



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA



MARIA EUCLÉCIA ALBUQUERQUE DA SILVA

**UM SISTEMA INTELIGENTE BASEADO EM ELETROENCEFALOGRAFIA PARA
APOIO AO DIAGNÓSTICO DO TRANSTORNO DO ESPECTRO AUTISTA**

Recife
2025

MARIA EUCLÉCIA ALBUQUERQUE DA SILVA

**UM SISTEMA INTELIGENTE BASEADO EM ELETROENCEFALOGRAFIA PARA
APOIO AO DIAGNÓSTICO DO TRANSTORNO DO ESPECTRO AUTISTA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de mestra em Engenharia Biomédica.

Área de concentração: Computação Biomédica.

Orientador: Prof. Dr. Wellington Pinheiro dos Santos

Recife
2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Silva, Maria Euclécia Albuquerque da.
Um sistema inteligente baseado em eletroencefalografia para apoio
ao diagnóstico do transtorno do espectro autista / Maria Euclécia
Albuquerque da Silva. - Recife, 2025.
50f.: il.

Dissertação (Mestrado) - Universidade Federal de Pernambuco,
Centro de Tecnologia e Geociências, Programa de Pós-Graduação em
Engenharia Biomédica, 2025.

Orientação: Wellington Pinheiro dos Santos.

Inclui referências.

1. Transtorno do espectro autista; 2. Aprendizado de máquina; 3.
Eletroencefalografia. I. Santos, Wellington Pinheiro dos. II.
Título.

UFPE-Biblioteca Central

MARIA EUCLÉCIA ALBUQUERQUE DA SILVA

**UM SISTEMA INTELIGENTE BASEADO EM ELETROENCEFALOGRAFIA PARA
APOIO AO DIAGNÓSTICO DO TRANSTORNO DO ESPECTRO AUTISTA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de mestra em Engenharia Biomédica. Área de concentração: Computação Biomédica.

Aprovado em:

BANCA EXAMINADORA

Prof. Dr. Wellington Pinheiro dos Santos (Orientador)

Universidade Federal de Pernambuco

Prof^o. Dr. Thiago Vasconcellos Modenesi (Examinador Interno)

Universidade Federal de Pernambuco

Prof^a. Dr^a. Magda Rhayanny Assunção Ferreira (Examinadora Externa)

Universidade Federal de Pernambuco

Recife
2025

Dedico este trabalho primeiramente a Deus, por ser minha força e guia em todos os momentos, e, em segundo lugar, à minha mãe e à minha família, cujo amor e apoio foram essenciais para a realização desta conquista.

AGRADECIMENTOS

Meus sinceros agradecimentos a Deus, primeiramente e a minha família pelo apoio em mais uma etapa.

Gostaria de agradecer também a FACEPE, pelo apoio e suporte financeiro para realização deste sonho.

"Ora, se algum de vós tem falta de sabedoria, peça-a a Deus, que a todos dá liberalmente e nada lhes impropria; e ser-lhe-á dada."

— *Tiago 1:5*

“É fundamental diminuir a distância entre o que se diz e o que se faz, de tal forma que, num dado momento, a tua fala seja a tua prática” (*Paulo Freire, Pedagogia da Esperança, 1994, p. 36*).

RESUMO

O trabalho aborda o uso de um sistema inteligente baseado em eletroencefalografia (EEG) para apoiar o diagnóstico do Transtorno do Espectro Autista (TEA). O TEA apresenta desafios no diagnóstico devido à variabilidade de sintomas e à ausência de biomarcadores objetivos, sendo o EEG uma alternativa promissora por identificar padrões de atividade cerebral associados ao transtorno. A metodologia incluiu a análise de um conjunto de dados com 56 indivíduos (28 com TEA e 28 controles neurotípicos), utilizando técnicas de aprendizado de máquina, como Redes Bayesianas, Naive Bayes e SVMs. Os sinais de EEG foram pré-processados e 34 atributos foram extraídos para análise. Os dados foram obtidos de base pública vinculada à Universidade de Sheffield, descrita por Dickinson et al. (2022). A análise contou com validação cruzada e repetição de experimentos para garantir robustez estatística. Os melhores resultados foram alcançados com o modelo SVM com kernel RBF (gama = 0,5), atingindo 98,22% de acurácia com 9 eletrodos. O estudo destaca o potencial do uso combinado de EEG e aprendizado de máquina para melhorar a precisão e a rapidez do diagnóstico clínico do TEA. Como desdobramentos, a pesquisa sugere o desenvolvimento de modelos mais generalizáveis, a ampliação da base de dados e a aplicação de sistemas inteligentes na prática clínica com vistas a diagnósticos precoces e intervenções personalizadas.

Palavras-chave: Eletroencefalografia. Transtorno do Espectro Autista. Aprendizado de Máquina. Redes Bayesianas. SVM.

ABSTRACT

This study addresses the use of an intelligent system based on electroencephalography (EEG) to support the diagnosis of Autism Spectrum Disorder (ASD). ASD presents diagnostic challenges due to symptom variability and the lack of objective biomarkers, with EEG emerging as a promising alternative by identifying brain activity patterns associated with the disorder. The methodology involved analyzing a dataset with 56 individuals (28 with ASD and 28 neurotypical controls), using machine learning techniques such as Bayesian Networks, Naive Bayes, and SVMs. EEG signals were preprocessed, and 34 features were extracted for analysis. The data was obtained from a public dataset linked to the University of Sheffield, as described by Dickinson et al. (2022). The analysis included cross-validation and multiple experiment repetitions to ensure statistical robustness. The best results were achieved with the SVM model using RBF kernel ($\gamma = 0.5$), reaching 98.22% accuracy with 9 electrodes. The study highlights the potential of combining EEG and machine learning to improve the accuracy and speed of ASD clinical diagnosis. As further developments, the research suggests the construction of more generalizable models, expanding the dataset, and applying intelligent systems in clinical practice to enable early diagnoses and personalized interventions.

Keywords: Electroencephalography. Autism Spectrum Disorder. Machine Learning. Bayesian Networks. SVM.

LISTA DE FIGURAS

FIGURA 1	Sinais de EEG	21
FIGURA 2	Diagrama de fluxo metodológico aplicado no presente trabalho	26
FIGURA 3	Diagrama de fluxo metodológico aplicado no presente trabalho	27
FIGURA 4	Análise dos Canais Ausentes	29
FIGURA 5	Atributos	30
FIGURA 6	Matriz de confusão do melhor modelo encontrado nas três abordagens	37
FIGURA 7	Matriz de confusão do melhor modelo encontrado nas três abordagens	39

LISTA DE TABELAS

Tabela 1	Resultados do treinamento para classificação de EEG para o cenário com 9 eletrodos.	36
Tabela 2	Métricas com valores correspondentes a fase teste de 9 eletrodos.	37
Tabela 3	Resultados do treinamento para classificação de EEG para o cenário com 15 eletrodos.	38
Tabela 4	Resultados da fase de testes dos melhores modelos	38

LISTA DE GRÁFICOS

Gráfico 1	Boxplot do Índice Kappa.	40
Gráfico 2	Boxplot da Acurácia	40
Gráfico 3	Boxplot da Especificidade	41
Gráfico 4	Boxplot da Sensibilidade	41
Gráfico 5	Boxplot da Área sob a curva ROC	42

LISTA DE ABREVIATURAS

APA	American Psychiatric Association
BN	Bayesian Networks
CPTs	Tabelas de Probabilidade Condicional
DSMV	Manual Diagnóstico e Estatísticos de Transtornos Mentais
ECA	Estatuto da Criança e do Adolescente
OMS	Organização Mundial da Saúde
PSO	Enxame de Partículas
SUS	Sistema Único de Saúde
SVMs	Support Vector Machines
TAE	Teoria de Aprendizagem Estatística
TEA	Transtorno do Espectro do Autismo

SUMÁRIO

1	INTRODUÇÃO	14
1.1	MOTIVAÇÃO E JUSTIFICATIVAS	14
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	TRANSTORNO DO ESPECTRO AUTISTA.....	17
2.2	ELETROENCEFALOGRAFIA	19
2.3	APRENDIZADO DE MÁQUINA.....	22
2.3.1	Redes bayesianas	23
2.3.2	Máquinas de vetor de suporte.....	23
3	METODOLOGIA	25
3.1	CONJUNTO DE DADOS EXISTENTES	27
3.1.1	Pré-processamento de dados EEG.....	27
3.1.2	Extração e seleção de atributos.....	29
3.1.3	Modelos de avaliação.....	31
3.1.4	Métricas de avaliação	32
3.1.5	Classificação e análise de resultados.....	33
4	RESULTADOS E DISCUSSÃO	35
5	CONCLUSÕES	44
	REFERÊNCIAS	47

1 INTRODUÇÃO

1.1 MOTIVAÇÃO E JUSTIFICATIVAS

O Transtorno do Espectro Autista (TEA) é uma condição neurológica que afeta o neurodesenvolvimento de indivíduos atípicos, com déficits na comunicação social e na interação social em diversos contextos (Manual Diagnóstico e Estatísticos de Transtornos Mentais - DSM V, 2014). De acordo com a *American Psychiatric Association* (2014), a preponderância dos diagnósticos pertencentes à categoria do TEA é eminentemente maior no sexo masculino, sendo cerca de quatro vezes mais prevalente em relação ao feminino, o que veio a ser considerado um dos achados mais consistentes na literatura, representando aproximadamente 70% dos diagnósticos.

Em amostras clínicas, observa-se que meninas apresentam maior propensão a manifestar deficiência intelectual concomitante. Isso sugere que casos sem comprometimento intelectual ou atrasos na linguagem podem não ser identificados, possivelmente devido à manifestação mais sutil das dificuldades sociais e de comunicação no sexo feminino. Ademais, a etiologia do TEA permanece incerta, embora exista consenso sobre a influência de fatores genéticos e ambientais.

Nesse contexto, os estudos analisam as diferenças entre homens e mulheres com TEA, com foco em aspectos relacionados ao diagnóstico, à imagiologia cerebral, aos principais comportamentos característicos do transtorno e às peculiaridades associadas ao sexo (Napolitano *et al.* 2022). Segundo Al-dewik *et al.* (2020) o TEA é uma das condições de neurodesenvolvimento mais prevalentes e desafiadoras no tratamento infantil, com uma taxa global superior a 1%. Em 2018, o CDC, citado pelo *Autism Speaks*, estimou que a prevalência nos Estados Unidos era de 1 em cada 59 crianças, o que motivou governos, profissionais de saúde e escolas a desenvolverem políticas e programas específicos para lidar com o transtorno.

As manifestações do TEA incluem déficits na comunicação e interação social (Critério A), além de padrões restritos e repetitivos de comportamento, interesses ou atividades (Critério B). Esses sintomas estão presentes desde a infância e impactam ou prejudicam significativamente o funcionamento cotidiano (Critérios C e D). O período em que ocorre o agravo funcional torna-se evidente, variando de acordo com as particularidades de cada indivíduo e do ambiente em que está inserido.

As características centrais do TEA são identificadas durante o desenvolvimento, mas intervenções, adaptações e suportes podem, em alguns contextos, mascarar as dificuldades. As

manifestações variam amplamente conforme a gravidade do transtorno, o nível de desenvolvimento e a idade, o que explica o uso do termo "espectro" (*American Psychiatric Association*, 2014). Quanto ao diagnóstico, o TEA continua sendo um desafio clínico, uma vez que envolve critérios amplos e subjetivos. A alta prevalência do TEA, que afeta cerca de 1% da população mundial, e sua manifestação heterogênea justificam a busca por ferramentas mais objetivas e acessíveis para auxiliar na identificação precoce e precisa da condição (*Centers for Disease Control and Prevention*, 2024).

Atualmente o diagnóstico do TEA é realizado através de instrumentos de rastreios observacionais que visam medir as capacidades sociais e cognitivas da criança. Os principais instrumentos de diagnósticos são o *DSM-V* e o *M-CHAT*, que investigam os déficits persistentes na interação e na comunicação social e examinam as respostas a itens "sim/não" que envolvem irregularidade no desenvolvimento para formular uma diagnose (Grabruker, 2021). De acordo com Hus e Segal (2021), o diagnóstico do TEA é desafiador, pois frequentemente é confundido com outras comorbidades médicas e disfunções do neurodesenvolvimento. As manifestações do TEA apresentam uma variação significativa em termos de desempenho e gravidade, podendo incluir a perda atípica de habilidades precoces e, em alguns casos, o desaparecimento do diagnóstico em determinadas crianças ao longo do tempo.

