



Departamento de  
**Meteorologia**

Universidade Federal do Rio de Janeiro  
Centro de Ciências Matemáticas e da Natureza  
Instituto de Geociências

**PREVISÃO DE EVENTOS METEOROLÓGICOS CONVECTIVOS  
UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA**

MARCELO PEREIRA NUNES

Rio de Janeiro

Abril de 2023

PREVISÃO DE EVENTOS METEOROLÓGICOS CONVECTIVOS  
UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA

MARCELO PEREIRA NUNES

Dissertação de Mestrado apresentada ao 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 à obtenção do título de Mestre em Ciências (área: Meteorologia).

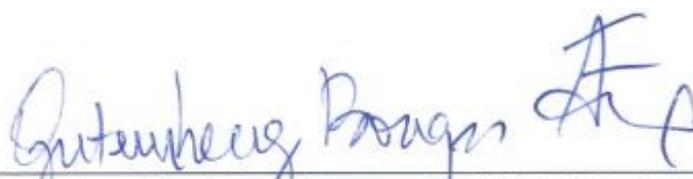
Orientadores: Gutemberg Borges França, *PhD.*  
Vinícius Albuquerque de Almeida, *DSc.*

Rio de Janeiro  
Abril de 2023

PREVISÃO DE EVENTOS METEOROLÓGICOS CONVECTIVOS  
UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA

MARCELO PEREIRA NUNES

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. Dr. Vinicius Albuquerque de Almeida - UFRJ - Orientador



Prof. Dr. Wallace Figueiredo Menezes - PPGM/UFRJ



Prof. Dr. Manoel Valdonel de Almeida - UFRJ



Prof. Dr. Enio Pereira de Souza - UFCG

RIO DE JANEIRO, RJ – BRASIL

Abril de 2023

Pereira-Nunes, Marcelo

Previsão de eventos meteorológicos convectivos utilizando técnicas de aprendizado de máquinas / Marcelo Pereira Nunes. - Rio de Janeiro: UFRJ/PPGM/IGEO/CCMN, 2023.

XVIII, 90 p.: il.; 29,7 cm.

Orientador: Gutemberg Borges França

Vinícius Albuquerque de Almeida

Dissertação (Mestrado) -  
UFRJ/PPGM/IGEO/CCMN Programa de Pós-graduação em Meteorologia, 2023.

Referências Bibliográficas: p. 74-86.

1. Eventos Meteorológicos Convectivos. 2. Índices de Instabilidade Termodinâmicos 3. Aprendizado de Máquina. 4. Meteorologia Aeronáutica. I. França, Gutemberg Borges *et al.* II. Universidade Federal do Rio de Janeiro, PPGM/IGEO/CCMN, Programa de Pós-graduação em Meteorologia. III. Título.

## AGRADECIMENTOS

Agradeço a minha família, meus pais Adão e Maria, minha esposa Rose, meus filhos Lucas e Leandro pela compreensão e paciência durante os momentos em que não pude me dedicar a eles.

Ao orientador Prof. Gutemberg Borges França pelo direcionamento, dedicação e confiança depositada que permitiram o desenvolvimento deste trabalho.

Ao orientador Prof. Vinícius Albuquerque de Almeida, pelos ensinamentos e incentivos transmitidos durante o curso e ao longo da pesquisa.

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.

Ao Instituto de Controle do Espaço Aéreo (ICEA) pela gentileza de disponibilizar os dados modelados dos índices de instabilidade termodinâmicos do modelo *Weahter Research and Forecasting* (WRF).

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

Ao amigo Jimmy Nogueira de Castro, meu colega de serviço no CIMAER, pelas valiosas sugestões para a pesquisa.

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

Resumo da Dissertação apresentada ao PPGM/IGEO/CCMN/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.).

## PREVISÃO DE EVENTOS METEOROLÓGICOS CONVECTIVOS UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA

Marcelo Pereira Nunes

Abril/2023

Orientadores: Gutemberg Borges França

Vinícius Albuquerque de Almeida

Programa de Pós-graduação em Meteorologia

Esta pesquisa desenvolveu modelos, baseados em aprendizado de máquina (AM), para previsão de 16 horas e 4 horas de ocorrência de evento meteorológico convectivo (EMC), assim como classificar a severidade do fenômeno meteorológico com 4 horas de antecedência, na Área de Controle Terminal de Brasília. No treinamento e teste dos algoritmos de AM foram utilizados os índices de instabilidade termodinâmicos (IIT) clássicos (entrada), obtidos dos perfis atmosféricos das radiossondagens de Brasília e descargas atmosféricas (DA) para caracterização do EMC (saída), considerando período de 2012 a 2017. Foi avaliada a aplicabilidade dos modelos ótimos desenvolvidos, utilizando como dados de entrada IIT modelados extraídos do modelo *Weather Research and Forecasting* (WRF) em duas simulações que utilizaram configurações diferentes de parametrização com capacidade de previsão de 30 horas. Para detecção de EMC, os resultados mostraram que os melhores modelos obtiveram POD, 1-FAR, F-MEASURE e KAPPA com valores respectivamente maiores que 0,90, 0,80, 0,90, 0,80 e BIAS variando de 0,89 e 1,12. Para a detecção da gravidade do evento, o modelo apresentou os seguintes valores estatísticos (entre parênteses): POD (0,82), 1-FAR (0,78), F-

MEASURE (0,82), KAPPA (0,59 ) e BIAS (0,97). Os modelos ótimos aplicados aos IIT modelados tiveram resultados promissores para previsão de EMC, indicando que podem ser o primeiro sinal de alerta para ocorrência do evento. Os resultados da aplicação dos modelos em trinta dias (*hindcast*), mostraram que, em geral, o conjunto de modelos com tempos de entrada de 16 horas e 4 horas foram capazes de detectar EMC (e severidade entre parênteses), no período de previsão entre 15 e 22h (local), com alta precisão e com taxas de probabilidade de acerto variando de 0,87 a 0,90 e 0,93 a 1,00 (0,97), respectivamente. Três erros do modelo na previsão do evento são discutidos, quando um baixo número de DA ocasionou falsos negativos e a ocorrência de EMC severos próximos à área de estudo ocasionou um falso positivo.

**Palavras-chave:** aprendizado de máquina, evento meteorológico convectivo, índices de instabilidade termodinâmicos, Weather Research and Forecasting, descargas atmosféricas.

Abstract of Dissertation presented to PPGM/IGEO/CCMN/UFRJ as a partial fulfillment of requirements for the degree of Master of Science (M.Sc.).

Marcelo Pereira Nunes

April/2023

Advisors: Gutemberg Borges França

Vinícius Albuquerque de Almeida

Department: Meteorology

This research developed models, based on machine learning (ML), for forecasting 16 hours and 4 hours of convective meteorological event (CME) occurrence, as well as classifying the severity of the meteorological phenomenon 4 hours in advance, in the Brasilia Terminal Control Area. In the training and testing of the ML algorithms, the classic thermodynamic instability indices (TII) (input) were used, obtained from the atmospheric profiles of radiosondes in Brasilia and atmospheric discharges (AD) to characterize the CME (output), considering the period from 2012 to 2017. The applicability of the developed optimal models was evaluated, using modeled TII data extracted from the Weather Research and Forecasting (WRF) model as input data in two simulations that used different parameterization settings with a 30-hour forecasting capacity. For CME detection, the results showed that the best models obtained POD, 1-FAR, F-MEASURE and KAPPA with values respectively greater than 0.90, 0.80, 0.90, 0.80 and BIAS ranging from 0, 89 and 1.12. To detect the severity of the event, the model presented the following statistical values (in parentheses): POD (0.82), 1-FAR (0.78), F-MEASURE (0.82), KAPPA (0.59 ) and BIAS (0.97). The optimal models applied to the modeled IIT had promising results for CME prediction, indicating that they may be the first warning sign for the occurrence of the event. The results of applying the models in thirty days (hindcast) showed

that, in general, the set of models with input times of 16 hours and 4 hours were able to detect CME (and severity in parentheses), in the forecast period between 15 and 22h (local), with high precision and with hit probability rates ranging from 0.87 to 0.90 and 0.93 to 1.00 (0.97), respectively. Three model errors in predicting the event are discussed, when a low number of ADs caused false negatives and the occurrence of severe CME close to the study area caused a false positive.

**keywords:** machine learning, convective meteorological event, thermodynamic instability indices, Weather Research and Forecasting , lightning.

## SUMÁRIO

<b>LISTA DE FIGURAS .....</b>	<b>XII</b>
<b>LISTA DE TABELAS .....</b>	<b>XIV</b>
<b>LISTA DE ABREVIATURAS E SIGLAS .....</b>	<b>XVI</b>
<b>1. INTRODUÇÃO .....</b>	<b>19</b>
1.1 Motivação .....	21
1.2 Objetivo .....	24
<b>2. REVISÃO BIBLIOGRÁFICA .....</b>	<b>25</b>
<b>3. FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>31</b>
3.1 Eventos Meteorológicos Convectivos .....	31
3.2 Descargas Atmosféricas .....	32
3.3 Índices de Instabilidade Termodinâmicos .....	34
3.3.1 ÍNDICE K (K) .....	35
3.3.2 SHOWALTER (SH) .....	36
3.3.3 TOTAL TOTALS (TT) .....	37
3.3.4 SWEAT (SW) .....	38
3.3.5 ENERGIA POTENCIAL DISPONÍVEL PARA CONVECÇÃO (CAPE) .....	39
3.4 Informe Meteorológico de Aeródromo (METAR) .....	40
3.5 Weather Research and Forecasting (WRF) .....	40
3.6 Inteligência Artificial .....	41
<b>4. DADOS E MÉTODO .....</b>	<b>44</b>
4.1 Área de Estudo .....	44

4.2 Dados.....	45
4.2.1 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS .....	46
4.2.2 DADOS DE DESCARGAS ATMOSFÉRICAS.....	46
4.2.3 DADOS DE METAR.....	46
4.2.4 DADOS MODELADOS DO WRF .....	46
4.3 Método .....	46
4.3.1 Métrica de Avaliação dos Resultados .....	51
<b>5. RESULTADOS E DISCUSSÕES.....</b>	<b>54</b>
5.1 Análise da distribuição dos dados de descargas atmosféricas .....	54
5.2 Análise dos dados de Radiossondagens .....	57
5.3 Definição dos dados modelados do WRF .....	59
5.4 Treinamento e teste dos algoritmos de AM.....	60
5.4.1 Treinamento e teste com os dados das radiossondagens das 00 UTC .....	61
5.4.2 Treinamento e teste com os dados das radiossondagens das 12 UTC .....	62
5.5 Análise estatística dos modelos ótimos testados com os dados de IIT modelados do WRF .....	63
5.6 Estudo de Casos.....	67
5.6.1 Análise sinótica dos dias 25 de outubro de 2013 e 26 de outubro de 2013.....	69
5.6.2 Análise sinótica do dia 28 de outubro de 2017 .....	70
<b>6. CONCLUSÕES.....</b>	<b>72</b>
<b>7. REFERÊNCIAS.....</b>	<b>74</b>

**ANEXO I..... 87**

## LISTA DE FIGURAS

<b>Figura 1:</b> Ranking de movimentos dos aeroportos em 2020. Fonte: adaptado do Anuário Estatístico de Tráfego Aéreo de 2020, disponível em <a href="https://www.decea.mil.br/">https://www.decea.mil.br/</a> .	23
<b>Figura 2:</b> Estágios de uma nuvem CB: (a)Cumulus; (b)Maduro; (c)Dissipação. Fonte: Adaptado de ELAT (2019).	32
<b>Figura 3:</b> Distribuição mundial das estações da STARNET. Fonte: STARNET (2022).	34
<b>Figura 4:</b> Exemplo de mensagem METAR com destaque para a ocorrência de chuva forte com trovoadas e visibilidade de 400 metros para SBBR (Aeroporto Internacional Juscelino Kubitschek) (em destaque sublinhado vermelho feito pelo autor). Fonte: REDEMET (2022).	40
<b>Figura 5:</b> Área de estudo em destaque. Está indicado o aeroporto de Brasília (SBBR).	44
<b>Figura 6:</b> Fluxograma dos passos metodológicos da pesquisa.	47
<b>Figura 7:</b> Domínios selecionados para os experimentos com IIT modelados extraídos do modelo WRF, com grades 27 km (D01), 9 km (D02) e 3 km (D03).	49
<b>Figura 8:</b> Número de DA por mês e a respectiva porcentagem de ocorrência mensal.	54
<b>Figura 9:</b> Ocorrências de DA por estação do ano e a respectiva porcentagem de ocorrência sazonal.	55
<b>Figura 10:</b> Mapa de calor dos meses de novembro, dezembro, janeiro, fevereiro e março das descargas atmosféricas para a área de estudo no período de 2012 a 2017.	56
<b>Figura 11:</b> Distribuição horária das DA detectadas entre 2012-2017 e a respectiva porcentagem de ocorrência horária.	57

- Figura 12:** Distribuição das radissondagens 00 UTC e 12 UTC realizadas no período de 2012 a 2017. .... 58
- Figura 13:** Planilha do Excel contendo parte dos dados dos IIT das 12Z..... 59
- Figura 14:** Imagens de satélite GOES-13, temperatura realçada, do dia 25 de outubro de 2013. Fonte: Adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>. .... 69
- Figura 15:** Imagens de satélite GOES-13, temperatura realçada, do dia 26 de outubro de 2013. Fonte: Adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>. .... 70
- Figura 16:** Imagens de satélite GOES-16, canal Infravermelho (CH13), do dia 28 de outubro de 2017. Fonte: Adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>. .... 71

