



Universidade Federal do Rio de Janeiro - UFRJ
Centro de Ciências Matemáticas e da Natureza
Instituto de Geociências
Departamento de Meteorologia

**Caracterização e Previsão de Eventos Convectivos Severos na Região Amazônica
usando Algoritmos de Aprendizado de Máquina**

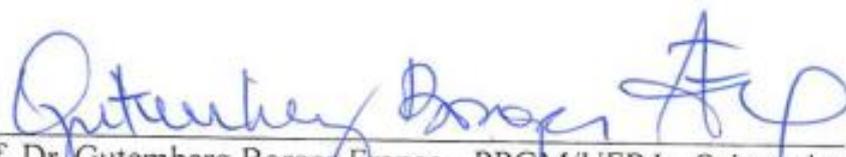
HUMBERTO DE CAMPOS BUENO

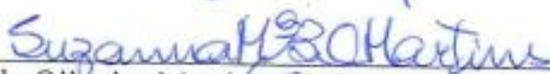
Rio de Janeiro, RJ
Outubro de 2024

**CARACTERIZAÇÃO E PREVISÃO DE EVENTOS CONVECTIVOS SEVEROS
NA REGIÃO AMAZÔNICA USANDO ALGORITMOS DE APRENDIZADO DE
MÁQUINA**

HUMBERTO DE CAMPOS BUENO

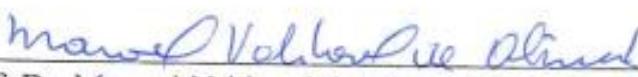
DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM METEOROLOGIA DO INSTITUTO DE GEOCIÊNCIAS DO CENTRO DE CIÊNCIAS MATEMÁTICAS E DA NATUREZA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (PPGM-IGEO-CCMN-UFRJ) COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA À OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS (ÁREA: METEOROLOGIA).


Prof. Dr. Gutemberg Borges França - PPGM/UFRJ - Orientador


Prof. Dra. Suzanna Maria Bonnet de Oliveira Martins - UFRJ/LMA - Coorientadora


Prof. Dr. Luiz Francisco Pires Guimarães Maia - PPGM/UFRJ


Prof. Dr. Hugo Abi Karam - PPGM/UFRJ


Prof. Dr. Manoel Valdonel de Almeida - LMA/UFRJ

RIO DE JANEIRO, RJ - BRASIL

OUTUBRO, 2024

AGRADECIMENTOS

Agradeço a minha amada família, meus pais Milton e Nilce, minhas amadas esposa Rosemeire e minha pequena filha Bárbara, a coisa mais importante da minha vida, pela compreensão e paciência durante os momentos em que não pude me dedicar a eles.

Ao orientador Prof. Gutemberg Borges França não só pelo direcionamento, dedicação e confiança depositada que permitiram o desenvolvimento deste trabalho, mas principalmente pela humanidade e compreensão que possui, por vezes orientando como se fora de pai para filho, realmente situações da vida.

À orientadora Profa. Suzanna Maria Bonnet de Oliveira Martins, pelos ensinamentos, incentivos transmitidos durante o curso e principalmente pela paciência e enorme ajuda ao longo da pesquisa, sem a qual esse trabalho não seria possível, incentivadora presente nos momentos de maior dificuldade.

Ao Departamento de Controle do Espaço Aéreo (DECEA) e ao Centro Integrado de Meteorologia Aeronáutica (CIMAER) por proporcionarem este Mestrado e acreditarem no caminho da capacitação profissional por meio da pesquisa.

À Universidade Federal do Rio de Janeiro (UFRJ) pela estrutura e oportunidade oferecida; um centro de excelência.

Ao Prof. Alan e ao Instituto Nacional de Pesquisas Espaciais (INPE) pela gentileza de disponibilizar os dados modelados dos índices de instabilidade termodinâmicos da Região Amazônica, para uma pesquisa futura.

Aos amigos do Laboratório de Meteorologia Aplicada (LMA), pelo convívio instrutivo e apoio constante.

Ao amigo Marcelo Pereira Nunes, meu primeiro colega de turma de Oficiais, de serviço no CIMAER e posteriormente de turma de Mestrado, pelas valiosas sugestões para a pesquisa.

E por fim, meus agradecimentos ao Professor Carlos Augusto Morales Rodriguez pela gentileza em fornecer e atualizar os dados de descargas atmosféricas extraídas da *Sferics Timing and Ranging Network* (STARNET), utilizados neste trabalho.

RESUMO

Este estudo apresenta uma ferramenta objetiva para previsão de tempestades convectivas severas (TSC) para a área Terminal de Manaus (TMA-Manaus), Amazonas, Brasil, utilizando algoritmos de Machine Learning (ML). A ocorrência e criticidade dos eventos de TSC foram caracterizados usando limiares de descargas atmosféricas (DA). Os modelos de ML utilizam limiares de DA como alvo e índices termodinâmicos extraídos da radiossonda das 12Z como preditores. A climatologia de AD entre 2012 e 2017 para a TMA-Manaus revelou que a DA ocorre em todos os meses do ano, e neste período ocorreram apenas 21 dias sem raios. Uma análise da importância das variáveis de entrada para classificar as TSC revelou que o índice Showalter, *Bulk Richardson*, *Convective Available Potential Energy*, Índice de Levantamento e Nível de Equilíbrio são os índices termodinâmicos mais relevantes para classificar o estado convectivo na região. Os resultados indicam que, para uma pequena quantidade de DA (69/dia), a média POD e FAR para os dez modelos selecionados foram $0,92 \pm 0,06$ e $0,19 \pm 0,01$, respectivamente. O algoritmo QDA apresentou o melhor desempenho com POD de 0,99 e FAR de 0,19. No entanto, como o aumento do limiar de DA até 5.000 DA por dia, foi observada uma diminuição no desempenho do modelo, pois a TSC mais crítica (com 5.000 DA) é rara em comparação com uma TSC com um limiar menor. Os resultados sugerem que os modelos de ML usando dados de radiossonda como preditores só são capazes de prever a ocorrência ou não de TSC com relativa precisão, mas os modelos não são capazes de classificar a criticidade. O desenvolvimento deste tipo de ferramenta marca um passo importante no sentido de antecipar e melhorar a previsão de TSC, melhorando a segurança e a eficiência dos voos na região.

Palavras-chave: tempestade convectiva; descargas atmosféricas; índices de instabilidade termodinâmicos; aprendizado de máquina; aviação; Região Amazônica.

Characterization and Forecasting of Severe Convective Events in the Amazon Region Using Machine Learning Algorithm

ABSTRACT

This study presents an objective tool for predicting severe convective storms (CS) in the Terminal Maneuverer Area of Manaus (TMA-Manaus), Amazon, Brazil, using Machine Learning (ML) algorithms. The occurrence and severity of CS events were characterized by atmospheric discharges (AD) using thresholds. The prediction models use AD thresholds as target and 12Z radiosonde data as predictors. The AD climatology between 2012 and 2017 for TMA-Manaus revealed that AD occurs in every month of the year, in this period there were only 21 days without lightning. The analysis of feature importance for classifying CS stages revealed that the Showalter index, Bulk Richardson Number, Convective Available Potential Energy, Lifted index, and Equilibrium Level are the most relevant thermodynamic indices for classifying the convective state in the region. Results indicate that, for a small amount of AD (69/day), the mean POD and FAR for the ten selected models were 0.92 ± 0.06 and 0.19 ± 0.01 , respectively. The QDA algorithm showed the best performance with a POD of 0.99 and a FAR of 0.19. However, as the AD threshold increased up to 5,000 AD per day, a decrease in model performance was observed, because Severe CS is rare compared to CS. The findings suggest that ML models using radiosonde data as predictors are only capable of predicting the occurrence or not of CS with relative accuracy, but the models are not capable of classifying whether it will be severe. The development of this tool marks a significant step towards improving the accuracy and timeliness of CS forecasts, thereby enhancing safety and efficiency of flights in the region.

keywords: convective storm; atmospheric discharges; thermodynamic instability indices; machine learning; Aviation impact; Amazon Region.

SUMÁRIO

1. INTRODUÇÃO	11
1.1 MOTIVAÇÃO.....	11
1.2 OBJETIVO	12
1.3 AMAZÔNIA	13
2. REVISÃO BIBLIOGRÁFICA	15
3. FUNDAMENTAÇÃO TEÓRICA	17
3.1 TEMPESTADES CONVECTIVAS SEVERAS	17
3.2 DESCARGAS ATMOSFÉRICAS.....	18
3.3 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS (IIT)	19
3.3.1 SHOWALTER (SH)	20
3.3.2 LEVANTAMENTO (LI).....	20
3.3.3 TOTAL TOTALS (TT)	21
3.3.4 SWEAT (SW)	22
3.3.5 ÍNDICE K	22
3.3.6 ÁGUA PRECIPITÁVEL (AP)	23
3.5 INTELIGÊNCIA ARTIFICIAL (IA)	24
3.5.1 SOFTWARE SCIKIT-LEARN	25
3.6 ÁREA TERMINAL DE AERÓDROMO (TMA).....	25
4. DADOS E MÉTODO	26
4.1 ÁREA DE ESTUDO	26
4.2 DADOS	27
4.2.1 DADOS DE DESCARGAS ATMOSFÉRICAS.....	27
4.2.2 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS	28
4.3 MÉTODO	28
4.3.1 DETALHES DO MÉTODO	28
4.3.2 ESTATÍSTICAS DE AVALIAÇÃO	30
5. RESULTADOS.....	31
5.1 ANÁLISE DOS DADOS.....	31

5.1.1	MODELOS DE ML: PREVISÃO DE TSC	31
5.1.2	DESCARGAS ATMOSFÉRICAS: CLASSIFICAÇÃO	36
6.	CONCLUSÃO	39
8.	REFERÊNCIAS.....	41
ANEXO I.....	48
ANEXO II.....	50

LISTA DE FIGURAS

Figura 1: Estágios de uma nuvem de tempestade.....	18
Figura 2: Representação da área de estudo	26
Figura 3: Gráfico de Dispersão de POD vs. FAR para 10 algoritmos de ML.....	34
Figura 4: Média e desvio padrão de POD e FAR para os 10 algoritmos de ML.....	35
Figura 5: Boxplot e Histograma ilustrando o DR-AD da TMA-Manaus	37
Figura 6: Distribuição de DA (%)	38
Figura 7: Densidade mensal de DA na TMA-Manaus.....	39

LISTA DE TABELAS

Tabela 3.1: Valores de SH ajustados para latitudes médias.....	20
Tabela 3.2: Valores do Índice de Levantamento.....	21
Tabela 3.3: Valores de TT para Tempestades	21
Tabela 3.4: Valores de SH ajustados para latitudes médias.....	22
Tabela 3.5: Valores de K para Tempestades	23
Tabela 4.1: Dados utilizados pela pesquisa	27
Tabela 4.2 Matriz de Confusão	30
Tabela 5.1: Discrepância percentual absoluta intra-variáveis para DR-AD > k,	33
Tabela 5.2: Desempenho dos 10 algoritmos ML	35

LISTA DE SIGLAS E ABREVIATURAS

AM - Aprendizado de Máquina

AP - Índice atmosférico termodinâmico denominado Água Precipitável

CGNA - Centro de Gerenciamento Nacional da Navegação Aérea

DA - Descarga Atmosférica

FAR - Probabilidade de Falsa Detecção ou alarme falso

FORTRACC – Forecasting and Tracking of Active Cloud Clusters

IA - Inteligência Artificial

IIT - Índices de Instabilidade Termodinâmicos

INPE - Instituto Nacional de Pesquisas Espaciais

K - Índice atmosférico termodinâmico K

LI - Índice atmosférico termodinâmico LIFT

NM – Milha Náutica

OACI - Organização de Aviação Civil Internacional

POD - Probabilidade de Detecção do Evento

RINDAT - Rede Integrada Nacional de Detecção de Descargas Atmosféricas

SBMN - Indicador de localidade da OACI para o Aeroporto de Manaus

SBEG - Indicador de localidade da OACI para o Aeroporto Eduardo Gomes

SH – Índice atmosférico termodinâmico SHOWALTER

SISCEAB - Sistema de Controle do Espaço Aéreo

SW - Índice atmosférico termodinâmico SWEAT

TMA - Área Terminal de Aeródromo

TT - Índice atmosférico termodinâmico TOTAL TOTALS

TSC - Tempestades Severas Convectivas

UTC - Tempo Universal Coordenado

1. INTRODUÇÃO

Vários fenômenos meteorológicos podem ter impacto na segurança das operações e na economia do transporte aéreo, incluindo neblina, turbulência, formação de gelo, baixa cobertura de nuvens, mudanças na pressão atmosférica, e especialmente Tempestades Severas Convectivas (TSC). Este último pode afetar significativamente todas as fases do voo, desde a decolagem, causando atrasos ou cancelamentos, nível de cruzeiro (causando necessidade de desvio das nuvens convectivas), até pouso, resultando em atrasos ou alterações no local de pouso.

O foco da Meteorologia Aeronáutica é melhorar a segurança e a eficiência do voo. Por meio do Departamento de Controle do Espaço Aéreo (DECEA), o Brasil é membro da Organização da Aviação Civil Internacional (OACI), e adota as recomendações no Anexo 3 da Convenção sobre Aviação Civil Internacional - Serviço Meteorológico de Navegação Aérea Internacional. O monitoramento e a previsão meteorológica para a aviação são gerenciados pelo Centro Integrado de Meteorologia Aeronáutica (CIMAER), que colabora com o Centro de Gerenciamento de Navegação Aérea (CGNA), ambos sediados no Rio de Janeiro-RJ, juntamente com empresas brasileiras representantes de companhias aéreas.