De acordo com o *American Psychiatric Association* (2014), às dificuldades associadas ao TEA estão concentradas em dois domínios principais: déficits persistentes na comunicação e interação social, e estilos limitados e frequentes de conduta, preferências ou ações. No primeiro domínio, destacam-se desafios na reciprocidade social, interpretação de gestos e expressões não verbais, e na formação e manutenção de relações sociais. No segundo, incluem-se comportamentos estereotipados, insistência em rotinas rígidas, interesses restritos ou fixações intensas, além de respostas atípicas a estímulos sensoriais, como hipersensibilidade ou hipossensibilidade. Essas características devem ser evidentes desde a infância e causar prejuízos significativos nas áreas social, educacional ou ocupacional, embora possam variar em intensidade e se manifestar de forma mais clara à medida que as demandas sociais aumentam (DSM-5, APA, 2014).

A utilização do EEG para diagnosticar o TEA destaca a capacidade de oferecer prognósticos precisos, atuando como um suporte fundamental ao diagnóstico. Nesse contexto, ressalta-se o potencial de transformar um processo tradicionalmente complexo e subjetivo em uma abordagem mais objetiva, ágil e confiável. Segundo Li et al., (2024) o diagnóstico do TEA enfrenta desafios relacionados à subjetividade e demora dos métodos tradicionais, baseados em observações clínicas. Tecnologias de aprendizado de máquina e aprendizado profundo surgem

como alternativas promissoras, proporcionando maior objetividade, eficiência e precisão ao processo diagnóstico.

A eletroencefalografia (EEG), como método de monitorização eletrofisiológica, destaca-se por registrar alterações na atividade elétrica do cérebro, permitindo identificar padrões anormais de atividade neuronal síncrona em crianças com TEA. Diante desse cenário, a presente pesquisa se orientou pela seguinte pergunta: É possível desenvolver e validar um modelo computacional baseado em aprendizado de máquina, utilizando sinais de EEG, que seja capaz de apoiar o diagnóstico do TEA com maior precisão e aplicabilidade clínica?

A utilização de tecnologias, como o EEG combinado a métodos de aprendizado de máquina, apresenta-se como uma abordagem promissora, permitindo identificar padrões neurológicos associados ao TEA. Dessa forma, buscamos contribuir para o avanço do conhecimento e para o desenvolvimento de estratégias mais eficazes no diagnóstico e acompanhamento de indivíduos com TEA (*Centers for Disease Control and Prevention, 2024*).

A análise automatizada de sinais de EEG reduz a subjetividade, acelera o diagnóstico e melhora a detecção de padrões sutis associados ao TEA, contribuindo para intervenções precoces e melhores desfechos terapêuticos. Essa abordagem não apenas fortalece a pesquisa sobre o transtorno, mas também amplia a acessibilidade e equidade no diagnóstico clínico, beneficiando crianças e suas famílias. Sendo assim, este trabalho desenvolveu e validou um modelo baseado em técnicas avançadas de aprendizado de máquina para analisar dados de EEG, com o intuito de apoiar o diagnóstico e o monitoramento de condições do TEA, garantindo maior precisão e aplicabilidade clínica.

Para isso, foi feita uma revisão da literatura sobre o uso de EEG e aprendizado de máquina no diagnóstico do TEA, com ênfase em abordagens baseadas em redes bayesianas; analisou uma base de dados a fim de identificar e selecionar padrões relevantes ao contexto do transtorno; aplicou modelos de aprendizado de máquina fundamentados na teoria de Bayes para a classificação de sinais cerebrais; avaliou o desempenho do modelo utilizando métricas como sensibilidade, especificidade e acurácia; comparou os resultados com abordagens existentes, discutindo seus diferenciais e potencial de aplicação na neurociência; e, por fim, apresentou insights e recomendações para futuras investigações e práticas clínicas que utilizem sinais de EEG no apoio ao diagnóstico do TEA.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 TRANSTORNO DO ESPECTRO AUTISTA

De acordo com o Manual de Diagnóstico e Estatístico de Transtornos Mentais - DSM V, consta que o TEA engloba em Transtornos do Neurodesenvolvimento criando assim um tópico favorecido a esse espectro. Os déficits caracterizam-se pelas limitações específicas na aprendizagem ou no controle de comportamento, pensamento e emoções. Além dos déficits na comunicação social, o diagnóstico do TEA exige a manifestação de padrões repetitivos e restritivos de comportamento. Atualmente o diagnóstico do Transtorno do Espectro Autista abrange diversos fatores, dentre eles o mais importante observações na criança, que se fundamenta na parte sensorial e aplicações de setores que correspondem a um nível que se inicia na infância à fase adulta.

A principal problematização no diagnóstico do TEA, seria a heterogeneidade dos sintomas, falta de biomarcadores, diagnóstico precoce e sistema de saúde e educação. Nesse contexto podemos aplicar o uso do EEG, para obtermos a informação precisa do apoio ao diagnóstico, com um intuito de apresentar os sinais importantes que elevam o conhecimento de profissionais adequados no conjunto em que se resume o diagnóstico e o cuidado familiar. Os setores apresentados são o apoio e suporte em que a criança e o adulto devem receber independente de sua fase e nível. O nível de suporte do Transtorno do Espectro Autista é escalado em três níveis consecutivos de acordo com a necessidade de cada espectro, as definições surgem com o decorrer de cada paciente, estabelecendo limites e critérios. Os níveis de gravidade podem ser definidos como: “Exigindo Apoio, Exigindo Apoio Substancial e Exigindo Apoio Muito Substancial” (Narrow *et al.* 2014).

Assim podemos classificar como Nível 1, Nível 2 e Nível 3. Nesse contexto pode-se estabelecer conexões entre os três níveis de suporte, definindo os principais pontos relacionados a cada um dos três níveis. O nível 1 é estabelecido por critérios na falta de apoio onde causam danos notáveis na comunicação social, nesse suporte há inflexibilidade na atividade em um ou mais contexto, referindo-se a comportamentos restritivos e repetitivos. O nível 2 é determinado por escassez grave na habilidade de comunicação social verbal e não verbal e danos sociais mesmo na presença de apoio, nisso pode-se relacionar um comportamento restritivo e repetitivo na dificuldade em lidar com a mudança. Com o nível 3 ocorrem danos graves na comunicação verbal e não verbal causando prejuízo no desempenho, limitações em dar início a interações sociais e aberturas sociais que partem de outros. Comportamentos restritivos e recorrentes

afetam o funcionamento na dificuldade em mudanças, foco ou ações que refletem o rigor do comportamento.

De acordo com a Organização Mundial da Saúde (OMS, 2023) o autismo afeta 1 em cada 100 crianças no mundo, com início na infância e persistência durante a adolescência e vida adulta, que vivem com alguma forma do TEA, que pode se apresentar de forma grave ou branda, e afeta principalmente a comunicação e a socialização.

No dia 26 de abril de 2017 é sancionada no Brasil a Lei nº 13.438, que altera a Lei nº 8.069, 13 de julho de 1990 - Estatuto da Criança e do Adolescente (ECA), tornando obrigatória a adoção pelo Sistema Único de Saúde (SUS) de protocolo que estabeleça padrões para a avaliação de riscos para o desenvolvimento psíquico das crianças. Nesta lei, fica decretada a obrigatoriedade de aplicação, a todas as crianças nos seus primeiros dezoito meses de vida, de protocolo ou outro instrumento criado com a finalidade de facilitar a detecção, em consulta pediátrica de acompanhamento da criança, de sinais que apontam para alterações voltadas ao atraso do seu neurodesenvolvimento ou que indiquem prejuízos psíquicos significativos, para possíveis encaminhamentos e intervenção precoce.

De acordo com T. Hirota e B. King (2023) o Transtorno do Espectro Autista abrange um total de 1,1% em 2008 para 2,3% em 2018 nos Estados Unidos, com essas estatísticas pode-se estabelecer um aumento no diagnóstico. É possível caracterizar esse percentual de forma branda relacionada às mudanças no critério de diagnóstico, melhor desempenho das ferramentas de triagem e diagnóstico e maior conscientização pública, com isso deve-se salientar a importância do prognóstico e investigação adequadas para assim suceder o apoio ao diagnóstico do TEA. Cerca de 2,3% de crianças com idade de 8 anos e 2,2% dos adultos nos EUA identificam-se com o déficit na comunicação social e pela presença de comportamentos ou interesses restritos e repetitivos, esses dados mostram a presença da faixa etária e de cada etapa do desenvolvimento infantil tendo suas particularidades estabelecidas as evidências sobre o diagnóstico e tratamento do TEA.

Com isso pode-se estabelecer relações sucintas entre a prevalência e o público alvo do diagnóstico e tratamento do TEA, com o intuito de prevalecer o apoio ao diagnóstico com base em dados e limitações na diagnose. Isso mostra a eficácia e a importância de estratégias baseadas em evidências para garantir um diagnóstico precoce e preciso, além de proporcionar intervenções adequadas e personalizadas, promovendo um impacto positivo no desenvolvimento e na qualidade de vida dos indivíduos diagnosticados e de suas famílias.

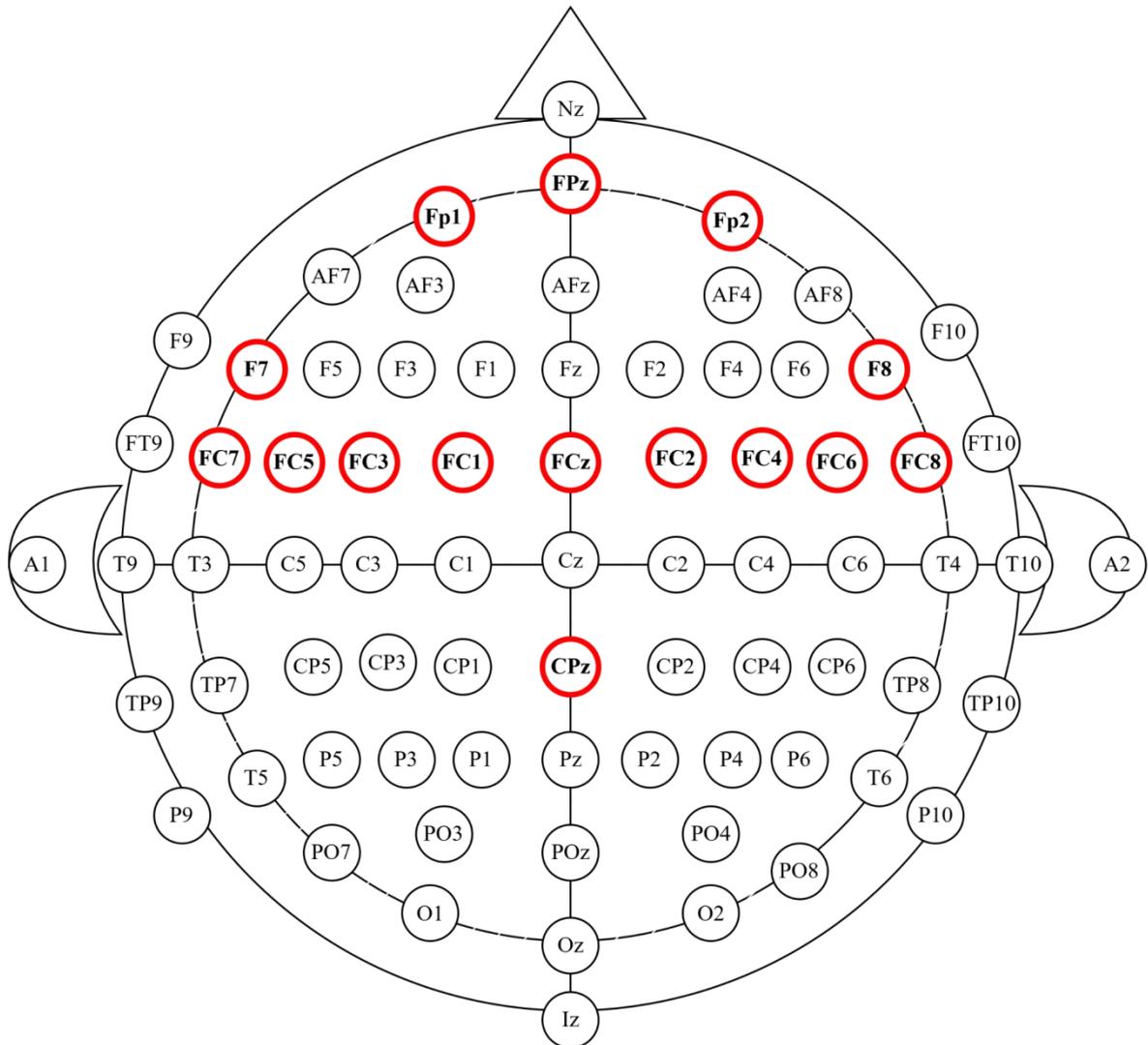
2.2 ELETROENCEFALOGRAFIA

O exame eletroencefalográfico serve para detectar algumas alterações elétricas no encéfalo, é de detecção simples e indolor. Esses sinais eletroencefalográficos que registram as diferentes áreas no cérebro são capazes de verificar as suas funções distintas associadas aos sinais e captação deles. Os sinais cerebrais são coletados utilizando a EEG onde a função é analisar a atividade elétrica do cérebro. O procedimento consiste em pôr uma touca no couro cabeludo com o intuito de verificar os sinais captados pela variação de potencial elétrico geradas pela atividade neuronal em conjunto com o gel condutivo para melhorar a qualidade do sinal e para reduzir a impedância entre o couro cabeludo e os eletrodos (Lopes da Silva; Niedermeyer, 2005).