## LISTA DE TABELAS

<b>Tabela 1:</b> Valores de índice K associados à possibilidade de ocorrências de tempestades. Fonte: adaptado de George (1960).....	36
<b>Tabela 2:</b> Valores de SH associados à previsão de tempestades. Fonte: Adaptado de Showalter (1947).....	37
<b>Tabela 3:</b> Valores de TT associados à previsão de tempestades. Fonte: Adaptado de Miller (1972). .....	38
<b>Tabela 4:</b> Valores de SW associados à previsão de tempestades. Fonte: Adaptado de Miller (1972). .....	39
<b>Tabela 5:</b> Valores de CAPE associados à previsão de tempestades. Fonte: Adaptado de Houze (2014). .....	40
<b>Tabela 6:</b> Conjunto de dados utilizado na pesquisa no período de estudo de 2012 a 2017. ....	45
<b>Tabela 7 -</b> Parametrizações utilizadas nas simulações I e II do modelo WRF.	49
<b>Tabela 8:</b> Matriz de Confusão. Fonte: Adaptado de Wilks (2006).....	51
<b>Tabela 9:</b> Valores de referência do índice KAPPA. <b>Fonte:</b> Adaptado de Landis e Koch (1977).....	53
<b>Tabela 10:</b> (a) e (b) representam respectivamente o número mensal de sondagens atmosféricas realizadas e total em porcentagem nos horários de 00 UTC e 12 UTC.....	58
<b>Tabela 11 –</b> Dias selecionados para estudo de caso no modelo WRF com o horário e o número de ocorrências de DA.....	60
<b>Tabela 12:</b> Análise estatística dos melhores algoritmos de AM para previsão de EMC, na TMA-BR, para o PP, por intervalo de DA e entrada com os dados dos IIT extraídos das radiossondagens 00 UTC. ....	62

- Tabela 13:** Análise estatística dos melhores algoritmos de AM para previsão de EMC, na TMA-BR, para o PP, por intervalo de DA e entrada com os dados dos IIT extraídos das radiossondagens 12 UTC. .... 63
- Tabela 14:** Resultado da detecção de DA para o PP de EMC usando os modelos ótimos das 12 UTC e, como dados de entrada, os valores dos IIT extraídos do modelo WRF em duas simulações com esquemas de parametrização distintos. .... 65
- Tabela 15:** Valores das estatísticas para os oito modelos da Tabela 14 . .... 66
- Tabela 16:** Resultado do estudo de caso com 30 dias de previsão de EMC (e severidade) para período entre 15 e 22h (local), usando modelos ótimos com os IIT (dados de entrada) das 00 UTC (modelos dos experimentos de números 2 a 7) e 12 UTC (modelos dos experimentos de números 22 a 29 e 36 para detecção de severidade)..... 68

**LISTA DE ABREVIATURAS E SIGLAS**

a	Verdadeiro Positivo
AD	Árvores de decisão
AMS	Sociedade Americana de Meteorologia
APP	Controle de Aproximação
ATC	Controle de Tráfego Aéreo
ATM	Gerenciamento do Tráfego Aéreo
b	Falso Negativo
B	Tendência de uma previsão
c	Falso positivo
CAPE	Energia potencial disponível para convecção
CB	<i>Cumulonimbus</i>
CGNA	Centro de Gerenciamento de Navegação Aérea
CIMAER	Centro Integrado de Meteorologia Aeronáutica
CSI	Índice de sucesso crítico
d	Verdadeiro negativo
DA	Descargas Atmosféricas
DECEA	Departamento de Controle do Espaço Aéreo
EMA	Estação Meteorológica de Altitude
EMC	Evento meteorológico convectivo
Excel	Editor de planilhas produzido pela Microsoft
F	F-Measure
FAR	Taxa de falso alarme
FIR	Região de Informação de Voo
FL	Nível de voo
ft	Altitude medida em pés

g	Aceleração gravitacional
GAMET	Previsão de tempo significativo em baixos níveis
GOES	Satélite ambiental operacional geoestacionário
h	Hora local
hPa	Hectopascal
IA	Inteligência artificial
IBM	Corporação de negócios de máquinas internacional
ICEA	Instituto de Controle do Espaço Aéreo
IIT	Índice de instabilidade termodinâmico
Jkg <sup>-1</sup>	Joule por quilo grama
K	Índice K
K'	KAPPA
KHZ	Quilohertz
km	Quilômetro
kt	Nó
METAR	Informe meteorológico de aeródromo
MICAPE	Energia potencial disponível para convecção calculada a partir de uma parcela de ar instável.
MLCAPE	Energia potencial disponível para convecção calculada a partir de uma parcela de ar média.
NASA	Administração Nacional da Aeronáutica e Espaço
NCAR	Centro Nacional para Pesquisa Atmosférica
NCE	Nível de convecção espontânea
NCEP	Centro Nacional de Previsão Ambiental
NE	Nível de equilíbrio
NOAA	Administração Nacional Oceânica e Atmosférica
OACI	Organização da Aviação Civil Internacional
POD	Probabilidade de detecção

PP	Período de Previsão
$\rho_0$	Taxa de aceitação relativa
$\rho_e$	Taxa hipotética de aceitação
QGIS	Software de informação geográfica livre e aberto
r	Razão de Mistura
REDEMET	Rede de Meteorologia do Comando da Aeronáutica
RMRJ	Região Metropolitana do Rio de Janeiro
SBCAPE	Energia potencial disponível para convecção calculada a partir de uma parcela de ar de superfície.
SBLI	Índice de Levantamento calculado a partir da parcela de ar de superfície
SH	Índice <i>Showalter</i>
SIGMET	Informe meteorológico em rota
STARNET	<i>Sferics Timing and Ranging NETWORK</i>
SW	Sweat
T	Temperatura do bulbo seco
TAF	Previsão de aeródromo
TD	Temperatura do ponto de orvalho
TMA	Área de controle terminal
TSRA	Trovoada com chuva
TT	<i>Total Totals</i>
Tv	Temperatura virtual
UTC	Tempo universal coordenado
VLF	Frequência muito baixa
WEKA	<i>Waikato Environment for Knowledge Analysis</i>
WRF	<i>Weather Research Forecasting</i>

## 1. INTRODUÇÃO

A convecção é um fator importante no controle da qualidade do ar por intermédio da dissipação das massas de ar poluídas na camada limite atmosférica (BARTHLOTT *et al.*, 2006), mas também pode dar origem a eventos meteorológicos convectivos (EMC) que têm potencial para causar grande influência na vida cotidiana humana, pois podem gerar tempestades severas que têm possibilidade de produzir chuvas fortes com ou sem granizo, turbulência, rajadas de vento, visibilidade baixa e trovoadas (ORVILLE, 1965; BANTA, 1990; BARTHLOTT *et al.*, 2006; LIMA & GOMES 2007; ALMEIDA, 2020), prejudicando a economia, causando danos materiais e perdas de vida. Mesmo eventos convectivos comuns produzem ventos fortes que podem resultar em alto custo financeiro e danos fatais (BHUIYAN *et al.* 2019).

A aviação é bastante sensível às condições atmosféricas. O aumento do tráfego aéreo global exige um melhor aproveitamento do espaço aéreo e, assim, com a devida segurança, diminuindo o espaçamento entre as aeronaves (ALMEIDA, 2020). Sendo assim, as informações meteorológicas passam a ser cada vez mais fundamentais, pois os procedimentos de pouso e decolagens nos aeroportos são frequentemente prejudicados por baixa visibilidade devido a fortes chuvas, cortante de vento e vento de rajada associado a eventos convectivos. Estes, ainda, em rota, podem causar desvios e, conseqüentemente, consumo extra de combustível pelas aeronaves, além dos transtornos para a tripulação devido a turbulência de ar claro induzida pela convecção (GULTEPE *et al.*, 2019; DA SILVA, 2021).

Nesse contexto, o Gerenciamento do Tráfego Aéreo (ATM, em inglês) segue regras rigorosas de segurança, tendo em vista garantir voos mais seguros, pontuais, regulares e eficazes, respeitando as condições meteorológicas atuantes e as limitações operacionais da infraestrutura aeronáutica.

Dentro do espaço aéreo controlado, área de dimensões e forma definidas, dentro do qual é prestado o serviço de controle de tráfego aéreo, há uma área denominada Controle de Aproximação (APP, em inglês) que visa, sobretudo, a separação entre as aeronaves que estejam executando procedimentos para chegar ou partir de um aeródromo (DECEA, 2016). Sob jurisdição do APP, em aeroportos com grande

densidade de tráfego aéreo, geralmente ao redor de aeroportos importantes, existe a Área de Controle Terminal (TMA, em inglês), onde a aeronave aguarda o pouso ou a decolagem e, portanto, sensível às condições meteorológicas.

Existem vários fatores que podem alterar o fluxo do tráfego aéreo dentro da TMA, como alto volume de tráfego, condições meteorológicas adversas, ativação do espaço aéreo militar, entre outros. Quando novas rotas aéreas são estabelecidas devido às condições meteorológicas adversas, dependendo do volume de tráfego aéreo e da evolução das condições meteorológicas, pode ocorrer aumento de separação entre as aeronaves, ocasionando atrasos, cancelamentos de voos, desvios de rota e, conseqüentemente, maior consumo de combustível.

Sendo assim, um EMC dentro da TMA exige vigilância constante por parte dos previsores, sendo de fundamental importância a caracterização desses fenômenos e o conhecimento do processo de sua formação com o intuito de melhorar a acurácia de suas previsões.

No Brasil, a responsabilidade de prestar o serviço meteorológico aeronáutico de vigilância e previsão é do Departamento de Controle do Espaço Aéreo (DECEA, <https://www.decea.mil.br/>) mediante o Centro Integrado de Meteorologia Aeronáutica (CIMAER, <https://www2.fab.mil.br/cimaer/>). Outro setor importante do DECEA é o Centro de Gerenciamento de Navegação Aérea (CGNA, <https://www2.fab.mil.br/cgna/>), que atua em conjunto com as companhias aéreas e precisa saber com antecedência as condições meteorológicas previstas para as áreas de aproximação para pouso dos aeroportos, TMA e rotas, no intuito de mitigar o impacto das condições de tempo adversas na segurança e navegação aérea.

Nesse cenário, faz-se necessário o desenvolvimento de ferramentas práticas de previsão de curto prazo de EMC que auxiliem o previsor a prestar uma informação mais rápida e assertiva.

Dessa forma, as pesquisas que buscam melhorar a qualidade e precisão dos métodos de previsão do tempo são imprescindíveis para a aviação, pois contribuirão, principalmente, para o aumento da segurança da navegação. O Laboratório de Meteorologia Aplicada (LMA), situado no Instituto de Geociências da Universidade Federal do Rio de Janeiro, vem desenvolvendo desde 2009 pesquisas aplicadas à

Meteorologia Aeronáutica a partir do grupo de pesquisa da Cátedra de Meteorologia Aeronáutica do qual um dos objetivos é auxiliar o desenvolvimento de técnicas de diagnóstico e prognóstico de fenômenos meteorológicos impactantes à navegação aérea.

### **1.1 Motivação**

O planejamento é um dos principais fatores para que o voo seja realizado de forma segura. Consiste na obtenção de uma ampla gama de informações prévias, incluindo as condições meteorológicas atuais e previstas do aeródromo de partida, da rota do voo, do aeródromo de chegada e do aeródromo de alternativa que é definido como o aeródromo para o qual uma aeronave poderá prosseguir quando for impossível ou desaconselhável dirigir-se ou efetuar o pouso no aeródromo de destino previsto (DECEA,2016). Para o piloto, uma das condições meteorológicas mais preocupantes é a presença de EMC no seu planejamento, pois além de poderem causar o cancelamento do voo, a fim de se evitar um acidente grave, podem causar desvios na rota e, por consequência, maior consumo de combustível. Nesse contexto, o previsor tem o papel primordial, pois deve fazer uma análise minuciosa do estado da atmosfera e fornecer aos aeronavegantes e aos órgãos de controle do espaço aéreo uma previsão meteorológica clara e mais precisa possível.

A previsão local de tempestades, que se formam em um curto intervalo de tempo e não oriundas de sistemas frontais, é um desafio constante para os previsores dos centros operacionais de meteorologia devido à grande variação espaço-temporal desse fenômeno meteorológico. Segundo França *et al.* (2016), os meteorologistas utilizam sua experiência para integrar, de forma subjetiva, as diferentes observações meteorológicas e/ou saídas de modelos atmosféricos para conseguirem estabelecer um modelo meteorológico conceitual e assim, realizarem previsões para os aeroportos sob sua responsabilidade.

De acordo com Organização da Aviação Civil Internacional (OACI), o tráfego aéreo global tem crescido constantemente desde as últimas décadas (ICAO, 2016), exigindo dos órgãos de gerenciamento de tráfego aéreo uma melhor otimização do espaço aéreo, e assim, com a devida segurança, diminuindo o espaçamento entre

elas (ALMEIDA, 2020). Nesse contexto, as informações meteorológicas passam a ser cada vez mais decisivas.

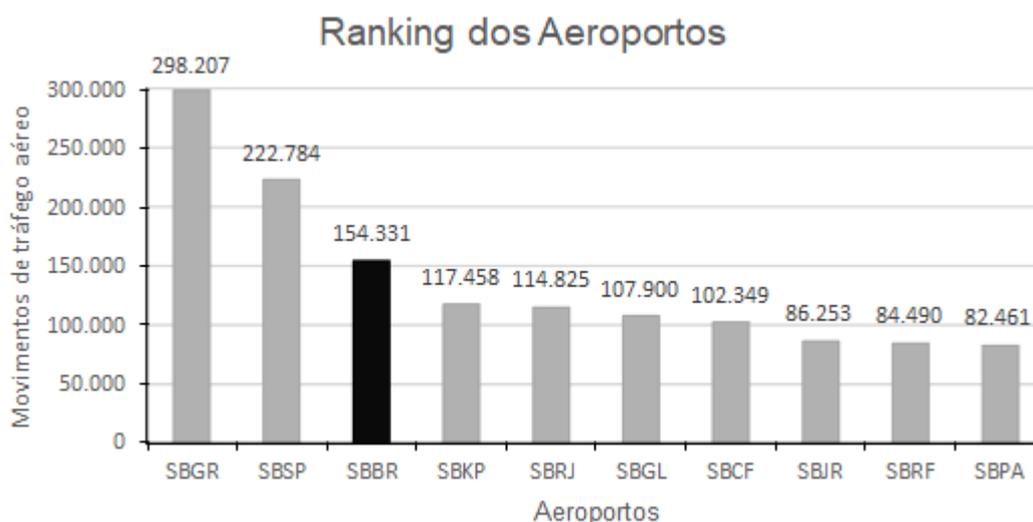
A fim de antecipar a adoção de estratégias e realizar a veiculação adequada de alertas com o objetivo de eliminar ou minimizar os impactos negativos causados pela atuação de sistemas convectivos severos, é necessário identificar, com antecedência de várias horas, condições favoráveis à formação desses fenômenos meteorológicos (NASCIMENTO, 2005). Este, ainda, declara que para definir uma estratégia de previsão de tempo convectivo severo é necessário conhecer melhor os ambientes atmosféricos sinóticos e de mesoescala propícios ao desenvolvimento de tempestades severas no Brasil; estudar procedimentos que maximizem a extração de informação relevante dos dados observados e de modelos de mesoescala que identifiquem estes ambientes atmosféricos, visando uma implementação operacional.

