

Análise de Características a partir de Algoritmos de Aprendizagem de Máquina para Auxílio ao Diagnóstico do Transtorno do Espectro Autista

Matheus Frota, Samuel Hericles, Gerônimo Aguiar, Pedro Renoir, Rayon Nunes,
Manoel Vilela, Denilson Gomes, Iális Cavalcante
Universidade Federal do Ceará - Campus de Sobral
Engenharia da Computação, Sobral-CE, Brasil

matheusfrota@hotmail.com.br, samuel.herilces@alu.ufc.br, geron@alu.ufc.br, renoir@alu.ufc.br,
rayonnunes@hotmail.com, manoel_vilela@engineer.com, denilsongomes@alu.ufc.br, ialis@sobral.ufc.br

ABSTRACT

Machine learning algorithms are being successfully applied in many areas of knowledge. In digital health, these ones permit for professionals to be helped to diagnose diseases and disorders in advance and more accurately, contributing to the effective treatment of their patients. In this context, the proposed methodology makes use of a public database on Autistic Spectrum Disorder and the most relevant characteristics presented by a patient. So, it analyzes with the Decision Tree, Support Vector Machine, Multi-Layer Perceptron and K-nearest neighbor algorithms to the construction of a model capable of simplifying a decision strategy to help the diagnosis of this disorder type.

keywords:Autism, Machine learning, Feature importance.

Resumo. Algoritmos de aprendizado de máquina estão sendo aplicados com sucesso em diversas áreas do conhecimento. No campo da saúde, estes mesmos abrem espaço para que profissionais possam ser auxiliados a diagnosticar doenças e transtornos antecipadamente e com maior precisão, contribuindo no tratamento eficaz de seus pacientes. Neste contexto, a metodologia proposta faz uso de uma base de dados pública sobre o Transtorno do Espectro Autista e as características mais relevantes apresentada por um paciente. Para isso, analisa com os algoritmos de Árvore de Decisão, *Support Vector Machine*, Perceptron de Múltiplas Camadas e K-vizinhos mais próximos a construção de um modelo capaz de simplificar uma estratégia de decisão no auxílio ao diagnóstico deste tipo de transtorno.

palavras-chaves:Autismo, Aprendizagem de máquina, Importância de características.

1. INTRODUÇÃO

Desde maio de 2013 os psicólogos e psiquiatras estão utilizando os critérios de avaliação do Manual de Diagnóstico e Estatístico de Transtorno Mentais (conhecido como DSM-5), elaborado pela *American Psychiatric Association* [21]. Os casos anteriormente analisados, utilizando o DSM-IV, que recebiam o diagnóstico de transtorno autista, transtorno de *Asperger* ou transtorno de desenvolvimento penetrante, que não seja especificado de outra forma, passam a ser nomeados de transtorno do espectro autista, podendo ainda ser caracterizado em 3 (três) diferentes níveis de gravidade. Os sintomas desses transtornos representam um contínuo único de prejuízos às pessoas portadoras, com intensidades variáveis, que vão de leve a grave, nos domínios de comunicação social e de comportamentos restritivos e repetitivos. Este estudo destaca que o transtorno do espectro do autismo (TEA), refere-se a uma ampla gama de condições caracterizadas por desafios com habilidades sociais, comportamentos repetitivos, fala e comunicação não-verbal [4].

O autismo afeta cerca de 1 em 59 crianças hoje nos Estados Unidos. Dessa forma, estima-se que o Brasil, com seus 200 milhões de habitantes, possui cerca de 3 milhões de autistas. O autista possui um conjunto distinto de traços comportamentais. As maneiras pelas quais as pessoas com autismo aprendem, pensam e resolvem problemas podem variar de muito qualificadas a severamente desafiadas pela situação [6].

Tendo em vista a problemática de alcançar o diagnóstico do autismo, o presente trabalho propõe um método para auxiliar especialistas na busca por um diagnóstico deste transtorno. Com base em dados público sobre um grupo crianças autistas, pretende-se analisar características descritas na base de dados junto as informações necessárias cedidas pelos responsáveis, acelerando o procedimento. Para colaborar com a análise de muitos dados, adota-se alguns algoritmos relevante na literatura sobre aprendizagem de máquina para buscar uma análise automática de auxílio aos especialistas [22, 24, 27]. Foram utilizados quatro tipos de algoritmos, visando a classificação das características comportamentais em crianças, com um determinado grau de autismo ou não. É importante destacar a ideia principal, que visa apenas fornecer um auxílio com a previsão ao diagnóstico. Caso o resultado seja positivo significa que a criança pode apresentar algum grau de autismo e, portanto, o especialista pode fazer uso dos resultados desta aplicação como base para o

laudo final.

O artigo está organizado como segue. Na seção 2 descreve-se os trabalhos relacionados à classificação e TEA. Na seção 3 é descrita toda estrutura da base de dados utilizada, apontando os atributos utilizados. A seção 4 aborda toda a fundamentação teórica do trabalho. Na seção 5 são apresentados os resultados obtidos, mostrando o gráfico de importância das variáveis utilizadas no modelo. Por fim, a seção 6 conclui o trabalho e sugere aplicações futuras.

2. TRABALHO RELACIONADOS

Teixeira *et al.* (2010) realiza uma revisão sistemática, nos anos de 2002 a 2009, a cerca da produção científica de autores brasileiros relacionados ao TEA. A pesquisa foi realizada partir das bases de dados públicas em 4 (quatro) diferentes repositórios, incluindo diversos descritores, tais como autismo e transtorno invasivo do desenvolvimento. Houve a análise de 93 artigos, em que aproximadamente um terço dos mesmo foi publicado em periódicos, mas a maioria dos artigos se baseava em amostras pequenas. Notou-se que o principal tema de pesquisa abordado estava ligado a programas de intervenção para TEA. Ao final, a referida revisão dispôs sobre o interesse de pesquisadores brasileiros na temática dos TEA, entretanto, a maior parte dessa produção científica se concentra em dissertações/teses em detrimento de poucas publicações em revistas relevantes. Esta revisão sistemática mostrava que há muito tempo faz-se necessário a abertura de novos estudos com amostras maiores que levariam a um maior impacto e visibilidade da produção científica brasileira relativa aos TEA [23].

Em Brito & Fernandes (2019) foi proposto o uso de redes neurais para classificar crianças com presença ou não de TEA. Utilizaram uma base de dados com 259 amostras e com as estratégias de *cross validation* e *k-fold* para auxílio no treinamento. Os dois modelos comportaram-se bem mas os autores deram como inconclusivo os seus resultados [6].

Processamento de imagens e algoritmos de aprendizado de máquina foram empregados por Liu *et al.* (2016) para classificar crianças com autismo. Os dados foram de rosto de crianças junto com um conjunto de dados do movimento ocular dessas, por fim obtiveram um acurácia de 88,51% nos resultados [15].