Com o objetivo de atender aos desafios da meteorologia aeronáutica, o DECEA e a Universidade Federal do Rio de Janeiro (UFRJ), especificamente o laboratório de Meteorologia Aplicada (LMA), têm colaborado através do Projeto da Cátedra de Meteorologia Aeronáutica. Esta parceria visa desenvolver estudos e implementar ferramentas para aplicações de previsões de curto a muito de curto prazo (*nowcasting*), melhorando a capacidade dos serviços de meteorologia aeronáutica em fornecer previsões oportunas e precisas para operações de voo. Vários resultados desta colaboração estão documentados em França, *et al.* (2023).

1.1 MOTIVAÇÃO

Na Região Norte do Brasil, particularmente na Região Amazônica, dois fatores chave influenciam significativamente as operações de voo em termos de segurança e economia de combustível: as distâncias entre aeroportos e o impacto das condições meteorológicas.

O Aeroporto Internacional Eduardo Gomes (SBEG), localizado em Manaus, está consideravelmente distante de outros grandes aeroportos urbanos, com distâncias aproximadas a Belém-PA (1.300 km), Boa Vista-RR (660 km), Porto Velho-RO (760 km) e Brasília-DF (1.930 km).

Além disso, a importância dos estudos na área de meteorologia aeronáutica para esta região é marcada pelas operações militares baseadas em Manaus, incluindo operações de aeronaves supersônicas no Aeroporto de Ponta Pelada (SBMN). A Base Aérea de Manaus (BAMN), criada em 1970, desempenhou um papel crucial no desenvolvimento da região, proporcionando apoio humanitário e contribuição para estratégias de defesa nacional. A experiência de pilotos e meteorologistas na BAMN apontam que a aviação militar é altamente sensível a condições climáticas adversas, particularmente precipitação convectiva e TSC.

A variabilidade da precipitação na Amazônia depende de fatores locais, como a floresta em si, e fatores remotos, como as circulações atmosféricas e oceânicas do Oceano Atlântico Tropical (por exemplo, a migração da Zona de Convergência Intertropical - ZCIT) e o Oceano Pacífico Tropical (no caso, o fenômeno El Niño Oscilação Sul - ENOS) (Marengo e Fisch 2021). Fatores antrópicos também devem ser considerados. Além disso, linhas de instabilidade frequentemente se desenvolvem na costa norte-nordeste (N-NE) da América do Sul e se propagam no interior em direção à Amazônia Central, causando precipitações apreciáveis (Ferreira e Cavalcanti 2022).

A prevalência de cisalhamento vertical do vento e outros fenômenos climáticos tropicais, como o Sistema de Monções da América do Sul (SMAS), influenciam significativamente as atividades convectivas e o desenvolvimento de tempestades na região. O SMAS é um fenômeno climático significativo que afeta padrões climáticos em grande parte da América do Sul, particularmente nas regiões tropicais e subtropicais.

Atualmente, a previsão de TSC para a Área Terminal (TMA) de Manaus (TMA-Manaus) envolve certo nível de subjetividade por parte dos meteorologistas, que dependem dos dados disponíveis como perfis atmosféricos, dados de satélite e modelos numéricos, e principalmente da experiência, observação e vivência do previsor com relação ao local.

1.2 OBJETIVO

Este estudo tem como objetivo desenvolver uma ferramenta objetiva para a previsão de TSC para a TMA-Manaus utilizando modelos de aprendizado de máquina

(ML). Dados de descargas atmosféricas (DA) detectadas por detector de raios que são utilizadas para determinar a localização e severidade da ocorrência de TSC, e índices termodinâmicos extraídos dos dados de radiossondagem de SBMN que aqui são usados como preditores dos modelos. Espera-se que esta ferramenta ajude melhorar as previsões de TSC, melhorando a segurança e a eficiência das operações de voo na região.

1.3 AMAZÔNIA

A Amazônia ocidental é composta pelos estados do Amazonas (o maior estado em área da Região Norte), Roraima, Acre e Rondônia, onde estão 42,97% da área da Amazônia legal e 57% da floresta amazônica brasileira; são estados que se destacam pelo extrativismo e pecuária, com foco também no potencial industrial. Boa Vista, Manaus e Porto Velho são as principais cidades da Amazônia ocidental. Boa Vista possui dois regimes de clima bem definidos, chuvoso de abril a setembro e seco de outubro a março (SUFRAMA, 2021).

A Amazônia oriental compreende o restante dos estados da região, local em que os padrões climatológicos apresentam regime de precipitação chuvoso entre janeiro e maio, regime seco de julho a novembro, e apresentam variabilidade espacial pronunciada, com características climáticas. A ZCIT e Zona de Convergência do Atlântico Sul (ZCAS) são os principais sistemas meteorológicos do regime chuvoso; já os aglomerados convectivos, linhas de instabilidade e as bandas frontais posicionadas no nordeste brasileiro são os principais sistemas atuantes no regime seco (Souza *et al.*, 2017).

Para Mendonça *et al.* (2007), o clima que predomina na Região Norte do Brasil, coincidente com o predomínio da floresta Amazônica e controlado por sistemas atmosféricos equatoriais como as Massas Equatoriais (Continental e Atlântica), a ZCIT, bem como sistemas tropicais, são denominados de “Climas de Latitudes baixas”, “Clima Equatorial Úmido da Frente Intertropical”, ou ainda, “Domínio Climático Quente”. O clima é quente e de homogeneidade térmica, pequena amplitude térmica diária e sazonal em decorrência da elevada umidade e da intensa nebulosidade, com média anual entre 24°C e 26°C; setembro e outubro são os meses mais quentes do ano, quando a temperatura pode atingir 40°C. Apresenta também homogeneidade pluviométrica, tanto espacial quanto temporal.

Três são os subtipos desse clima:

1. Clima equatorial Superúmido (sem seca): predomina no extremo oeste do estado do Amazonas. Clima com elevadas temperaturas e umidade/pluviosidade o ano inteiro; São Gabriel da Cachoeira-AM é uma localidade em que predomina esse clima;

2. Clima Equatorial com Subseca (1 a 2 meses secos): predomina na parte central e nordeste da Região Norte do Brasil. Apresenta elevada temperatura ao longo do ano inteiro, e pluviosidade reduzida (Subseca) em até 2 meses do ano. Nesse clima estão as duas metrópoles da Amazônia: Manaus apresenta homogeneidade térmica anual, e subseca em agosto (mês de maior temperatura) e maior pluviosidade em março e abril; Belém apresenta maior regularidade térmica que Manaus, com seis meses chuvosos (destaque para março) e 6 meses de redução nas chuvas (junho a novembro). Belém recebe também a influência da maritimidade.

3. Clima Equatorial com Subseca (3 meses secos): Nesse subtipo vemos claramente o efeito da continentalidade, e possui índices térmicos e pluviométricos bastante elevados. Porto Velho (RO) possui clima quente e boa regularidade térmica anual, com queda de junho a agosto (os meses menos chuvosos, com destaque para julho); já o verão é bastante chuvoso.

Para Marengo e Fisch (2021), a Amazônia tem por representatividade ser uma das principais áreas verdes do planeta, e que vêm sofrendo importantes desequilíbrios, provocando a degradação do seu solo e alterações de seus ecossistemas, que é decorrente do desmatamento provocado pelas atividades agrícolas e de pecuária.

A floresta amazônica é importante no equilíbrio dinâmico entre sua vegetação e o clima, e reciclagem dos processos de vapor d'água. Ainda, de 30 a 50% das precipitações pluviométricas da bacia amazônica consistem em evapotranspiração reciclada pela vegetação. A Amazônia ainda é responsável por parte das precipitações no sul e sudeste do Brasil, através dos chamados "rios voadores" ou "jatos de baixos níveis". Os ventos alísios transportam umidade do Atlântico Equatorial para a Amazônia, devido a um aumento de pressão no verão e outono, podendo configurar-se em um jato de baixos níveis (850 hpa) pelo deslocamento de seu fluxo para o sudeste, após o encontro com os Andes.

Em invernos em anos de El Niño (La Niña), observam-se anomalias negativas (positivas) de precipitação dominantes no Norte do Brasil e que são persistentes até o outono do ano seguinte.

Segundo Cavalcanti (2021), basicamente a variabilidade interanual do clima na Amazônia é influenciada por 3 fatores: o fenômeno ENOS, a migração latitudinal da ZCIT e as circulações atmosféricas e oceânicas do Atlântico tropical; além disso, os fatores

antrópicos também devem ser considerados. Ainda, para Ferreira (2022), Linhas de Instabilidade (LIs) desenvolvem-se frequentemente na costa norte-nordeste (N-NE) da América do Sul e podem se propagar para o interior do continente, causando quantidades apreciáveis de precipitação; como esses sistemas têm uma escala temporal associada ao ciclo diurno (brisa marítima e aquecimento terrestre), sua máxima atividade convectiva pode ser notada nas imagens de satélite no final da tarde ou no dia posterior à sua formação, quando estão se propagando em direção à Amazônia Central.

Alguns projetos que trabalham com dados na Amazônia são:

1. Experimento de Biosfera-Atmosfera em Grande Escala na Amazônia (LBA): "LBA Amazônia" refere-se ao "LBA" (LBA - *Large-Scale Biosphere-Atmosphere Experiment in Amazonia*), que é uma iniciativa de pesquisa significativa que visa entender as interações biosfera-atmosfera na floresta amazônica; esse projeto envolve vários estudos, incluindo clima, ecologia e efeitos do uso da terra, para reunir dados sobre como as mudanças na Amazônia impactam os padrões climáticos globais (KELLER *et al.*, 2004), e;

2. Projeto *Green Ocean Amazon* (GOAmazon): tem como objetivo desenvolver pesquisas relacionadas com a dinâmica da floresta e sua interação com a atmosfera, envolve experimentos e modelos climáticos e tem o Instituto Nacional de Pesquisas da Amazônia (INPA) como parceiro do projeto. É financiado pelo Governo do Amazonas via Fundação de Amparo à Pesquisa do Estado do Amazonas (Fapeam), bem como a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) e o "Departamento de Energia dos Estados Unidos da América" (DOE/EUA) (INPA, 2024).

2. REVISÃO BIBLIOGRÁFICA

House e Beebe (1955) descreveram métodos gerais para previsão de TSC, visando familiarizar os previsores com gráficos, ferramentas, métodos, parâmetros, definições e técnicas usadas em operações de previsão do tempo.

Mueller *et al.* (2003) propuseram um sistema de previsão com prazo de até 1 hora para localização de TSC, utilizando observações de superfície, dados de radar meteorológico, dados de satélite e modelos de previsão numérica do tempo (NWP), juntamente com as etapas de desenvolvimento de tempestade.

Nascimento (2005) apresentou uma metodologia para previsão de condições climáticas severas, incluindo TSC, no Brasil, estudando variáveis termodinâmicas de

radiossondas e resultados de um modelo NWP de mesoescala. Este estudo mostrou que certos parâmetros utilizados em previsões de latitudes médias para o Hemisfério Norte podem ser aplicados a sistemas de previsão no Brasil.

Isaac *et al.* (2006) desenvolveram um sistema de previsão para a aviação no Canadá, integrando dados de modelos NWP, observações de superfície, radar meteorológico e dados de satélite, e um radiômetro de micro-ondas para gerar previsões meteorológicas com prazo de até 6 horas para os principais aeroportos .

Queiroz (2009) criou um índice de severidade para previsão de curto prazo (*nowcasting*) da SC utilizando dados de radar meteorológico e o sistema FORTRACC.

França, *et al.* (2016) desenvolveram um modelo de rede neural para prever locais de ocorrência de instabilidade atmosférica significativa na TMA-Rio de Janeiro, Brasil, utilizando observações de superfície, radiossondas, dados de precipitação e AD.

Paulucci (2017) analisou a distribuição espacial e temporal das DA nuvem-solo na Região Metropolitana do Rio de Janeiro (RMRJ) no período de 2000 a 2016, identificando um pico de concentração entre 18h e 19h (horário local).

Hermisdorff (2018) tentou aumentar a previsibilidade do TSC na RMRJ estudando a predição de DA a partir de árvores de decisão e dados de múltiplas fontes, incluindo a rede RINDAT, observações de superfície do Aeroporto Internacional do Rio de Janeiro (SBGL), e variáveis termodinâmicas de radiossondas de SBGL.

Gulpepe *et al.* (2019) destacaram o reconhecido impacto dos processos atmosféricos na aviação desde 1900.

Bonnet, *et al.* (2020) exploraram a usabilidade de um método de previsão de imagens de radar meteorológico usando algoritmos de *Deep Learning* para auxiliar na previsão qualitativa da precipitação com horizonte de previsão de até 1 hora.

Soares *et al.* (2021) propuseram dois modelos de previsão de TSC com horizonte de 6 a 8 horas, usando índices termodinâmicos extraídos do GOES-R e modelos ML para a rota aérea entre os estados do Rio de Janeiro e São Paulo.