Sendo assim, os previsores levam em consideração um conjunto de ações, análise de dados e informações que sugerem razoável subjetividade e probabilidade. Essas informações podem sofrer mudanças dependendo do ponto de vista do profissional e da maneira como se utilizam as ferramentas disponíveis.

No âmbito do CGNA e CIMAER, as solicitações de previsões do tempo são constantes. No CGNA, devido à preocupação do órgão de Controle de Tráfego Aéreo (ATC, em inglês) e das empresas aéreas a respeito de sistemas convectivos que estejam atuando sobre alguma aerovia ou aeroporto, esse profissional elabora diversos briefings meteorológicos, além de atender solicitações frequentes de previsões meteorológicas de curto prazo ou *nowcasting* de 30 minutos a 2 horas. No CIMAER, além de manter vigilância constante dos fenômenos meteorológicos que possam impactar a navegação aérea no Brasil, uma das principais atribuições dos previsores é a confecção de códigos meteorológicos específicos para as Regiões de Informação de Voo (*Flight Information Region – FIR*), que são as áreas do espaço aéreo dentro das quais são prestados serviços de informação de voo e alerta, como,

por exemplo, as mensagens GAMET<sup>1</sup> e SIGMET<sup>2</sup>, assim como informar as condições meteorológicas previstas para diversos aeródromos do território brasileiro através do código de previsão de aeródromo (TAF).

Devido ao impacto que os EMC causam no fluxo de navegação área e os briefings e mensagens meteorológicas rotineiras confeccionadas pelo CIMAER e CGNA, fica evidenciada a preocupação dos previsores com a ocorrência ou previsão desse fenômeno meteorológico, caracterizado pela ocorrência de descargas atmosféricas (DA). Vários trabalhos, sobre esse tema, foram desenvolvidos para a TMA de São Paulo e TMA do Rio de Janeiro, mas há poucos estudos que abordam esse assunto na região Centro-Oeste. Este trabalho, de forma pioneira, busca estudar o comportamento das DA na TMA de Brasília (TMA-BR), que abrange o Aeroporto Internacional Juscelino Kubitschek, localizado na cidade de Brasília no Distrito Federal. Segundo o anuário de tráfego aéreo (DECEA, 2021), esse aeroporto foi o terceiro mais movimentado do Brasil em 2020, conforme a Figura 1.



**Figura 1:** Ranking de movimentos dos aeroportos em 2020. Fonte: adaptado do Anuário Estatístico de Tráfego Aéreo de 2020, disponível em <https://www.decea.mil.br/>.

<sup>1</sup> O GAMET (*General Aviation Meteorological Forecasting*) é uma previsão de área, em linguagem clara e abreviada, contendo informações relativas a fenômenos meteorológicos em rota, entre a camada de superfície e o FL100 (*Flight Level* de 10000ft) ou FL150 (*Flight Level* de 15000ft), em regiões montanhosas, perigosos para voos em níveis baixos (DECEA, 2020).

<sup>2</sup> O SIGMET (*Signification Meteorological Information*) é uma mensagem que contém a previsão ou ocorrência de determinados fenômenos meteorológicos em rota que possam afetar a segurança das operações aéreas, e o desenvolvimento desses fenômenos no tempo e no espaço (DECEA, 2020).

Neste cenário, o desenvolvimento de ferramentas de apoio à tomada de decisão do meteorologista, visando a padronização de procedimentos, diminuindo a subjetividade e tornando as previsões mais rápidas e precisas poderão contribuir com um melhor planejamento do voo, proporcionando à escolha de rotas mais curtas e seguras.

## 1.2 Objetivo

Este trabalho tem dois objetivos principais, a saber: 1) Desenvolver um modelo de previsão de eventos meteorológicos convectivos baseado em aprendizado de máquina (AM) para a TMA-BR, considerando como preditores os índices de instabilidade termodinâmicos (IIT) extraídos dos perfis atmosféricos de radiossondagens e, como variável objetivo, a densidade dos registros de DA para identificação de áreas com potencial convectivo e, caso exista EMC, verificar sua severidade; e 2) Avaliar a aplicabilidade dos modelos desenvolvidos no item anterior, utilizando como dados de entrada IIT modelados obtidos do modelo *Weather Research and Forecasting* (WRF).

Para atingir os objetivos principais, foram delimitados os seguintes objetivos específicos:

- 1: Analisar a climatologia das descargas atmosféricas na área de estudo;
- 2: Avaliar o uso de algoritmos de AM como modelos de previsão de EMC; e
- 3: Avaliar o uso dos melhores algoritmos de AM com os dados modelados de IIT extraídos do WRF.

No Capítulo 2 deste trabalho será apresentada a revisão bibliográfica, no capítulo 3 será apresentada a fundamentação teórica, no capítulo 4 será apresentado os dados e métodos, no capítulo 5 será apresentado os resultados e discussões e no capítulo 6 será apresentada as conclusões.

## 2. REVISÃO BIBLIOGRÁFICA

O impacto dos processos atmosféricos na aviação é reconhecido desde o ano 1900 (GULTEPE *et al.*, 2019). Nas últimas décadas, diversos estudos a respeito do comportamento e previsão de tempestades foram realizados. Nesses estudos, o uso de radares meteorológicos foram amplamente usados como fonte primária na análise do comportamento de tempestades convectivas, como Merritt (1969), Rasmussen *et al.* (1992), Ryzhkov *et al.* (2002), Schuur *et al.* (2012), Smith *et al.* (2016) e Hubbert *et al.* (2018) que usaram radares Doppler de dupla polarização com objetivo de detectar a intensidade e a atividade convectiva de tempestades, bem como as condições de formação de gelo dentro das nuvens para identificar as características microfísicas das partículas. Alguns estudos pioneiros também utilizaram dados de radar meteorológico, como Battan (1973) que usou dados observados de radar para estudar a consistência da taxa de precipitação; e Wilson e Wilk (1981) que realizaram pesquisas com radar Doppler e técnicas gráficas interativas na previsão de curto prazo de fenômenos meteorológicos.

Outros estudos à respeito de EMC severos foram desenvolvidos nas décadas de 1980 e 1990, utilizando dados de DA detectados remotamente e amostragens de perfis atmosféricos obtidos por radiossondagens. Por exemplo, Weisman e Klemp (1986) usaram índices de instabilidade termodinâmicos (IIT) adquiridos de radiossondagens para distinguir os tipos de tempestades observadas. Turman e Tettelbach (1980), Williams *et al.* (1989) e Kane (1991) aplicaram um sistema semelhante com capacidade de detectar DA em tempo real e, assim, usaram-no para analisar o comportamento de tempestades convectivas.

Ainda na década de 1990, o algoritmo TITAN (DIXON & WIENER, 1993) demonstrou uma metodologia para identificação automatizada em tempo real, rastreamento e previsão de curto prazo de tempestades utilizando dados de radar. Da mesma forma, Johnson *et al.* (1998) apresentaram o algoritmo *Storm Cell Identification and Tracking* (SCIT) para a identificação e o rastreamento preciso de tempestades convectivas.

Além do uso de radares meteorológicos e detectores de DA, as imagens de satélite têm sido utilizadas para identificar EMC (WILSON *et al.*, 1998). Estudos anteriores analisaram a capacidade de vários índices atmosféricos na previsão de eventos

meteorológicos convectivos, incluindo ponto de orvalho, índice K, energia potencial disponível para convecção (CAPE), inibição convectiva (CIN), entre outros (SCHULTZ, 1989; LEE & PASSNER, 1993; FUELBERG & BIGGAR, 1994; HUNTRIESER *et al.*, 1997; HAKLANDER & VAN DELDEN, 2003; e MANZATO, 2005). Estes índices têm sido utilizados em combinação com a experiência do meteorologista para prever a probabilidade e a gravidade de uma tempestade iminente (BHUIYAN *et al.* 2019).

Durante a Segunda Guerra Mundial, em 1940, houve a necessidade de se desenvolver uma tecnologia voltada para análise de balística, quebras de códigos e cálculos para projetos de armas nucleares, assim surgiram os primeiros estudos sobre inteligência artificial (IA) com os primeiros projetos de construção de computadores (LIMA *et al.*, 2014). Uma das personagens mais marcantes dessa época foi Alan Turing, que desenvolveu várias técnicas para acelerar a quebra do código da máquina Enigma, a fim de entender como os alemães se comunicavam e auxiliar as Forças Aliadas na Guerra (Franco, 2014).

Em 1950, Alan Turing publicou um artigo chamado “*Computing Machine and Intelligence*” e seu subsequente Teste de Turing, com o qual se pretendia descobrir se uma máquina podia ou não emular o pensamento humano, estabelecendo os princípios básicos da IA. No entanto, somente após a Conferência de Dartmouth, em 1956, foi definido o conceito de Inteligência artificial (IA), no qual uma máquina pode ser criada para simular as ações humanas (Moor, 2006).

Com o término da Segunda Guerra Mundial, o computador foi gradualmente sendo utilizado por outras áreas, além da militar e científica, estimulando pesquisas de software, hardware e linguagens de programação (LIMA *et al.*, 2014).

A partir dos anos 1970, houve uma expansão do uso da IA para a solução de problemas reais. No começo, os problemas eram tratados pela IA por meio de aquisição de conhecimento de especialistas de determinada área que era então codificado, frequentemente por regras lógicas, em um programa de computador, mas esse processo possuía várias limitações, como subjetividade e pouca cooperação por parte do especialista, por causa do receio de ser demitido após repassar seu conhecimento (FACELI *et al.*, 2021).

O aumento da complexidade dos problemas a serem computacionalmente tratados e a maior disponibilidade de dados gerados por diferentes setores, motivou o desenvolvimento de ferramentas computacionais mais sofisticadas e autônomas para aquisição de conhecimento. A maioria dessas ferramentas é baseada em Aprendizado de Máquina, um ramo da IA que faz parte de várias das tecnologias atualmente utilizadas (FACELI *et al.*, 2021), tendo como base de execução os algoritmos, que buscam detectar automaticamente padrões nos dados e, em seguida, fazer suposições sobre eles.

Na meteorologia, começaram a surgir trabalhos utilizando modelos numéricos de previsão do tempo e técnicas de aprendizado de máquina com o objetivo de prever a ocorrência de eventos convectivos. Assim, Mccann (1992) apresentou um método de previsão de tempestades severas de 3 a 7 horas, utilizando duas redes neurais a partir de variáveis baseadas na convergência de umidade da superfície. Combinou operacionalmente as duas redes neurais e obteve um único produto de hora em hora que aprimora o reconhecimento de padrões e que demonstrou potencial significativo do uso dessas redes para se fazer previsões de eventos meteorológicos convectivos. Mueller *et al.* (2003) desenvolveram um método de previsão de 1 hora para localizar tempestades com base em dados de estações meteorológicas de superfície, radares, satélites e modelagem numérica. Posteriormente, Isaac *et al.* (2006, 2011, 2012 e 2014) propuseram vários trabalhos relacionados à previsão do tempo para os principais aeroportos do Canadá, utilizando dados de modelos numéricos, observações de estações de superfície, radar, satélite e um radiômetro de micro-ondas para gerar previsões de até aproximadamente seis horas.

Recentemente, modelos não paramétricos baseados em aprendizagem de máquina estão se tornando bastante difundidas em vários estudos científicos, aplicados, por exemplo, ao estudo do clima (GUIKEMA *et al.*, 2010, 2014; GUIKEMA & QUIRING, 2012), na previsão de mudanças climáticas (BROWN & ROBINSON, 2011) e na previsão da velocidade do vento a 10 metros da superfície (BHUIYAN *et al.* 2019). Modelos não paramétricos de aprendizagem de máquina, como *random forests* (BREIMAN, 2001), foram populares na previsão do tempo e aplicação hidrometeorológica (BHUIYAN *et al.* 2017).

No Brasil, diversos trabalhos começaram a avaliar a previsão de EMC, analisando-se o comportamento dos índices de instabilidade termodinâmicos. Desse modo, Nascimento (2005) fez uma descrição atualizada de parâmetros atmosféricos úteis no auxílio à identificação de ambientes favoráveis à ocorrência de tempestades convectivas severas. Destacou que o cálculo de tais parâmetros atmosféricos fazendo o uso, dentre outras fontes de informação, de saídas de modelos de mesoescala, representa uma estratégia operacional de possível adoção no Brasil. Concluiu que alguns IIT, originalmente concebidos para as latitudes médias do hemisfério norte, podem ser úteis para a previsão de sistemas convectivos no Brasil, em particular para a região sul do país, e servir de base conceitual para a elaboração de índices mais adequados para as regiões tropicais do país.

Teixeira e Satyamurty (2007) sugeriram que variáveis dinâmicas e termodinâmicas obtidas através de modelagem numérica, como, por exemplo, o índice instabilidade por levantamento e a água precipitável, podem ser analisadas a fim de se detectar o potencial da atmosfera para a formação de EMC. Kunz (2007) avaliou a capacidade das variáveis meteorológicas obtidas de radiossondagens das 12 UTC na previsão de sistemas convectivos no sudoeste da Alemanha. Entre os melhores resultados encontrados pelo referido autor, o índice de instabilidade por levantamento apresentou 80% de probabilidade em prever se um dia de tempestade era esperado ou não. Nesse contexto, Polifke *et al.* (2016) usaram dados de precipitação do município do Rio de Janeiro, associados com quatorze indicadores atmosféricos (variáveis meteorológicas) obtidos do modelo WRF, entre eles, o índice K, índice *Total Totals*, água precipitável e a energia potencial disponível para convecção, com o objetivo de avaliar a capacidade de cada indicador na identificação e previsão das condições atmosféricas favoráveis à formação de chuvas severas. Verificou-se que os melhores indicadores foram a divergência do vento em 300 hPa, umidade específica em 850 hPa, energia total das tempestades severas em 500 hPa e água precipitável.