A proposta de Halibas *et al.* (2018) analisa cinco modelos de aprendizado de máquina para estudar o problema da desordem do espectro autista em uma base específica. Os autores fazem um comparativo entre os modelos com as métricas de acurácia, precisão e revocação. A base de dados utilizada foi definida a partir de 1100 amostras com 10 atributos comportamentais, sendo crianças, adolescentes e adultos. As amostras já possuem a classe conhecida: "Pertencente" ou "Não Pertencente" ao espectro autista. Este estudo apontou o classificador baseado em *Deep Learning* como o modelo que alcança melhor desempenho em quase todas as métricas, com acurácia de 96,38% [11].

O trabalho de Dutta *et al.* (2017) mostra um modelo de aprendizagem de máquina como método para auxiliar o diagnóstico do autismo em crianças, mas também mostra sua aplicação na previsão de outras síndromes e transtornos, mas tendo seu melhor funcionamento para a previsão do autismo. Sua proposta envolve uma Regra de Associação (AR) baseado em um método de valor mínimo de Redundância e Máxima Relevância (mRMR), usados para extrair os sintomas que precisam ser testados para o processo de previsão, a

partir de uma base de dados. Foi utilizado o método de Diferença de Informação Mútua (MID) junto ao maior ganho de informação (HIG) para selecionar os sintomas adicionais, através de casos anteriores, e que podem estar vinculados aos sintomas principais [7]. O método de diagnóstico recebe informações de vários sensores e smartphones. Os dados do sensor são buscados através de um aplicativo por um aparelho mobile. Executando o aplicativo, os sintomas preliminares tentam puxar os sintomas associados em várias iterações do banco de dados através da Internet. O banco de dados utilizado contém registros de pacientes anteriores com sintomas padrão. O tal sistema proposto precisa de um ou dois sintomas primários como entradas para seu funcionamento, os sintomas são ainda testados e confirmados. O procedimento foi feito em um conjunto de 200 crianças portadoras de transtornos variados, com 10 tipos de autismos e 50 sintomas distintos. Os resultados obtidos na aplicação de tal método para todos os casos envolveram taxas entre 70 a 86% de precisão. No geral, os experimentos demonstraram funcionamento bem melhor para casos específicos, mais voltados ao desenvolvimento intelectual das crianças, onde pode ainda se ter uma melhor recuperação no desenvolvimento cognitivo. Ao mesmo tempo, não se mostra tão eficiente para casos que demonstram uma maior dificuldade na socialização da criança, por exemplo. Por fim, o artigo faz ressalvas sobre os exames de sintomas preliminares que geralmente não são suficientes para uma boa previsão do início do autismo em crianças, e que sintomas preliminares acabam indicando diferentes diagnósticos [7].

Este transtorno também foi avaliado em algoritmos de aprendizagem de máquina não supervisionados. Em Elbattah *et al.* (2019), a idéia principal é aprender agrupamentos de dados com base na representação visual de percursos de rastreamento ocular. Levando em consideração que o TEA é uma condição ao longo da vida caracterizada por prejuízos sociais e de comunicação, analisou o rastreamento ocular de 59 crianças organizadas em dois grupos: autistas e não-autistas. Aplicou-se a tecnologia de eye-tracking para rastreamento das variações de direcionamento das visões dos pacientes que se deparavam com imagens de formas e geravam caminhos de visadas entre os mesmos. Avaliou-se três hipóteses principais: Os padrões visuais dos caminhos de rastreamento de rastreamento ocular indicariam uma estrutura subjacente de grupos? Se sim, os aglomerados descobertos poderiam revelar possíveis conexões relacionadas à dinâmica do comportamento do olhar (por exemplo, velocidade, aceleração)? Além disso, como os agrupamentos variariam em relação às características dos participantes (por exemplo, idade)? O modelo de agrupamento foi treinado usando representações compactadas aprendidas por um autoencoder profundo [3]. Seus resultados atingiram uma tendência promissora da estrutura de agrupamento. Além disso, os grupos são explorados para fornecer informações interessantes sobre as características do comportamento do olhar envolvido no autismo. Confirmou-se empiricamente que os caminhos de rastreamento ocular podiam ser agrupados em grupos coerentes, que se assemelhavam amplamente ao agrupamento original de amostras (ou seja, autistas ou não-autistas). Isso traz um significado nas visualizações do caminho de digitalização que podem efetivamente discriminar as amostras diagnosticadas com autistas e outros grupos [8].

No artigo de Sartipi, Shayesteh & Kalbkhani (2018) faz-se uso de algoritmos de aprendizado de máquina em apli-

cações médicas a fim de permitir um diagnóstico rápido e preciso de doenças, em específico para o caso de TEA. Seu objetivo é classificar entre autistas e não-autistas usando as informações de dados de ressonância magnética funcional incluindo os de em estado de repouso (do inglês *Resting State functional Magnetic Resonance Imaging*, rs-fMRI). Em sua proposta, cada região de interesse (do inglês *Region of Interest*, ROI) de dados de cada sujeito é decomposta usando a transformada wavelet discreta *double-density dual-tree* (D3TDWT) dem sub-bandas de frequência de tempo. Num momento seguinte, o modelo de heteroscedasticidade condicional auto-regressiva generalizada (do inglês *generalized autoregressive conditional heteroskedasticity*, GARCH) é usado para a extração de recursos dessas sub-bandas. Posteriormente, as características mais discriminantes são selecionadas pelo teste-t de duas amostras e, finalmente, os dados são classificados por *Support Vector Machine*. Este foi testado em vários conjuntos de dados e os resultados atingiram uma precisão de classificação de 71,6% para indivíduos do sexo masculino e, para o sexo feminino, 93,7%. Considerando os ROIs significativos, há uma redução nas conexões ântero-posteriores entre os autistas, e que pode ser considerado em abordagens clínicas [20].

A metodologia apresentada em Akter *et al.* (2019) destaca que (TEA) é um grupo de deficiências no desenvolvimento neurológico que não são curáveis, mas que podem ser amenizado por intervenções precoces. O referido trabalho abordou um conjuntos de dados de autistas detectados precocemente em grupos de bebês, crianças, adolescentes e adultos e aplicou vários métodos de transformação de recursos. Várias técnicas de classificação foram implementadas com esses conjuntos de dados, que foram transformados e avaliados em seu desempenho. Notou-se que o SVM mostrou o melhor desempenho para o conjunto de dados da criança, enquanto que o GLMboost para os adolescentes, e o Adaboost apresentou os melhores resultados para o conjunto de dados da criança e de adultos. As transformações de recursos que resultam na melhores classificações foram função seno para a criança e *z-score* para conjuntos de dados de crianças e adolescentes. Após estas análises, várias técnicas de seleção de recursos foram usadas com esses conjuntos de dados transformados em *z-score* para identificar os fatores de risco significativos de TEA nos grupos de diferentes faixas etárias. Os resultados destas abordagens analíticas indicam que, quando adequadamente otimizados, os métodos de aprendizado de máquina podem fornecer boas previsões de estado de autismo. Isso sugere que pode ser possível aplicar esses modelos para a detecção de autismo em seus estágios iniciais [2].