De Castro *et al.* (2022) desenvolveram uma ferramenta usando modelos de ML e dados de DA para prever TSC na TMA-São Paulo com até 5 horas de antecedência, determinando o local, período e gravidade do TSC.

Almeida, *et al.* (2020) e Nunes, *et al.* (2023) utilizaram dados de DA e radiossondagem para prever TSC na TMA-Rio de Janeiro e TMA-Brasília, respectivamente, usando modelos de ML.

Com base nesta última pesquisa, o presente estudo utiliza dados de DA para caracterizar dados de TSC e radiossonda de SBMN para treinar algoritmos de ML

3. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta uma fundamentação teórica, abordando os assuntos desta dissertação: como tempestades convectivas severas, descargas atmosféricas, índices de instabilidade termodinâmicos, Inteligência Artificial e a Área Terminal de Aeródromo, conforme descritos nas seções a seguir.

3.1 TEMPESTADES SEVERAS CONVECTIVAS (TSC)

Segundo Ynoue *et al.* (2017), uma tempestade é um fenômeno formado por uma única nuvem cumulonimbus (Cb) com grande extensão vertical ou por um aglomerado de nuvens desse tipo. Quando a tempestade é severa, produz muita chuva, raios, trovões, ventos fortes, granizo e, às vezes, tornados. Alguns dos fenômenos de tempo mais perigosos, como tornados, podem ser produzidos por uma supercélula, que é uma única nuvem Cb, porém que pode ser maior do que um aglomerado de nuvens desse tipo, e que pode se estender a mais de 20 km na atmosfera.

O ciclo de vida de uma tempestade é composto de três fases: estágio de cumulus, estágio maduro e estágio de dissipação; no primeiro estágio, o aquecimento da superfície favorece a ocorrência de movimentos ascendentes, que vão contribuir para a formação de nuvens Cb. Nesse estágio, o ar dentro da nuvem é dominado por correntes ascendentes (YNOUE *et al.*, 2017).

Ainda de acordo com Ynoue *et al.* (2017), no estágio maduro, começam a ocorrer, ao lado das correntes ascendentes, correntes descendentes de ar que saem da base da nuvem, bem como precipitação. O estágio de dissipação acontece quando as correntes descendentes de ar dominam a nuvem. Note-se que a própria nuvem se dissipa, pois as correntes descendentes inibem a ascensão de ar úmido, que é o seu “combustível”.

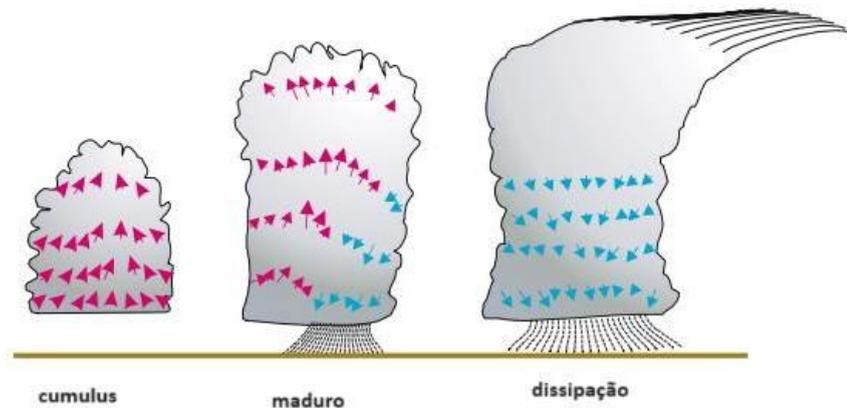


Figura 1 – Estágios de uma nuvem de tempestade.
 Fonte: Varejão-Silva (figura adaptada pelo autor).

Para Nascimento (2005), as tempestades severas ocorrem com baixa frequência, em comparação com outros sistemas de meteorologia (por exemplo, chuvas decorrentes de sistemas frontais) e geralmente podem ocasionar sérios riscos para a população e para a economia. Já para Ferreira (2006), o aquecimento diferencial da superfície terrestre é o principal mecanismo para o desenvolvimento de tempestades, portanto as nuvens Cb se desenvolvem quando a atmosfera está suficientemente instável devido ao aquecimento da superfície ou outro mecanismo atmosférico.

As nuvens Cb são as nuvens de tempestade e que podem produzir todas as formas de precipitações, incluindo grandes gotas de chuvas, granizos, pancadas fortes, raios e rajadas violentas; no interior dessas nuvens ocorrem intensas correntes ascendentes e descendentes e, devido a elas, pode-se encontrar granizo em quase toda a sua extensão, como também, em seu entorno as rajadas de vento podem atingir valores de aproximadamente 54 nós (100 km/h), a uma distância de até 12 km além da nuvem.

Para Tory e Thurston (2015), uma nuvem pirocumulus é uma nuvem cumuliforme densa associada ao fogo ou à atividade vulcânica; é produzida por um intenso aquecimento do ar.

Laurel (2016) foi um dos primeiros pesquisadores que apresentou uma evidência de que o nível de condensação nas chamadas nuvens pirocumulonimbus pode ser significativamente mais elevado do que o nível de condensação por elevação padrão do ambiente, e que o perfil termodinâmico ambiental, as variações diárias de umidade e o cisalhamento do vento exercem influência significativa sobre o início e o desenvolvimento de nuvens piroconvectivas.

3.2 DESCARGAS ATMOSFÉRICAS (DA)

Descargas atmosféricas são descargas elétricas de grande extensão (alguns

quilômetros) e de grande intensidade (picos de intensidade de corrente acima de um kiloampere); ocorrem devido ao acúmulo de cargas elétricas em certas regiões localizadas da atmosfera, em geral dentro de tempestades (ELAT, 2023).

Ainda, a descarga inicia quando o campo elétrico produzido por estas cargas excede a capacidade isolante, também conhecida como rigidez dielétrica, do ar em um dado local na atmosfera, que pode ser dentro da nuvem ou próximo ao solo; e quebrada a rigidez, tem início um rápido movimento de elétrons de uma região de cargas negativas para uma região de cargas positivas (ELAT, 2023).

O raio é um fenômeno elétrico dentro das nuvens Cb que pode atingir até 12 Km de altura e vários quilômetros de diâmetro. As tempestades com trovoadas ocorrem quando determinadas condições particulares (temperatura, pressão, umidade do ar, velocidade dos ventos, entre outras) afetam um tipo específico de nuvem. O mecanismo de produção de cargas elétricas cresce de tal maneira que gera uma onda elétrica (raio) que parte da base da nuvem em direção ao solo, buscando locais com menor potencial, resultando em uma trajetória ramificada e aleatória. Esta primeira onda atinge entre 20 e 100 m do solo, então o primeiro raio entre a nuvem e o solo permitirá a passagem de uma quantidade de cargas com corrente de pico em torno de 20.000 Amperes; após isso, ocorre um violento aquecimento do meio atmosférico, de até 30.000°C, provocando assim a expansão do ar (TROVÃO). Após isso, os elétrons retornam, fazendo com que a energia seja devolvida sob a forma de luz (RELÂMPAGO). As descargas atmosféricas podem ser de 3 formas, principalmente: ascendentes (da terra para a nuvem), descendentes (da nuvem para a terra) ou ainda entre nuvens (UNESP, 2000).

3.3 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS (IIT)

Os Índices de Instabilidade Termodinâmicos (IIT) são valores determinados a partir dos perfis verticais de temperatura e umidade da atmosfera obtidos usando as sondagens atmosféricas (radiosondagens). O grau de instabilidade atmosférica pode ser identificado por meio de diagramas termodinâmicos específicos (chamados de SKEW-T), analisando-se as curvas de estado da atmosfera em comparação à evolução de uma parcela de ar, levando-se em conta os gradientes adiabáticos.

Dessa forma, os IIT foram idealizados para traduzir numericamente o grau de instabilidade atmosférica e auxiliam na determinação do potencial convectivo da atmosfera, sendo utilizados como uma das principais ferramentas para os prognósticos do TSC, a exemplo do estudo desenvolvido por Mourão e Menezes (2006) na previsão de tempestades convectivas no Estado do Rio de Janeiro.

Nesta pesquisa os IIT foram utilizados como variáveis preditoras dos algoritmos de ML. A seguir são descritos sucintamente seis IIT: SHOWALTER (SH), LIFT (LI), TOTAL TOTALS (TT), SWEAT(SW), K e Água Precipitável (PA).

3.3.1 SHOWALTER (SH)

O método de SH é empregado para avaliar o potencial e a intensidade de uma tempestade. O SH foi um dos primeiros índices criados, utilizando a temperatura do ar e o ponto de orvalho em 850 hPa e a temperatura do ar em 500 hPa (SHOWALTER, 1947)

Quanto mais negativos forem os valores deste índice, maior será o potencial da atmosfera para o surgimento de tempestades.

O SH é calculado a partir da Eq. 1,

$$SH = T_{500} - T_{parcela}; \quad (1)$$

onde, T_{500} representa a temperatura do ar (em graus Celsius) em 500 hPa e $T_{parcela}$ representa a temperatura em 500 hPa (em Celsius) de uma parcela que subiu desde o nível de 850 hPa.

Os valores utilizados para a classificação da possibilidade de ocorrências da formação de tempestades são descritos na Tabela 3.1.

Tabela 3.1: Valores de SH ajustados para latitudes médias. Fonte: (SHOWALTER, 1947).

SH	Possibilidade de Tempestades
SH > +3	Sem Atividade Convectiva
+1 ≤ SH ≤ +3	Possíveis Pancadas de Chuva/Tempestades Isoladas
-2 ≤ SH ≤ +1	Tempestades Prováveis
-6 ≤ SH ≤ -2	Possibilidade de Tempestades Severas
SH ≤ -6	Tempestades Severas Prováveis/Possibilidade de Tornados

3.3.2 LEVANTAMENTO (LI)

O LI, do inglês *LIFT*, é determinado pela diferença da temperatura do ar e da temperatura de uma parte de ar levantada adiabaticamente a uma determinada altura de pressão na troposfera, geralmente 500 hPa. Se o valor for

positivo, a atmosfera (na altura apropriada) permanece estável; se o valor for negativo, a atmosfera é instável. Quanto menor for esse índice, maior será o risco de ocorrência de tempestades.

O LI é representado na Eq. 2:

$$LI = T_{500} - T_{p500} [^{\circ}C]; \quad (2)$$

onde, T é a temperatura do ambiente e T_p é a temperatura da parcela em 500 hPa. No LI a parcela é levantada da superfície. Este é definido na Tabela 3.2.

Tabela 3.2: Valores do Índice de Levantamento (LI) (GALWAY, 1956).

LI	Nível de Instabilidade
$0 \leq LI \leq +3$	Estável. Possível convecção fraca na presença de forte levantamento ou algum mecanismo forçante
$-3 \leq LI \leq 0$	Marginalmente Instável
$-6 \leq LI \leq -3$	Marginalmente Instável
$-9 \leq LI \leq -6$	Muito Instável
$LI < -9$	Extremamente Instável

3.3.3 TOTAL TOTALS (TT)

TT é usado para analisar o entranhamento de ar frio na troposfera média, sendo importante para prever tempestades severas até com a ocorrência de granizo.

O TT pode ser calculado a partir da Eq.3:

$$TT = T_{d850} + T_{850} - 2(T_{500}); \quad (3)$$

onde T é a temperatura do bulbo seco e T_d a temperatura do ponto de orvalho (MILLER, 1972).

A representação dos valores do índice TT é apresentada conforme a Tabela 3.3.

Tabela 3.3: Valores de TT para Tempestades (MILLER, 1972).

TT	Potencial para Tempestades
$40 \leq TT \leq 45$	Isoladas, algumas poucas moderadas

$46 \leq TT \leq 47$	Esparsas, algumas poucas pesadas
$48 \leq TT \leq 49$	Esparsas moderadas, algumas pesadas e algumas isoladas severas
$50 \leq TT \leq 51$	Esparsas pesadas, algumas severas e tornados isolados
$52 \leq TT \leq 55$	Esparsas a numerosas pesadas, poucas a esparsas pesadas, alguns tornados
$TT > 55$	Esparsas a numerosas pesadas, poucas a esparsas pesadas, alguns tornados

3.3.4 SWEAT (SW)

Utiliza dados sobre as estruturas termodinâmica e cinemática da atmosfera. Sua utilização é indicada para o diagnóstico do potencial de tempestades severas. Valores inferiores a 300 são considerados indicativos de condições propícias para tempestades severas, enquanto os valores inferiores a 400 indicam maior probabilidade de eventos de tornados (MILLER, 1972)

O SW é calculado a partir da umidade nos baixos níveis, do ponto de orvalho em 850 hPa, da instabilidade, pelo índice TT, pelo jato de baixos níveis, obtido do vento em 850 hPa, pelo jato de altos níveis, obtido pelo vento em 500 hPa e pela advecção quente, entre 850 e 500 hPa. A Eq. 4 apresenta o SW de forma matemática..

$$SW = 12Td_{850} + 20(TT-49) + 2V_{850} + V_{500} + 125f(C); \quad (4)$$

onde, $C = \text{sen}(dir_{500} - dir_{850}) + 0,2$, dir_{500} representa a direção do vento em 500 hPa e dir_{850} representa a direção do vento em 850 hPa.

Tabela 3.4: Valores de SW para latitudes médias (MILLER, 1972).