Outro aspecto importante no estudo das tempestades é a caracterização espaço-temporal das descargas atmosféricas (DA). Paulucci *et al.* (2017) realizaram um estudo climatológico sobre as características espaço-temporais de descargas atmosféricas ocorridas na região metropolitana do Rio de Janeiro no período de 2001 a 2016. De acordo com os autores, o nível de pico da atividade elétrica foi observado

aproximadamente entre 18h e 19h (hora local), quando há mais energia potencial disponível para convecção e o resultado da análise espacial revelou que a maioria das DA observadas sobre o Rio de Janeiro deriva do efeito orográfico, que estimula a formação de tempestades convectivas ao longo da parte sul da encosta.

Na tentativa de desenvolver um sistema de previsão a curtíssimo prazo para a TMA-RJ, França *et al.* (2016) realizaram uma investigação sobre a capacidade de uma rede neural em produzir sinais de alerta precoces ou ser uma ferramenta de previsão para EMC severos. Foram utilizados doze anos de dados de superfície, altitude, precipitação e descargas atmosféricas. O teste mostrou que o modelo proposto (ou rede neural) pode capturar o conteúdo físico dentro do conjunto de dados, e seu desempenho é bastante satisfatório para a primeira e segunda horas para previsão de EMC severos para a área de estudo.

Freitas *et al.* (2019) analisaram as descargas atmosféricas e sua previsão através de Inteligência Artificial, nesse caso, as Árvores de Decisão (AD), na Região Metropolitana do Rio de Janeiro (RMRJ), entre abril de 2000 a dezembro de 2016. A seleção das variáveis foi realizada através de análises e correlações. No geral, os resultados indicaram taxas de acertos acima de 80% e taxa de erro abaixo de 20% nos eventos. Nos estudos de casos, foi verificado que a representatividade espacial do conjunto de índices utilizado e, conseqüentemente, a previsão da AD ótima obtida é representativa além de 100 km dos limites da RMRJ definida para teste.

Almeida *et al.* (2020) apresentaram um método para a previsão de eventos meteorológicos convectivos na TMA-GL, utilizando dados de radiossondagem e sensoriamento remoto de descargas atmosféricas entre os anos de 2001 a 2016. Seis modelos de aprendizado de máquina foram treinados e validados por 10 anos (2001-2010), e um teste foi produzido por 6 anos (2011-2016). Os resultados mostraram que o algoritmo *deep learning* alcançou os melhores resultados para previsão de ocorrência e severidade de tempestades com base nas seguintes estatísticas: probabilidade de detecção (POD) (0,91 e 0,85), BIAS (1,03 e 1,07) e proporção de alarme falso (FAR) (0,12 e 0,20), respectivamente. A análise do teste de 6 anos mostrou que o modelo tem desempenho crescente para eventos de alto impacto, e esse desempenho diminui gradualmente à medida que os eventos se tornam mais fracos e mais frequentes.

Usando uma abordagem diferente para obtenção dos IIT, Soares *et al.* (2021) propuseram modelos de previsão de 6-8h dos eventos meteorológicos convectivos, baseado em algoritmos de AM, para duas áreas I e II correspondendo à rota aérea Rio de Janeiro – São Paulo. Os modelos tiveram como dados de entrada (input) cinco índices de instabilidade termodinâmicos extraídos de dados de sensoriamento remoto do satélite GOES-R, considerando os períodos de janeiro a março de 2018 e 2019. Os resultados indicaram que os algoritmos denominados *Multilayer Perceptron* e *Simple Logistic* foram os de melhor desempenho para a previsão de 8h e 6h com a probabilidade de detecção (0,86 e 0,94), taxa de falso alarme (0,14 e 0,08), viés (1,01 e 1,01), *F-measure* (0,86 e 0,94) e KAPPA (0,72 e 0,85), nas áreas I e II, respectivamente. O algoritmo *Random Forest* apresentou o melhor desempenho para detecção da severidade de 6h e 8h dos eventos e os valores das estatísticas foram superiores as anteriores para a área 2.

Continuando a pesquisa no uso das DA, associadas com os IIT obtidos de perfis atmosféricos das radiossondagens e algoritmos de aprendizado de máquina, De Castro *et al.* (2022) desenvolveram modelos para previsões de tempo convectivo severo na TMA-SP. A distribuição estatística definiu os locais, horários e a gravidade dos eventos convectivos. As estatísticas POD, 1-FAR, BIAS, KAPPA e *F-measure* foram usadas para avaliar a previsão de 5 horas de detecção de eventos convectivos (e entre parênteses se é severo quando a ocorrência de raios é maior ou igual a 1.000), obtendo valores de 0,91 (0,85), 0,95 (0,94), 0,92 (0,89), 0,74 (0,77) e 0,88 (0,95), respectivamente. Os resultados da aplicação do modelo por trinta dias (*hindcast*), mostraram que ele é eficaz, pois atingiu 96,7% de ocorrência e 86,7% se foram severos.

Com o objetivo de desenvolver modelos híbridos de previsão de eventos convectivos, utilizando IA e dados do modelo atmosférico regional WRF, Da Silva *et al.* (2022) aplicaram algoritmos de aprendizado de máquina associados ao modelo WRF, para prever a ocorrência de eventos convectivos na TMA-RJ. Nove modelos de aprendizado de máquina foram treinados e validados. Os resultados mostraram que a combinação dos algoritmos *AdaBoost.M1* e *Random Forest* alcançou os melhores resultados para previsão de tempestades com base nas seguintes estatísticas: POD (1,00), FAR (0,03) e CSI (0,98). O modelo teve um desempenho satisfatório na detecção e severidade dos EMC.

### 3. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresenta-se de forma sucinta a teoria sobre eventos meteorológicos convectivos, descargas atmosféricas, índices de instabilidade termodinâmicos, modelo WRF e os princípios da inteligência artificial enfatizando o uso dos algoritmos de AM para se fazer previsões.

#### 3.1 Eventos Meteorológicos Convectivos

Os EMC podem dar origem a tempestades convectivas severas e segundo Johns e Doswell (1992) e Moller (2001). Esses sistemas convectivos são capazes de produzir granizo grande (com pedras de 1,9 cm ou mais de diâmetro ao atingirem a superfície) e/ou rajadas de vento intensas (velocidade acima de 50 kt) e/ou tornados.

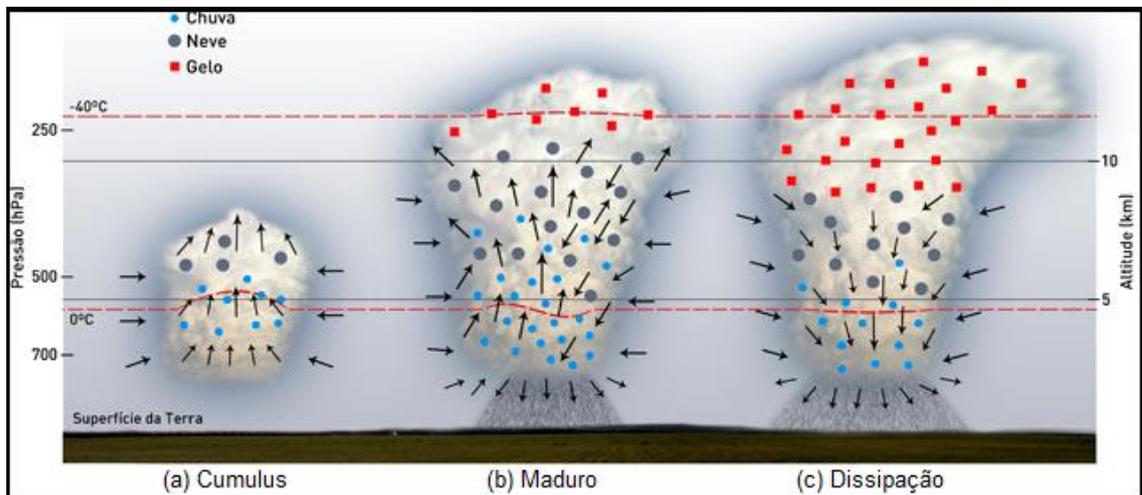
Muitos eventos convectivos têm ciclos de vida curtos e não causam danos significativos. Porém, se uma tempestade crescer, tornando-se severa, podem ocorrer inundações nos grandes centros urbanos, descargas elétricas que causam problemas nas linhas de energia, ventos fortes e tornados que derrubam casas e outras construções. Além disso, precipitações intensas e/ou rajadas de vento tornam o procedimento de pouso e decolagem das aeronaves mais perigosos.

Conforme Nascimento (2005), as tempestades severas são de baixa frequência em relação aos outros sistemas meteorológicos (como chuvas frontais, por exemplo), mas têm potencial de causar impactos significativos na sociedade.

As nuvens que provocam tempestades convectivas severas são chamadas de nuvens *cumulonimbus* (CB). O autor Byers (1951) identificou três estágios que compõe o ciclo de vida de uma nuvem CB, conforme descritos abaixo:

- Estágio *Cumulus*: Fase inicial, no qual as correntes ascendentes ocorrem em toda a nuvem, caracterizando o início da sua formação de acordo com a Figura 2(a);
- O Estágio de Maturidade: Inicia-se quando a precipitação atinge o solo, existindo na nuvem áreas de correntes ascendentes, descendentes e descargas elétricas. É nessa fase que ocorrem os fenômenos severos associados, como chuvas fortes, ocorrência de granizo e ventos intensos (Figura 2(b)); e

- Estágio de Dissipação: Fase final. Devido as fracas correntes ascendentes, a nuvem começa a se dissipar. Devido as correntes descendentes frias, a chuva resfria o ar nos níveis mais baixos e a instabilidade cessa como demonstrado na Figura 2(c).



**Figura 2:** Estágios de uma nuvem CB: (a)Cumulus; (b)Maduro; (c)Dissipação. Fonte: Adaptado de ELAT (2019).

### 3.2 Descargas Atmosféricas

Segundo o Glossário de Meteorologia da Sociedade Americana de Meteorologia [*American Meteorological Society (AMS)*], as DA, também conhecidas como raios ou relâmpagos, são descargas elétricas transitórias de alta corrente com percursos medidos em quilômetros. A fonte mais comum de relâmpagos é a carga elétrica separada em nuvens CB e mais da metade das DA ocorrem dentro dessas nuvens. Segundo Da Silva (2021), globalmente, esse tipo de descarga corresponde à aproximadamente 70% do valor total. Essa porcentagem varia com a latitude geográfica e é de aproximadamente 80-90% nas áreas próximas ao equador geográfico e em torno de 50-60% nas regiões de latitudes médias (Da Silva, 2021).

Conforme Nascimento (2005), a detecção de DA é fundamental para a previsão *nowcasting* de EMC e monitoramento do tempo (previsão de médio e longo prazos). É essencial prever com antecedência condições favoráveis à formação de tempestades severas, de forma a enviar mensagens de alerta que permitam que os órgãos ATC e empresas áreas antecipem a adoção de estratégias que eliminem ou mitiguem o impacto negativo destes fenômenos meteorológicos.

A detecção de descargas atmosféricas tem sido um assunto de pesquisa desde o final dos anos 1940 (CHRONIS & ANAGNOSTOU, 2003). Os primeiros estudos sobre DA foram realizados, principalmente, para investigar a propagação de ondas eletromagnéticas agitadas por descargas atmosféricas (SCHUMANN & KONIG, 1954). Hoje a pesquisa sobre DA está focada mais precisamente no monitoramento contínuo da precipitação de tempestades (GRECU *et al.*, 2000; GOODMAN, 1990), concentração atmosférica de ozônio (CHAMEIDES, 1986), segurança da aviação, previsão de tempo de curto prazo (ALEXANDER *et al.*, 1999) e estudos climáticos globais (WILLIAMS, 1992). Com o avanço da tecnologia para monitorar DA em grandes regiões (além de continentes) com base em receptores de rádio de frequência muito baixa (LEE, 1986<sup>a</sup>, b), as informações sobre descargas atmosféricas puderam ser usadas para entender os recursos de alta resolução de estimativa de chuva a partir de observações de satélite (Morales & Anagnostou, 2003).

A *Sferics Timing and Ranging NETWORK* (STARNET), versão experimental da rede de detecção de descargas atmosféricas surgiu em 1997, conforme informações disponibilizadas em <http://www.starnet.iag.usp.br/sistema.php>. A STARNET foi concebida pela empresa Resolution Display Inc a partir do programa de desenvolvimento de pequenas empresas da NASA. Essa rede é capaz de detectar os ruídos de rádio produzidos pelas descargas atmosféricas, com alta resolução temporal (5 minutos), através de um conjunto de antenas receptoras de VLF (*Very Low Frequency*, 7-15 KHz). A Figura 3 mostra a disposição espacial das antenas do sistema.



**Figura 3:** Distribuição mundial das estações da STARNET. Fonte: STARNET (2022).

Os dados de descargas atmosféricas coletados pela STARNET, caracterizam cada ocorrência, indicando a localização (latitude e longitude), quantidade e horário (UTC) dos eventos.

Neste trabalho, as descargas atmosféricas são utilizadas para caracterização espaço-temporal da severidade dos EMC.

### **3.3 Índices de Instabilidade Termodinâmicos**

Conforme descrito pela AMS: “um índice de instabilidade é qualquer quantidade que tente avaliar o potencial da atmosfera para atividade convectiva e que possa ser prontamente analisada a partir de dados obtidos por sondagens operacionais”.