Algumas estratégias de uso de algoritmos de aprendizagem de máquina foram centradas em comitê de classificadores. O mesmo pode-se tratar do trabalho de Yuan & Ma (2012), que já destacava que as tentativas existentes de melhorar o desempenho do AdaBoost em conjuntos de dados desequilibrados têm se concentrado amplamente na modificação de sua regra de atualização de peso ou na incorporação de amostras ou técnicas de aprendizado sensíveis a custos. Os autores buscaram modelar um comitê de classificadores com uma nova perspectiva. Inicialmente, o conjunto de dados passa por uma superamostragem e o AdaBoost padrão é aplicado para criar uma série de classificadores básicos. Em seguida, os pesos dos classificadores são novamente treinados por algoritmos genéticos ou técnicas de otimização

comparáveis, nas quais medidas de desempenho mais direcionadas podem ser usadas diretamente como função objetivo. Consequentemente, essa estratégia de amostragem unida à reponderação pode propositalmente ajustar o AdaBoost em direção a uma determinada medida de desempenho de interesse. Resultados experimentais em dez conjuntos de dados de benchmark mostram que essa estratégia pode aumentar o desempenho do AdaBoost de maneira confiável e tem superioridade consistente sobre outros métodos associados e que são apresentados na literatura [27].

O trabalho de Gok (2019) também destaca que o diagnóstico de TEA nos estágios iniciais é muito desejável para estabelecer o tratamento rapidamente. Do ponto de vista do aprendizado de máquina, a tarefa de previsão de risco de autismo é uma classificação binária dos genes de risco de autismo, ou seja, se um RNA longo não codificado (do inglês *long non-coding RNA*, lncRNA) causa o transtorno ou não. Buscou-se desenvolver um modelo de aprendizado de máquina, treinado a partir de dados de expressão gênica no desenvolvimento cerebral, para a classificação binária dos genes de risco de autismo. O modelo foi composto de duas partes principais: extração de recursos com transformada de wavelet de Haar, métodos de discretização e classificação com o algoritmo de aprendizado de rede bayesiana. Apresentou-se uma comparação o modelo proposto em outros algoritmos de classificação independentes que abordaram dados sobre o gene lncRNA. Os resultados experimentais obtiveram uma importante eficiência do modelo, principalmente com a sensibilidade, área sob curva ROC e pontuações da medida F1-Score de 90,2%, 83,9%, e 80,6%, respectivamente [9].

O estudo de Usta *et al.* (2019) segue uma linha de pesquisa junto a outros trabalhos que mostram melhorias parciais em alguns sintomas principais dos TEA com o decorrer do tempo. No entanto, os fatores preditivos (por exemplo, pré-tratamento, distúrbios psiquiátricos comórbidos, habilidades adaptativas e de linguagem etc.) para melhorar o resultado não foram estudados com métodos de aprendizado de máquina. O objetivo dos autores envolveu examinar os cálculos preditores de resultados com métodos de aprendizado de máquina, incluindo estimativa estatística, teorias da informação e aprendizado matemático, descobrindo automaticamente padrões úteis em grandes quantidades de dados. Adotou-se um grupo de amostras que compreendeu 433 crianças (na faixa de 3 a 6 anos de idade) com diagnóstico de TEA. Os sintomas das mesmas foram avaliados pela Lista de Verificação do Comportamento do Autismo, Lista de Verificação do Comportamento Aberrante e escalas de Impressão Clínica Global (do inglês *Clinical Global Impression*, CGI)[26]. Foi avaliado o desempenho de algoritmos de aprendizado de máquina (Naive Bayes, Modelo Linear Generalizado, Regressão Logística, Árvore de Decisão) nos referidos dados, incluindo os 254 itens nos formulários de linha de base. Pacientes com menos de 2 pontos de CGI nos sintomas de TEA em 36 meses foram aceitos como o grupo que apresentou de melhor resultado para a classe de previsão. Uma proporção significativa dos casos apresentou melhora significativa nos sintomas de TEA. No modelo de aprendizado de máquina assumido em uma escala CGI mostrou-se que o grupo de diagnóstico afetou o prognóstico. No grupo de autismo, pai e mãe mais velhos, menor peso ao nascer e maior idade no diagnóstico apresentam pior resultado. Na idade do transtorno de Asperger no diagnóstico, a idade da primeira avaliação e os pilares do desenvolvimento

afetaram o prognóstico. Acompanhando outros trabalhos na literatura, identificou-se diagnóstico precoce, reabilitação precoce e severidade dos sintomas de TEA na avaliação inicial do resultado previsto. Além disso, observou-se que diagnósticos psiquiátricos comórbidos estão afetando o resultado dos sintomas de TEA na observação clínica. Os modelos de aprendizado de máquina revelam que vários outros são mais significativos (por exemplo, idade dos pais, peso ao nascer, variáveis sociodemográficas, etc.) em termos de informações prognósticas e também no planejamento do tratamento de crianças com TEA [26].

A pesquisa de Abbas *et al.* (2018) engloba dois algoritmos treinados para identificar o autismo. Um destes é baseado em questionários estruturados e curtos, relatados pelos pais. E o outro algoritmo usa a identificação de comportamentos-chave a partir de vídeos caseiros semiestruturados de crianças. Um algoritmo de combinação é então usado para combinar os resultados em uma única avaliação de maior precisão. Para superar a escassez e desequilíbrio dos dados de treinamento, aplicou-se técnicas de seleção de características, *feature engineering* e técnicas de codificação de características. Permitiu-se uma determinação inconclusiva, quando apropriado, a fim de aumentar a precisão da triagem quando conclusiva. O desempenho é então validado em um estudo clínico controlado, que se seguiu com 162 crianças para verificar o desempenho desses algoritmos e sua combinação. Atingiu-se uma melhoria significativa da precisão em relação às ferramentas de triagem padrão nas medições de acurácia, sensibilidade e especificidade. O estudo assegura que um processo de aprendizado de máquina é um método confiável para a detecção de autismo fora dos ambientes clínicos. Uma variedade de fatores de confusão na análise clínica é discutida juntamente com as soluções projetadas nos algoritmos. Os resultados finais são estatisticamente limitados e demandam futuros estudos clínicos para ampliar o tamanho da amostra [1].