SW	Potencial para Tempestades Severas
$SW \leq 300$	Sem potencial para tempestades severas
$300 \leq SWEAT \leq 400$	Potencial para tempestades severas
$SWEAT > 400$	Potencial para tornados

3.3.5 ÍNDICE K

Este expressa o potencial para a ocorrência de tempestades e fundamentando-se na taxa de variação vertical da temperatura, no conteúdo de umidade na baixa troposfera e na extensão da camada úmida.

O índice K é definido como na Eq. 5 (GEORGE, 1960):

$$K = (T_{850} - T_{500} + T_{d850} - (T_{700} - T_{d700})); \quad (5)$$

onde T_{850} é temperatura do nível de 850 hPa, T_{d850} expressa a temperatura do ponto de orvalho do nível de 850 hPa, T_{700} representa a temperatura do nível de 700 hPa, T_{d700} indica a temperatura do ponto de orvalho do nível de 700 hPa e T_{500} representa a temperatura do nível de 500 hPa ($^{\circ}\text{C}$);

Segundo Henry (1987), o índice K com valores abaixo de 30 são considerados com grande potencial para a ocorrência de tempestades e, acima de 40, extremamente altos.

O termo "K" é empregado para prever trovoadas de massas de ar, que ocorrem como resultado do movimento vertical de uma parte de ar, bem como devido ao aquecimento local ou ao efeito orográfico que essa parcela sofre ao subir a encosta de uma montanha. A tabela 3.5 apresenta os níveis de K para tempestades.

Tabela 3.5: Valores de K para Tempestades (GEORGE, 1960).

K	Possibilidade de Tempestades
$K < 20$	Sem Atividade Convectiva
$20 < K < 25$	Tempestades Isoladas
$25 < K < 30$	Tempestades amplamente dispersas
$30 < K < 35$	Tempestades dispersas
$35 < K < 40$	Muitas Tempestades

3.3.6 ÁGUA PRECIPITÁVEL (AP)

O índice "AP" representa a quantidade de vapor de água atmosférico contido em uma coluna vertical de área transversal unitária, medindo o volume líquido e supondo a condensação de todos os valores de água medidos; é expresso pela Eq. 6, onde p representa o nível de pressão atmosférica e g a aceleração da gravidade,

$$AP = \frac{1}{\rho g} \int_{p_1}^{p_2} x dp; \quad (6)$$

onde, $x(p)$ é a razão de mistura no nível de pressão (p) e AP representa o vapor de água precipitável, contido em uma camada limitada pelas pressões p_1 e p_2 .

3.4 INTELIGÊNCIA ARTIFICIAL (IA)

A Inteligência Artificial (IA) é uma tecnologia computacional ou um conjunto de tecnologias, tais como redes neurais artificiais, algoritmos e sistemas de aprendizado, com o objetivo de imitar as habilidades mentais humanas, tais como raciocínio, percepção do ambiente e tomada de decisões. A ideia é que máquinas possam resolver diversos problemas, desde os mais complexos até os mais simples, utilizando uma grande quantidade de dados (ICMC, 2021).

De acordo com Munakata (1994), há diversas definições distintas para o conceito de inteligência artificial, principalmente devido às diversas possibilidades de uso. Segundo o dicionário World College Dictionary, IA é o estudo dos sistemas computacionais que executam tarefas que os seres humanos necessitam de inteligência para realizar. Conforme Haykin (2002), a Inteligência Artificial foi criada com o objetivo de simular a execução de tarefas cognitivas semelhantes às realizadas por seres humanos.

No início da Segunda Guerra Mundial, vários pesquisadores começaram a pesquisar a Inteligência Artificial para simular a maneira como os seres humanos pensam e tomam decisões. Com a evolução da IA, foram desenvolvidas ferramentas computacionais para aplicação do *Machine learning* (ML), que se constitui do aprendizado de padrões por intermédio de sistemas computacionais, permitindo que os computadores aprendam automaticamente “sem” intervenção humana.

Yao (2013) aponta que o ML é um dos setores mais importantes da Inteligência Artificial dedicados à criação de modelos computacionais de aprendizagem. O ML pode ser usado em diversas áreas, como: ciências cognitivas, ciência da computação, estatística, complexidade computacional, teoria da informação, teoria do controle, filosofia e biologia. O ML, de forma computacional, se refere à capacidade de uma máquina melhorar seu desempenho com base em resultados anteriores. O ML é o estudo de como criar computadores que aprendem com a experiência e modificam suas tarefas de acordo com esse aprendizado, ao contrário dos computadores tradicionais, que não mudam a menos que os programadores realizem as mudanças de forma explícita.

Um estudo feito por Grewal (2014), diferentes definições de IA em todas as perspectivas sobre o tema e suas aplicações ao longo do tempo foram feitas; e, dessa forma, definiu que a IA tem como objetivo desenvolver técnicas para resolver problemas, simulando o cérebro humano, através de sistemas e algoritmos computacionais.

Para o uso de ML, existem recursos que simplificam a identificação de padrões, como as redes neurais artificiais, que são sistemas computacionais com nós interconectados que funcionam como os neurônios do cérebro humano; já para o treinamento dos modelos preditivos, esta pesquisa utilizou os algoritmos de ML classificatórios, que funcionam como ferramenta para o reconhecimento de padrões, e que são um modelo prático para a inferência indutiva.

3.4.1 SOFTWARE SCIKIT-LEARN

Neste trabalho foi utilizado o scikit-learn, que é uma biblioteca da linguagem Python e uma das melhores opções para aplicação prática de ML, desenvolvida especificamente para aplicação prática de ML. Esta biblioteca dispõe de ferramentas simples e eficientes para análise preditiva de dados, é reutilizável em diferentes situações, possui código aberto, sendo acessível a todos e foi construída sobre os pacotes NumPy, SciPy e matplotlib.

Um dos fatores que traz destaque a linguagem são, justamente, suas bibliotecas e pacotes, que proporcionam muita simplicidade às aplicações, além de garantir *scripts* descomplicados e eficientes. Dentre estes pacotes, temos o NumPy e o Pandas como os principais para a preparação dos dados, e o scikit-learn, ou apenas sklearn, sendo o mais utilizado para efetiva criação de modelos de ML. O scikit-learn está organizado em muitos módulos, cada um desenvolvido para uma finalidade específica. Nestes módulos encontram-se funções para as mais diferentes aplicações.

3.5 ÁREA TERMINAL DE AERÓDROMO (TMA)

Em aviação, TMA é definida como área de controle terminal, por designar uma região do espaço aéreo, normalmente situada na confluência de rotas de tráfego aéreo e nas imediações de um ou mais aeródromos, onde haja um alto volume de tráfego aéreo (DECEA, 2024).

As TMA, que têm limites laterais e verticais variáveis, são áreas críticas

Esta área abrange a cidade de Manaus, incluindo o Aeroporto Internacional Eduardo Gomes (SBEG) e o Aeroporto Militar de Ponta Pelada (SBMN). A área considerada é representada pelo quadrado circunscrito ao círculo TMA (Figura 2), entre as latitudes de 02° a 04° S, e longitudes de 59° a 61° W.

4.2 DADOS

Os dados utilizados são de duas fontes diferentes:

- 1) Dados de descargas atmosféricas (DA) Sferics Timing and Ranging Network - STARNET (seção 4.2.1): compreende janeiro de 2012 a dez-2017, totalizando 2.192 dias;
- 2) Índices de instabilidade provenientes dos dados de radiossondagem do aeroporto SBMN (seção 4.2.2): compreende junho de 2012 a julho de 2016, totalizando 1.216 dias devido a ausência de radiossondas em vários dias, inclusive o período de 23 de fevereiro de 2015 a junho de 2015.

Assim, para utilização em modelos de ML, o período dos dados foi equalizado. A tabela 4.1 sumariza os dados utilizados.

Tabela 4.1: Dados utilizados pela pesquisa.

Fonte	Frequência	Qtd	Dados	Finalidade	Período	Área
STARNET (www.starnet.iag.usp.br)	variável (~ms)	2,422,173	Localização (lat., lon.), tempo ocorrência e quantidade de AD	Determinar a ocorrência de TSC e estudar a sua climatologia	2012-2017	Quadrado circunscrito ao círculo da TMA na Fig. 1
Radiosonda SBMN (weather.uwyo.edu/)	uma vez ao dia às 12Z	1,216	24. índices termodinâmicos	Variáveis Predictoras de TSC em algoritmos de ML	2012-2016	Localização : 59.98°W, 3.15°S

4.2.1 DADOS DE DESCARGAS ATMOSFÉRICAS (DA)

Os dados utilizados neste trabalho compreendem duas fontes (Tabela 4.1):

DA detectadas pela rede *Sferics Timing and Ranging NETwork* (STARNET) (Morales *et al.*, 2002), disponível em www.starnet.iag.usp.br. A STARNET consiste em um conjunto de antenas receptoras que operam na frequência VLF (7 - 15 kHz) e detectam ruídos de rádio emitidos pelas DA na atmosfera. A partir disso, o sistema localiza

a DA dentro da área de cobertura da antena usando o método de diferença do horário de chegada.

A rede STARNET possui antenas receptoras instaladas em diversas regiões do Brasil, América Central e países africanos. Neste estudo, foi utilizada a antena baseada no Amazonas. Os dados têm frequência variável, pois depende da ocorrência de raios, portanto a detecção pode ocorrer em uma escala de milissegundos. Para identificar a ocorrência e classificar a gravidade da TSC, os dados de DA foram acumulados por dia, denominados Taxa Diária de DA (TX-DA);

4.2.2 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS

Os dados da radiossonda da estação de altitude de SBMN estão disponíveis duas vezes ao dia, 00Z e 12Z. Para avaliar o clima em cenário pré-convectivo, apenas a radiossondagem das 12Z foi considerada. A hipótese é que os índices termodinâmicos podem capturar o estado atmosférico local e, portanto, podem ser usados como preditores para algoritmos de ML para gerar previsões de TSC para a TMA-Manaus.

Foram utilizados 24 índices termodinâmicos das radiossondagens das 12Z: índice Showalter (SHOW), índice Lifted (LIFT), LIFT calculado usando temperatura virtual (LFTV), índice SWEAT (SWET), índice K (KINX), Cross Totals (CTOT), Índice Vertical Totals (VTOT), Índice Total Totals (TTOT), Energia Potencial Convectiva Disponível (CAPE), CAPE usando temperatura virtual (CAPV), Inibição Convectiva (CINS), CINS usando temperatura virtual (CINV), Nível de Equilíbrio (EQLV), Nível de Equilíbrio usando temperatura virtual (EQTV), Nível de Convecção Livre (LFCT), LFCT usando temperatura virtual (LFCV), Bulk Richardson Number (BRCH), Bulk Richardson Number usando CAPV (BRCV), Temp [K] do nível de condensação por elevação (LCLT), Pres [hPa] do Nível de Condensação de Elevação (LCLP), Temperatura Potencial média da camada mista (MLTH), proporção média de mistura de camada mista (MLMR), Espessura de 1000 hPa a 500 hPa (THTK) e Água precipitável [mm] para toda a sondagem (PWAT).

4.3 MÉTODO

4.3.1 DETALHES DO MÉTODO

As etapas do método utilizado estão descritas a seguir:

a) Climatologia de descargas atmosféricas (DA): Foram analisadas as AD nas escalas anual, mensal e horária, para identificar o período com maior frequência de ocorrência de DA na TMA-Manaus, utilizando todo o conjunto de dados de DA.

b) Taxa Diária de DA (TX-DA): Os dados de AD foram acumulados por dia para identificar a ocorrência do Tempo Severo Convectivo (TSC) e definir os limiares de severidade.

c) Equalização de dados: As séries temporais foram equalizadas, identificando-se e removendo, no conjunto TX-DA, os dias com dados de radiossondagem faltantes.

d) Variáveis Binárias: Criaram-se variáveis binárias para a ocorrência ou não de TSC (SIM/NÃO), para cada limiar de severidade obtido na etapa (b), resultando em uma variável binária para cada limiar.

e) Seleção de Variáveis Predictoras: Selecionaram-se variáveis predictoras dentre os 24 índices termodinâmicos, a serem usadas como entrada para os modelos de ML. Para isso, foi feita uma análise de correlações entre-variáveis e a discrepância percentual absoluta intra-variáveis quando representam SIM, para TSC com DA acima de um limiar definido e Não, caso contrário, conforme Eq. Zz:

$$A = \text{abs} \left(\frac{\text{média_variável}(\text{sim}) - \text{média_variável}(\text{não})}{\text{média_variável}(\text{não})} \right) \times 100 \quad (7)$$

f) Conjuntos de treinamento e de teste: Para cada limiar (etapa (b)), os dados foram divididos aleatoriamente em conjuntos de treinamento e de teste (50-50), tomando o conjunto das variáveis predictoras escolhidas na etapa (e) e usando as variáveis binárias criadas na etapa (d) como o alvo.

g) Normalização: Para cada limiar (etapa (b)), o conjunto de treinamento foi normalizado e os mesmos coeficientes de normalização foram aplicados ao conjunto de teste.

h) Balanceamento de classes: Para cada limiar (etapa (b)), foi feito o equilíbrio das classes aplicando a técnica de sobreamostragem aleatória ao conjunto de treinamento. A classe SIM foi replicada aleatoriamente de 1 a 5 vezes, isto é, foi mantida a quantidade da classe SIM do menor limiar e a quantidade da classe SIM do limiar superior foi replicada 5 vezes.

i) Algoritmos de ML: Por fim, foram selecionados 10 algoritmos categóricos de ML: *Logistic Regression*, *Random Forest*, *Extra Trees*, *Decision Tree*, *Gradient Boosting* (GBoosting), *Bagging*, *Adaptive Boosting* (AdaBoost), *K-Nearest Neighbors* (KNN), *Linear Discriminant Analysis* (LDA) e *Quadratic Discriminant Analysis* (QDA). Esses algoritmos foram selecionados por sua eficácia no tratamento de dados categóricos e por sua capacidade de capturar padrões intrincados no conjunto de dados. Os algoritmos foram implementados e executados no ambiente virtual *Google Colab* utilizando o pacote *Python Scikit-Learn* (Pedregosa et al., 2011), versão 1.2.2, com hiperparâmetros padrão.