De acordo com o Sistema Internacional 10-10, os eletrodos transferem as variações de voltagem resultantes da dinâmica elétrica dos neurônios corticais para um amplificador de sinais, permitindo a sua interpretação com maior densidade de eletrodos e precisão. Com o Sistema Internacional 10-10 e 10-20 são utilizados como amplificador de sinais, sendo a distinção entre eles o espaço geométrico dos eletrodos no couro cabeludo, tendo como referência os intervalos percentuais anatômicos no crânio. As vantagens aplicadas no uso do EEG, são de aplicação e detecção de padrões neurais específicos, avaliação da resposta sensorial e estudo do impacto de comorbidades. Esses são fatores que integram um dos principais fundamentos que pode-se levar em consideração a aplicação e o uso do EEG no apoio ao diagnóstico do TEA (Lopes da Silva; Niedermeyer, 2005).

Os sinais coletados foram tipicamente de baixa amplitude, variando entre 1 a 100 μV , e podem ser analisados em diferentes bandas de frequência (delta, teta, alfa, beta, e gama). Os eletrodos utilizados neste projeto foram de 9 e 15 são eles respectivamente os Fp1, FPz, Fp2, F7, F8, FC7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FC8 e o CPz. A sua nomenclatura tem origem em suas respectivas localizações no couro cabeludo, sendo eles utilizados na região cerebral associada, posição anatômica no crânio e referências específicas. Como significado podemos associar em: P= Parietal, O= Occipital, F= Frontal, C= Central, T= Temporal e Z= Linha média. São utilizados uma variância nos sinais para definir exatamente a captação desses eletrodos onde surgem a capacidade de interação entre cada um deles em uma respectiva média. A Figura 1, a seguir, exemplifica a distribuição de sinais de cada hemisfério do cérebro, a fim de mostrar que os eletrodos têm a mesma distância entre si.

Figura 1: Sinais de EGG



Fonte: Autoria própria (2024)

De acordo com a Figura 1, os sinais de EGG destacados, possibilitam a detecção adequada dos eletrodos utilizados nos indivíduos que possuem TEA com um intuito de apresentar a percepção dos 9 e 15 eletrodos. No hemisfério esquerdo foram utilizados os eletrodos Fp1, F7, FC7, FC5, FC3 e FC1. No hemisfério direito foram utilizados os eletrodos Fp2, F8, FC2, FC4, FC6, e FC8. Com uma ênfase no FPz, FCz e o CPz que estão posicionados nas regiões fronto-centrais (FC) e centrais (C). Esse modelo seguiu o sistema internacional 10-10. Os lobos do cérebro usufruídos, de acordo com a Figura 1, foram os lobos frontal, parietal e temporal. Cada lobo destaca-se a sua função cognoscitiva, responsável pela integração das informações sensoriais.

Os microestados EEG constituem padrões temporários e estáveis de atividade elétrica cerebral, manifestando-se em intervalos de milissegundos. Esses microestados são indicativos

de estados transitórios de funcionamento cerebral, evidenciando a organização funcional do cérebro em resposta a uma variedade de estímulos e condições. A análise desses microestados permite uma compreensão mais aprofundada da dinâmica cerebral, destacando como diferentes redes neurais interagem e se adaptam a mudanças no ambiente ou em tarefas cognitivas. Assim, os microestados EEG fornecem informações valiosas sobre os mecanismos subjacentes ao processamento cognitivo e emocional, contribuindo para a elucidação de funções cerebrais complexas. A análise de microestados cerebrais tem sido utilizada como uma ferramenta importante na pesquisa sobre dinâmica cerebral, com aplicações em distúrbios psiquiátricos e estudos sobre atenção e cognição, fornecendo insights sobre a atividade cerebral em alta resolução temporal (Koenig et al., 2002).

Segundo Milovanović e Grujičić (2021), a comorbidade entre epilepsia e TEA varia de 5% a 46%, embora os mecanismos causais, a relação temporal e a interação com deficiências ainda não sejam totalmente compreendidos. Nesse contexto, as técnicas quantitativas de EEG se destacam como ferramentas promissoras para detectar e medir disfunções em regiões cerebrais específicas, além de regular a atividade neuronal associada ao TEA. Estudos de coerência de EEG de conectividade funcional identificaram padrões de superconectividade local e subconectividade de longo alcance entre diferentes áreas cerebrais, fornecendo uma compreensão mais aprofundada das características neurofisiológicas do transtorno.

Além disso, estima-se que o sistema nervoso seja composto por mais de 100 bilhões de neurônios (Guyton & Hall, 2017). No TEA, o crescimento e as conexões dos interneurônios GABAérgicos nos córtices pré-frontal e temporal são prejudicados, afetando o equilíbrio excitatório/inibitório. Avanços na análise quantitativa de EEG e a comorbidade entre epilepsia e TEA sugerem um papel da transmissão GABAérgica aberrante, com impactos na organização e conectividade neuronal, especialmente no córtex frontal.

2.3 APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina ou “*Machine Learning*” como conhecido mundialmente é uma ferramenta para o apoio ao diagnóstico do TEA. Torna-se fundamental a existência desse termo como referência a “*artificial intelligence*”. O aprendizado de máquina faz-se o subconjunto da IA, Inteligência Artificial, em que envolve algoritmos computacionais. Esse trajeto percorre desde análise computacional a conjuntos de entradas e saídas de dados onde busca recomendações ou decisões autônomas. Com algumas repetições e modificações do algoritmo a máquina torna-se capaz de receber uma entrada e prever uma saída. As saídas são relacionadas a um conjunto de resultados conhecidos, com o objetivo de presumir a clareza do algoritmo de trabalho, que ajusta de forma frequente a habilidade de prever novos resultados (Mitchell, 1997).

O aprendizado de máquina tem sido utilizado em áreas que vão desde diagnósticos mais precisos até a personalização de tratamentos. Além de sua aplicação no TEA, pesquisas sugerem que o ML tem sido crucial no aprimoramento de exames de imagem, como a detecção precoce de câncer por meio de mamografias, e também no monitoramento contínuo de pacientes com doenças crônicas, como diabetes e hipertensão (Souza & Alves, 2022; Oliveira & Cunha, 2019). Outros estudos enfatizam o uso do aprendizado de máquina em sistemas de suporte à decisão clínica, que podem recomendar tratamentos baseados em grandes volumes de dados históricos de pacientes (Almeida & Rocha, 2021).

O potencial do aprendizado de máquina na saúde tem sido amplamente explorado, com indicações de que essa tecnologia pode revolucionar o setor nos próximos anos. Especialistas apontam que, com a capacidade de processar e analisar grandes volumes de dados em tempo real, o ML pode ajudar a prever surtos de doenças e otimizar a alocação de recursos hospitalares, promovendo uma gestão de saúde mais eficiente (Martins & Sousa, 2020). Além disso, a personalização de tratamentos, baseada no perfil genético de cada paciente, promete ser uma das maiores contribuições dessa tecnologia no futuro (Costa & Ribeiro, 2021).

2.3.1 Redes bayesianas

A rede bayesiana é uma função de como um rótulo e a instâncias são adquiridos por um conjunto de atributos. Redes Bayesianas são modelos probabilísticos que representam um conjunto de variáveis e suas relações de dependência condicional através de um grafo dirigido acíclico. Elas combinam conceitos de teoria de probabilidade com grafos e são úteis para inferência em situações de incerteza.

As redes bayesianas são modelos probabilísticos que lidam com a incerteza nos dados ao modelar relações de dependência condicional entre variáveis. Esses modelos são particularmente úteis para inferência probabilística e tomada de decisão em domínios complexos, como o diagnóstico médico, onde variáveis interdependentes influenciam mutuamente suas probabilidades. A partir da estrutura gráfica de uma rede bayesiana, é possível estimar a probabilidade de determinados eventos com base na evidência disponível, facilitando a interpretação dos dados e a tomada de decisões informadas (Pearl, 1988).

2.3.2 Máquinas de vetor de suporte

As Máquinas de Vetores de Suporte (*Support Vector Machines, SVMs*) são baseadas na Teoria de Aprendizagem Estatística (TAE) proposta por Vapnik Chernovemkis nas décadas de 1960 e 1970 (Vapnik, 1995). A TAE está relacionada com as condições matemáticas, que tem como função principal prever modelos de aprendizado de forma correta, em novos dados que não foram utilizados no treinamento do modelo. Esse fracionamento ocorre a partir de uma categorização onde considera o menor erro e a maximização da capacidade de generalização de um classificador, a priori considera o erro de treinamento ao mesmo tempo que maximiza a habilidade de difusão geral de um classificador para aprendizagem supervisionada.

Essa teoria aborda duas questões principais: (1) o problema da *generalização*, que é a capacidade de um modelo de se adaptar bem a dados novos e não vistos, e (2) o problema da *complexidade do modelo*, que busca equilibrar a simplicidade e a precisão para evitar o *overfitting* (quando o modelo se ajusta excessivamente aos dados de treinamento, tornando-se menos eficaz em dados novos) (Vapnik, 1998; CIN, UFPE).

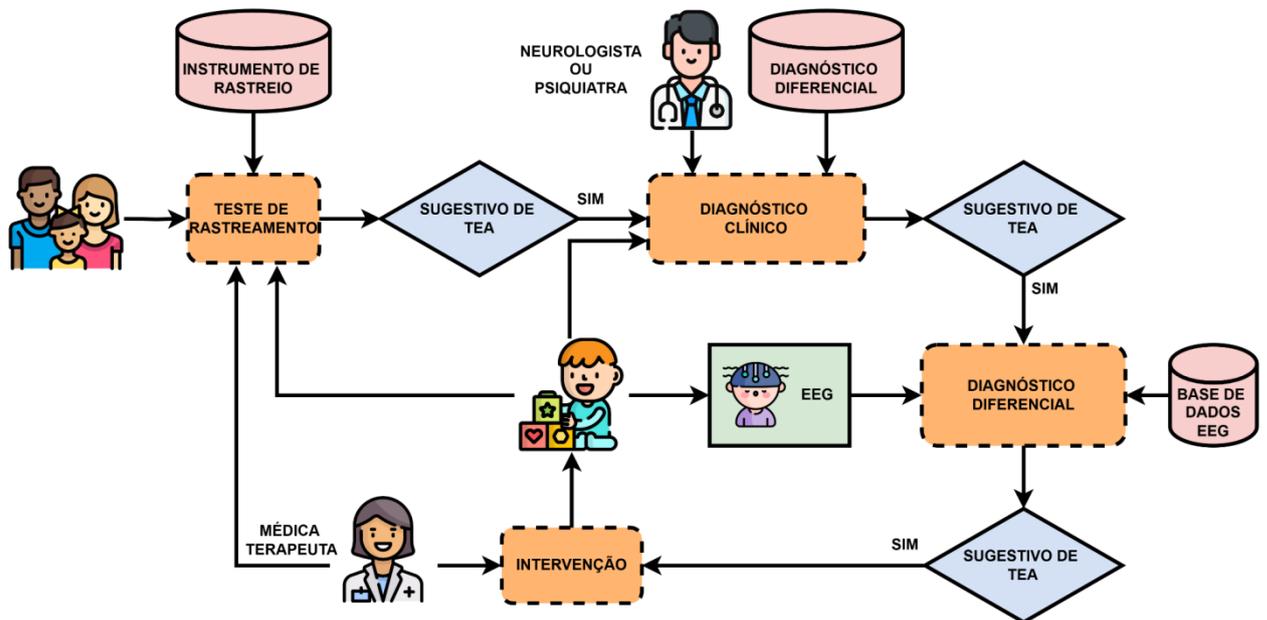
As máquinas de vetor de suporte (SVM) são algoritmos de aprendizado supervisionado amplamente aplicados em tarefas de classificação binária. Desenvolvidas para problemas onde os dados podem ser separados linearmente, as SVMs encontram um hiperplano de máxima margem que separa duas classes no espaço de características. Esse modelo matemático determina a posição do hiperplano por meio de vetores de suporte, que representam as

observações mais próximas da fronteira de separação, maximizando, assim, a margem entre as classes e reduzindo o risco de erro de generalização (Vapnik, 1995).