Portanto, os valores assumidos pelos IIT permitem uma avaliação rápida da situação termodinâmica e do cisalhamento do vento em relação às condições de convecção no momento da coleta de dados. Devido aos valores numéricos dos IIT estarem associados à ocorrência de EMC (FREITAS *et al.*, 2019, ALMEIDA *et al.*, 2020, SOARES *et al.*, 2021 e DE CASTRO *et al.*, 2022), eles são utilizados com frequência

nos centros de previsão operacional auxiliando no prognóstico de ocorrência de eventos convectivos severos.

Os IIT podem ser calculados através dos dados de perfis atmosféricos obtidos através de radiossondagens realizadas pelas estações meteorológicas de altitude (EMAs). Conforme está descrito no Manual de Estações Meteorológicas de Altitude (DECEA, 2015), as EMAs destinam-se a coletar e processar os dados meteorológicos de temperatura, de umidade e de pressão, desde a superfície até níveis superiores da atmosfera, utilizando-se de sinais enviados por radiossonda lançada acoplada a um balão meteorológico.

Os valores de temperatura, direção e velocidade do vento, nos diversos níveis da atmosfera, são calculados a partir do posicionamento do balão em função do tempo e das coordenadas vertical e horizontal. Os dados coletados a partir de EMAs podem ser utilizados em diversas aplicações, a saber: previsão de ventos e temperaturas em altitude, turbulência, formação de gelo em aeronaves, cálculo de probabilidade de trovoadas, formação de nuvens, trilhas de condensação, bem como para emprego na área de pesquisa e de modelagem numérica.

Na sequência, serão descritos sucintamente cinco IIT utilizados neste trabalho como variáveis preditoras dos algoritmos da AM a serem treinados e avaliados, isto é: K, SHOWALTER (SH), SWEAT (SW), TOTAL TOTALS (TT) e Energia Potencial Disponível para Convecção (CAPE).

### 3.3.1 ÍNDICE K (K)

Indica o potencial para ocorrência de tempestades baseando-se no *Lapse Rate* (diminuição da temperatura com o aumento da altura) entre 850 hPa e 500 hPa ( $T_{850} - T_{500}$ ), a umidade em 850 hPa ( $TD_{850}$ ) e a depressão do Ponto de Orvalho em 700 hPa.

Este é definido conforme a Equação (01) (GEORGE, 1960):

$$K = (T_{850} + TD_{850}) - (T_{700} + TD_{700}) - T_{500}, \quad (01)$$

onde T e TD referem-se respectivamente à temperatura do bulbo seco e à temperatura do ponto de orvalho respectivamente e os números do lado indicam o

nível de pressão (em hPa) correspondente. Na Tabela 1, estão os valores de K de referência e a possibilidade de ocorrência de tempestades.

**Tabela 1:** Valores de índice K associados à possibilidade de ocorrências de tempestades. Fonte: adaptado de George (1960).

Índice K	Possibilidade de Tempestades
K < 20	Sem atividade Convectiva
20 < K < 25	Tempestades isoladas
25 < K < 30	Tempestades muito isoladas
30 < K < 35	Tempestades esparsas
K > 35	Muitas Tempestades

### 3.3.2 SHOWALTER (SH)

O SH indica o excesso de temperatura de uma parcela de ar em relação ao ambiente em 500 hPa e foi proposto por Showalter (1947) conforme a equação (02):

$$SH = T_{500} - T_{500pl850hPa}, \quad (02)$$

onde T500 é a temperatura ambiente em 500 hPa e T500pl850hPa é a temperatura de uma parcela de ar que sobe adiabaticamente desde o nível de 850 hPa até o nível de 500 hPa.

A Tabela 2 ilustra os valores de Showalter associados com a possibilidade de ocorrência de tempestades.

**Tabela 2:** Valores de SH associados à previsão de tempestades. Fonte: Adaptado de Showalter (1947).

Índice Showalter	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

### 3.3.3 TOTAL TOTALS (TT)

O índice TT foi descrito por Miller (1972) de acordo com a Equação (03). Leva em conta o *Lapse Rate* entre 850 hPa e 500 hPa ( $T_{850} - 2T_{500}$ ) e a umidade em 850 hPa ( $TD_{850}$ ).

$$TT = T_{850} + TD_{850} - 2(T_{500}), \quad (03)$$

onde a equação representa a soma das temperaturas do bulbo seco (T) e do ponto de orvalho (TD) em 850 hPa, subtraída pelo dobro da temperatura em 500 hPa. Deve-se ter atenção quando comparamos os valores de K e TT. Conforme mencionado por Dias (2000), quando temos um índice K e TT com valores elevados a tendência é que ocorra EMC mais prolongados com grandes acumulados de precipitação, mas com fenômenos associados menos intensos, já quando temos K com valores baixos e TT com valores altos é provável que tenhamos EMC severos com possibilidade de queda de granizo e rajadas de vento devido à existência de uma camada de ar seco logo acima da camada úmida, assim o entranhamento de ar ambiente em níveis acima de 700 hPa provoca forte evaporação que intensifica as correntes descendentes que produzem rajadas de vento em superfície.

A Tabela 3 exemplifica os valores de TT favoráveis a formação de tempestades.

**Tabela 3:** Valores de TT associados à previsão de tempestades. Fonte: Adaptado de Miller (1972).

Índice Total Totals (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 poucas pesadas; algumas isoladas severas
$50 \leq TT \leq 51$	Esparsas pesadas, algumas severas; tornados isolados
$52 \leq TT \leq 55$	Esparsas a numerosas pesadas, poucas a esparsas pesadas, alguns tornados
$TT > 45$	Numerosas a pesadas, pancadas de chuva esparsas; tornados esparsos.

### 3.3.4 SWEAT (SW)

O índice SW foi sugerido por Miller (1972) e é representado pela Equação (04):

$$SW = 20(TT - 49) + 12TD_{850} + 2V_{850} + V_{500} + 125[\sin(\Delta V_{500 - 850}) + 0.2] , \quad (04)$$

onde TT é definido como zero se inferior a 49 e indica a instabilidade atmosférica,  $TD_{850}$  é o ponto de orvalho em 850 hPa que indica a umidade em baixos níveis,  $V_{850}$  e  $V_{500}$  são as velocidades do vento de 850 hPa e 500 hPa que indicam a velocidade do jato de baixos níveis. A expressão  $\Delta V_{500 - 850}$  é a diferença entre a direção do vento em 500 hPa e 850 hPa. Valores acima de 250 são significativos (Dias, 2000). Nesse índice apenas contribuições positivas de cada termo são consideradas para o cálculo. Os termos negativos são igualados a zero.

A Tabela 4 apresenta os valores de índice SW associados à possibilidade de tempestades.

**Tabela 4:** Valores de SW associados à previsão de tempestades. Fonte: Adaptado de Miller (1972).

Índice Sweat	Possibilidade de Tempestades
$SW \leq +300$	Sem Potencial para Tempestades Severas
$+300 < SW \leq +400$	Potencial para Tempestades Severas
$SW > +400$	Potencial para tornados

### 3.3.5 ENERGIA POTENCIAL DISPONÍVEL PARA CONVECÇÃO (CAPE)

A CAPE mede a energia potencial disponível para a convecção, ou seja, analisa a capacidade da atmosfera para desenvolver convecção de acordo com o aquecimento de superfície. Como as tempestades são sustentadas pelo empuxo que as correntes ascendentes sofrem a partir da base da nuvem, integrando-se esse empuxo obtemos a CAPE e tem unidade de  $Jkg^{-1}$ :

$$CAPE = g \int_{NCE}^{NE} \frac{Tv(p) - Tv(a)}{Tv(a)} dz , \quad (05)$$

onde o  $g$  é a gravidade, NCE é o nível de convecção espontânea, NE é o nível de equilíbrio da parcela de ar (correspondente ao topo da nuvem),  $Tv(p)$  é a temperatura virtual da parcela e  $Tv(a)$  é a temperatura virtual do ambiente. A temperatura virtual é a temperatura que o ar seco deveria ter para ter a mesma densidade que o ar úmido, com a mesma pressão. Pode-se calcular a  $Tv$  usando a seguinte fórmula:

$$Tv = T \cdot (1 + 0,61 \cdot r) , \quad (06)$$

onde  $r$  é a razão de mistura definida como a relação entre a massa de vapor d'água em gramas existente em um quilograma de massa de ar seco.

Segundo Craven *et al.* (2002) há três tipos de CAPE: a CAPE calculada a partir de uma parcela à superfície (SBCAPE), a CAPE calculada a partir de uma parcela média ou MLCAPE (100 hPa a partir da superfície) e a CAPE calculada a partir de uma parcela de ar mais instável nos primeiros 300 hPa a partir da superfície (MICAPE).

A Tabela 5 apresenta os valores de referência de CAPE associados ao potencial para ocorrências de tempestades severas.

**Tabela 5:** Valores de CAPE associados à previsão de tempestades. Fonte: Adaptado de Houze (2014).

CAPE (Jkg <sup>-1</sup> )	Potencial para Tempestades Severas
1000 < CAPE ≤ 2500	Alto Potencial
2500 < CAPE ≤ 4000	Potencial para instabilidade acumulada
CAPE > 4000	Potencial para instabilidade extrema

### 3.4 Informe Meteorológico de Aeródromo (METAR)

O METAR (Figura 4) é um informe meteorológico regular de aeródromo, disponibilizado a cada hora, utilizado para a descrição completa das condições meteorológicas observadas em um aeródromo específico (DECEA, 2020).

```
METAR SBBR 062000Z 01019G29KT 0400 R29R/0800 R11L///// R29L/0900
R11R///// +TSRA SCT014 BKN030CB 18/17 Q1017=
```

**Figura 4:** Exemplo de mensagem METAR com destaque para a ocorrência de chuva forte com trovoadas e visibilidade de 400 metros para SBBR (Aeroporto Internacional Juscelino Kubitschek) (em destaque sublinhado vermelho feito pelo autor). Fonte: REDEMET (2022).

### 3.5 Weather Research and Forecasting (WRF)

Conforme descrito por Da Silva (2021) e Polifke (2019), os modelos são classificados em globais, os quais buscam aproximações numéricas de equações matemáticas representativas de leis da física que regem os movimentos e sua interação com a superfície à nível mundial e regionais, modelos com maior resolução, que conseguem solucionar equações matemáticas sobre uma área menor e representar melhor, os processos de turbulência, convecção, microfísica e as condições de contorno, como a topografia, cobertura do solo, entre outros. Um exemplo de modelo regional amplamente utilizado na Meteorologia é o WRF.

O modelo WRF foi projetado tanto para a pesquisa quanto para a previsão numérica do tempo (POWERS *et al.*, 2017). É muito portátil pois é passivo de ser instalado em diversas plataformas computacionais, de domínio público e disponibilizado gratuitamente pela internet. Este vem sendo desenvolvido através de cooperação do *National Center for Atmospheric Research* (NCAR) e o *National Center for Environmental Prediction* (NCEP/NOAA), (SKAMAROCK *et al.*, 2019). Segundo Powers *et al.* (2017), apesar de ser oficialmente apoiado pelo NCAR, o WRF se tornou um verdadeiro “modelo comunitário”, pelo fato de seu desenvolvimento de longo prazo ser auxiliado por meio dos interesses e contribuições de uma base de usuários em todo o mundo e, com isso, o modelo cresceu para fornecer recursos especializados para uma vasta variedade de aplicações de previsão do sistema terrestre, da evolução da química atmosférica, hidrologia, incêndios florestais, furações e clima regional.

Para representar as equações dinâmicas de processos físicos de pequena escala, o modelo WRF usa esquemas de parametrização, que consiste em construir fórmulas para resolver os termos associados ao fluxo turbulento de momentum, calor e umidade (Morales, 2013).

### **3.6 Inteligência Artificial**

Durante a segunda guerra mundial, a IA começou a ser estudada com o objetivo de criar um mecanismo capaz de agir e pensar igual ao ser humano. Sistemas com habilidades de generalizar e aprender próximas ao cérebro humano ainda são metas distantes. De forma ampla, a IA tem como uma de suas premissas de longo prazo o desenvolvimento de máquinas que podem desempenhar funções tão bem quanto os humanos ou até melhor (NILSSON,1998).

Mcculloch e Pitts (1943), apresentaram um artigo que fala pela primeira vez à respeito de redes neurais (estruturas de raciocínio artificiais em forma de modelo matemático que imitam nosso sistema nervoso). Outro artigo importante dessa época é o trabalho de Shannon (1950) sobre como programar uma máquina para jogar xadrez utilizando cálculos de posição simples mais eficientes. Em 1950, o matemático Allan Turing desenvolveu uma forma de avaliar se uma máquina consegue imitar um humano em

uma conversa por escrito que é lida por um avaliador. Esse é o teste de Turing originalmente conhecido como jogo da imitação.

Em 1951, Marvin Minsky criou a *Snark*, uma calculadora de operações matemáticas simulando sinapses que são as ligações entre os neurônios. Em 1952, Arthur Samuel criou um jogo de damas no IBM (*International Business Machines Corporation*) 701 (computador da época) que consegue aprender por conta própria e virou um desafio para jogadores amadores.

Todos esses trabalhos mencionados foram relevantes, mas vieram antes do período considerado como o início da inteligência artificial. O marco zero foi em 1956 na chamada Conferência de *Darhmouth* (MOOR, 2006). Nesse evento, o campo de pesquisa foi denominado IA e a máxima do setor foi definido: “Cada aspecto de aprendizado ou outra forma de inteligência pode ser descrita de forma tão precisa que uma máquina pode ser criada para simular isso”.

A inteligência artificial é o que algoritmos exibem quando realizam uma tarefa de forma inteligente. A tarefa pode ser simples ou tão complexa que dependa de habilidades próxima ao cérebro humano. Um algoritmo é um conjunto de passos para realizar uma determinada tarefa e habilitá-los a tomarem decisões que resolvam um problema proposto (SZWARCFITER & MARKENZON, 1994).

Conforme Soares (2021), a IA abrange, de modo geral, diversas áreas da ciência como, por exemplo, Sistemas Especialistas, Aprendizado de Máquina, Robótica, Algoritmos Genéticos, entre outros.