Mais detalhes sobre cada algoritmo podem ser encontrados no Anexo 1.

- j) **Treinamento e Teste:** foi realizado o treinamento e teste dos algoritmos para cada limiar (etapa (b)).
- k) **Avaliação e Discussão:** Os resultados foram avaliados e discutidos utilizando as métricas Probabilidade de Detecção (POD) e Taxa de Alarmes Falsos (FAR). Um evento positivo implica a ocorrência de TSC (ou classe SIM) e um evento negativo implica a não ocorrência de TSC (ou classe NÃO).

4.3.2 ESTATÍSTICAS DE AVALIAÇÃO

Para a avaliação do desempenho dos algoritmos classificatórios foram utilizadas as métricas estatísticas abaixo que são baseadas na matriz de confusão, conforme Tabela 4.2.

Tabela 4.2: Matriz de Confusão. Fonte: (WILKS, 2006).

		Verdade	
		Condição Positiva	Condição Negativa
Prognóstico	Verdadeiro Positivo (a)		Falso Positivo (c)
	Falso Negativo (b)		Verdadeiro Negativo (d)

Para a avaliação dos testes dos modelos, foram utilizados os seguintes índices estatísticos baseados na tabela de contingência:

- i) Probabilidade de Detecção (POD) – Mede a fração de eventos observados que foram previstos corretamente. Uma pontuação perfeita é 1. Matematicamente é descrito de acordo com a equação 8.

$$POD = \frac{a}{a+b} \quad (8)$$

- ii) Razão de Falso Alarme (FAR) – Medida da fração de previsões “SIM” em que o evento não ocorreu. Uma pontuação perfeita é 0. É calculado usando a equação 9.

$$FAR = \frac{c}{c+a} \quad (9)$$

5. RESULTADOS

Esta seção está estruturada em duas partes distintas: a primeira traz um diagnóstico da análise espaço-temporal da AD na TMA-Manaus, enquanto a segunda tem como foco a avaliação dos resultados de um conjunto de algoritmos de ML na previsão de TSC na TMA-Manaus.

5.1 ANÁLISE DOS DADOS

Esta seção trata do processamento dos dados para a aplicação na construção dos modelos preditivos de eventos convectivos e no treinamento dos algoritmos de ML.

Os dados da climatologia da região revelam, em grande parte, a climatologia esperada da região amazônica, pois revelam os períodos sazonais de maior atividade convectiva e aqueles que possuem menor atividade, bem como aqueles períodos do dia em que se tem (ou não), a maior atividade convectiva.

5.1.1 MODELOS DE ML: PREVISÃO DE TSC

Os dados de radiossondagens foram obtidos conjuntamente com vinte e quatro IIT no endereço <http://wyoweb.uwyo.edu/>, conforme Anexo II. Os índices foram armazenados e organizados cronologicamente por data e hora em uma base de dados (planilha eletrônica), e seguindo os passos do método.

Da radiossonda SBMN 12Z, cada um dos 24 índices termodinâmicos encapsula aspectos críticos das condições atmosféricas, oferecendo informações valiosas sobre o potencial de classificação da atividade convectiva. Para otimizar a eficiência dos modelos de ML e ao mesmo tempo minimizar a redundância e a complexidade computacional, conduzimos uma análise de correlações entre variáveis (etapa 3.3e da metodologia). Este procedimento melhora a interpretabilidade do modelo treinado e melhora seu desempenho preditivo ao mitigar a multicolinearidade e o sobreajuste.

Ao identificar e eliminar variáveis altamente correlacionadas, o conjunto de dados foi simplificado para focar nos preditores mais informativos e discriminativos. O subconjunto de variáveis resultante representa uma seleção equilibrada que captura a termodinâmica atmosférica essencial relevante para a previsão de TSC. Após esse processamento, restaram 16 índices dos 24 iniciais:

SWET, BRCH, CAPE, LIFT, SHOW, KINX, EQLV, CINS, CINV, LFCT, VTOT, LFCV, MLTH, PWAT, THTK e LCLT.

A Tabela 5.1 apresenta a discrepância percentual absoluta entre as 16 variáveis para eventos de SC e não SC de acordo com o limiar de severidade definido, com $DR-AD > k$ identificado como SIM onde $k = 69, 283,5, 1.063,5, 2.555,75, 5.000$, e NÃO caso contrário. Espera-se que as características com maior discrepância entre os eventos SIM e NÃO sejam particularmente influentes. Os maiores valores de discrepância estão vinculados às 5 variáveis a seguir: SHOW, BRCH, CAPE, LIFT e EQLV. Esses 5 índices são indicadores fundamentais da ocorrência de TSC na TMA-Manaus, dada a sua capacidade de avaliar a estabilidade atmosférica e o cisalhamento do vento.

Por exemplo, o índice SHOW avalia o potencial de trovoadas na previsão do tempo, com valores positivos indicando condições atmosféricas estáveis menos propícias à formação de trovoadas, enquanto valores negativos sugerem maior instabilidade, aumentando a probabilidade de atividade convectiva e trovoadas (Doswell III, Davies-Jones, e Keller 1993). Da mesma forma, os valores BRCH denotam instabilidade elevada, promovendo o desenvolvimento de tempestades através da formação de nuvens convectivas e precipitação intensa (Brooks e Craven 2002). Além disso, o BRCH considera o cisalhamento vertical do vento (Holton 2004), predominante em regiões tropicais como Manaus, influenciando a estrutura e a intensidade das tempestades.

CAPE representa a energia disponível que impulsiona o desenvolvimento de nuvens convectivas, com valores elevados indicando potencial convectivo aumentado devido a fatores como umidade, temperaturas quentes e aquecimento solar, facilitando a instabilidade atmosférica e a formação de nuvens convectivas (Moncrieff 2010). O LIFT avalia a estabilidade atmosférica comparando a temperatura de uma parcela de ar elevada com o seu ambiente a uma altitude específica, normalmente 500 hPa acima do nível do solo. Valores negativos de LIFT sugerem parcelas mais quentes e menos densas, indicando instabilidade favorável à atividade convectiva, enquanto valores positivos implicam estabilidade, inibindo o desenvolvimento convectivo (Emanuel 1994). O EQLV marca a altitude onde uma parcela de ar elevada se torna mais fria do que o ambiente circundante, deixando de subir livremente (Wallace e Hobbs 2006).

Juntos, esses recursos oferecem insights cruciais sobre a termodinâmica atmosférica, essenciais para a compreensão e previsão de TSC na TMA- Manaus. A discussão a seguir aborda especificamente a relevância desses recursos durante o treinamento de ML.

Tabela 5.1 - Discrepância percentual absoluta intra-variáveis para DR-AD > k, identificado como SIM onde k = 63, 283,5, 1.063,5, 2.555,75, 5.000 e NÃO caso contrário.

Variável		Discrepância percentual de acordo com o limiar DR-AD				
		69	283.5	1063.5	2555.75	5000
1	SHOW	114.59	139.03	171.85	173.26	258.96
2	BRCH	42.90	4.98	39.08	39.44	50.11
3	CAPE	34.89	33.29	26.37	26.59	30.60
4	LIFT	47.18	37.84	27.64	19.63	20.90
5	EQLV	18.85	15.49	12.35	11.74	14.12
6	CINS	27.24	21.95	7.89	9.34	5.79
7	KINX	8.75	7.00	6.20	5.45	5.91
8	CINV	20.61	14.17	0.53	0.11	2.95
9	SWET	2.09	2.31	1.21	1.16	1.69
10	LFCT	3.57	2.79	1.65	1.89	1.38
11	VTOT	0.80	0.81	1.32	0.71	0.90
12	LFCV	2.38	1.94	0.81	0.96	0.55
13	MLTH	0.12	0.05	0.00	0.05	0.12
14	PWAT	6.12	4.33	1.90	0.24	0.07
15	THTK	0.06	0.04	0.08	0.11	0.07
16	LCLT	0.28	0.21	0.08	0.02	0.03

Cada algoritmo de ML (etapa 3.3i da metodologia) foi executado três vezes, utilizando 5, 16 e 24 variáveis como entrada, de acordo com a ordem de relevância estabelecida no índice da Tabela 5.1. A Figura 3 apresenta cinco gráficos dos algoritmos de ML de dispersão de POD vs. FAR, para os 5 limiares de DR-AD. A Figura 4 resume o comportamento médio dos resultados POD e FAR usando as 16 variáveis selecionadas. Para mostrar o desempenho de predição de TSC para o limiar mais baixo de DR-AD (69), a Tabela 5.2 apresenta os resultados de POD e FAR, para cada modelo e o número de variáveis de entrada usados na execução, classificados por valor de POD.

Na Figura 4, podemos observar a diminuição no desempenho do modelo e o aumento na dispersão à medida que o limiar de DR-AD aumenta. Limiares mais baixos proporcionam os melhores resultados, provavelmente devido a um maior número de eventos SIM. Além disso, há uma alta taxa de falsos positivos (falso negativos) para limiares baixos (altos), uma vez que os eventos SIM (NÃO)

têm um número maior (Figura 3b).

Avaliando os resultados de DR-AD=69, o conjunto de modelos ML demonstra desempenhos bastante semelhantes e aceitáveis, com valores médios de POD e FAR de $0,92 \pm 0,06$ e $0,19 \pm 0,01$, respectivamente, conforme ilustrado nas Figuras 3a e 4. Da Tabela 5.2, o melhor desempenho foi alcançado pelo algoritmo QDA, com POD de 0,99 e FAR de 0,19. Ainda da Tabela 5.2, os modelos LDA e Regressão Logística apresentam resultados comparativos ao QDA para um menor número de variáveis. A Árvore de Decisão apresentou os piores resultados.

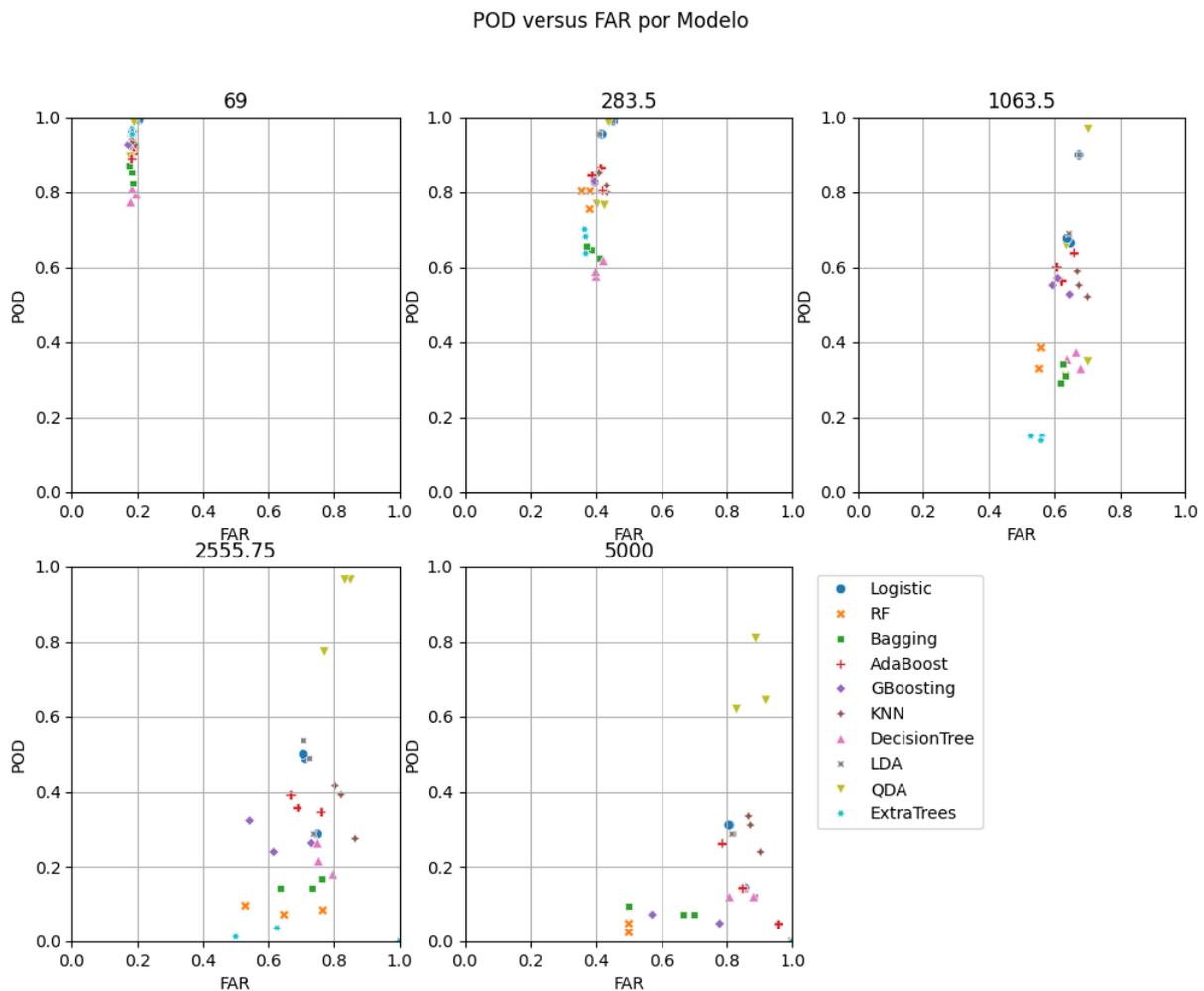


Figura 3 - Gráfico de Dispersão de POD vs. FAR para 10 algoritmos de ML usando a 5, 16, e 24 recursos selecionados como entrada, para cada limite de DR-AD: (a) 69, (b) 283.5, (c) 1063.5, (d) 2555.75 e (e) 5000.