3 METODOLOGIA

Nesta seção analisamos a metodologia para o desenvolvimento de uma pesquisa exploratória em que se busca a natureza das indagações em forma de experimento e em técnica de coleta de dados. A metodologia do trabalho foi fundamental para caracterizar e averiguar a respeito do TEA que visa a melhor forma de apoio para diagnóstico, a vantagem em prever os métodos estabelecidos é a partir do fluxograma, que representa as diferentes etapas de um procedimento. O processo que envolve o modo para fluir o apoio ao diagnóstico, iniciou-se pela maneira em separar os dados coletados de indivíduos com TEA e a forma clínica em analisar esses dados. Esses dados são estabelecidos através de forma sucinta em que ocorre a anamnese desses pacientes e transferidos para uma clínica própria qual buscamos estabelecer métodos em interpretar o melhor caminho para investigar uma possível solução. Essa metodologia está detalhada no diagrama da Figura 2.

Figura 2 – Diagrama do fluxo metodológico aplicado no presente trabalho



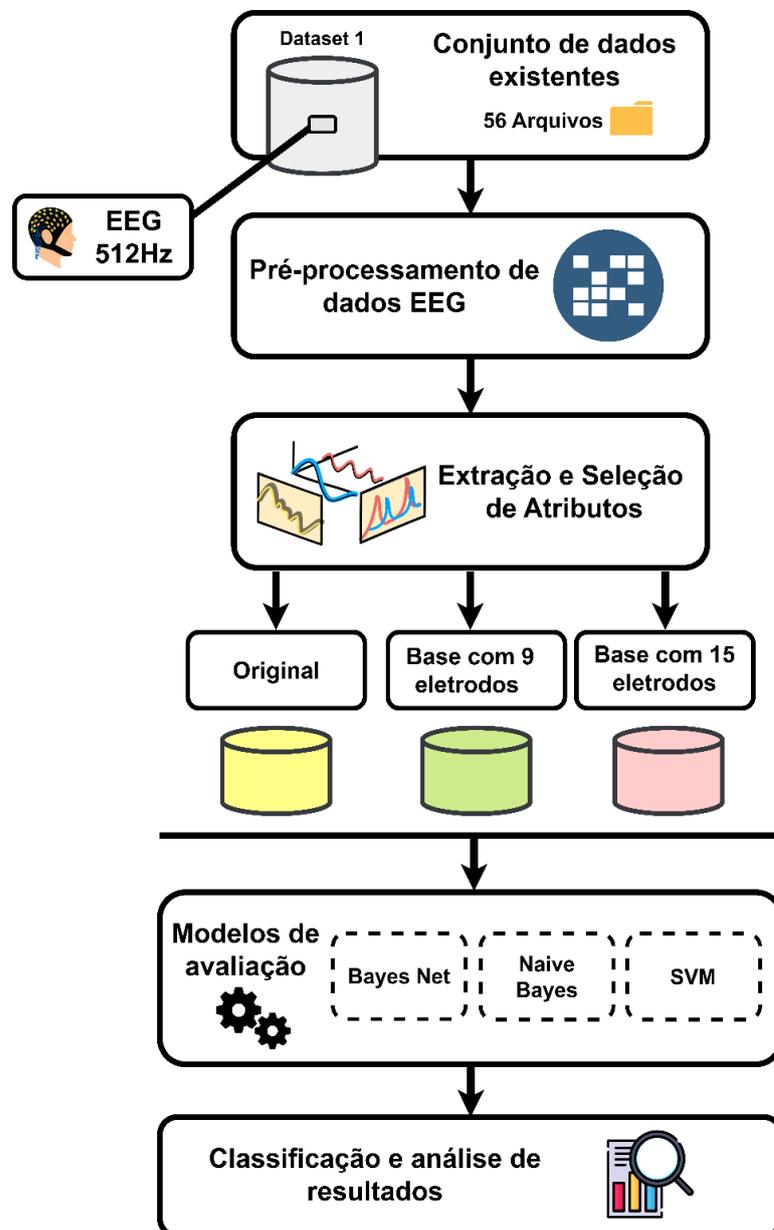
Fonte: Autoria própria (2024).

Com a Figura 2, exemplificamos que atualmente o diagnóstico clínico do TEA é realizado por neurologistas e neuropsiquiatras. O teste de rastreio é supervisionado por familiares e terapeutas, possibilitando a identificação de um diagnóstico sugestivo. Os instrumentos de diagnóstico para o autismo baseiam-se em uma análise de dados, dentre eles questionários, exames e documentos que possam identificar o TEA. Para aprimorar o desempenho e proporcionar um melhor custo-benefício no apoio ao diagnóstico do TEA,

propõe-se a utilização de medidas da atividade elétrica no cérebro, como o EEG, juntamente com técnicas de aprendizado de máquina, que podem fornecer um diagnóstico diferencial para o TEA.

Com o aprimoramento do uso do EEG, que identifica um padrão para o diagnóstico, destacamos e caracterizamos a facilidade em obter tais resultados de uma forma semelhante e simplificada, sem as dificuldades e processos que podem surgir de maneira não mencionada. Como apresentado na Figura 3, a seguir, destacamos o método utilizado para a obtenção de dados e confirmação dos resultados do EEG.

Figura 3 – Diagrama de fluxo metodológico aplicado no presente trabalho



Fonte: Autoria própria (2024)

Na Figura 3, apresenta-se o fluxograma da pesquisa. Esta pesquisa testou modelos de aprendizado de máquina para classificação binária e detecção do autismo. Cada uma das etapas de pré-processamento, extração e seleção de atributos e, por fim, classificação, estão detalhadas nas seções a seguir.

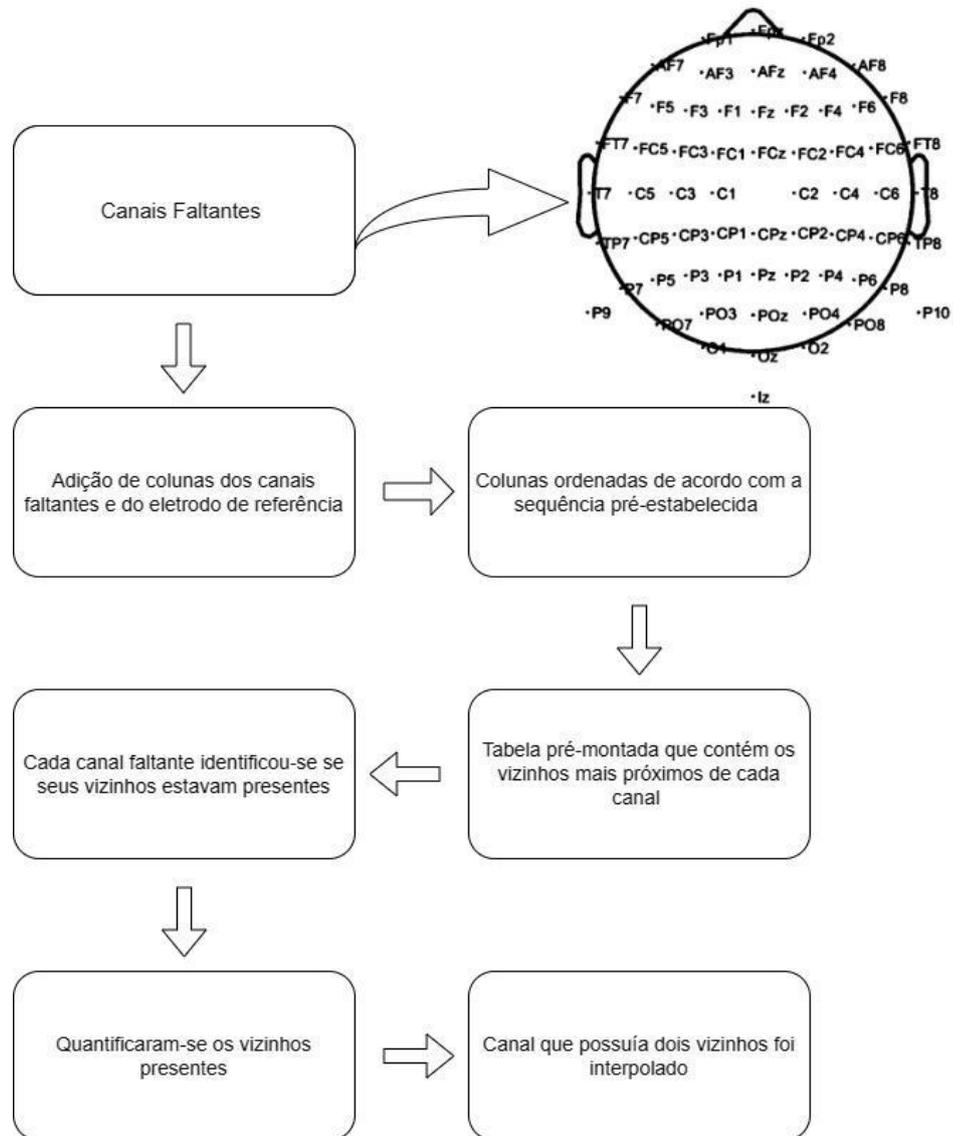
3.1 CONJUNTO DE DADOS EXISTENTES

O conjunto de dados existentes foram obtidos através de uma base de dados pública, tal qual é obtida pela Universidade de Sheffield, esses dados estão ligados a publicação "*Electrophysiological signatures of brain aging in autism spectrum disorder*" por Dickinson, Jeste e Milne (2022), na qual são referenciados como Dataset 1. A série de dados adquiridos consistiu em 56 indivíduos, 28 com diagnóstico do TEA e 28 controles neurotípicos com idades entre 18 e 68 anos. O conjunto de dados foi registrado em um período de 2,5 minutos (150 segundos) de repouso com os olhos fechados para obter o maior número possível de dados EEG sem artefatos. Os sinais de EEG foram amostrados a uma taxa de 140 Hz, reduzidos para 512 Hz, utilizando eletrodos ativos via *Biosemi Active two EEG system* (Dickinson *et al.*, 2022). O sistema de aquisição de dados contém 64 canais no total, que são rotulados com base no sistema de aquisição padrão internacional 10-10.

3.1.1 Pré-processamento de dados EEG

Uma vez que a base de dados foi obtida, pré-processamos a base de dados original. O conjunto de dados obtidos tem a forma de uma matriz em que é organizado da seguinte forma: canal (número de eletrodos) x tempo (s). Este passo foi efetuado utilizando o software GNU Octave 8.2.0. As janelas utilizadas foram de 2s com sobreposição de 0,5s e frequência de amostragem 512Hz. Os arquivos existentes eram em “.SET” e “.FDT” e foram convertidos em “.EDP” e “.CVS”. Realizou-se uma análise dos canais ausentes e desenvolveu-se um *script* usando o código Python 3.11.1 para o processamento desses dados, seguindo as etapas como mostrado na Figura 4, a seguir.

Figura 4: Análise dos Canais Ausentes



Fonte: Autoria própria (2024).

A média aritmética dos sinais dos vizinhos presentes foi usada para interpolar os canais ausentes. Esse processo foi repetido quatro vezes, pois a cada etapa de interpolação, mais vizinhos se tornavam disponíveis para os canais inicialmente não interpolados. Após esses passos, a primeira versão da base de dados tratada foi obtida. Em seguida rotulamos o conjunto de dados para pessoas atípicas e típicas e o algoritmo utilizado na base de dados.

3.1.2 Extração e seleção de atributos

Para cada janela, foram extraídos 34 atributos distintos do sinal, incluindo, entre outros, o valor médio, a amplitude máxima e a frequência média. Após o pré-processamento, foram extraídos 34 atributos de cada janela e cada canal de EEG, totalizando 2176 atributos (34 atributos explícitos x 64 canais). Na Figura 5 encontram-se as fórmulas para a realização da extração de atributos.