A abordagem de Ben-Sasson, Robins & Yom-Tov (2018) possui o objetivo de testar a viabilidade da avaliação do risco de desordem do espectro do autismo nas preocupações dos pais a partir de fontes baseadas na Web, usando ferramentas de análise de texto automatizadas e questionamento de padrão mínimo. Os participantes desta pesquisa eram 115 pais, preocupados com o desenvolvimento da comunicação social de seus filhos. As crianças tinham entre 16 e 30 meses de idade e 66 das mesmas tinham histórico familiar de transtorno do espectro do autismo. Os pais relataram suas preocupações na internet e completaram um questionário para um rastreador específico em TEA, a Lista de Verificação Modificada e Revisada para Autismo em Crianças (do inglês *Modified Checklist for Autism in Toddlers-Revised*). Fizeram uso também de um amplo rastreador de desenvolvimento, denominado Questionário de Idades e Estágios (do inglês *Ages and Stages Questionnaire*, ASQ). Um algoritmo previa o risco de distúrbio do espectro do autismo usando uma combinação do texto dos pais e uma única pergunta de triagem, selecionada pelo algoritmo para aprimorar a precisão da previsão. As medidas de triagem identificaram 58 a 88% das crianças em risco de desordem do espectro autista. Crianças com histórico familiar de TEA apresentaram três vezes mais chances de apresentar risco de transtorno do espectro do autismo nas medidas de triagem. A previsão do risco de uma criança no ASQ ou na lista de verificação foi significativamente mais precisa quando prevista a partir do

texto combinado com uma pergunta selecionada automaticamente do que apenas no texto. As perguntas da lista de verificação frequentemente selecionadas automaticamente que previam risco eram: seguir um ponto, brincadeira de faz de conta e preocupação com surdez. Observou-se que a internet pode ser usada para pré-rastrear distúrbios do espectro do autismo, usando as preocupações dos pais, administrando algumas perguntas de triagem padronizadas para aumentar esse processo [5].

Com essa revisão dos trabalhos relacionados foi possível entender a problemática de análise de crianças e outros grupos etários com características de TAE, além da complexidade de se aplicar técnicas de aprendizagem de máquina nesse âmbito das bases de dados disponíveis. Este trabalho assume uma análise das características que mais caracterizam o TAE em um paciente sobre o resultado de um algoritmo de classificação de dados.

3. BASE DE DADOS

A base de dados utilizada neste trabalho, disponível em [25], conta com um total de 292 observações e 21 atributos. Dentre esses atributos, tem 10 variáveis comportamentais que são descritas na Tabela 1.

| Variável | Descrição |
|-----------|---|
| A1_Score | Alta percepção em baixos ruídos que geralmente outros não percebem |
| A2_Score | Maior concentração na visão do todo em comparação a pequenos detalhes |
| A3_Score | Facilidade de comunicação com várias pessoas diferentes ao mesmo tempo |
| A4_Score | Facilidade de fazer múltiplas tarefas simultaneamente |
| A5_Score | Dificuldade de manter uma conversa com seus colegas |
| A6_Score | Facilidade em manter conversas informais |
| A7_Score | Dificuldade de percepção de intenções e sentimentos em histórias |
| A8_Score | Dificuldade de brincar com a imaginação fugindo da realidade |
| A9_Score | Facilidade de reconhecer sentimentos a partir de expressões faciais alheias |
| A10_Score | Dificuldade de fazer novas amizades |
| result | Soma das características comportamentais avaliadas relacionados ao TEA. |
| Class/ASD | Decisão do algoritmo sobre a criança pertencer ou não ao espectro autista |

Table 1: Variáveis comportamentais

4. FUNDAMENTAÇÃO TEÓRICA

Nesta seção são destacados os principais métodos que foram utilizados para a caracterização das variáveis em estudo. Destacam-se aqui o método de classificação de Árvore de Decisão, K-Vizinhos Mais Próximos (do inglês *K-Nearest Neighbors* - KNN), Perceptron de Múltiplas Camadas, Support Vector Machine (SVM) e a função de avaliação (*Feature Importance*).

4.1 Árvore de Decisão

O algoritmo de Árvore de Decisão se mostrou uma técnica bem adequada para o trabalho por apresentar na base de dados variáveis assumidamente binárias. Esta técnica pode ser definida como uma estrutura de dados que determina uma classe como um nó folha, um nó decisão que contém algum teste sobre um atributo e, a cada resultado, uma aresta para uma subárvore [17].

Este algoritmo tem algumas vantagens, como reduzir a complexidade do problema em regiões de decisão. É possível aproximar espaços de alta-dimensionalidade em vários níveis da árvore e conjunto de dados com muitas características. Permite-se ainda estimar um agrupamento menor desses em cada nó da árvore para testar com outros subconjuntos para melhorar a performance do algoritmo. Porém, com grandes bases de dados, os erros podem se acumular ao longo dos níveis da árvore de decisão e não podendo otimizar precisão e eficiência ao mesmo tempo [18].

4.2 K-Vizinhos Mais Próximos

O algoritmo dos vizinhos mais próximos é um dos modelos preditivos mais simples e com diversas aplicações [7, 10, 11]. Cada amostra de uma base de dados D é considerada a partir de um vetor X com l dimensões, onde $X = [x_1, x_2, \dots, x_l] \in R^l$. A saída é determinada de acordo com o parâmetro K , ou seja, o rótulo de maior incidência nos k vizinhos mais próximos. Em caso de empate existem várias técnicas, mas as mais usadas são escolhas aleatórias do vencedor, ou reduzir o valor de K até achar um vencedor absoluto.

4.3 Perceptron de Múltiplas Camadas

Um modelo de rede neural bastante aplicado na literatura é o Perceptron de Múltiplas Camadas, do inglês *Multi-Layer Perceptron* (MLP) [13, 24, 28]. Resume-se esse algoritmo de redes neurais como uma das várias técnicas inspiradas em um neurônio biológico, sendo este representado matematicamente por uma unidade que recebe dados de entrada x_i e somadas junto com pesos w_i que dão mais importância a certos valores de entradas do que outros. Após este efeito, realiza-se a soma de todos esses valores e subtrai-se do total um valor que filtra o contingente comum a todos, que é denominado *bias* Θ . Segue-se com o cálculo do resultado u que é passado por uma função de ativação $g(u)$ que, dependendo do resultado, ativa ou não a saída do neurônio [12], conforme pode-se visualizar na Figura 1.

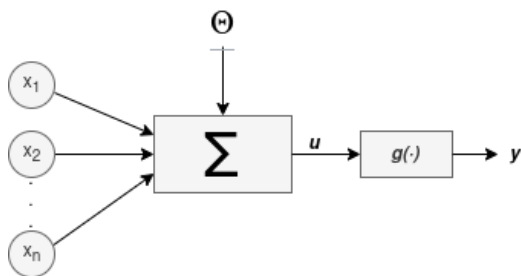


Figure 1: Estrutura matemática de um neurônio em redes neurais.

Dessa maneira, uma rede MLP utiliza as unidades matemáticas dos neurônios em determinadas camadas: a camada de entrada em que os padrões são apresentados à rede; a oculta onde é feita a maior parte do processamento através das conexões ponderadas, e que podem ser consideradas

para extração de características; e a camada de saída onde o resultado final é concluído e apresentado [28].