O AM usa algoritmos em um computador para que ele aprenda, a partir de dados históricos, a realizar tarefas, como prognósticos, classificação e previsão (SOARES, 2021). De acordo com a *Britannica Encyclopedia*, AM é uma disciplina voltada para a implementação de softwares que possam aprender de forma autônoma. Por exemplo, inserindo uma planilha de dados num programa específico de computador que utiliza algoritmos para fazer previsões ou classificações, basicamente, o aprendizado de máquina é um processo de treinamento desses algoritmos para que os mesmos possam analisar e aprender a respeito dos dados inseridos. Assim, quando novos dados do mesmo tipo forem inseridos na máquina (computador), os

algoritmos treinados serão capazes de fazer classificações ou previsões usando esses dados.

De acordo com Brownlee (2016), os dados devem ser divididos em um conjunto de treinamento com a função de treinar o algoritmo e um conjunto de teste que usa os novos dados para prever resultados ou verificar a precisão do modelo. Para entender melhor o processo de aprendizagem de máquina, imagine um ser humano que aprende a andar de bicicleta. Ele não precisa assimilar uma série de instruções, mas sim, aprender a manter o equilíbrio na bicicleta por tentativa e erro, ocasionando algumas quedas no processo. O conceito de AM funciona de forma parecida. Em vez de escrever o código de cada instrução que a máquina deverá executar, a máquina vai aprender o que deve ser feito por tentativa e erro, ou seja, os algoritmos de aprendizagem de máquina são técnicas para descobrir a função de destino com o propósito de prever a variável de saída, a partir de variáveis de entrada (BROWNLEE, 2016).

Segundo Soares (2021), como diferentes algoritmos de AM fazem suposições distintas sobre a forma e estrutura da função de destino que está sendo estimada, bem como o melhor ajuste de otimização, é importante testar um conjunto de algoritmos diferentes em um problema de aprendizado de máquina, com o objetivo de encontrar a estrutura da função ótima.

Um aspecto importante no uso de algoritmos de AM para se fazer previsões ou classificações é seu baixo custo computacional, podendo ser usado tanto em pesquisas como nos centros operacionais, por exemplo, o CIMAER ou CGNA.

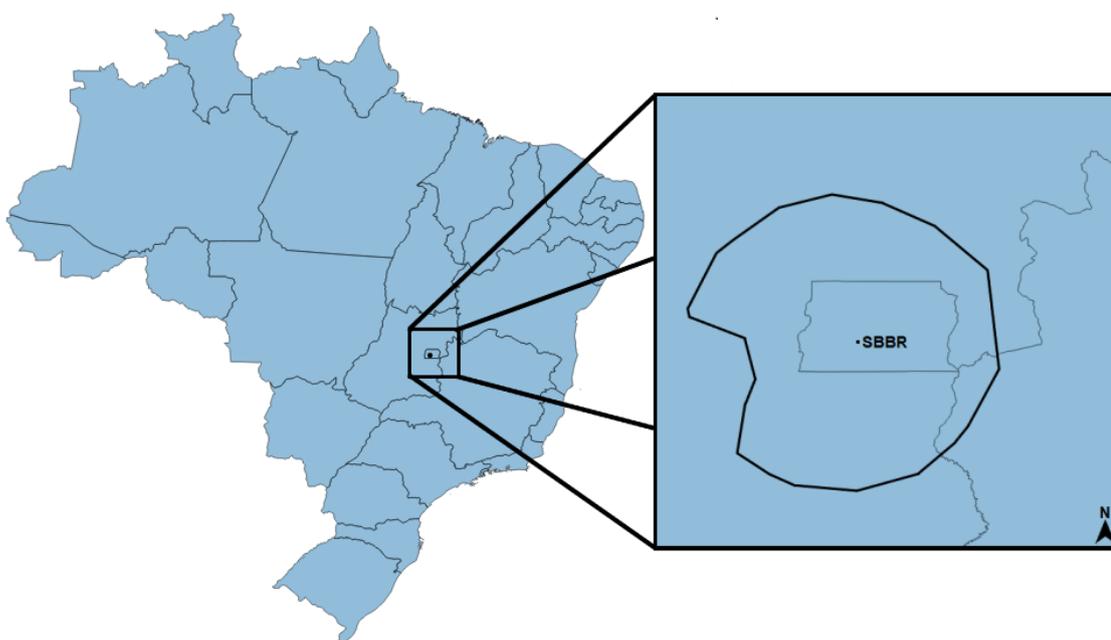
Neste trabalho será utilizado os classificadores pertencentes à versão 3.8.6 do WEKA (*Waikato Environment for Knowledge Analysis*), um software livre, desenvolvido pela Universidade de Waikato, Nova Zelândia. O WEKA usa a linguagem de programação Java, cuja principal característica é sua portabilidade, assim é possível utilizá-la em diversos sistemas operacionais. Para o treinamento e teste dos modelos preditivos, foram utilizados 25 algoritmos disponíveis no WEKA, listados no Anexo I, porém a seleção foi baseada nos que obtiveram melhor desempenho.

## 4. DADOS E MÉTODO

Neste Capítulo, apresenta-se a área de estudo, o conjunto de dados utilizados e a metodologia empregada nesta pesquisa.

### 4.1 Área de Estudo

A área de estudo selecionada está representada na Figura 5. Essa área é limitada pelas coordenadas geográficas de latitudes de  $17^{\circ} 08' 21''$  a  $14^{\circ} 19' 15''$  S e longitudes  $49^{\circ} 10' 21''$  a  $46^{\circ} 18' 49''$  W (quadrante na cor preta). O aeroporto internacional Juscelino Kubitschek (SBBR) localizado na cidade de Brasília – DF, que possui altitude média de 1.172 metros, está inserido no domínio da TMA-BR (área destacada na figura pelo polígono na cor preta dentro do quadrilátero).



**Figura 5:** Área de estudo em destaque. Está indicado o aeroporto de Brasília (SBBR).

É importante destacar que a massa equatorial continental (quente e de umidade específica alta), cuja origem está na planície amazônica, tende a avançar para a região centro-oeste, área da região de estudo, no verão, onde provoca aumento das temperaturas, umidade e precipitações (Barros, 2002).

Frequentemente a região centro-oeste também é influenciada pela massa tropical atlântica que, devido ao anticiclone do Atlântico Sul, tem uma atividade bastante relevante durante o ano todo (MONTEIRO, 1963). Segundo Barros e Balero (2012),

no verão a massa tropical atlântica torna-se instável devido ao aquecimento que sofre ao entrar em contato com o continente quente e no período do inverno, em virtude do continente estar mais frio, a massa torna-se estável. Esse processo contribui para a ocorrência de tempo instável no verão e estável no inverno.

## 4.2 Dados

Neste trabalho são utilizadas quatro fontes de dados relativos ao período de 2012 a 2017 conforme indicados na Tabela 6, com detalhes sobre a sua fonte, frequência, tipo e finalidade.

**Tabela 6:** Conjunto de dados utilizado na pesquisa no período de estudo de 2012 a 2017.

Tipo de Dados	Finalidade	Frequência	Fonte
Índices de Instabilidade Termodinâmicos (K, SH, SW, TT e MLCAPE)	Variáveis preditoras de EMC dos algoritmos de AM	12 horas	Universidade do Wyoming (weather.uwyo.edu/)
Descargas Atmosféricas	Definição de ocorrência e intensidade de EMC	----	Descargas Atmosféricas (www.starnet.iag.usp.br)
METAR	Detectar os horários em que houve ocorrência de EMC em SBBR	1h	REDEMET (https://redemet.decea.gov.br)
Índices de Instabilidade Termodinâmicos (K, SH, SW, TT e SBCAPE) extraídos do modelo WRF	Variáveis preditoras de EMC dos algoritmos de AM	1h	Instituto de Controle do Espaço Aéreo (ICEA)

#### 4.2.1 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS

Os IIT foram determinados utilizando os dados de perfil atmosférico (radiossondagem) das 00 e 12 UTC da EMA de Brasília, localizada na cidade de Brasília – DF, na latitude de 15° 52' 16"S e longitude de 047° 55' 07"W.

#### 4.2.2 DADOS DE DESCARGAS ATMOSFÉRICAS

Os dados de descargas atmosféricas coletados pela STARNET, caracterizam cada ocorrência, indicando a localização (latitude e longitude), quantidade e horário dos eventos.

Neste trabalho, as descargas atmosféricas são utilizadas para caracterização espaço-temporal da ocorrência de EMC e a classificação se o evento é severo ou não.

#### 4.2.3 DADOS DE METAR

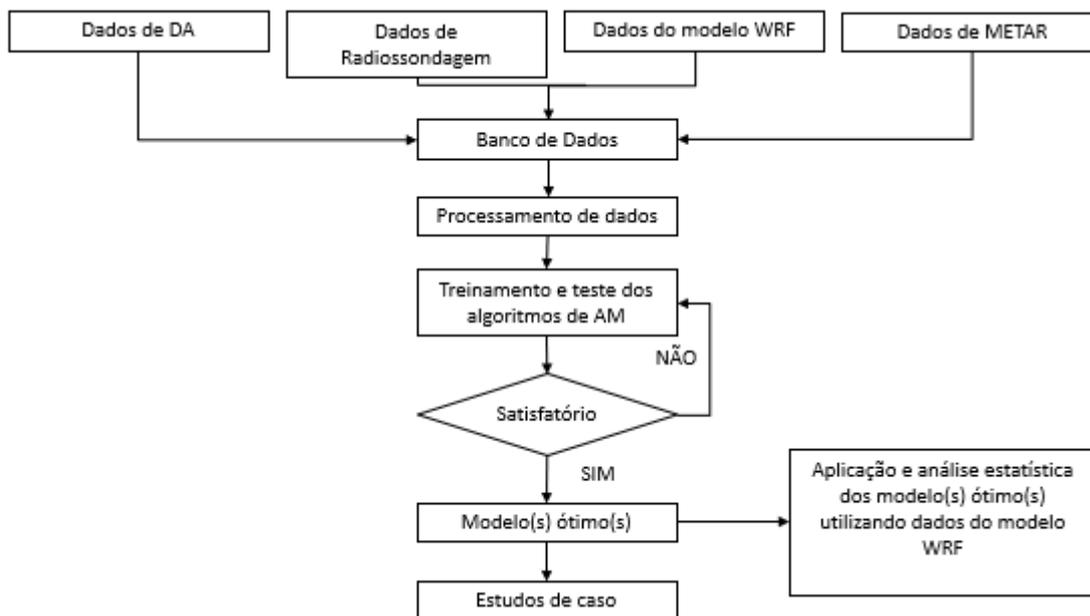
Os dados de METAR de SBBR foram utilizados para identificar a ocorrência de EMC, caracterizado pela ocorrência de trovoadas (TS) no aeroporto. Juntamente com os dados de DA mais significativos na área de estudo, os dados de METAR de SBBR serviram para determinar os dias em que os IIT seriam obtidos através de modelagem numérica com o uso do modelo WRF.

#### 4.2.4 DADOS MODELADOS DO WRF

Os IIT K, SH, SW, TT e CAPE, gerados pelo modelo WRF, foram disponibilizados pelo Instituto de Controle do Espaço Aéreo (ICEA), com saída horária, para serem utilizados como variáveis (ou preditores) de entrada (*input*) para os modelos baseado em inteligência computacional que foram desenvolvidos e testados.

### 4.3 Método

Nesta seção é apresentada a metodologia desenvolvida para este estudo. Ela está esquematizada no diagrama de blocos da Figura 6, no qual são mostradas as etapas do desenvolvimento metodológico.



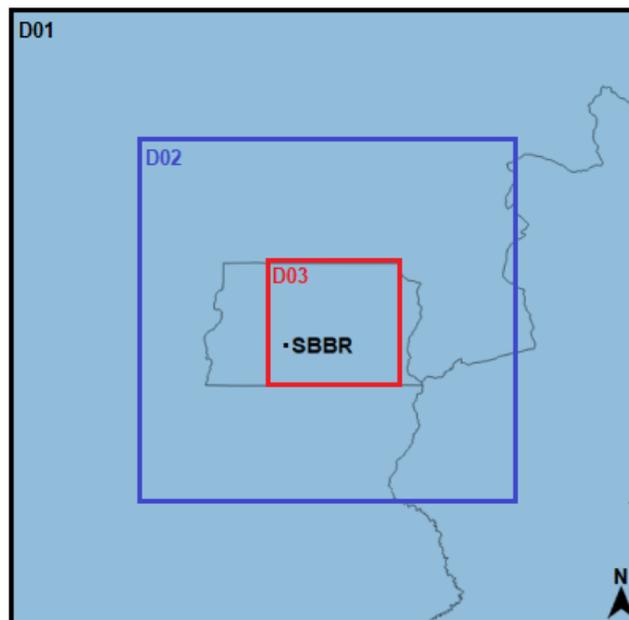
**Figura 6:** Fluxograma dos passos metodológicos da pesquisa.

Para atingir os objetivos delimitados na Seção 1.2, o método divide-se em 5 partes com 12 passos, a saber:

#### a) Processamento de dados

- i. Analisa-se mensalmente e anualmente, o número de radiossondagens realizadas para o período de estudo;
- ii. Organizam-se os dados das DA em uma planilha do tipo Excel e analisa-se a distribuição desses valores através de gráficos horário, mensal e sazonal, assim define-se o período de previsão (PP) de EMC;
- iii. Utilizando-se os meses de maior ocorrência de DA, analisa-se a distribuição espacial das descargas atmosféricas com o pacote computacional, software livre, denominado QGIS (versão 3.8.5, disponível em <https://qgis.org/en/site/>;
- iv. Divide-se, em duas planilhas do tipo Excel (00 UTC e 12 UTC), os valores de cada IIT (K, SH, SW, TT e CAPE), *input* (preditores) dos algoritmos de aprendizagem de máquina, associando-se, cronologicamente, o conjunto de IIT de cada dia, ao respectivo número de DA que ocorreram durante o PP de EMC definido no passo ii. Analisa-se a consistência estatística desses índices, excluindo do trabalho valores inconsistentes. Por fim, define-se o vetor de *output* dos algoritmos de AM, se “SIM”, caso de ocorrência de DA (no PP), ou “NÃO”, caso contrário.