Figura 5: Atributos

Parâmetro	Equação	Parâmetro	Equação
Média (μ)	$\mu = \frac{1}{N} \sum_{n=1}^N x_n$	Comprimento de Onda	$WL = \sum_{n=1}^{N-1} x_{n+1} - x_n $
Variância	$var = \frac{1}{N-1} \sum_{n=1}^N (x_n - \mu)^2$	Taxa de Cruzamentos	$ZCR = \sum_{n=1}^{N-1} [sgn(x_n \times x_{n+1}) \cap x_n - x_{n+1} \geq \text{limiar}]$ $sgn(x) = \begin{cases} 1, & \text{if } x \geq \text{limiar} \\ 0, & \text{caso contrário} \end{cases}$
Desvio-padrão (σ)	$\sigma = \sqrt{\frac{1}{N-1} \sum_{n=1}^N x_n - \mu ^2}$	Variações do Sinal da Curva	$SSC = \sum_{n=1}^{N-1} [f(x_n - x_{n-1}) \times (x_n - x_{n+1})]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \text{limiar} \\ 0, & \text{caso contrário} \end{cases}$
Raiz Quadrada Média	$RMS = \sqrt{\frac{\sum_{n=1}^N (x_n)^2}{N}}$	Atividade do parâmetro Hjorth	$Hjorth_{ativ} = \frac{1}{N-1} \sum_{n=1}^N (x_n - \mu)^2$
Variações Médias de Amplitude	$AAC = \frac{1}{N} \left(\sum_{n=1}^N \left \frac{dx(t)}{dt} \right \right)$	Mobilidade do parâmetro Hjorth	$Hjorth_{mobilidade} = \sqrt{\frac{\text{var} \left(\frac{dx(t)}{dt} \right)}{\text{var}(x(t))}}$
Desvio Diferencial Absoluto	$DASDV = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{dx(t)}{dt} \right)^2}$	Complexidade do parâmetro Hjorth	$Hjorth_{complexidade} = \frac{Hjorth_{mobilidade} \left(\frac{dx(t)}{dt} \right)}{Hjorth_{mobilidade}(x(t))}$
Valor Absoluto Integrado	$IAV = \sum_{n=1}^N x_n$	Frequência Média	$MNF = \frac{\sum_{j=1}^M f_j P_j}{\sum_{j=1}^M P_j}$ Onde f_j, P_j são as frequências e energia do espectro, respectivamente, e M é o comprimento das frequências
Detector Logarítmico	$LOGD = e^{\left(\frac{1}{N} \sum_{n=1}^N \log(x_n) \right)}$	Frequência Mediana	$MDF = \frac{1}{2} \sum_{j=1}^M P_j$

Integral Quadrada Simples	$SSI = \sum_{n=1}^N x_n^2$	Potência Média	$MNP = \sum_{j=1}^M \frac{P_j}{M}$
Valor Absoluto Médio	$MAV = \frac{1}{N} \sum_{n=1}^N x_n $	Pico de Frequência	$PKF = \max(P_j)$
<i>Kernel</i> Logarítmico Médio	$MLOGK = \frac{1}{N} \left \sum_{n=1}^N x_n \right $	Razão do Espectro de Potência	$PSR = \frac{PKF}{\sum_{j=1}^M P_j}$
Skewness (s)	$s = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^3}{\sigma^3}$	Energia Total	$TP = \sum_{j=1}^M P_j$
Kurtosis	$kurtosis = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^4}{\sigma^4}$	Primeiro Momento Espectral	$SM1 = \sum_{j=1}^M f_j P_j$
Amplitude Máxima	$MAX = \max(x_n)$	Segundo Momento Espectral	$SM2 = \sum_{j=1}^M f_j^2 P_j$
Terceiro Momento	$M3 = \left \frac{1}{N} \sum_{n=1}^N (x_n)^3 \right $	Terceiro Momento Espectral	$SM3 = \sum_{j=1}^M f_j^3 P_j$
Quarto Momento	$M4 = \left \frac{1}{N} \sum_{n=1}^N (x_n)^4 \right $	Variância da Frequência Central	$VCF = \frac{SM2}{TP} - \left(\frac{SM1}{TP} \right)^2$
Quinto Momento	$M5 = \left \frac{1}{N} \sum_{n=1}^N (x_n)^5 \right $		

Fonte: ESPINOLA, C. W., Análise Computacional da Voz como uma Ferramenta de Auxílio Diagnóstico de Transtornos Mentais. Recife, 2021 (p.86-87).

Diante disto, os atributos mais relevantes foram selecionados utilizando métodos de otimização, incluindo enxame de partículas (PSO) e busca evolucionária. O propósito dessa seleção era identificar quais atributos são mais pertinentes para tarefas de classificação. Como resultado, foram produzidos três conjuntos de dados finais: a base de dados original, a base de dados otimizada com PSO e a base de dados otimizada com busca evolucionária, todas devidamente tratadas. Na parte de pré-processamento das três bases foram divididos em base de treinamento e validação com 80% dos dados e a base teste com 20% dos dados.

3.1.3 Modelos de avaliação

Os modelos de aprendizado testados foram os: Bayes Net, Naive Bayes e SVM (Kernel Linear, Kernel Polinomial de grau 2, Kernel Polinomial de grau 3 e Kernel RBF, variando os valores de gama: 0,5; 0,2 e 0,1).

As Redes Bayesianas (*Bayesian Networks* - BN), também conhecidas como redes de crença, pertencem à família de modelos gráficos probabilísticos (Ben-gal, 2008). Esse tipo de rede pode ser definida como um grafo direcionado acíclico, no qual os nós correspondem às variáveis do domínio e as arestas correspondem às dependências probabilísticas diretas entre elas (Friedman *et al.*, 2003). A força da correlação entre essas variáveis é definida nas Tabelas de Probabilidade Condicional (CPTs) anexadas a cada nó, as quais especificam o grau de crença do nó em um estado específico, dado os estados dos nós pais (Chen, 2012).

O algoritmo *Naive Bayes* é derivado do teorema de Bayes (Vijay *et al.*, 2023). Ele é a forma mais simples de rede bayesiana, na qual todos os atributos são independentes, dado o valor da variável de classe, caracterizando o que é conhecido como independência condicional (Zhang, 2004). O Classificador *Naive Bayes* tem a vantagem de exigir uma pequena quantidade de dados de treinamento para estimar os parâmetros (média e variância das variáveis) necessários para a classificação (Aditya *et al.*, 2022).

As Máquinas de Vetores de Suporte (SVMs) foram desenvolvidas por Vapnik como uma implementação do SRM (Burbidge *et al.*, 2001). Essa abordagem pode ser descrita como uma ferramenta de predição que visa encontrar um hiperplano específico, uma linha ou limite de decisão, capaz de separar eficientemente conjuntos de dados ou classes, minimizando o problema de sobreajuste de dados (Somvanshi *et al.*, 2016). As SVMs têm demonstrado ampla eficácia em aplicações de aprendizado de máquina envolvendo conjuntos de dados volumosos e de alta dimensionalidade (Ruping, 2001).

Por meio da aplicação do software Weka 3.9.6, cada grupo de dados foi separado em conjuntos de treinamento (80%) e teste (20%). Posteriormente, foram empregados os modelos e parâmetros mencionados anteriormente. Cada experimento foi repetido 30 vezes, e utilizou-se validação cruzada com 10 divisões para garantir uma base estatística sólida. Esse procedimento resultou na obtenção das métricas fundamentais para avaliar o desempenho de cada classificador:

3.1.4 Métricas de avaliação

- **Acurácia:** é uma das métricas mais simples e comuns usadas para avaliar a performance de modelos de classificação. Ela fornece uma medida geral da taxa de acertos do modelo, ou seja, a proporção de previsões corretas em relação ao total de previsões feitas pelo modelo (Srivastava, 2014). A acurácia é especialmente útil quando as classes estão balanceadas, ou seja, têm aproximadamente o mesmo número de amostras em cada classe. Nesses casos, a acurácia fornece uma boa indicação do desempenho geral do modelo. Para calcular a acurácia se usa a seguinte fórmula:

$$\text{Acurácia} = \frac{\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}}{\text{Total de Amostras}} \quad (1)$$

- **Índice kappa:** também conhecido como Kappa de Cohen, é uma métrica de avaliação de concordância usada para medir o grau de acordo entre as previsões de um modelo de classificação e as classes reais dos dados (Merlini e Rossini, 2022). Diferentemente da acurácia, o Kappa leva em consideração o acordo que pode ser esperado ao acaso, o que o torna especialmente útil quando se lida com classes desbalanceadas. O índice Kappa é calculado comparando as frequências observadas de concordância entre o modelo e os dados reais com as frequências que seriam esperadas ao acaso. Ele fornece um valor que varia de -1 a 1. A fórmula para calcular esse índice é:

$$\text{Índice Kappa} = \frac{(p_o - p_e)}{(1 - p_e)} \quad (2)$$

Onde P_o é a proporção de concordância observada entre o modelo e os dados reais. É calculada como a soma dos verdadeiros positivos e verdadeiros negativos dividida pelo número total de amostras. P_e é a proporção de concordância esperada ao acaso. É calculada como a multiplicação das proporções marginais correspondentes das previsões do modelo e das classes reais, somadas e divididas pelo número total de amostras ao quadrado.

- **Sensibilidade:** também conhecida como recall ou taxa de verdadeiros positivos, é uma métrica de avaliação de modelos de classificação que mede a capacidade do modelo de

identificar corretamente as amostras positivas (classe positiva) em relação ao total de amostras reais que pertencem à classe positiva (Mun e Jumadi, 2020). A fórmula para calcular a sensibilidade é a seguinte:

$$\text{Sensibilidade} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}} \quad (3)$$

- Especificidade: é outra métrica de avaliação usada em modelos de classificação. Ao contrário da sensibilidade, que se concentra nas amostras positivas, a especificidade avalia a capacidade do modelo de identificar corretamente as amostras negativas (classe negativa) em relação ao total de amostras reais que pertencem à classe negativa (Abuhaija *et al.*, 2023). A fórmula para calcular a especificidade é a seguinte:

$$\text{Especificidade} = \frac{\text{Verdadeiros Negativos}}{\text{Verdadeiros} + \text{Falsos Positivos}} \quad (4)$$

- Área ROC: a área sob a curva ROC (*Receiver Operating Characteristic*) ou AUC-ROC é uma métrica de avaliação usada em modelos de classificação binária para medir a qualidade geral do modelo e sua capacidade de distinguir entre as duas classes (positiva e negativa) (Smelser, 2001). A Curva ROC é uma representação gráfica que mostra a taxa de verdadeiros positivos (sensibilidade) em função da taxa de falsos positivos (1 - especificidade) para diferentes pontos de corte de decisão do modelo.

Cada ponto na curva ROC corresponde a um limiar diferente usado para classificar as amostras como positivas ou negativas. O ponto (0,0) representa o ponto de corte em que todas as amostras são classificadas como negativas, enquanto o ponto (1,1) representa o ponto de corte em que todas as amostras são classificadas como positivas. A AUC-ROC mede a área sob essa curva ROC e varia de 0 a 1. Quanto maior o valor da AUC-ROC, melhor é o desempenho do modelo.

Finalmente, foram avaliadas as métricas mencionadas anteriormente, possibilitando a comparação de cada modelo para a escolha do melhor classificador e configuração.

3.1.5 Classificação e análise de resultados

Após concluir esse processo, as métricas cruciais para avaliar o desempenho de cada classificador foram adquiridas, incluindo acurácia, índice kappa, sensibilidade, especificidade e área sob a curva ROC. Em seguida, essas métricas foram examinadas e possibilitaram a comparação entre os diferentes modelos. A matriz de confusão é uma tabela que é usada para avaliar o desempenho de um modelo de classificação em problemas de aprendizado supervisionado. Ela mostra a distribuição dos valores previstos pelo modelo em comparação com os valores reais do conjunto de dados. A matriz é especialmente útil quando se trabalha com problemas de classificação com várias classes, como será utilizada no modelo das Tabelas 1 e 2. Os valores foram úteis para avaliar o desempenho do modelo de classificação, permitindo calcular as métricas.

4 RESULTADOS E DISCUSSÃO

Após a etapa de extração e seleção de atributos, testaram-se modelos de máquinas de aprendizado na classificação dos sinais. Primeiro foi realizado o treinamento e validação, utilizando validação cruzada de 10 folds e 30 rodadas para cada arquitetura. Em seguida, foi realizada a etapa de teste. Os melhores resultados da base de 9 e 15 eletrodos foram no modelo do SVM com kernel RBF e parâmetro gama de 0,5. Os resultados das tabelas a seguir (Tabela 1 e 2), são resultados de uma base com 9 eletrodos. Na parte de pré-processamento da base original foram divididos em base de treinamento e validação com cerca de 80% e a base teste com 20%.

Nas tabelas a seguir observamos as métricas como: Acurácia, Índice Kappa, Sensibilidade, Especificidade e Área ROC, através dos modelos: Bayes Net, Naive Bayes, SVM (Polinomial linear, de grau 2 e 3) e (RBF 0,5, RBF 0,2 e RBF 0,1).

Tabela 1: Resultados do treinamento para classificação de EEG para o cenário com 9 eletrodos.