4.4 Support Vector Machine

Outro algoritmo de treinamento para aprendizado de máquina abordado é o Support Vector Machine (SVM). Este é utilizado para reconhecimento de padrões baseado na Teoria de Aprendizagem Estatística [19], utilizado comumente para classificação, mas também é possível utilizá-lo para regressão. Objetos de entradas são vistos como pontos pertencentes ao hiperplano R^l , em que l é a quantidade de atributos. O SVM traça um plano separador entre esses pontos dividindo-os em duas classes. Um problema é dito linearmente separável se for possível separar as classes usando um plano de dimensão $l - 1$. Isso pode ser facilmente estendido à classificação de k classes construindo k classificadores de duas classes cada. A interpretação geométrica do SVM traz uma busca da separação ideal de uma superfície. Ou seja, formar o hiperplano que seja equidistante das duas classes em questão [22].

4.5 Feature Importance

A medida *Feature Importance* trata-se do processo de busca das melhores variáveis que descrevem um determinado modelo. A Equação 1 destaca o comportamento dessa medida, com base nas técnicas de *impureza de Gini*, ou *redução média das impurezas*, que calcula a importância de cada característica de entrada no resultado do modelo [18]. Desse modo, seu valor é calculado pela soma das frequências das classes sobre o número total delas.

$$G = \sum_{j=1}^J P_j(1 - P_j) \quad (1)$$

Destaca-se na Equação 1 que G varia de (0,1) e representa o grau de importância da variável j no modelo, J é a quantidade de classes e P representa a frequência de uma categoria.

Para o caso específico da Árvore de Decisão, deve-se atentar ao cálculo de $\hat{\Gamma}(t)$. As classes são divididas em nós que são características para classificar as amostras do conjunto de dados. Assim, seja $\hat{\phi}_j(t)$ a frequência de uma classe j em um nó t , a impureza de Gini [14] aplicada ao nó t é definida na Equação 2. Logo, cada nó é avaliado a partir dessa métrica.

$$\hat{\Gamma}(t) = \sum_{j=1}^J \hat{\phi}_j(t)(1 - \hat{\phi}_j(t)) \quad (2)$$

4.6 Matriz de Confusão

A matriz de confusão permite analisar de forma detalhada o comportamento de um classificador, e para isso ela divide os resultados em quatro tipos [22]:

- **Verdadeiro Positivo (VP):** Se refere aos casos que o modelo previu corretamente a classe positiva.
- **Verdadeiro Negativo (VN):** Se refere aos casos que o modelo previu corretamente a classe negativa.
- **Falso Positivo (FP):** Se refere aos casos onde o modelo previu que era de uma classe positiva mas na verdade era da classe negativa.

Table 2: Exemplo de Matriz de Confusão
Valor Previsto

| | | Valor Previsto | |
|------------|----------|-----------------------|-----------------------|
| | | Positivo | Negativo |
| Valor Real | Positivo | Verdadeiros Positivos | Falsos Negativos |
| | Negativo | Falsos Positivos | Verdadeiros Negativos |

- **Falso Negativo (FN):** Se refere aos casos onde o modelo previu que era de uma classe negativa mas na verdade era da classe positiva.

A matriz de confusão em um problema binário, como o que está sendo abordado na base em questão (autista ou não-autista), pode ser configurado de acordo com a Tabela 2.

4.7 Métricas de Desempenho

A partir da matriz de confusão é possível utilizar métricas que podem ser assumidas para analisar o desempenho do classificador. Dentre estas métricas, tem-se a *Accuracy*, que define qual a porcentagem de acertos do classificador. Outra medida é a *Precision* que identifica quantos casos classificados como corretos estavam realmente corretos. Há ainda a *Recall*, que avalia com que frequência o classificador aborda os exemplos de uma classe. E por último, destaca-se a *F1-Score* (F1) que combina Precisão e Revocação em uma média harmônica e indica a qualidade geral do modelo. As definições matemáticas de *Accuracy*, *Precision*, *Recall* e *F1-Score* estão dispostas nas Equações 3 a 6, respectivamente.

$$Accuracy = \frac{VP + FP}{VP + VN + FP + FN} \quad (3)$$

$$Precision = \frac{VP}{VP + FP} \quad (4)$$

$$Recall = \frac{VP}{VP + FN} \quad (5)$$

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \quad (6)$$

5. RESULTADOS E TESTES

Ao aplicar os classificadores destacados à base de dados, foi possível observar a taxa de erros, a performance de cada um e extrair as características mais influentes na decisão do modelo. O trabalho avalia ao final qual a melhor classificação e análise dos dados contidos na base mencionada.

5.1 Matriz de Correlação

Para a construção de um modelo consistente, faz-se necessário a seleção de variáveis descorrelacionadas entre si. Destaca-se que dadas duas variáveis análogas, há uma redundância de informações que poderá afetar o desempenho do modelo. Dessa forma, para analisar a correspondência entre as variáveis, desenvolveu-se a matriz de correlação como visto na Figura 2.

Em seguida, pode-se observar uma correlação de no máximo que 0,35 em um índice no intervalo absoluto [0, 1] das variáveis de entrada entre si. E ainda, um índice mais alto na

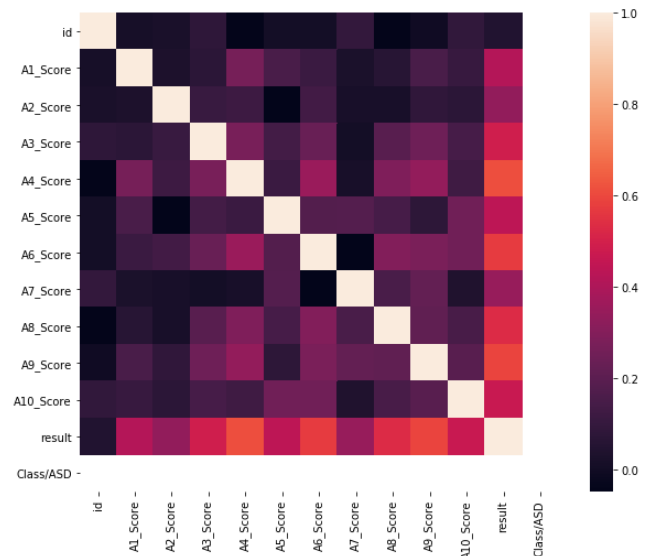


Figure 2: Matriz de Correlação

correlação entre as variáveis de entrada e a variável resposta. Isso indica que as variáveis de entrada não são correlacionadas entre si, e possuem correlação relevante com a variável resposta.

5.2 Avaliação das características

Os gráficos dessa seção apresentam a classificação das importâncias das variáveis de acordo com o algoritmo de aprendizado de máquina. Em cada um dos mesmos é mostrado também um *boxplot* por característica, em que se definem os valores máximo, mínimo, *outliers*, mediana e medidas de tendências de dispersão.