- v. Selecionam-se dez dias para estudo de caso, quando deverão ser obtidos dados dos IIT extraídos do modelo WRF, no período de 2012 a 2017, sendo que cada dia representa um evento de ocorrência ou não de EMC em SBBR conforme a seguinte metodologia:
- Selecionam-se sete dias com número significativo de descargas atmosféricas, durante o PP, na área de estudo e que tenha ocorrência de EMC em SBBR, detectado através da presença de trovoadas (TS) no código METAR de SBBR;
  - Selecionam-se três em que não houve ocorrência de DA na área de estudo; e
  - Define-se as configurações do modelo WRF com resolução de grade horizontal de 27 km x 27 km, 9 km x 9 km e 3 km x 3km (D01, D02, D03), respectivamente) com domínios inserido nas coordenadas geográficas 17° 52' 16"S-13° 52' 16"S e 49° 55' 07 "W-45° 55' 07"W, centrado em 15° 52' 16"S e 047° 55' 07"W (SBBR) e 48 níveis verticais conforme Figura 7. Solicitam-se os IIT (Tabela 6) modelados ao ICEA com duas simulações, com esquemas de parametrizações diferentes, conforme Tabela 7. Na primeira simulação, usou-se as parametrizações do ICEA. Para a segunda simulação, adotou-se o resultado do teste de sensibilidade de melhor previsibilidade de EMC com as parametrizações cumulus e microfísica na região da TMA-RJ por Da Silva (2021). A inicialização do WRF começou sempre às 00 UTC do dia anterior, mas as primeiras 12 horas foram descartadas;



**Figura 7:** Domínios selecionados para os experimentos com IIT modelados extraídos do modelo WRF, com grades 27 km (D01), 9 km (D02) e 3 km (D03).

**Tabela 7 -** Parametrizações utilizadas nas simulações I e II do modelo WRF.

Esquema de Parametrização				Simulação	
Categoria	Esquema	Opção	Referência	I	II
Cumulus	Grell-Freitas	3	Grell e Freitas, 2014		X
Cumulus	Grell-Devenyi	93	Grell e Devenyi, 2002	X	
Microfísica	WRF Double-moment 6-class	6	Hong e Lim, 2006		X
Microfísica	Morrison	10	Morrison <i>et al.</i> , 2009	X	
Radiação	New Goddard lw/sw	5	Chou e Suarez, 1999	X	X
Camada de Superfície	MM5	1	Jimenez <i>et al.</i> , 2012	X	X
Cobertura do Solo	Noah Land	2	Tewari <i>et al.</i> , 2004	X	X
Camada Limite Planetária	Yonsei University	1	Hong <i>et al.</i> , 2006	X	X

## **b) Treinamento e teste dos algoritmos de AM**

- vi. Divide-se os vetores de entrada em dois grupos, IIT extraídos das radiossondagens das 00 UTC e 12 UTC, com 20 classes cada (sem DA, > 10, > 20, > 30, > 40, > 50, > 100, > 200, > 300, > 400, > 500, > 600 > 700, > 800, > 900, > 1000, > 2000, > 3000, > 4000 e > 5000), de acordo com o número de DA, com *input* e *output* definidos no passo iv e, na sequência, realiza-se experimentos, separadamente, para os dados das 00 UTC e 12 UTC, utilizando os algoritmos de AM disponíveis no pacote WEKA via validação cruzada<sup>3</sup>.
- vii. Analisa-se os resultados do treinamento e teste, considerando o conjunto de dados do passo anterior, utilizando as métricas definidas na Seção 4.3.1.
- viii. Define-se como modelo(s) ótimo(s) aqueles que possuem as estatísticas de POD  $\geq 0,90$ , 1-FAR  $\geq 0,80$ , F-MEASURE  $\geq 0,90$ ; KAPPA  $\geq 0,81$  e BIAS  $\geq 0,80$  para “SIM” ou “NÃO” de ocorrência de EMC e POD  $\geq 0,80$ , 1-FAR  $\geq 0,75$ , F-MEASURE  $\geq 0,80$ ; KAPPA  $\geq 0,51$  e BIAS  $\geq 0,80$  para detecção de eventos severos;

## **c) Classificação da severidade do EMC**

- ix. A severidade é definida comparando a variação percentual da distribuição total de DA durante o PP por tentativa e erro. Definiu-se, nesse trabalho, que o EMC é classificado como severo quando a taxa de DA não faz parte do percentil 70.

## **d) Aplicação do(s) modelo(s) ótimo(s) nos IIT extraídos do modelo WRF**

- x. Aplica-se os modelos ótimos baseado em aprendizado de máquina, desenvolvidos com os dados das radiossondagens das 12 UTC, nos dados dos IIT extraídos do modelo WRF (IIT modelados numericamente); e
- xi. Analisa-se os resultados considerando as métricas estabelecidas na Seção 4.3.1.

---

<sup>3</sup> Método que divide o conjunto de dados, em k subconjuntos mutuamente exclusivos do mesmo tamanho, e cada subconjunto é utilizado para teste e os k-1 restantes são utilizados para treinamento. Este procedimento é executado k vezes alternando o subconjunto de testes até que todos os subconjuntos sejam utilizados para treinamento e avaliação de teste (KOHAVI, 1995).

### e) Estudo de caso

- xii. Consistiu-se em realizar um *hindcast* de eventos considerando uma amostra independente de eventos convectivos e não convectivos retirados do conjunto de treinamento;

#### 4.3.1 Métrica de Avaliação dos Resultados

Avalia-se os modelos classificatórios nas fases de treinamento e testes, utilizando-se como base a matriz de confusão (Tabela 8). O resultado ideal ocorre apenas quando os valores se concentram na diagonal principal, significando que todos os valores previstos e observados coincidem. Quando há valores fora da diagonal principal da matriz de confusão quer dizer que o modelo possui um erro, que deve ser conhecido por quem utilizará os resultados.

**Tabela 8:** Matriz de Confusão. Fonte: Adaptado de Wilks (2006).

		Observado	
		Evento	Não-evento
Previsto	Evento	a (Verdadeiro Positivo)	c (Falso Positivo)
	Não-evento	b (Falso Negativo)	d (Verdadeiro Negativo)

Foram utilizados os seguintes índices estatísticos baseados na Matriz de Confusão:

- a) 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 07.

$$POD = \frac{a}{a+b} \quad (07)$$

- b) 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 08.

$$FAR = \frac{c}{c+a} \quad (08)$$

c) Bias (B) – Medida da proporção da frequência de eventos previstos pela frequência de eventos observados. Uma previsão não tendenciosa exibe  $B = 1$ . Um valor maior que 1 indica uma previsão superestimada e menor que 1, uma previsão subestimada. É descrito na Equação 09.

$$B = \frac{a+c}{a+b} \quad (09)$$

d) Coeficiente KAPPA ( $K'$ ) – É a medida do grau de concordância na classificação entre o previsto e o observado, utilizado para medir o desempenho de um algoritmo. Os valores de KAPPA situam-se entre -1 e 1, onde 1 é perfeita concordância, 0 é exatamente o que seria esperado por acaso, e valores negativos indicam que a concordância encontrada foi menor daquela esperada por acaso. Pode ser calculado pela Equação 10.

$$K' = \frac{p_0 - p_e}{1 - p_e} \quad (10)$$

onde  $p_0$  é taxa de aceitação relativa (Equação 11).

$$p_0 = \frac{a+d}{a+b+c+d} \quad (11)$$

e  $p_e$  é a taxa hipotética de aceitação (Equação 12).

$$p_e = \frac{[(a+b)(a+c)] + [(c+d)(b+d)]}{(a+b+c+d)^2} \quad (12)$$

A Tabela 9 demonstra o nível de concordância que valor de KAPPA indica.

**Tabela 9:** Valores de referência do índice KAPPA. **Fonte:** Adaptado de Landis e Koch (1977).

Coeficiente KAPPA	Nível de Concordância
< 0	Insignificante
0-0,2	Fraca
0,21-0,4	Razoável
0,41-0,6	Moderada
0,61-0,8	Forte
0,81-1	Quase Perfeita

e) *F-Measure* (F) – É a medida da acurácia (precisão) de um teste. Ele considera a precisão e a revocação (sensibilidade) para estimar a pontuação. Segundo Sasaki (2007), o *F-Measure* é a média harmônica da precisão e revocação (Equação 13), onde obtém o seu melhor valor em 1 (precisão perfeita).

$$F = \frac{2P' \cdot POD}{P' + POD} \quad (13)$$

onde POD e P' representam revocação e precisão, respectivamente. E P' pode ser expresso como na Equação 14.

$$P' = \frac{a}{a+c} \quad (14)$$

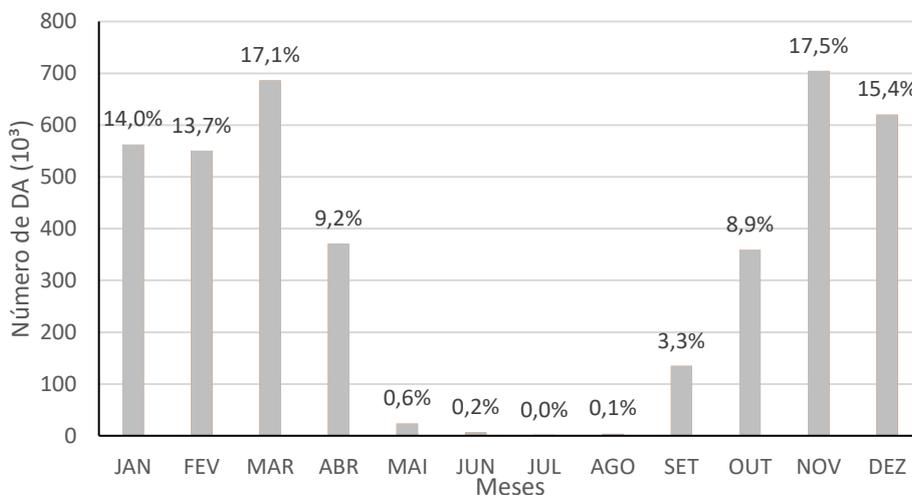
## 5. RESULTADOS E DISCUSSÕES

A seguir são apresentados os resultados e discussões de acordo com os passos estabelecidos na Seção 4.3.

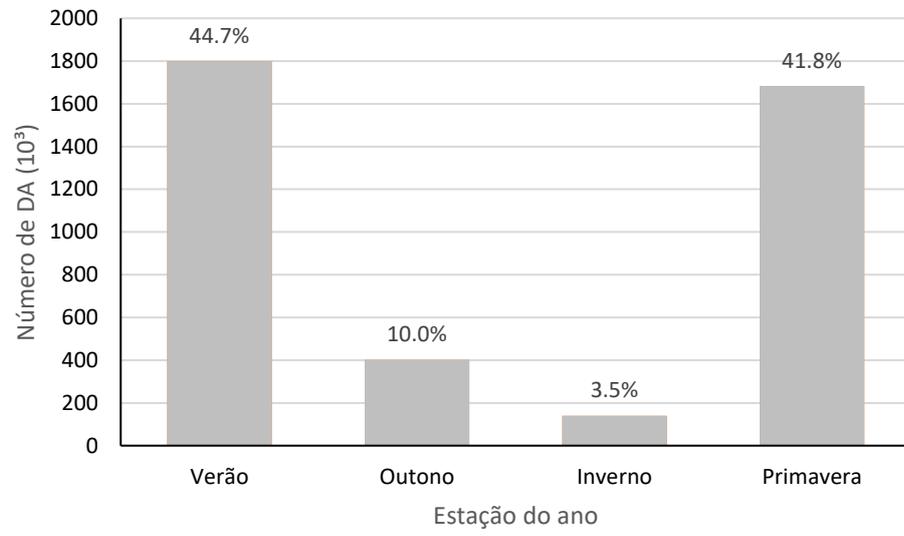
### 5.1 Análise da distribuição dos dados de descargas atmosféricas

As Figuras 8 e 9 representam, respectivamente, as ocorrências mensais e sazonais das descargas atmosféricas para o período de 2012 a 2017. Fica evidenciado que nos meses de novembro, dezembro, janeiro, fevereiro e março, ocorrem o maior número de descargas atmosféricas (77,6%) e nos meses de maio, junho e julho e agosto a quantidade de DA são insignificantes (0,9%), assim como o verão e o inverno são, respectivamente, os períodos de maior (44,7% dos eventos) e menor (3,5% dos eventos) ocorrências de DA. Essas informações estão de acordo com Barros e Balero (2012), mencionados na seção 4.1.