<i>Modelo</i>		<i>Métrica</i>				
		<i>Acurácia</i>	<i>Índice Kappa</i>	<i>Sensibilidade e</i>	<i>Especificidade e</i>	<i>Área ROC</i>
<i>Bayes Net</i>		73,26 ± 2,09	0,47 ± 0,04	0,75 ± 0,03	0,72 ± 0,03	0,83 ± 0,02
<i>Naive Bayes</i>		61,01 ± 0,65	0,20 ± 0,03	0,26 ± 0,02	0,93 ± 0,01	0,74 ± 0,02
<i>SV M</i>	<i>Poli (Linear)</i>	84,48 ± 1,60	0,69 ± 0,03	0,84 ± 0,02	0,85 ± 0,02	0,84 ± 0,02
	<i>Poli 2</i>	96,39 ± 0,85	0,93 ± 0,02	0,96 ± 0,01	0,96 ± 0,01	0,96 ± 0,01
	<i>Poli 3</i>	96,57 ± 0,88	0,93 ± 0,02	0,96 ± 0,01	0,97 ± 0,01	0,97 ± 0,01
<i>SV M</i>	<i>RBF 0,5</i>	96,71 ± 0,80	0,93 ± 0,02	0,96 ± 0,01	0,97 ± 0,01	0,97 ± 0,01
	<i>RBF 0,2</i>	93,81 ± 1,10	0,88 ± 0,02	0,92 ± 0,02	0,95 ± 0,01	0,94 ± 0,01
	<i>RBF 0,1</i>	89,26 ± 1,37	0,79 ± 0,03	0,86 ± 0,02	0,93 ± 0,02	0,89 ± 0,01

Fonte: Autoria própria (2024).

Na Tabela 1, constam os valores relacionados a classificação de EEG com 9 eletrodos. Destacamos o melhor modelo em negrito, sendo ele escolhido por apresentar uma melhor acurácia média. Para termos uma leitura clara e objetiva da tabela acima, destacamos que o valor superior das métricas estão no modelo citado como SVM (Polinomial 2, Polinomial 3 e RBF 0,5). Vale salientar que a acurácia do pior modelo consta no Naive Bayes e o melhor modelo no SVM RBF 0,5. Com esses valores relacionados distinguimos entre as métricas: Acurácia, Índice Kappa, Sensibilidade, Especificidade e Área ROC os valores relacionados a cada um deles.

Na Tabela 2, a seguir, exemplifica os valores relacionados à classificação dos melhores modelos, segundo as métricas citadas como a acurácia, índice kappa, sensibilidade,

especificidade e área ROC. Nesta tabela contém informações da fase de teste da classificação de EEG com 9 eletrodos, utilizando modelo escolhido na etapa de treinamento/validação.

Tabela 2: Métricas com valores correspondentes a fase teste de 9 eletrodos.

Métrica	SMO/SVM RBF 0,5
Acurácia (%)	98,22
Índice Kappa	0,96
Sensibilidade	0,98
Especificidade	0,98
Área ROC	0,98

Fonte: Autoria própria (2024).

A Tabela 2 informa os resultados do teste, com o melhor modelo escolhido na fase de treino do modelo selecionado, como citado o (SVM) de grau 0,5. Nele retratamos os valores de cada métrica, gerada na fase de teste, posterior a fase treino. Nessa linha de raciocínio estabelece primeiro a fase treino e logo em seguida utilizou-se os modelos selecionados para a fase teste, como podemos citar as Tabelas 1 e 2, a distinção entre eles está relacionada pela quantidade de eletrodos gerados em cada fase. Inferiu-se que o modelo que se aproxima da maior proporção ou relacionado ao valor 1, gerará a melhor classificação.

Na Figura 6, mostra a matriz de confusão gerada a partir da fase teste de 9 eletrodos. Essas informações provêm do valores previsto x real, onde a classificação de indivíduos com TEA se localiza nas proporções citadas a seguir.

Figura 6: Matriz de confusão do melhor modelo encontrado nas três abordagens: SVM com Kernel Rbf de gamma=0,5, após PSO.

		Previsto	
		TEA	C
Real	TEA	2317	50
	C	34	2340

Fonte: autoria própria (2024).

Na matriz de confusão exemplificamos os resultados de forma clara e objetiva, os indivíduos com TEA e do grupo controle. Podemos destacar a precisão de pacientes com TEA em uma escala de aproximadamente 0,98%. Nessas características podemos apontar o valor

sugestivo de que as previsões positivas estão ditas como corretas. Na revocação indica que o modelo detectou aproximadamente 0,97% dos casos positivos.

Tabela 3: Resultados do treinamento para classificação de EEG para o cenário com 15 eletrodos.

Modelo		Métrica				
		Acurácia	Índice Kappa	Sensibilidade	Especificidade	Área ROC
Bayes Net		74,41 ± 2,01	0,49 ± 0,04	0,77 ± 0,03	0,72 ± 0,03	0,82 ± 0,02
Naive Bayes		60,27 ± 1,68	0,20 ± 0,03	0,27 ± 0,02	0,92 ± 0,01	0,74 ± 0,02
SV M	Poli (Linear)	90,52 ± 1,32	0,81 ± 0,03	0,90 ± 0,02	0,91 ± 0,02	0,91 ± 0,01
	Poli 2	97,77 ± 0,67	0,96 ± 0,01	0,98 ± 0,01	0,98 ± 0,01	0,98 ± 0,01
	Poli 3	97,66 ± 0,67	0,95 ± 0,01	0,97 ± 0,01	0,98 ± 0,01	0,98 ± 0,01
SV M	RBF 0,5	98,41 ± 0,55	0,97 ± 0,01	0,99 ± 0,01	0,98 ± 0,01	0,98 ± 0,01
	RBF 0,2	97,52 ± 0,69	0,95 ± 0,01	0,97 ± 0,01	0,98 ± 0,01	0,98 ± 0,01
	RBF 0,1	95,50 ± 0,90	0,91 ± 0,02	0,94 ± 0,01	0,97 ± 0,01	0,96 ± 0,01

Fonte: autoria própria (2024).

Na Tabela 3, constam os valores relacionados a classificação de EEG com 15 eletrodos, onde os melhores modelos estão com resultados específicos para a base, em negrito. Visto que dentre eles podemos evidenciar o SVM RBF 0,5, com valor primordial para o estudo classificatório. Dentre as métricas estabelecidas, a Especificidade e a Área ROC, sabe-se que nos modelos (Polinomial 2, Polinomial 3, RBF 0,5 e RBF 0,2) os valores são os mesmos. O pior modelo continua sendo o Naive Bayes, possuindo resultados menores que os demais, trazendo não menos importante a especificidade do Bayes Net continua sendo menor que o Naive Bayes. Visto que citamos também na Tabela 1, isso é um resultado que deve ser levado em consideração em ambos os casos, pois estabelece uma classificação negativa em relação ao total de amostras reais.

Tabela 4: Resultados da fase de testes dos melhores modelos

Métrica	SMO/SVM RBF 0,5
Acurácia (%)	54,54
Índice Kappa	0,09
Sensibilidade	0,54
Especificidade	0,63
Área ROC	0,54

Fonte: autoria própria (2024).

A Tabela 4 informa os resultados do teste, com o melhor modelo escolhido na fase de treino do modelo selecionado, como citado no SVM de grau 0,5. Nele retratamos os valores de cada métrica, gerada na fase de teste, posterior a fase treino. Nessa linha de raciocínio estabeleceu-se primeiro a fase treino e logo em seguida utilizou-se a fase teste, como podemos citar as Tabelas 3 e 4, a distinção entre eles está relacionada pela quantidade de eletrodos gerados em cada fase.

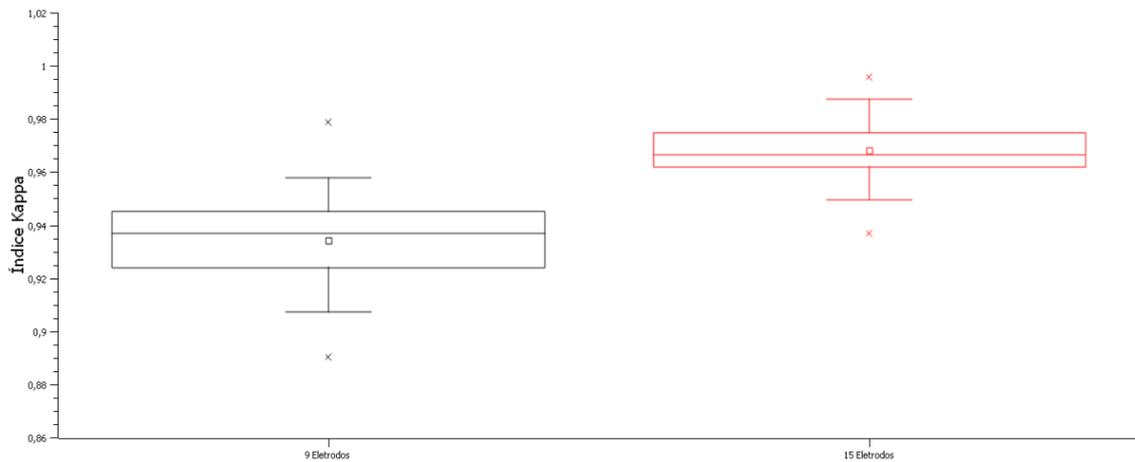
Figura 7: Matriz de confusão do melhor modelo encontrado nas três abordagens: SVM com Kernel Rbf de gamma=0,5, após PSO.

		Previsto	
		TEA	C
Real	TEA	2258	109
	C	2046	328

Fonte: autoria própria (2024).

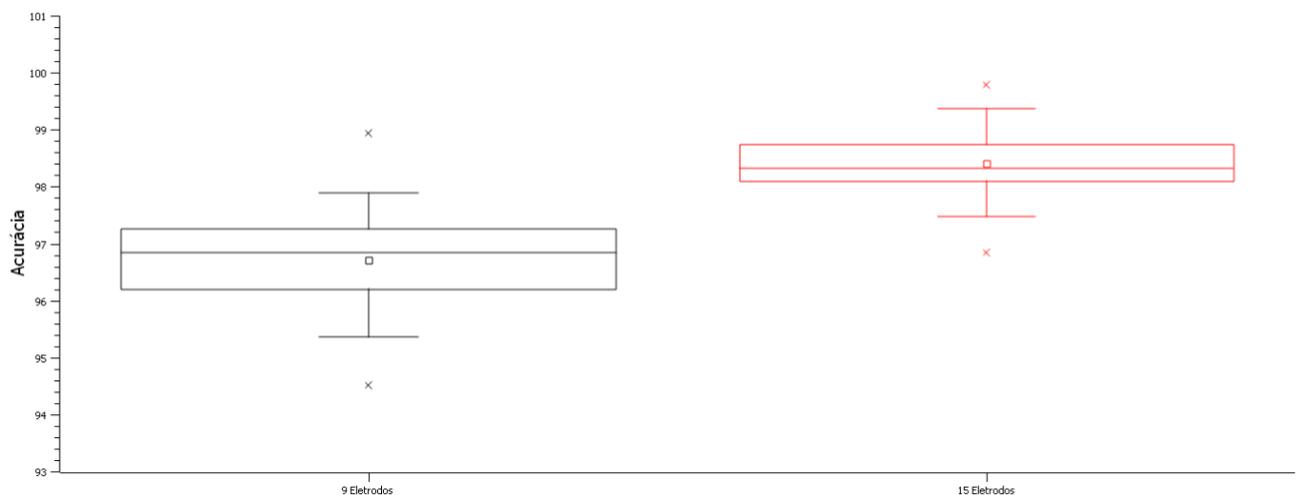
Na matriz de confusão ela exemplifica os resultados de forma clara e objetiva, os indivíduos com TEA e do grupo controle. Podemos destacar a precisão de pacientes com TEA em uma escala de aproximadamente 0,52%. Nessas características podemos apontar o valor sugestivo de que as previsões positivas estão ditas como razoáveis. Na revocação indica que o modelo detecta aproximadamente 0,97% dos casos positivos.

O gráfico boxplot é uma ferramenta eficaz para identificar a dispersão e assimetria dos dados, além de destacar a presença de valores atípicos. Ele permite uma comparação visual entre diferentes conjuntos de dados e ajuda na detecção de padrões, tendências e variações. É amplamente utilizado em estatísticas descritivas, análise exploratória de dados e na identificação de discrepâncias em experimentos ou estudos comparativos.

Gráfico 1: Boxplot do Índice Kappa

Fonte: Autoria própria (2024).