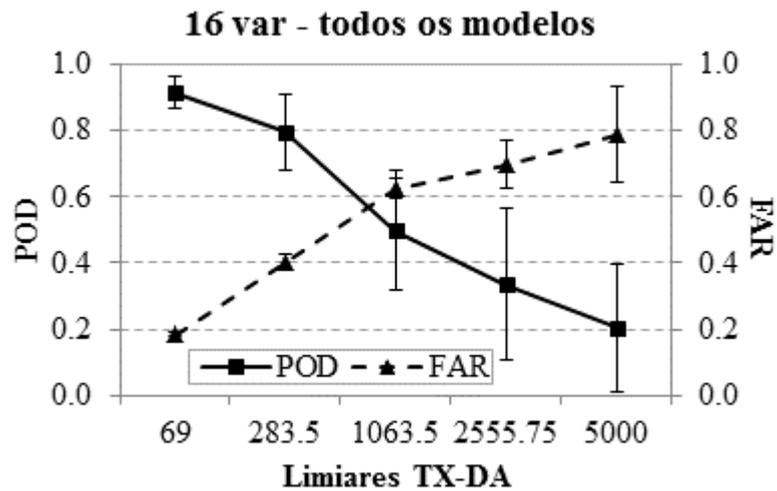


Figura 4 – Média e desvio padrão de POD e FAR para os 10 algoritmos de ML usando 16 características, por limite DR-AD.

Tabela 5.2 – Desempenho dos 10 algoritmos ML para DR-AD=69 usando os recursos 5, 16 e 24. Os resultados são classificados pelo valor POD.

Modelo	# Recursos	POD	FAR
QDA	24	0.99	0.19
LDA	5	0.99	0.20
Logistic	5	0.99	0.20
ExtraTrees	24	0.96	0.18
ExtraTrees	16	0.96	0.18
Logistic	24	0.96	0.18
QDA	5	0.96	0.19
Logistic	16	0.96	0.19
RF	24	0.95	0.18
LDA	24	0.95	0.18
LDA	16	0.95	0.18
RF	16	0.94	0.18
KNN	24	0.94	0.19
GBoosting	24	0.93	0.17
ExtraTrees	5	0.93	0.19

KNN	5	0.93	0.19
KNN	16	0.93	0.19
GBoosting	5	0.93	0.20
GBoosting	16	0.92	0.18
AdaBoost	16	0.92	0.19
RF	5	0.92	0.19
AdaBoost	5	0.91	0.18
QDA	16	0.90	0.18
AdaBoost	24	0.89	0.18
Bagging	16	0.87	0.17
Bagging	5	0.86	0.18
Bagging	24	0.85	0.17
DecisionTree	16	0.80	0.19
DecisionTree	5	0.80	0.20
DecisionTree	24	0.76	0.18

5.1.2 DESCARGAS ATMOSFÉRICAS: CLASSIFICAÇÃO DA SEVERIDADE DOS EVENTOS CONVECTIVOS

Para avaliar a distribuição da quantidade de DA e estabelecer os limites de gravidade da TSC, são apresentados na Figura 5 o *boxplot* e o histograma da quantidade de DA-DR da TMA-Manaus. A distribuição dos quartis foi a seguinte: o 1º Quartil (Q1) em 69,0, a mediana (Q2) em 283,5, o 3º quartil (Q3) em 1.063,5 e uma cerca superior [$Q3 + (1,5 * IQR)$] em 2.555,75. Essas quantidades definem os limiares de criticidade. De 2012 a 2017, houve apenas 21 dias com DR-AD = 0, enquanto 248 dias ultrapassaram o limite superior, indicando eventos significativos. Determinamos que uma taxa de 5.000 AD está relacionada à ocorrência de SC mais intenso, com impactos potenciais consideráveis no tráfego aéreo dentro da TMA-Manaus. Foram 109 dias excedendo 5.000 DR-AD.

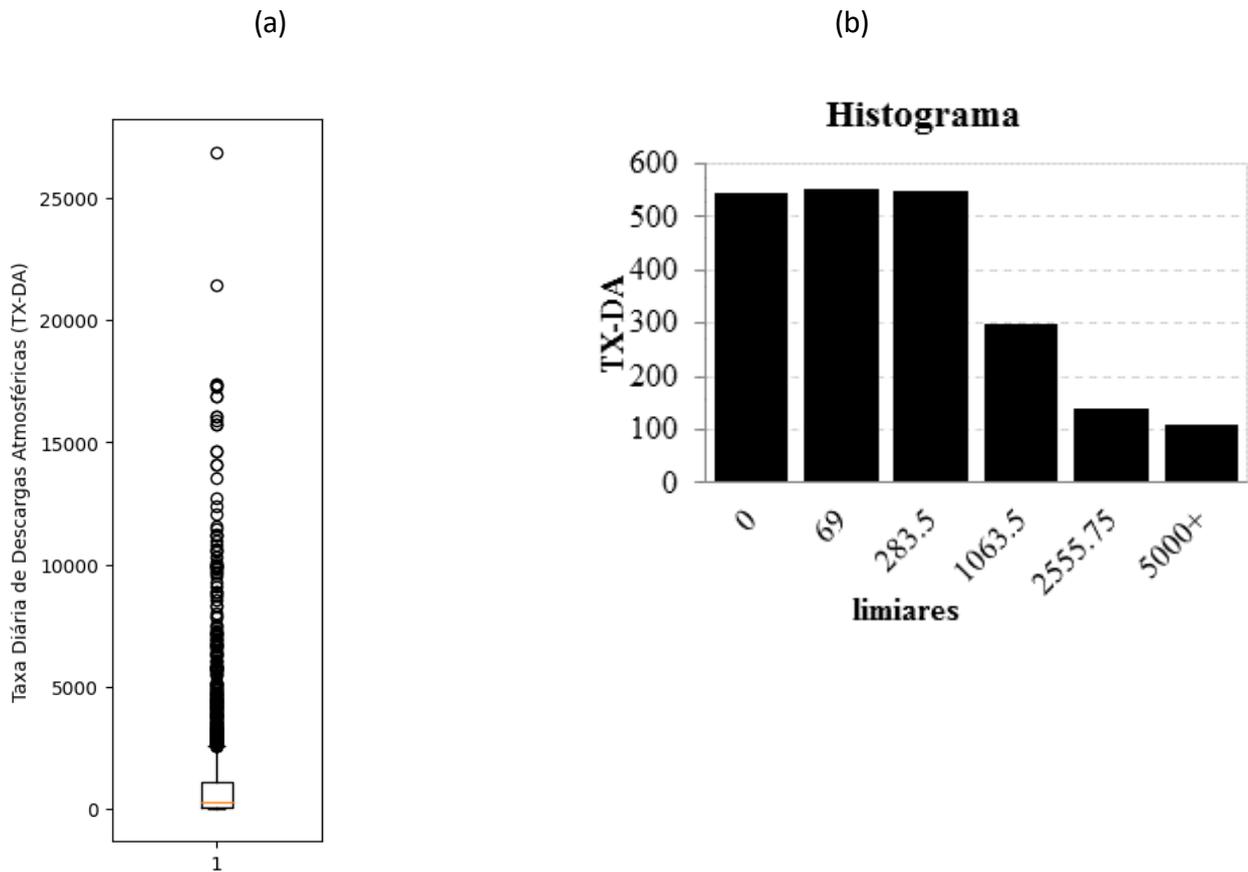


Figura 5 – (a) Boxplot ilustrando o DR-AD da TMA-Manaus para 2.192 dias (de 2012 a 2017): média 1105,0; mínimo 0; Q1 69,0; Q2 283,5; Q3 1063,5; limite superior 2555,75; máximo 26886; (b) Histograma do DR-AD com limiars de caixas.

A Figura 6 mostra a distribuição percentual de eventos AD de dois conjuntos de dados: o conjunto completo de AD de 2012 a 2017, composto por 2.422.173 eventos (rotulado como "Todos"), e o conjunto de dados obtido após a etapa 3.3c da metodologia, que equaliza os dados de AD com os dados de radiossondagem, consistindo em 1.692.333 eventos (rotulados como "Utilizados") que foram usados para treinar e testar os algoritmos de ML. A Figura 6 inclui (a) a distribuição horária de AD, (b) a distribuição mensal e (c) a distribuição anual. Incluímos a distribuição "Utilizados" para demonstrar sua semelhança com a distribuição "Todos".

Na Figura 6a, vemos que 73,68% dos eventos de AD ocorrem entre 15Z e 21Z, com 51,2% ocorrendo entre 16Z (20:00 Hora Local) e 19Z (23:00 Hora Local), indicando que o período noturno é o mais impactante, como esperado. Como os dados de entrada são adquiridos às 12Z, o horizonte da previsão é definido como 4 horas. A Figura 6b

mostra que as maiores ocorrências de DA ocorrem durante os meses de primavera, de agosto a novembro. A Figura 6c indica que 56,1% dos eventos de DA ocorreram entre 2013 e 2014. Não foi observada correlação com o fenômeno ENOS; outros parâmetros de variabilidade climática, como a temperatura da superfície do mar e a variação na posição da ZCIT, não foram avaliados.

A Figura 7 ilustra a distribuição espacial mensal da densidade de DA na área estudada no período de 2012 a 2016. Conforme visto na Figura 6b, os meses com maior DA apresentaram distribuição espacial mais homogênea, enquanto os demais meses apresentaram núcleos concentrados de DA. Mas é interessante notar que a DA ocorre em todos os meses do ano.

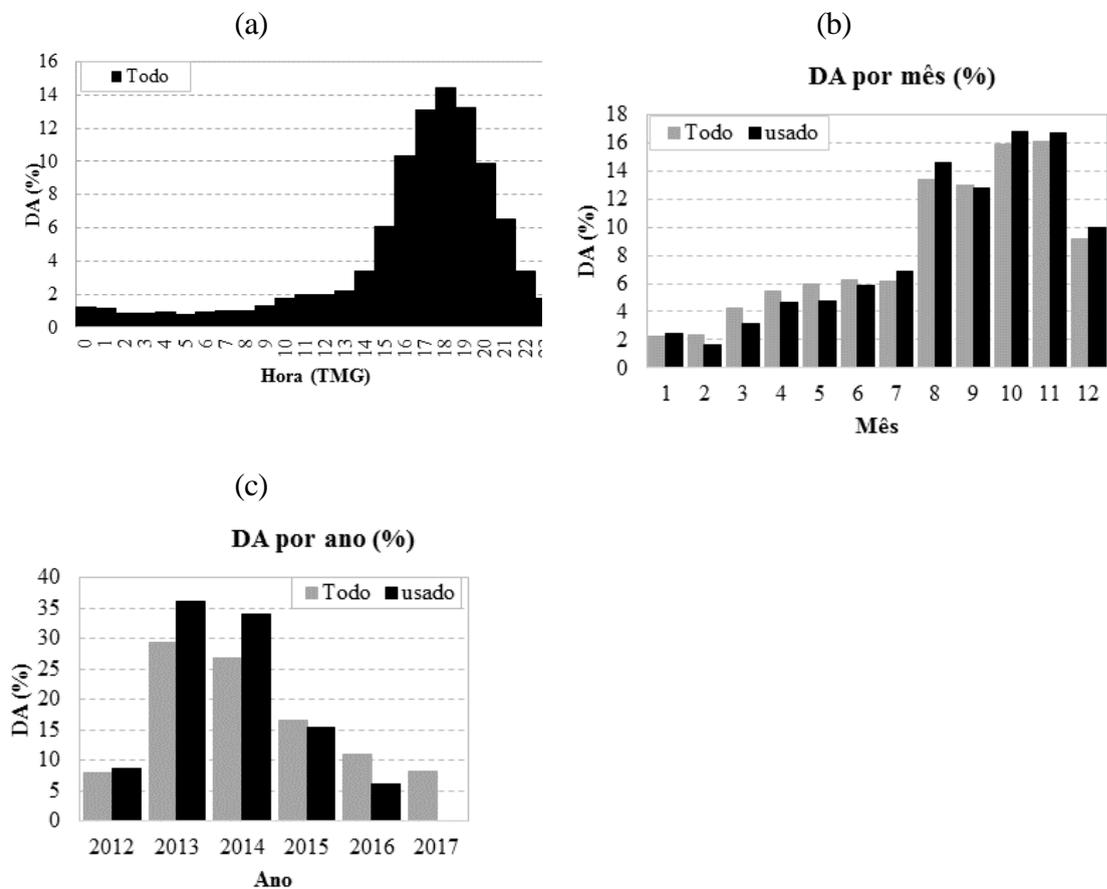


Figura 6 - Distribuição de DA (%) considerando o conjunto de dados “All” (2.422.173 eventos) e o conjunto de dados “Used” (1.692.333 eventos): (a) por hora, (b) mensal e (c) anual. Os dados “Used” são obtidos aplicando a etapa 3.3c da metodologia.

