O.; MILHAM, M. P. Network centrality in the human functional connectome. **Cerebral cortex**, Oxford University Press, v. 22, n. 8, p. 1862–1875, 2011. Citado na página [90](#).