Identificamos com os gráficos boxplots as análises comparativas paramétricas, onde utiliza-se a comparação de grupos através da mediana, primeiro e terceiro quartis, destacando assim, uma medida de tendência central, mediana, e duas medidas de variabilidade, primeiro e terceiro quartis. No gráfico 1, indica a dispersão dos valores estabelecidos tanto no de 9 eletrodos quanto no de 15 eletrodos, a distinção entre eles indica a classe de eletrodos, e os parâmetros estabelecidos da caixa. Quanto mais estreita a caixa for mais próxima ao valor médio e possui uma menor dispersão. São grandezas diretamente proporcionais. Enquanto na caixa mais larga há variância nos valores e as caixas são mais dispersas. A seguir consta a forma estabelecida da Acurácia, onde as medidas e valores são basicamente os mesmos, mas com possíveis diferenças.

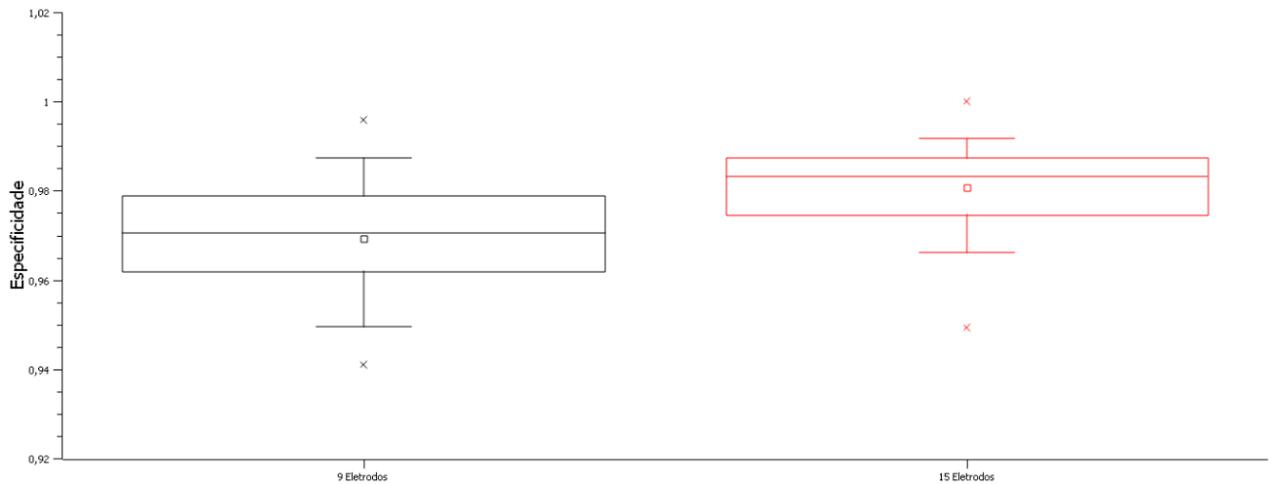
Gráfico 2: Boxplot da Acurácia

Fonte: Autoria própria (2024).

No gráfico 2, podemos observar os valores dispersos na caixa com 9 eletrodos em que ocorre dispersão de dados em torno da mediana. Já a caixa com o valor de 15 eletrodos está

compacta e estreita possibilitando assim uma menor dispersão. O primeiro quartil relacionado a caixa de 9 eletrodos está acima da mediana cerca de 25%, enquanto os outros 75% estão entre a média possibilitando assim um pior desempenho. Relacionado a caixa de 15 eletrodos acontece obrigatoriamente o inverso do valor obtido em 9 eletrodos, na qual o valor está acima da linha da mediana indicando um melhor desempenho.

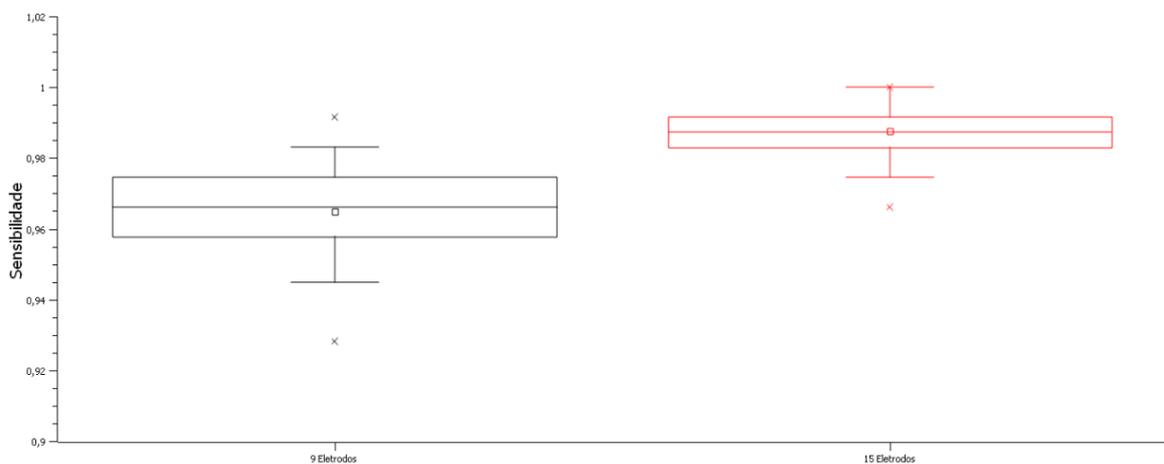
Gráfico 3: Boxplot da Especificidade



Fonte: Autoria própria (2024).

O gráfico 3, está diretamente relacionado com a possibilidade em obter uma dispersão em ambas as caixas com 9 e 15 eletrodos, em que a linha da mediana está acima da média. Mas, a quantidade de dispersão entre as caixas da direita e da esquerda estão em formatos diferentes, a primeira (direita) está exatamente no meio da caixa possibilitando assim uma variabilidade de valores, exatamente 50% em cada lado. A segunda (esquerda) está com uma menor dispersão, possuindo assim próximos ao valor médio.

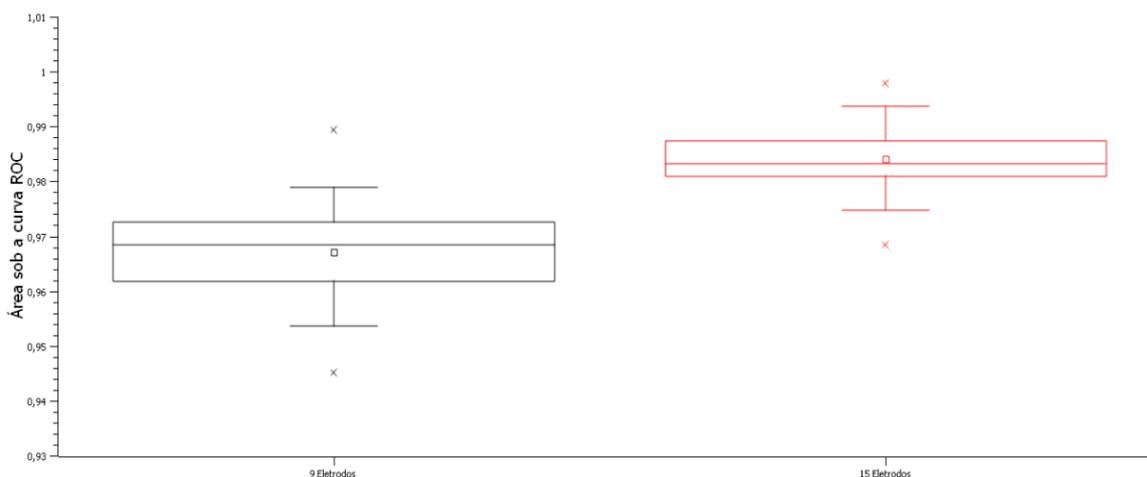
Gráfico 4: Boxplot da Sensibilidade



Fonte: Autoria própria (2024).

O gráfico 4, representa a dispersão de valores entre 9 e 15 eletrodos. Na caixa da direita que contém 9 eletrodos, a linha mediana está abaixo da média possibilitando assim uma equivalência de valores entre 50%. A caixa da esquerda está diretamente ligada à média e a mediana, e possui uma compactação menor do que a da direita, significando assim uma menor dispersão próximo ao valor médio. Na caixa da direita, apesar da igualdade entre ambos os lados, o que difere é a localização da média, ela varia os valores e são mais dispersos.

Gráfico 5: Boxplot da Área sob a curva ROC



Fonte: Autoria própria (2024).

No gráfico 5, a caixa da direita possui uma maior dispersão em relação a caixa da esquerda, possibilitando assim uma variedade de valores. A caixa com 9 eletrodos (direita), está com a linha mediana acima da média, significa que o quartil 3 está com 25% acima da média variando os valores e dispersão. Enquanto a caixa com 15 eletrodos (esquerda) está com a linha mediana abaixo da média, significa que o quartil 3 está com 75% e contém a média nesse intervalo de valores. Possuindo assim uma menor dispersão em comparação a caixa da direita.

Com isso podemos salientar sobre os seguintes aspectos relacionados aos dados resultantes sobre o apoio ao diagnóstico do TEA, o resultado de classificação é suficiente para o projeto, mas com alguns pontos de vista selecionados são um tanto positivos quanto negativos, devido a alguns critérios apresentados na base de dados, como exemplo a faixa etária especificando os indivíduos com ou sem TEA. Os pontos positivos são as métricas que estão atingindo 98,22% de acurácia na base com 9 eletrodos, porém na base com 15 eletrodos estão atingindo um valor de 54,54% de acurácia, um parâmetro negativo mas com bastante relevância, para que seja comparado objetivando o melhor modelo aplicado na prática clínica precisa além de ser realizado novos dados com mais especificações do público alvo.

Segundo Silva, Adrielly, Et.al (2022) utilizou-se a mesma base de dados como “Sheffield”, esse estudo buscou enfatizar o diagnóstico de perturbações do TEA com EEG de

baixa densidade e aprendizagem automática, com um intuito de buscar informações precisas para o apoio ao diagnóstico do TEA torna-se um estudo comparativo, que traz referências precisas que auxiliaram essa dissertação. A metodologia consiste em aprimorar os dados de indivíduos com TEA, neste artigo proposto empregou-se 56 pessoas adultas no total com idade entre 18-68 anos, sendo eles submetidos a um EGG de 150 segundos de gravação, as quais mantiveram os olhos fechados em repouso, sendo expostos a estímulos visuais. Os dados de EEG foram filtrados para remover ruídos. Alguns participantes apresentaram dados de uma matriz com 64 sensores e outros com 128 sensores que foram ajustados para garantir a consistência da análise.

O *Software* EEGLAB foi usado para filtrar os segmentos de EGG e remover segmentos contaminados. Um estudo comparativo entre outros modelos da base de dados como exemplo o Random Forest com 500 árvores de decisão com 99,75% de acurácia obteve um alto índice em comparação aos modelos anteriores, isso mostra a eficácia e a comprovação da base “Dataset 1”, com métricas equivalentes a 9 eletrodos. A distinção entre o artigo proposto deste artigo são as métricas e modelos com resultados relevantes para apoiar o diagnóstico do TEA e os eletrodos utilizados. Reforçamos assim que mesmo com alguns critérios da base pré-definidos temos resultados promissores e um tanto desejáveis para uma continuidade dessa metodologia na prática clínica, o que podemos melhorar nesse aspectos citados anteriormente é a busca de novas investigações a serem feitas.

Tendo em vista que Salankar, Nilima et al (2022), relacionam o contexto entre processamento automatizado de sinais biomédicos e cuidados com a saúde, e trazem uma precisão de 99,99% de acurácia utilizando o “*Support Vector Machine*”. Nesse contexto, é apresentada uma abordagem original para a categorização de classes de stress e não stress através do processamento de sinais de EGG multi-canal. A metodologia utilizada foi referente a duas etapas iniciais do pré-processamento de dados: (1) o “*Empirical Mode Decomposition*” resultados da função modo intrínseco e (2) “*Variable Mode Decomposition*” resultados em modelos, dando prosseguimento foram construídos gráficos de seguimento de segunda ordem, extração de características, seleção de canais e classificação. Além disso, houve uma boa acurácia nos modelos de SVM e *Multilayer Perceptron Neural Network* (MLPNN). Este método tem um desempenho adequado para o conjunto de dados considerados. As precisões mais elevadas de subconjunto 99,89%, de canal 98,89% e de lobo 99,99% são obtidas, respectivamente. O método concebido demonstrou um potencial de incorporação para a aplicação pretendida. Com isso, salientamos a eficácia do modelo SVM, como concebido no

trabalho. Isso mostra uma capacidade em aprimorar os dados futuros e a utilização da base de dados da pesquisa proposta.