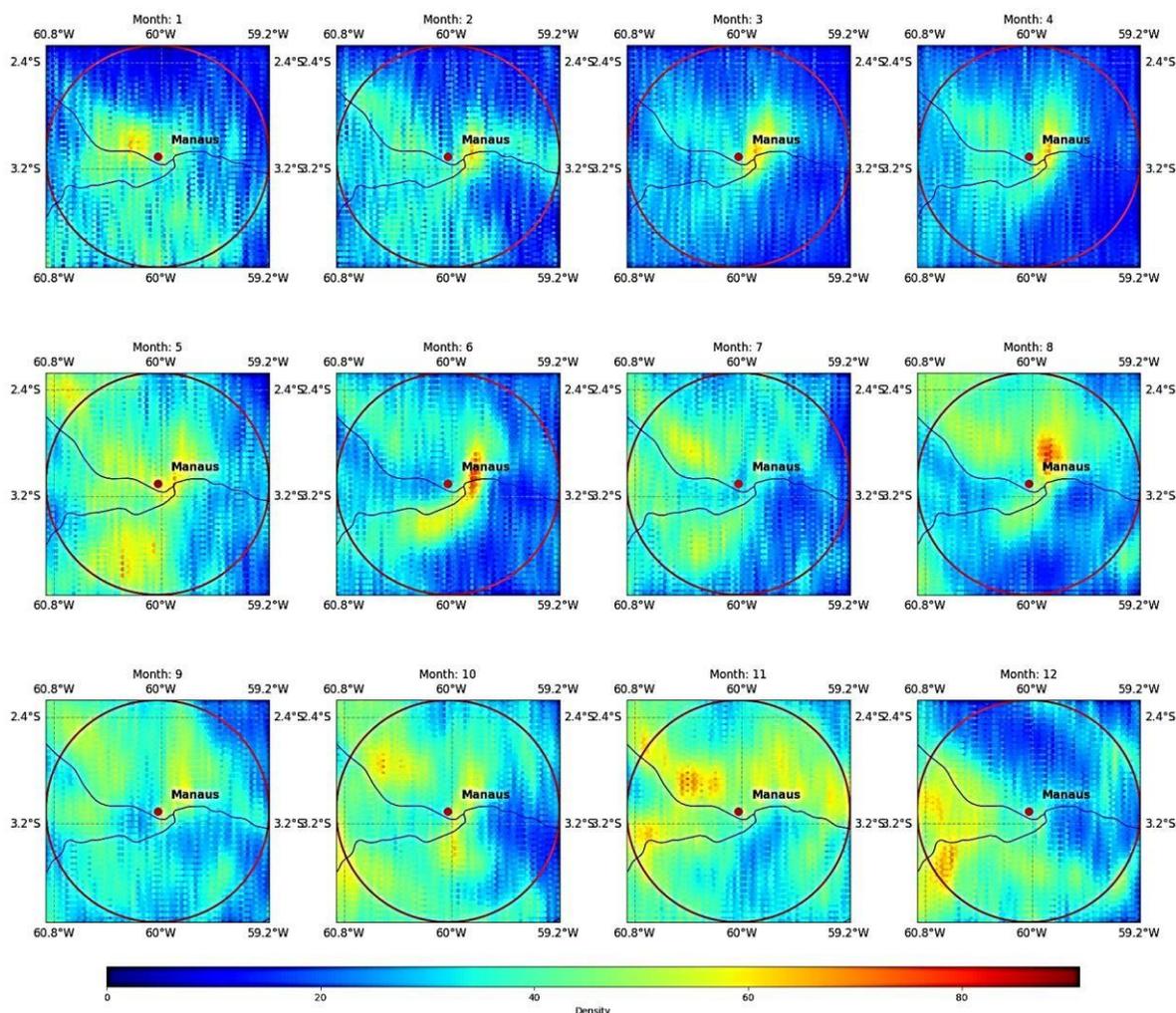


Figura 7 – Densidade mensal de DA na TMA-Manaus, representada por um círculo vermelho com raio aproximado de 100 km, no período de 2012 a 2016.

6. CONCLUSÃO

Este estudo avaliou o desempenho de modelos 10 de ML na predição da ocorrência de TSC usando índices termodinâmicos derivados de dados de radiossonda SBMN das 12Z e DA para caracterizar TSC. A climatologia da DA entre 2012 e 2017 revelou que a DA ocorre em todos os meses do ano, neste período houve apenas 21 dias sem DA. Ao empregar um processo sistemático de seleção de variáveis, identificamos um subconjunto de 16 dos 24 índices iniciais que eram mais informativos para a previsão de TSC, reduzindo assim a redundância e a complexidade computacional do modelo. A análise revelou que 5 índices, SHOW, BRCH, CAPE, LIFT e EQLV foram particularmente influentes na distinção entre eventos TSC e não-TSC, destacando a sua importância na avaliação da estabilidade atmosférica e do cisalhamento do vento.

A avaliação dos modelos de ML, que incluiu múltiplas execuções com números variados de variáveis selecionadas e diferentes limiares de DR-AD, demonstrou que esses modelos só são capazes de prever a ocorrência ou não de TSC com relativa precisão. O algoritmo QDA apresentou o melhor desempenho com POD de 0,99 e FAR de 0,19 para DR-AD=69, utilizando as 24 variáveis como entrada. Os modelos LDA e Regressão Logística apresentaram resultados comparativos ao QDA para um menor número de variáveis. A diminuição no desempenho à medida que o limiar de DR-AD aumenta indica desafios na previsão de eventos de TSC mais severos, o que é uma limitação crítica para aplicações práticas em regiões como a TMA-Manaus.

Nesse trabalho houve o desenvolvimento de um modelo específico para o período de agosto a dezembro devido a maior frequência de descargas atmosféricas, visando a possibilidade de um aprimoramento da previsão de Eventos Convectivos Severos e sua severidade na Região Amazônica.

Os resultados destacam o potencial dos modelos de ML em melhorar as previsões de TSC usando dados de radiossondagens e AD. No entanto, para melhorar a sua utilidade prática, especialmente na meteorologia operacional, e uma investigação ou pesquisa futura deve buscar a integração de fontes de dados adicionais, tais como perfiladores infravermelhos atmosféricos de alta frequência, e em melhorar a capacidade dos modelos de prever a gravidade dos eventos, bem como os dados modelados dos índices de instabilidade termodinâmicos da Região Amazônica, disponibilizados pelo INPE.

8. REFERÊNCIAS

ANAC: condicoes meteorologicas adversas para o voo. condicoes meteorologicas adversas para o voo. Disponível em: <https://www.gov.br/anac/pt-br/assuntos/seguranca-operacional/meteorologia-aeronautica/condicoes-meteorologicas-adversas-para-o-voo/trovoadas#a05>. Acesso em: 16 set. 2024.

Almeida, V. A. de, G. B. França, and H. F. de. Campos Velho. 2020. Short-range forecasting system for meteorological convective events in Rio de Janeiro using remote sensing of atmospheric discharges. *International Journal of Remote Sensing*, 41(11), 4372–4388. <https://doi.org/10.1080/01431161.2020.1717669>

Bonnet, S. M., A. Evsukoff, and C. A. Morales Rodriguez. 2020. Precipitation Nowcasting with Weather Radar Images and Deep Learning in São Paulo, Brasil. *Atmosphere*, 11, 1157. <https://doi.org/10.3390/atmos11111157>

Breiman, L. 1996. Bagging predictors. *Machine Learning*, 24(2), 123-140.

Breiman, L. 2001. Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and regression trees*. Wadsworth International Group.

Brooks, H. E., and J. P. Craven. 2002. A database of proximity soundings for significant severe thunderstorms, 1957–1993. *Bulletin of the American Meteorological Society*, 83(1), 88-100.

Cover, T. M., and P. E. Hart. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27. <https://doi.org/10.1109/TIT.1967.1053964>

Cox, D. R. 1958. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 215-232.

<https://doi.org/10.1111/j.2517-6161.1958.tb00292.x>

De Castro, J. N., G. B. França, V. A. de Almeida, and M. V. de Almeida. 2022. Severe convective weather forecast using machine learning models. *Pure and Applied Geophysics*, 179(8), 2945-2955. <https://doi.org/10.1007/s00024-022-03088-8>.

DECEA. TMA - Área de Controle Terminal. 2024. Disponível em:

<https://www.decea.mil.br/index.cfm?i=utilidades&p=glossario&single=2357#:~:text=%C3%81rea%20de%20controle%20situada%20geralmente,%2C%20Marte%2C%20Santos%2C%20etc..> Acesso em: 04 out. 2024.

De Souza, Everaldo & Ferreira, Douglas & Guimarães, José & Franco, Vania & de Azevedo, Franklin & Souza, Paulo. (2017). *_Revista Brasileira de Climatologia_PADRÕES CLIMATOLÓGICOS E TENDÊNCIAS DA PRECIPITAÇÃO NOS REGIMES CHUVOSO E SECO DA AMAZÔNIA ORIENTAL*. *Revista Brasileira de Climatologia*. 21. 81. 10.5380/abclima.v21i0.41232.

Domingos, Pedro. (2012). A Few Useful Things to Know About Machine Learning. *Commun. ACM*. 55. 78–87. 10.1145/2347736.2347755.

Doswell III, C. A., R. Davies-Jones, and D. L. Keller. 1993. Thunderstorm forecasting: A review. *Meteorological Monographs*, 21, 27-59.

ELAT: Grupo de Eletricidade Atmosférica. Disponível no site da Instituto Nacional de Pesquisas Espaciais, 2023. Disponível em: <http://www.inpe.br/webelat/rindat/menu/desc.atm/>. Acesso em: 05 dez. 2023.

Emanuel, K. A. 1994. *Atmospheric Convection*. Oxford University Press.

Ferreira, N., and I. Cavalcanti. 2022. *Sistemas meteorológicos atuantes no Brasil*. São Paulo: Oficina de Textos. ISBN 978-65-86235-52-4.

- Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. In *Annals of Eugenics*, 7(2), 179-188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- França, G. B., F. L. Albuquerque Neto, and H. F. Campos Velho. 2023. *Nowcasting using machine learning and deterministic models: a Brazilian initiative to improve aviation meteorology*. Rio de Janeiro, RJ: Editora da Universidade da Força Aérea. 280 p. <https://hdl.handle.net/11422/22155>
- França, G. B., M. V. Almeida, and A. Rossete. 2016. An Automated Nowcasting Model of Significant Instability Events in the Flight Terminal Area of Rio De Janeiro, Brazil. *Atmospheric Measurement Techniques*. 9, 2335–2344. <https://doi.org/10.5194/amt-9-2335-2016>.
- Freund, Y., and R. E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. In *Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- GALWAY, J. G. 1956. The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc.* p. 528–529.
- GEORGE, J. J. 1960. *Weather Forecasting for Aeronautics*. Academic Press. p. 673.
- Geurts, P., D. Ernst, and L. Wehenkel. 2006. Extremely randomized trees. *Machine Learning*, 63(1), 3-42. <https://doi.org/10.1007/s10994-006-6226-1>
- Gultepe, I., R. Sharman, P. D. Williams, B. Zhou, G. Ellrod, P. Minnis, ... and F. A. Neto. 2019. A review of high impact weather for aviation meteorology. *Pure and applied geophysics*, 176, 1869-1921. <https://doi.org/10.1007/s00024-019-02168-6>

GREWAL, D. S., 2014. A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering. IOSR Journal of Computer Engineering, v. 16, i 1, p. 09-13.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: Data mining, inference, and prediction. *Springer*.
<https://doi.org/10.1007/978-0-387-84858-7>

HAYKIN, S. (2002). Neural Networks. A Comprehensive Foundation, Macmillan, New York, NY.

HENRY, W. The Skew T Log P Diagram. National Weather Service Training Center, EUA, 1987.68p.

Hermesdorff, J. 2018. Previsão de instabilidade atmosférica significativa usando árvore de decisão na região metropolitana do Rio de Janeiro. Dissertation, Universidade Federal do Rio de Janeiro, 109 f.

Holton, J. R. 2004. An Introduction to Dynamic Meteorology. *Elsevier Academic Press*.

House, D.C., and R.G. Beebe. 1955. SELS Forecasting Procedures. *US Weather Bureau*.

ICAO (International Civil Aviation Organization). 2018. Annex 3 to the Convention on International Civil Aviation, Meteorological Service for International Air Navigation. 20th Edition. ICAO.