No modelo de vizinhos mais próximos (*k-NN*), o primeiro parâmetro configurado é *k* que representa o número de vizinhos a serem considerados, valor este que apresentou melhor desempenho foi o valor 1 (um). A análise das características mais relevantes com esse modelo está disposto na Figura 3. Destacam-se as características A1, A9, A4, A10 e A7 com importância maior que 0,04. As duas primeiras características, apesar de apresentarem alguns *outliers*, dispõem de maior relevância dentre as demais de acordo com suas variabilidades. A4 possui valores superiores a A10 e A7 (estas praticamente empatadas) mas com alta variabilidade em suas ocorrências.

No modelo de rede neural *MLP* foi utilizado apenas uma camada oculta com 20 neurônios e 150 iterações. Além disso, aplicou-se o otimizador de pesos Broyden-Fletcher-Goldfarb-Shanno (BFGS) que é um método baseado em derivadas que realiza as atualizações usando as informações do gradiente do modelo [16]. Por fim, definimos em 0,1 o valor do termo de regularização ou penalidade dos pesos para prevenir *overfitting*, penalizando pesos com grandes valores. O gráfico sobre a importância das características pode ser visualizado na Figura 4. Evidenciam-se variabilidades quase que em mesma escala para a maioria das características, mas informações das medianas de A1, A10 e A4, com importância maior que 0,08 apesar da presença de *outliers*.

Com o modelo SVM foi adotado *kernel* linear que atua como o produto interno das entradas com uma função não

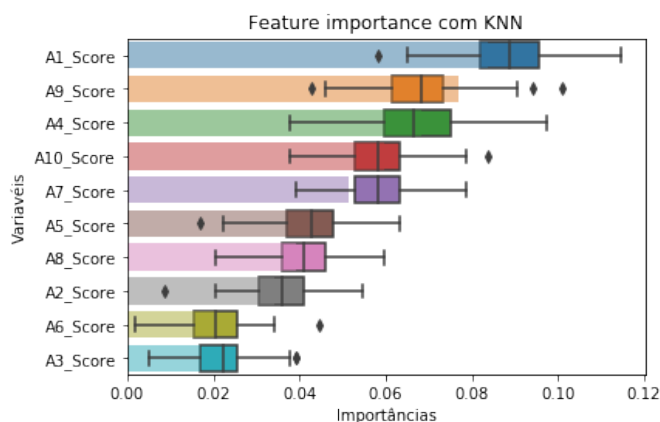


Figure 3: Importância das variáveis no modelo KNN

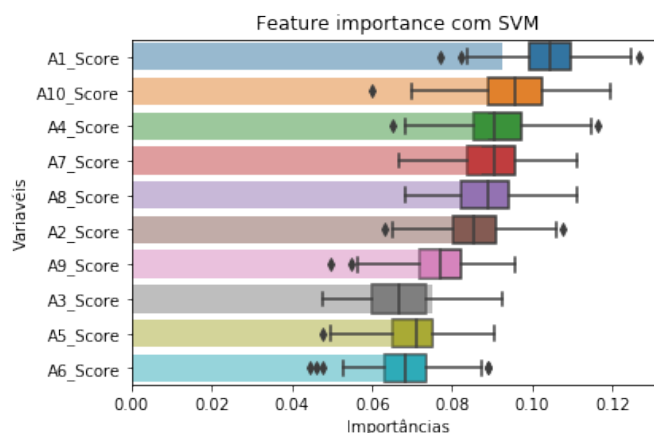


Figure 5: Importância das variáveis no modelo SVM

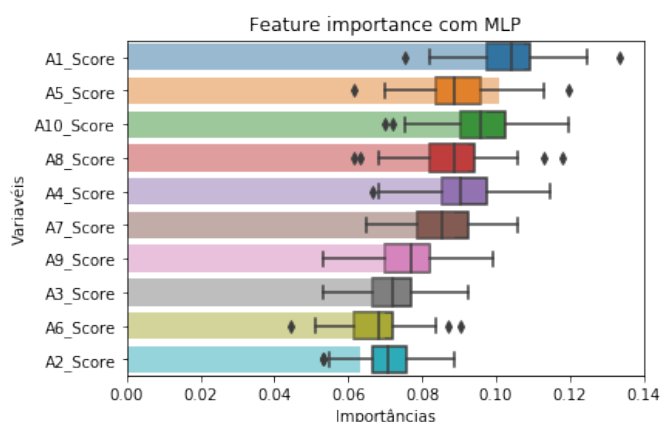


Figure 4: Importância das variáveis no modelo MLP

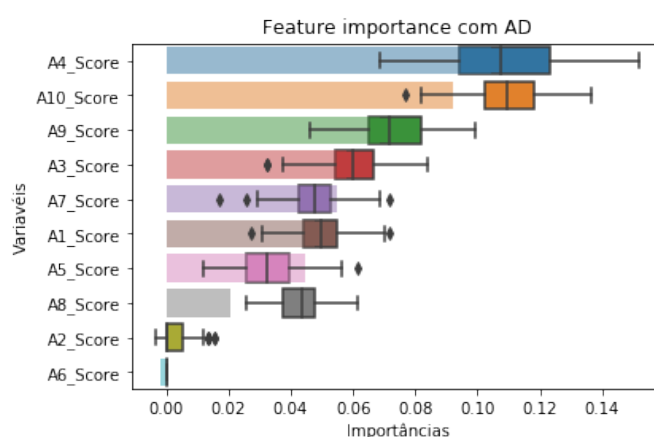


Figure 6: Importância das variáveis no modelo de árvore de decisão

linear [19]. A análise das características mais relevantes com esse modelo está disposto na Figura 5. Destacam-se as características A1, A10, A4 e A7 com importância maior que 0,09. Mesmo que as duas primeiras características apresentem alguns *outliers*, estas se destacam com maior relevância dentre as demais de acordo com suas variabilidades. Para esse modelo, todas as variáveis obtiveram relevância com mediana acima de 0,06.

Para o método de AD assumiu-se alguns parâmetros nos experimentos. Primeiramente definiu-se um valor inteiro igual a 100 (cem) que representa justamente a profundidade máxima da árvore em questão. O segundo parâmetro envolve a estratégia usada para escolher a divisão em cada nó, que foi assumida como aleatória. Por fim, aplicou-se um valor inteiro igual a 2 (dois) para o número mínimo de amostras necessárias em um nó folha. Assim, um ponto de divisão em qualquer profundidade terá esse valor mínimo em cada um dos ramos. A Figura 6 é exibida com a análise das características mais relevantes para a Árvore de Decisão. As características A4, A10, A9, A3, A1 e A7 com importância maior que 0,05. As duas primeiras características apresentam maior relevância dentre as demais de acordo com suas variabilidades.