5 CONCLUSÕES

A partir da observação dos aspectos analisados utilizou-se uma abordagem metodológica aplicada na exploração de sinais EEG e em técnicas de aprendizado de máquina para apoiar o diagnóstico do TEA. A metodologia abrange o pré-processamento de dados, extração e seleção de atributos, e a utilização de modelos de classificação, com ênfase em Redes Bayesianas SVM.

Os experimentos utilizados com dados da base “*Dataset 1*” mostraram que o modelo SVM com kernel RBF demonstrou a melhor eficácia, alcançando uma acurácia de 98,22% na configuração com 9 eletrodos. Esses resultados mostram que é possível obter uma classificação precisa dos sinais de EEG para diferenciar indivíduos com TEA e controle neurotípico. A realização de métodos automatizados, como as exploradas neste estudo, representa um avanço significativo no apoio ao diagnóstico.

Além disso, a análise dos resultados indicou que a combinação de atributos relevantes extraídos dos sinais cerebrais com técnicas de aprendizado de máquina pode reduzir significativamente a subjetividade no processo de diagnóstico, aumentando a confiabilidade das avaliações. A utilização de métricas robustas, como acurácia, sensibilidade e especificidade, reforça a eficácia do modelo proposto para aplicações clínicas.

Outro ponto relevante foi a confirmação de que modelos baseados em EEG possuem grande potencial como ferramentas de apoio no diagnóstico do TEA, permitindo intervenções precoces que podem impactar positivamente o desenvolvimento e a qualidade de vida dos pacientes. Além disso, o estudo destacou a importância do uso de sistemas automáticos para otimizar o tempo e a eficiência de profissionais da área de saúde.

Durante o desenvolvimento desta pesquisa, enfrentou-se vários desafios. A primeira dificuldade foi a limitação na disponibilidade de bases de dados específicas para o problema do TEA. Além disso, a base de dados utilizada apresentou limitações relacionadas à faixa etária dos participantes e ao volume reduzido de amostras, o que exigiu técnicas robustas de pré-processamento para lidar com artefatos e dados incompletos.

Além disso, outra dificuldade foi a necessidade de ajustar e comparar diferentes modelos de aprendizado de máquina, garantindo que o desempenho fosse avaliado de forma justa e estatisticamente consistente. Ademais, o uso de equipamentos de EEG de baixa densidade gerou limitações na resolução espacial dos sinais cerebrais coletados, impactando potencialmente a qualidade dos resultados. Essa característica demandou um cuidado extra no tratamento dos dados para minimizar os efeitos das limitações tecnológicas.

As contribuições foram realizadas de acordo com a cronologia associada aos semestres. Os resultados obtidos tiveram ampla divulgação em veículos de relevância acadêmica:

- Apresentação em Conferência: “Motirõ: Aplicativo para Testes de Rastreamento do Transtorno do Espectro Autista,” apresentado no VI Simpósio de Inovação em Engenharia Biomédica 2022. Este trabalho desempenhou como uma proposta de ferramenta de apoio para profissionais de educação na aplicação dos respectivos testes com certificado de 12h/a.
- Apresentação em Conferência: “Um sistema inteligente baseado em eletroencefalografia para apoio ao diagnóstico do Transtorno do Espectro Autista: uma revisão de escopo,” apresentado no VII Simpósio de Inovação em Engenharia Biomédica 2023. Este trabalho destacou o desempenho das atividades relacionadas à revisão de literatura com certificado de 12h/a.
- Apresentação em Conferência: “Um Sistema Inteligente Baseado em Eletroencefalografia para Apoio ao Diagnóstico do Transtorno do Espectro Autista,” apresentado no VII Simpósio de Inovação em Engenharia Biomédica 2024. Este trabalho evidenciou o apoio ao diagnóstico precoce do TEA com certificado de 20h/a.
- Capítulo de Livro: “Infrared Thermography: Principles and Applications in Healthcare,” publicado no livro *Biomedical Imaging Principles and Advancements*. Este capítulo tem como objetivo explorar os fundamentos e aplicações da termografia infravermelha (IR) no setor de saúde.
- Participação do Artigo: “Development of an Electrode Reduction Model to Support ASD Diagnosis Based on EEG and Machine Learning,” publicado no IEE *Latin American Conference on Computational Intelligence (LA-CCI)*. Este artigo tem como objetivo principal desenvolver um modelo de redução de eletrodos para EEG, utilizando aprendizado de máquina, para facilitar e aprimorar o diagnóstico do TEA.

A despeito de que este estudo tenha obtido resultados promissores, há diversas lacunas que podem ser exploradas em pesquisas futuras:

- Ampliar a base de dados utilizada, incluindo participantes de diferentes faixas etárias e condições clínicas, para aumentar a generalização dos modelos.
- Investigar a aplicação de redes neurais profundas e técnicas de aprendizado por reforço para classificação de sinais de EEG.
- Avaliar a integração de outras modalidades de imagem e sinais biomédicos para diagnóstico multimodal do TEA.

- Desenvolver ferramentas de software baseadas em aprendizado de máquina que possam ser utilizadas em contextos clínicos por profissionais da área de saúde.
- Explorar o impacto do uso de EEG de alta densidade para melhorar a resolução espacial e a precisão dos modelos propostos.
- Investigar a viabilidade de utilizar modelos preditivos em tempo real para auxiliar na tomada de decisões clínicas durante avaliações de TEA.
- Ampliar o uso de técnicas de seleção de atributos para identificar padrões cerebrais específicos que possam diferenciar subtipos de TEA.
- Implementar estudos longitudinais que acompanhem pacientes diagnosticados com TEA para avaliar a evolução das análises de EEG ao longo do tempo e sua relação com intervenções terapêuticas.
- Com estas direções, espera-se que os avanços no campo contribuam ainda mais para o diagnóstico precoce e a melhoria da qualidade de vida de indivíduos com TEA e suas famílias.

REFERÊNCIAS

- ADITYA, S.; SHARMA, A.; KUMAR, V. **Applications of Naive Bayes in health informatics: a survey.** *Health Technology*, [S.l.], v. 12, p. 1–14, 2022. DOI: 10.1007/s12553-022-00601-7.
- AL-DEWIK, N. et al. **Overview and Introduction to Autism Spectrum Disorder (ASD).** *Advances in Neurobiology*, p. 3–42, 2020.
- ALMEIDA, F., & ROCHA, L. (2021). **Machine Learning in Clinical Decision Support Systems.** *Health Informatics Journal*.
- BEN-GAL, I. **Bayesian networks.** In: RUGGERI, F.; FALTIN, F.; KENETT, R. (Org.). *Encyclopedia of Statistics in Quality and Reliability*. Chichester: John Wiley & Sons, 2008. p. 1–6.
- BURBIDGE, R.; TROTTER, M.; BUISSERET, P.; BUZEK, J. **A comparison of feature selection methods and their application to computer aided diagnosis.** *Medical Informatics and the Internet in Medicine*, London, v. 26, n. 1, p. 25–35, 2001.
- CENTERS FOR DISEASE CONTROL AND PREVENTION. **Data & Statistics on Autism Spectrum Disorder.** 2020. Disponível em: <https://www.cdc.gov/ncbddd/autism/data.html>. Acesso em: 14 dez. 2024.
- CHEN, J. **Bayesian networks and decision graphs.** 2. ed. New York: Springer, 2012.
- CIN. Universidade Federal de Pernambuco. **Máquinas de Vetor de Suporte (SVM).** Disponível em: <www.cin.ufpe.br>. Acesso em: 12 de dezembro de 2024
- COS. Universidade Federal do Rio de Janeiro. **Redes Bayesianas.** Disponível em: <www.cos.ufrj.br>. Acesso em 12 de dezembro de 2024
- COSTA, J., & RIBEIRO, M. (2021). **Personalized Medicine through Machine Learning.** *Journal of Medical Informatics*.
- DIREITO, P., & Et al. (2021). **Artificial Intelligence and Its Applications in Modern Technology.** *Technology Review*.
- FAROOQ, M.S., TEHSEEN, R., SABIR, M. *Et al.* **Detection of autism spectrum disorder (ASD) in children and adults using machine learning.** *Sci Rep* **13**, 9605 (2023). <https://doi.org/10.1038/s41598-023-35910-1>
- FRIEDMAN, N., GEIGER, D. & GOLDSZMIDT, M. BAYESIAN NETWORK CLASSIFIERS. *Machine Learning* **29**, 131–163 (1997). <https://doi.org/10.1023/A:1007465528199>
- GRABRUCKER, A. M. (ED.). **Autism Spectrum Disorders.** [s.l.] Exon Publications, 2021.
- GUYTON, Arthur C. ; HALL, John E. . **Fundamentos de Fisiologia Médica:** Tradução da 13ª

- Edição. 13. ed. Rio de Janeiro: Elsevier Editora Ltda, 2017. 267 p. ISBN 978-1-4557-7006-9.
- HAYKIN, S. (1999). **Neural Networks – A Comprehensive Survey**. PrenticeHall, New Jersey, second edition.
- HELM, J. M., SWIERGOSZ, A. M., HAEBERLE, H. S., KARNUTA, J. M., SCHAFFER, J. L., KREBS, V. E., ... RAMKUMAR, P. N. (2020). **Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions**. *Current Reviews in Musculoskeletal Medicine*. doi:10.1007/s12178-020-09600-8
- HIROTA, T.; KING, B. H. **Autism Spectrum Disorder: A Review**. *Jama*, v. 329, n. 2, p. 157–168, 10 jan. 2023.
- HUS, Y.; SEGAL, O. **Challenges Surrounding the Diagnosis of Autism in Children**. *Neuropsychiatric Disease and Treatment*, v. 17, n. 17, p. 3509–3529, 3 dez. 2021.
- KOENIG, T.; PRICHEP, L.; LEHMANN, D.; VALDÉS-SOSA, P.; BRAEKEN, J.;
- HARMER, R. E. **Millisecond by millisecond, year by year: Normative EEG microstates and developmental stages**. *NeuroImage*, v. 16, p. 41-48, 2002.
- LI, J. et al. **Identification of autism spectrum disorder based on electroencephalography: A systematic review**. *Computers in Biology and Medicine*, v. 170, p. 108075–108075, 1 mar. 2024.
- LOPES DA SILVA, F. H.; NIEDERMEYER, E. **Electroencephalography: basic principles, clinical applications, and related fields**. 5. ed. Philadelphia: Lippincott Williams & Wilkins, 2005.
- MARTINS, A., & SOUSA, R. (2020). **Predictive Analytics in Healthcare: A Review of Machine Learning Applications**. *Journal of Healthcare Management*.
- MILOVANOVIC, M.; GRUJICIC, R. **Electroencephalography in Assessment of Autism Spectrum Disorders: A Review**. *Frontiers in Psychiatry*, v. 12, 29 set. 2021.
- MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.
- MMWR. **Surveillance Summaries**, v. 69, n. 4, 27 mar. 2020.
- NAPOLITANO, A. et al. **Sex Differences in Autism Spectrum Disorder: Diagnostic, Neurobiological, and Behavioral Features**. *Frontiers in Psychiatry*, v. 13, 13 maio 2022.
- NARROW, William E. *et al.* **Manual Diagnóstico e Estatístico de Transtornos Mentais: DSM-5**. 5. ed. Porto Alegre: Artmed, 2014. 96 p.
- OLIVEIRA, S., & CUNHA, P. **Machine Learning for Cancer Detection: Advances and Challenges**. *Oncology Reports*, 2019.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Autism**. 2022. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>. Acesso em: 14 dez. 2024.

PEARL, J. **Probabilistic reasoning in intelligent systems: networks of plausible inference.** San Francisco: Morgan Kaufmann, 1988.

RIBEIRO, T., & SILVA, M. (2020). **The Evolution of Machine Learning Algorithms.** Journal of Computer Science.

RÜPING, S. **Incremental learning with support vector machines.** Dortmund: University of Dortmund, 2001. (Technical Report, No. 2001-10).

SMOLA, A. J., BARLETT, P., SCHÖLKOPF, B., & SCHUURMANS, D. (1999). **Advances in Large Margin Classifiers.** The MIT Press. Disponível em: <<http://www.kernelmachines.org/nips98/lmc-book.pdf>>.

SOMVANSHI, M.; CHOUDHARY, A.; KUMAR, A. **A review of SVM and other machine learning algorithms.** *International Journal of Computer Applications*, New York, v. 133, n. 8, p. 7–10, 2016. DOI: 10.5120/ijca2016908561.

SOUZA, F., & ALVES, D. (2022). **Machine Learning in Chronic Disease Management: A Comprehensive Study.** Health Data Science.

VAPNIK, V. N. (1995). **The Nature of Statistical Learning Theory.** SpringerVerlag.
WORLD HEALTH ORGANIZATION. **Autism Spectrum Disorders.** Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>>.