ICMC JÚNIOR,. Inteligência Artificial. 2021. Disponível em:

https://icmcjunior.com.br/inteligenciartificial/?gad_source=1&gclid=Cj0KCQjwjNS3BhChARIsAOxBM6rCrvi38W0spiYADzx7tGPm6lVJzOndrC6Vc-IKW0tjCNBFVkyDeH0aAnvpEALw_wcB. Acesso em: 04 out. 2024.

Isaac, G. A., M. Bailey, S. G. Cober, N. Donaldson, N. Driedger, A. Glazer, A., ... and F. Fabry. 2006. Airport Vicinity Icing and Snow Advisor (AVISA). In *44th AIAA Aerospace Sciences Meeting and Exhibit*. <https://doi.org/10.2514/6.2006-1219>

Keller, M., Alencar, A., Asner, G.P., Braswell, B., Bustamante, M., Davidson, E., Feldpausch, T., Fernandes, E., Goulden, M., Kabat, P., Kruijt, B., Luizão, F., Miller, S., Markewitz, D., Nobre, A.D., Nobre, C.A., Priante Filho, N., da Rocha, H., Silva Dias, P., von Randow, C. and Vourlitis, G.L. (2004), ECOLOGICAL RESEARCH IN THE LARGE-SCALE BIOSPHERE– ATMOSPHERE EXPERIMENT IN AMAZONIA: EARLY RESULTS. *Ecological Applications*, 14: 3-16. <https://doi.org/10.1890/03-6003>

Lareau, N. P. and Clements, C. B.: Environmental controls on pyrocumulus and pyrocumulonimbus initiation and development, *Atmos. Chem. Phys.*, 16, 4005–4022, <https://doi.org/10.5194/acp-16-4005-2016>, 2016.

Marengo, J., and G. Fisch. 2021. Clima da Região Amazônica. In Cavalcanti, I. F. A.; Ferreira, N. J. Clima das regiões brasileiras e variabilidade climática (1st ed., pp. 9-24). São Paulo: Oficina de Textos. ISBN 978-65-86235-24-1.

MENDONÇA, F.; DANNI-OLIVEIRA, I. M. Climatologia: noções básicas e climas do Brasil. São Paulo: Oficina de Texto, 2007. 206 p.

MILLER, R. C., 1972. Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central. Tech. Rept. 200(R). Headquarters, Air Weather Service, USAF. p. 190.

Moncrieff, M. W. 2010. The multiscale organization of moist convection and the intersection of weather and climate. *Meteorological Monographs*, 37, 407-418.

Morales, C. A., J. A. Weinman, M. S. Branch, E. N. Anagnostou, S. Goodman, G. H. Center, and E. Williams. 2002. Continuous Long-Range Thunderstorm Monitoring by a VLF Receiver Network. Part II: Cloud-to-Ground and Intra-Cloud Detection Efficiency. Colorado State University.

MOURÃO, C. E. F.; MENEZES, W. F., 2006. “Estudo do comportamento de indicadores de tempo severo em casos de tempestades sobre o Rio de Janeiro”. Anais do XIV Congresso Brasileiro de Meteorologia. Florianópolis-SC.

Mueller, C., T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and J. Yee. 2003. NCAR auto-nowcast system. *Weather Forecast*, 18, 545–561.

MUNAKATA, TOSHINORI, 1994. Fundamentals of the new Artificial Intelligence: neural, evolutionary, fuzzy and more, 2nd ed. Springer-Verlag London. Cleveland State University.

Nascimento, E.L. 2005. Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: uma estratégia operacional adotável no Brasil? *Revista Brasileira de Meteorologia*, 20, 121-140.

Nunes, M. P., G. B. França, and V. A. Almeida. 2023. Convective Meteorological Events Forecast using Machine Learning models. PREPRINT (Version 1) available at Research Square. <https://doi.org/10.21203/rs.3.rs-2802341/v1>

Paulucci, T. B. 2017. Caracterização Espaço-Temporal de Descargas Atmosféricas e Tempestades Elétricas na Região Metropolitana do Rio de Janeiro entre 2001 e 2016. Monography, Universidade Federal do Rio de Janeiro, 73p.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, ... and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

PROJETO GoAmazon. 2024. Instituto Nacional de Pesquisas da Amazônia - INPA. Disponível em: <https://www.gov.br/inpa/pt-br/Pesquisa/projeto-goamazon>. Acesso em: 16 set. 2024.

Queiroz, A. P. 2009. Monitoramento e previsão imediata de tempestades severas usando dados de radar. Dissertation, Instituto Nacional de Pesquisas Espaciais: São José dos Campos.

Soares, C. M., G. B. França, M. A. Valdonel, and V. A. Almeida. 2021. Study of GOES- R Thermodynamic Indices for Short-Term Forecasting of Convective Weather Events Using Machine Learning. *Pure and Applied Geophysics*, 178, 4651-4663. <https://doi.org/10.1007/s00024-021-02889-7>.

SHOWALTER, A. K., 1947. A stability index for forecasting thunderstorms. *Bull. Amer. Meteor. Soc.* 34. p. 250– 252.

SUFRAMA. Amazônia Ocidental. 2021. Disponível em: <https://www.gov.br/suframa/pt-br/assuntos/amazonia-ocidental>. Acesso em: 25 set. 2024.

TECH, Didática. A biblioteca scikit-learn – Python: o que é, para que serve. 2024. Disponível em: A biblioteca scikit-learn – Python: o que é, para que serve. Acesso em: 04 out. 2024.

Tory KJ and Thurston W (2015) *Pyrocumulonimbus: A Literature Review*, Bushfire and Natural Hazards CRC, Australia.

UNESP. DESCARGAS ATMOSFÉRICAS (RAIOS). 2000. Informativo AREIA, Ano 1 nº 4, novembro de 2000. Disponível em: <https://www2.feis.unesp.br/irrigacao/raios.htm>. Acesso em: 04 out. 2024.

Wallace, J. M., and P. V. Hobbs. 2006. *Atmospheric science: An introductory survey*. Academic Press.

WILKS, S. D., 2006. *Statistical Methods in the Atmospheric Sciences*. 2ª ed., New York, USA, Academic Press.

YAO, X. and LIU, Y., 2013. *Machine Learning*. School of Computer Science, University of Birmingham, UK, C. 12, p. 341-343.

Ynoue, R. Y., Ambrizzi, T., Reboita, M. S., & Silva, G. A. M. da. (2017). *Meteorologia: noções básicas*. São Paulo: Oficina de Textos.

ANEXO I

LISTA DE ALGORITMOS UTILIZADOS NA PESQUISA¹

ALGORITMOS	DESCRIÇÃO DO ALGORITMO
Randon Forest (Breiman 2001)	É um método de aprendizado de conjunto que constrói múltiplas árvores de decisão e combina suas saídas por meio de média ou votação. Ele introduz aleatoriedade usando um subconjunto aleatório de recursos para cada árvore e treinando cada árvore em uma amostra de inicialização dos dados..
ExtraTrees (Geurts, Ernst, e Wehenkel 2006)	É como Random Forest, mas constrói cada árvore de decisão a partir de todo o conjunto de dados de treinamento em vez de uma amostra bootstrap e seleciona limites aleatórios para cada recurso.
DecisionTree (Breiman <i>et al.</i> , 1984)	As árvores de decisão são estruturadas hierarquicamente, com nós internos significando decisões com base em recursos e nós folha indicando rótulos de classe. Os hiperparâmetros encapsulam variáveis como profundidade máxima, amostras mínimas por folha, critérios de divisão e recursos máximos.
GBoosting (Friedman 2001)	O Gradient Boosting constrói sequencialmente um conjunto de aprendizes fracos, geralmente árvores de decisão, com cada aprendiz subsequente focado em retificar erros dos anteriores. Seus hiperparâmetros abrangem fatores como o número de estimadores (árvores), taxa de aprendizado e profundidade máxima de cada árvore.
Bagging (Breiman 1996)	Bagging, ou Bootstrap Aggregating, é uma técnica de aprendizado de conjunto que cria vários modelos, geralmente árvores de decisão, treinando cada um em subconjuntos distintos dos dados de treinamento. Essa variação reduz o overfitting e melhora a estabilidade e generalização do conjunto. Os modelos são treinados independentemente em subconjuntos aleatórios dos dados, introduzindo diversidade e garantindo que capturem diferentes aspectos da distribuição de dados. As previsões desses modelos são combinadas por meio de média ou votação. Os hiperparâmetros do Bagging incluem a escolha do estimador base, número de estimadores e amostras máximas consideradas, permitindo a personalização para diferentes conjuntos de dados e tarefas de aprendizagem.
AdaBoost (Freund e Schapire 1997)	AdaBoost, ou Adaptive Boosting, é um método de aprendizagem de conjunto que constrói sequencialmente um classificador robusto adicionando aprendizes fracos ao conjunto. Ele atribui pesos mais altos a pontos de dados classificados incorretamente em cada iteração, focando em instâncias desafiadoras para aumentar o desempenho geral e a generalização. Hiperparâmetros como o estimador base (normalmente árvores de decisão), número de estimadores e taxa de aprendizagem são essenciais. Parâmetros adicionais podem incluir a complexidade do aprendiz fraco (por exemplo, profundidade máxima das árvores de decisão) e configurações de estratégia de amostragem como peso de amostragem ou taxa de subamostragem, permitindo o ajuste fino do comportamento do modelo.
KNN (Cover e Hart 1967)	K-Nearest Neighbors (KNN) atribui rótulos a pontos de dados com base na classe majoritária entre seus k vizinhos mais próximos no espaço de recursos. Seu desempenho depende de hiperparâmetros importantes, como o número de vizinhos (k), a escolha da métrica de distância (por exemplo, Euclidiana ou Manhattan) e a função de ponderação ('uniforme' ou 'distância'). Ajustar esses parâmetros permite que os profissionais adaptem o KNN para se adequar às características do conjunto de dados, melhorando sua precisão preditiva.
QDA (Hastie, Tibshirani, e Friedman 2009)	A Análise Discriminante Quadrática (QDA) é um algoritmo de classificação que assume que cada classe tem sua própria matriz de covariância, oferecendo flexibilidade na captura de relacionamentos complexos de recursos. Seus hiperparâmetros incluem o parâmetro de regularização para evitar overfitting e a especificação de probabilidades anteriores para cada classe, influenciando as previsões do modelo. Ajustar esses parâmetros garante um desempenho robusto em diferentes conjuntos de dados e cenários.

LDA (Fisher 1936)	A Análise Discriminante Linear (LDA) é um método estatístico usado para reconhecimento de padrões e aprendizado de máquina. É usado principalmente para redução de dimensionalidade, preservando o máximo possível de informações discriminatórias de classe. A LDA funciona encontrando as combinações lineares de recursos que melhor separam duas ou mais classes. Isso é obtido maximizando a razão da variância entre classes para a variância dentro da classe, garantindo que as combinações resultantes forneçam a melhor separação possível entre as diferentes classes.
Logistic Regression (Cox 1958)	A regressão logística é um método estatístico usado para problemas de classificação binária. Ela modela a probabilidade de que uma determinada entrada pertença a uma classe específica. A regressão logística usa a função logística (também conhecida como função sigmoide) para transformar a combinação linear de recursos de entrada em um valor de probabilidade entre 0 e 1. Ela é particularmente útil para situações em que a variável dependente é categórica, permitindo a previsão de resultados binários.

¹ Dados obtidos do sítio de Internet do Manual do Software WEKA, disponível em <https://weka.sourceforge.io/doc.dev/allclasses-noframe.html>.

ANEXO II

LISTA DE ÍNDICES DE INSTABILIDADE TERMODINÂMICOS²

SIGLA	NOMECLATURA DOS ÍNDICES
SH	Índice de instabilidade SHOWALTER utilizado para previsão de tempestades
LI	Índice de Levantamento (<i>Lifted</i>) semelhante ao SH
LITV	Cálculo do índice LI usando a temperatura virtual
SW	SWEAT que representa o Índice de ameaça de Tempo Severo
K	Índice K aplicado para estimar a probabilidade de chuva e tempestades
CT	Cross Totals aplicado para prognosticar tempestades severas
VT	Vertical Totals aplicado para prognosticar o potencial desenvolvimento de tempestade severa
TT	Total Totals aplicado para prognosticar o potencial desenvolvimento de tempestade severa
CAPE	Energia Potencial Convectiva Disponível (J/kg)
CAPV	CAPE calculado utilizando temperatura virtual
CINS	Inibição de Convecção (J/kg)
CINV	CINS calculado utilizando temperatura virtual
EQLV	Nível de Equilíbrio (hPa)
EQTV	EQLV calculado utilizando temperatura virtual
LFCT	Nível de Convecção Livre (hPa) por comparação da temperatura entre uma parcela e a superfície
LFCV	LFCT calculado utilizando temperatura virtual
BRCH	Número de Bulk Richardson
BRCV	BRCH calculado utilizando CAPV
LCLT	Temperatura (K) no LCL, o nível de condensação em elevação, de uma média dos 500 metros mais baixos
LCLP	Pressão (hPa) no LCL, o nível de condensação em elevação, de uma média dos 500 metros mais baixos
MLTH	Camada média de mistura THTA (K)
MLMR	Camada média de mistura MIXR (g/kg)
THTK	Espessura entre 1000 mb e 500 (m)
PA	Água Precipitável (mm) em toda sondagem

² Dados obtidos do sítio de Internet da Universidade do Wyoming, acessado em 21 de julho de 2024, no endereço: <http://weather.uwyo.edu/upperair/indices.html>.