Levando em consideração as Figuras 3 a 6, percebe-se que as variáveis A1, A10 e A4 alcançaram maior destaque na

maioria dos experimentos apresentados nos quatro modelos. Os algoritmos *SVM* e *MLP* expressaram um perfil mais parecido entre eles no comportamento das importâncias dentre as características. Todas com um grau de importância maior que nos demais métodos, em termos gerais. As variáveis nos modelos *KNN* e *AD* que trouxeram maior variância de relevância entre as mesmas. As diferenças entre as importâncias das características tiveram comportamento mais similar que entre as demais abordagens. Estes dois últimos métodos possuem menor demanda de aproximação dos dados em decorrência dos ajustes de seus parâmetros.

5.3 Curva de Aprendizagem

A Curva de Aprendizagem é um gráfico que dispõe o desempenho do modelo de Aprendizagem de Máquina ao longo dos ciclos de experimentos. A partir dos gráficos produzidos é possível diagnosticar problemas com o aprendizado, além de identificar se os dados de treinamento e validação são adequadamente representativos [10, 12, 29].

Na Figura 7 pode-se acompanhar o comportamento da curva de aprendizagem do modelo de Árvore de Decisão. Nota-se que quando o tamanho da base de dados de treino aumenta o desempenho nos dados de treino e de teste vai convergindo em um ponto de alto desempenho.

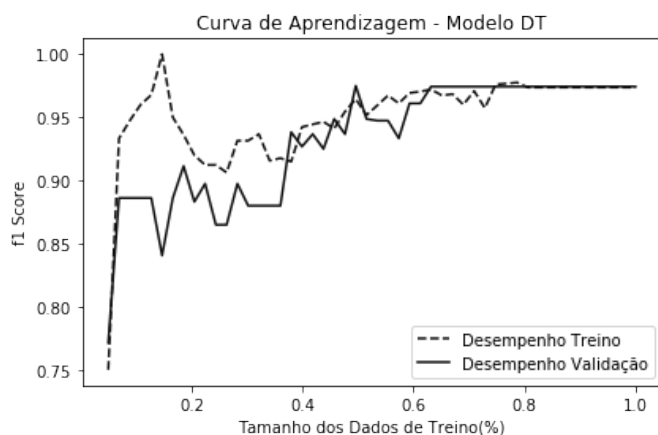


Figure 7: Curva de Aprendizagem do modelo de Árvore de Decisão

Table 3: Matriz de Confusão

| | | Rótulo Previsto | |
|-------------|-------------|-----------------|-------------|
| | | Autismo | Não-Autismo |
| Rótulo Real | Autismo | 26 | 1 |
| | Não-Autismo | 2 | 27 |

Com este mesmo gráfico é possível notar que o modelo proposto não foi afetado com *overfitting* ou *underfitting* [12, 29]. Ademais, destaca-se que o desempenho do modelo nos dados de validação é satisfatório. Esse comportamento aponta que o modelo está se ajustando corretamente aos dados. A mesma percepção foi atingida nos experimentos dos outros modelos analisados.

5.4 Modelo e Análise dos dados

Ao submeter a base de dados ao algoritmo de Árvore de Decisão na intenção de prever a variável *Class/ASD*, obtém-se a descrição do modelo de dados como definido na Figura 8. O modelo ilustrado relaciona a classificação obtida pela variável *result* distribuída pela frequência de casos. Essa relação vem a indicar que os indivíduos que atendem a 7 (sete) ou mais características esperadas da TEA necessariamente foram clinicamente diagnosticados com o transtorno.

5.5 Avaliação da Performance do Modelo

Para evitar que o modelo crie um viés às condições iniciais de implementação, utilizou-se uma validação cruzada do tipo *k-fold* com $k = 5$. Além deste tratamento, é necessário avaliar a quantidade de falso-positivos e falso-negativos em uma amostra não observada no treinamento do modelo. Para tal utilizou-se a Matriz de Confusão e levou-se em consideração uma amostragem de 20% dos indivíduos separados para teste do modelo. Dessa maneira, repetiram-se os experimentos k vezes de forma com que toda a população possa ser testada. Os valores alcançados estão na Tabela 3 e a média de falso-positivos foi de 7,692% enquanto a média de falso-negativos obtida foi de 3,704%.

6. CONCLUSÃO

Os algoritmos utilizados atingiram um bom resultado nesse problema, chegando a superar valores de 95% na medida F1

Score relacionada ao desempenho dos modelos. Os resultados obtidos foram influenciados por basicamente as mesmas variáveis. Assim reforça que mesmo com a natureza distinta dos algoritmos, houve uma mesma tendência de comportamento dos classificadores.

Portanto, a metodologia proposta torna possível medir que as variáveis mais discriminantes e que apontam ao diagnóstico de TEA consistem em: facilidade de fazer múltiplas tarefas simultaneamente (A4), na dificuldade de fazer novas amizades (A10) e na percepção de baixos ruídos (A1), sendo relacionadas pela ausência e presença destes traços comportamentais, respectivamente. Ainda é possível concluir que o indivíduo que apresenta correspondência em sete ou mais dos dez traços comportamentais propostos pelo questionário possui fortes indícios de possuir o transtorno em algum espectro do autismo.

Futuramente este trabalho se propõe a aplicar modelos mais robustos na bases de dados, como algoritmos com otimização de gradiente e Redes Neurais Artificiais Profundas. Além disso, pretende-se construir e coletar de uma base de dados junto a especialistas para analisar e desenvolver novos modelos computacionais com maior eficiência neste auxílio ao diagnóstico de TEA.

Referências

- [1] H. Abbas, F. Garberson, E. Glover, and D. P. Wall. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *Journal of the American Medical Informatics Association*, 25(8):1000–1007, August 2018.
- [2] T. Akter, M. S. Satu, M. M. I. Khan, M. H. Ali, S. Uddin, P. Lió, J. M. W. Quinn, and M. A. Moni. Machine learning-based models for early stage detection of autism spectrum disorders. *IEEE Access*, 7:166509–166527, 2019.
- [3] A. Albuquerque, T. Amador, R. Ferreira, A. Veloso, and N. Ziviani. Learning to rank with deep autoencoder features. In *International Joint Conference on Neural Networks (IJCNN 2018)*, pages 1–8, July 2018.
- [4] A. P. A. APA. *DSM-5 - Manual Diagnóstico e Estatístico de Transtornos Mentais*. Artmed, 2014.
- [5] A. Ben-Sasson, D. L. Robins, and E. Yom-Tov. Risk assessment for parents who suspect their child has autism spectrum disorder: Machine learning approach. *Journal of Medical Internet Research*, 20(4):1–10, April 2018.
- [6] R. X. de Brito, C. A. R. Fernandes, and M. A. B. Amora. Análise de desempenho com redes neurais artificiais, arquiteturas MLP e RBF para um problema de classificação de crianças com autismo. *iSys-Brazilian Journal of Information Systems*, 12(1), 2019.
- [7] S. R. Dutta, S. Giri, S. Datta, and M. Roy. A machine learning-based method for autism diagnosis assistance in children. In *International Conference on Information Technology (ICIT 2017)*, pages 36–41, December 2017.
- [8] M. Elbattah, R. Carette, G. Dequen, J.-L. Guérin, and F. Cilia. Learning clusters in autism spectrum disorder: Image-based clustering of eye-tracking scanpaths with deep autoencoder. In *41st Annual International*

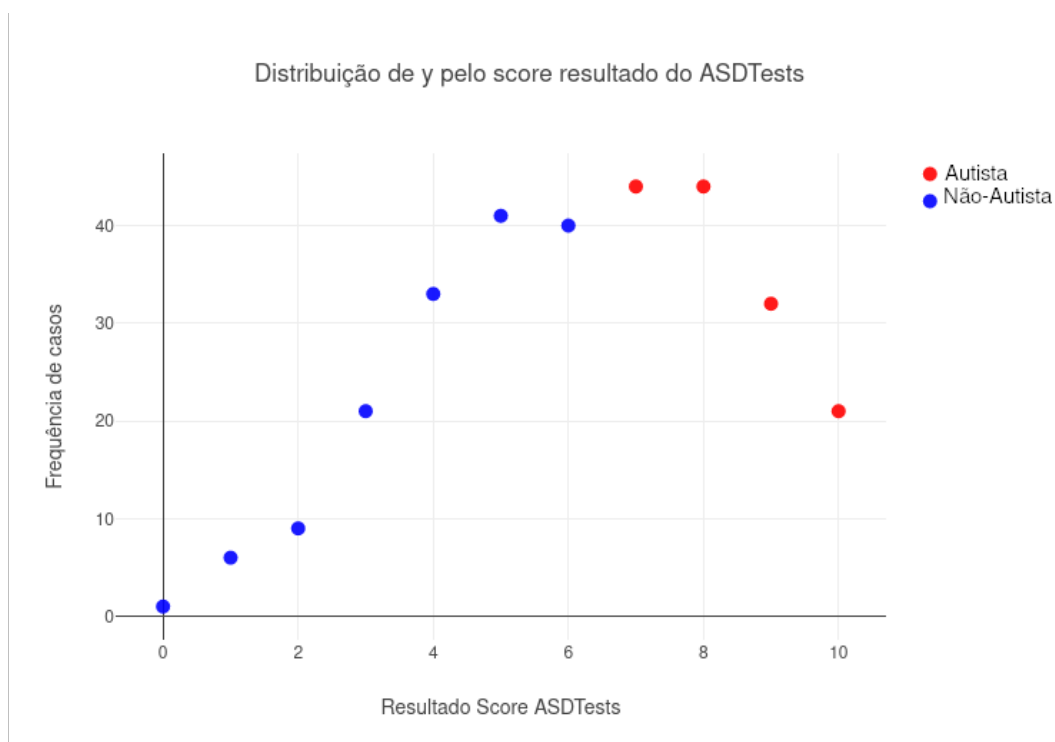


Figure 8: Modelo de análise dos dados

Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 1417–1420, July 2019.

- [9] M. Gök. A novel machine learning model to predict autism spectrum disorders risk gene. *Neural Computing and Applications*, 31(10):6711–6717, October 2019.
- [10] J. Grus. *Data Science do Zero*. Alta Books, 1th edition, 2016.
- [11] A. S. Halibas, L. B. Reazol, E. G. T. Delvo, and J. C. Tibudan. Performance analysis of machine learning classifiers for asd screening. In *2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pages 1–6, November 2018.
- [12] S. Haykin. *Redes Neurais: Princípios e prática*. Bookman, Porto Alegre, 2 edition, 2001.
- [13] K. K. Hyde, M. N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D. R. Dixon, and E. Linstead. Applications of supervised machine learning in autism spectrum disorder research: a review. *Review Journal of Autism and Developmental Disorders*, 6(2):128–146, June 2019.
- [14] H. Ishwaran. The effect of splitting on random forests. *Machine Learning*, 99(1):75–118, 2015.
- [15] W. Liu, M. Li, and L. Yi. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, 9(8):888–898, 2016.
- [16] O. M. Maciel. Algoritmos quase-newton para otimização multiobjetivo. Master’s thesis, Universidade Federal do Amazonas, Manaus-AM, 2016.
- [17] S. O. Rezende, M. C. Monard, and A. C. P. d. L. Carvalho. Sistemas inteligentes para engenharia: pesquisa e desenvolvimento. *Anais III Workshop de Sistemas Inteligentes para Engenharia*, 1999.
- [18] S. R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674, 1991.
- [19] R. Samsudin, P. Saad, and A. Shabri. A hybrid least squares support vector machines and gmdh approach for river flow forecasting. *Hydrology and Earth System Sciences Discussions*, 7(3):3691–3731, 2010.
- [20] S. Sartipi, M. G. Shayesteh, and H. Kalbkhani. Diagnosing of autism spectrum disorder based on garch variance series for rs-fmri data. In *9th International Symposium on Telecommunications (IST)*, pages 86–90, December 2018.
- [21] A. Speaks. Autism speaks. <https://www.autismspeaks.org/>, 2019. Acesso em: 05/04/2019.
- [22] H. Talabani and E. Avci. Performance comparison of svm kernel types on child autism disease database. *International Conference on Artificial Intelligence and Data Processing (IDAP 2018)*, pages 1–5, 2018.
- [23] M. C. T. V. Teixeira, T. P. Mecca, R. d. L. Velloso, R. B. Bravo, S. H. B. Ribeiro, M. T. Mercadante, and C. S. d. Paula. Literatura científica brasileira sobre

- transtornos do espectro autista. *Revista da Associação Médica Brasileira*, 56:607–614, 2010 2010.
- [24] F. A. Thabtah, F. Kamalov, and K. Rajab. A new computational intelligence approach to detect autistic features for autism screening. *International Journal of Medical Informatics*, 117:112–124, 2018.
- [25] F. F. Thabtah. Autistic Spectrum Disorder Screening Data for Children Data Set - UC Irvine Machine Learning Repository. <https://bit.ly/36DHgxC>, 2017. Acesso em: 08/07/2019.
- [26] M. B. Usta, K. Karabekiroglu, B. Sahin, M. Aydin, A. Bozkurt, T. Karaosman, A. Aral, C. Cobanoglu, A. D. Kurt, N. Kesim, I. Sahin, and E. Ürer. Use of machine learning methods in prediction of short-term outcome in autism spectrum disorders. *Psychiatry and Clinical Psychopharmacology*, 29(3):320–325, 2019.
- [27] B. Yuan and X. Ma. Sampling + reweighting: Boosting the performance of adaboost on imbalanced datasets. *International Joint Conference on Neural Networks (IJCNN 2012)*, pages 1–6, 2012.
- [28] A. H. Zaji, H. Bonakdari, H. Z. Khameneh, and S. R. Khodashenas. Application of optimized artificial and radial basis neural networks by using modified genetic algorithm on discharge coefficient prediction of modified labyrinth side weir with two and four cycles. *Measurement*, page 107291, 2019.
- [29] H. Zhang, L. Zhang, and Y. Jiang. Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. IEEE, 2019.