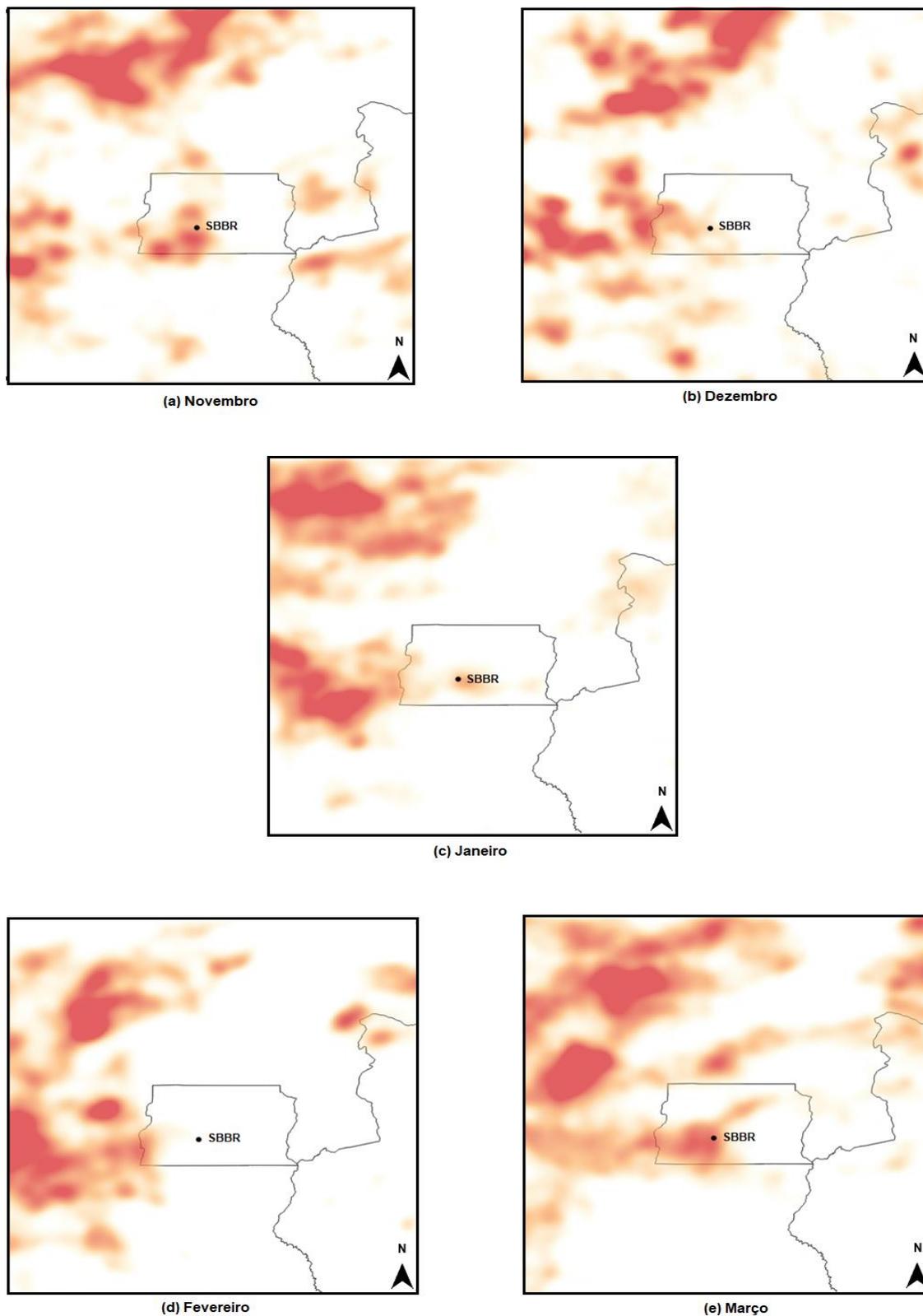
Conforme demonstrado na Figura 8, as DA ocorrem com maior frequência entre os meses de novembro a março. Dessa forma, delimitaram-se esses meses para verificar a distribuição espacial das DA, demonstrada na Figura 10 por meio de um mapa de calor, da densidade das DA, para a área de estudo. Pode-se observar que há três setores da TMA-BR, preferencialmente, de maior densidade de DA, os setores oeste, noroeste e norte.



**Figura 8:** Número de DA por mês e a respectiva porcentagem de ocorrência mensal

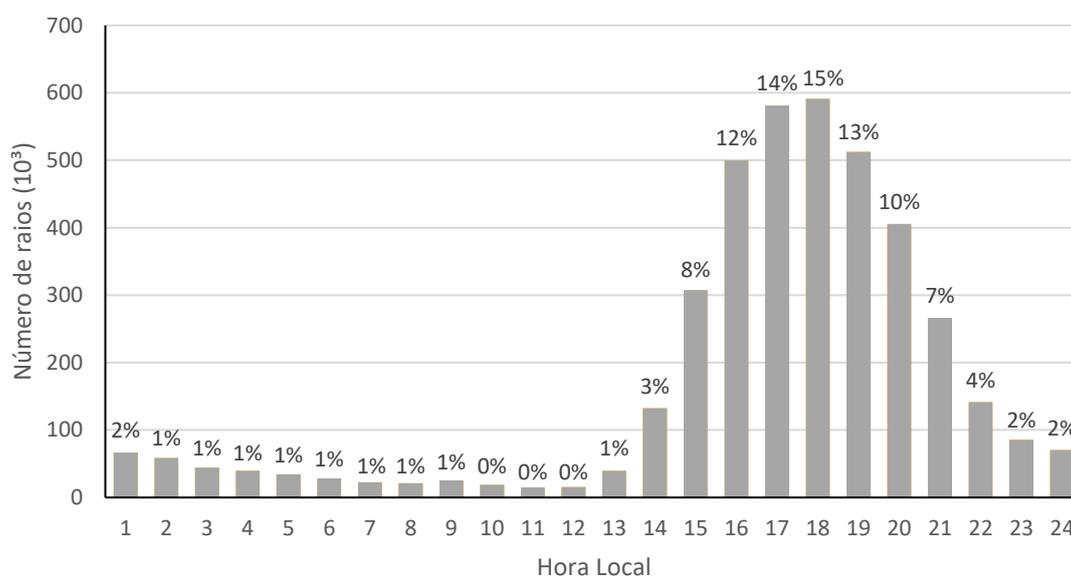


**Figura 9:** Ocorrências de DA por estação do ano e a respectiva porcentagem de ocorrência sazonal



**Figura 10:** Mapa de calor dos meses de novembro, dezembro, janeiro, fevereiro e março das descargas atmosféricas para a área de estudo no período de 2012 a 2017.

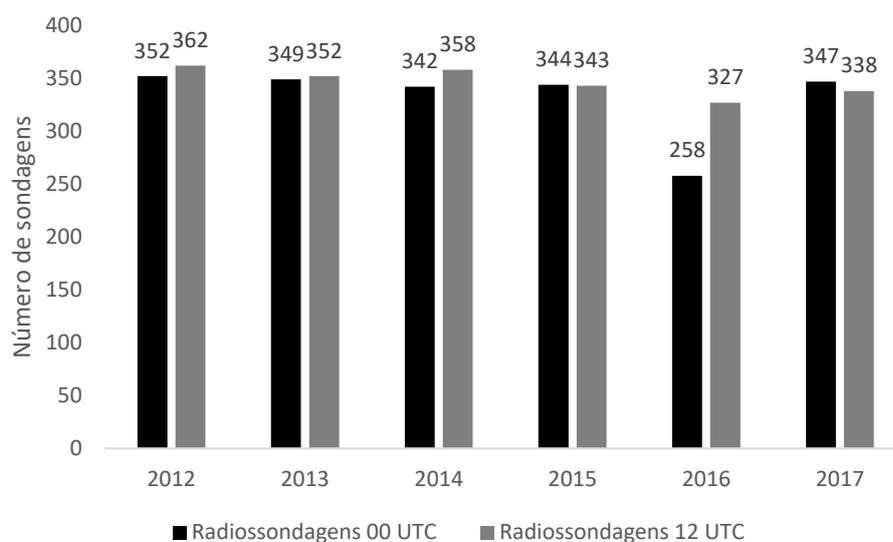
Elaborou-se a Figura 11 com a distribuição horária das ocorrências de descargas atmosféricas entre os anos de 2012 a 2017. Identifica-se que a atividade convectiva se inicia às 15h (local) e, considerando o período de uma hora, tem seu pico entre às 18 e 19h (local) (15% das DA). O período entre as 15 e 22h (local) (DA ocorridas entre às 15 e 21:59:59h) representa 79% do total de descargas atmosféricas detectadas pela STARNET, assume-se este o período da previsão de EMC. Para a classificação da severidade do EMC, adotou-se o critério especificado no passo ix do método, o qual classifica como severo os eventos que não fazem parte do percentil 70 de ocorrências de DA durante o PP, e assim, chegou-se ao valor aproximado de 1.000 DA, ou seja, os EMC com mais de 1.000 DA serão classificados como severos, assim como De Castro *et al.* (2022).



**Figura 11:** Distribuição horária das DA detectadas entre 2012-2017 e a respectiva porcentagem de ocorrência horária

## 5.2 Análise dos dados de Radiossondagens

Foram realizadas pela EMA de Brasília 1.992 e 2080 radiossondagens das 00 e 12 UTC, respectivamente (Figura 12), no período de 2012 a 2017. A Tabela 10(a) e 10(b) representam, respectivamente, as radiossondagens realizadas as 00 e 12 UTC. Observa-se que o percentual de radiossondagens realizadas as 00 e 12 UTC corresponde a 91% e 95%, respectivamente conforme destacados na Tabela 10.



**Figura 12:** Distribuição das radissondagens 00 UTC e 12 UTC realizadas no período de 2012 a 2017.

**Tabela 10:** (a) e (b) representam respectivamente o número mensal de sondagens atmosféricas realizadas e total em porcentagem nos horários de 00 UTC e 12 UTC.

(a)

00Z	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
2012	30	28	31	30	29	29	27	31	30	27	29	31	96%
2013	30	28	31	28	30	30	29	29	27	31	26	30	96%
2014	31	27	30	24	31	30	30	28	29	23	29	30	94%
2015	29	28	28	30	26	28	29	31	29	30	26	30	94%
2016	27	11	31	15	0	0	29	29	30	27	28	31	70%
2017	29	25	31	30	30	30	31	27	25	30	29	30	95%
Total	95%	86%	98%	87%	78%	82%	94%	94%	94%	90%	93%	98%	91%

(b)

12Z	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
2012	30	29	31	30	31	30	31	31	30	28	30	31	99%
2013	30	28	31	30	31	30	29	29	29	31	25	29	96%
2014	29	27	30	29	31	30	31	30	29	31	30	31	98%
2015	31	27	30	29	26	27	27	29	29	29	29	30	94%
2016	28	13	30	28	29	28	29	28	28	28	28	30	89%
2017	30	24	30	29	28	29	30	27	24	30	28	29	93%
Total	96%	87%	98%	97%	95%	97%	95%	94%	94%	95%	94%	97%	95%

Os IIT (Tabela 6) das 00 e 12 UTC, extraídos das radiossondagens realizadas pela EMA de Brasília e obtidos no endereço <http://wyoweb.uwyo.edu/>, foram

armazenados e organizados cronologicamente em planilhas do Excel, associando-se a cada conjunto de IIT o número de DA diária detectadas no período entre as 15 e 22h (local) (período de previsão definido na Seção 5.2). Seguindo os passos do método, excluiu-se do trabalho valores de índices inconsistentes. Conforme a Figura 13, os IIT foram organizados de acordo com intervalos de DA, classificando a ocorrência de EMC como “SIM” ou “NÃO” de acordo com o número de DA dentro do intervalo e, assim, organizando-se os dados a serem utilizados pelos algoritmos de AM nos passos de treinamento e teste para identificar o ambiente atmosférico favorável à formação ou não de EMC.

Dia	Mês	Ano	SH	SW	K	TT	CAPE	DA	Ocorrência	Intervalo
13	9	2012	4.74	104.39	28.3	40.2	2.35	11	S	11-20
28	9	2012	5.83	176.42	20.6	36.7	0	11	S	11-20
29	10	2012	2.41	182.19	25.5	42.8	0	11	S	11-20
5	5	2013	6.24	134.79	3.9	36	0	11	S	11-20
14	6	2013	6.48	163.22	15.7	36.2	0	11	S	11-20
28	7	2013	8.94	120.63	-17.93	32.45	0	11	S	11-20
14	10	2013	5.94	164.82	19.3	36.8	0	11	S	11-20
20	10	2015	5.91	141.81	17.9	37	0	11	S	11-20
9	5	2016	4.03	154.21	19.5	40.8	0	11	S	11-20
24	10	2016	2.78	191.22	21.9	40.6	0	11	S	11-20
16	12	2016	2.5	221.79	31.5	39.6	68.37	11	S	11-20
9	10	2017	7.2	80.61	20.5	36.4	0	11	S	11-20
25	10	2017	5.62	129.99	19.5	38	0	11	S	11-20
20	4	2012	6.98	145.8	21.7	34.6	0	12	S	11-20
5	10	2012	5.97	132.82	8.3	37.8	0	12	S	11-20
20	6	2013	3.55	180.8	18	39.7	0	12	S	11-20
10	7	2013	4.9	204.2	1.1	37.6	0	12	S	11-20
2	9	2013	3.03	166.59	24.9	42.2	0	12	S	11-20
9	9	2013	4.03	146.19	-5.7	41.4	0	12	S	11-20
13	9	2014	4.93	157.79	22.7	39.6	0	12	S	11-20

Figura 13: Planilha do Excel contendo parte dos dados dos IIT das 12Z.

### 5.3 Definição dos dados modelados do WRF

Foram selecionados dez dias para estudo de caso, conforme a Tabela 11, entre o período de 2012 a 2017. Desses dez eventos, em sete (analisados levando-se em consideração o número de DA mais significativos), houve ocorrência de EMC em SBBR, identificada através da presença de TS no código METAR. Nos três dias restantes, não houve detecção de DA na área de estudo.

**Tabela 11** – Dias selecionados para estudo de caso no modelo WRF com o horário e o número de ocorrências de DA.

Dias selecionados	Intervalo do período crítico (maior quantidade de DA) na área de estudo (Hora Local)	Número de DA no intervalo do período crítico
12/04/2012	-	0/0
12/05/2013	-	0/0
10/04/2014	16-17	2146
07/03/2015	21-22	2664
09/03/2015	17-18	2191
11/03/2015	19-20	2720
26/05/2016	-	0/0
04/12/2016	15-16	2413
06/12/2016	16-17	2197
12/01/2017	15-16	1653

O ICEA realizou duas simulações com o modelo WRF, com as configurações definidas no passo v do método e esquemas de parametrização conforme Tabela 7, considerando os eventos da Tabela 11. Foram selecionados os dados dos IIT da resolução de grade horizontal de maior resolução (3 km x 3 km) para serem utilizados como dados de entrada no teste de previsão de EMC usando o(s) modelo(s) ótimo(s) de AM.

#### 5.4 Treinamento e teste dos algoritmos de AM

O treinamento e teste dos algoritmos de AM é um processo que leva tempo e a seleção da variáveis preditoras (*input*) é crítica. É bem conhecido que treinar os algoritmos de AM com um número excessivo de variáveis de entrada dificulta esse processo. Dessa forma, a exemplo de Freitas *et al.* (2019), Almeida *et al.* (2020) e De Castro *et al.* (2022), foram escolhidos os ITT clássicos K, SH, SW, TT e CAPE (a

CAPE não foi utilizada para o treinamento dos dados do grupo das 00 UTC) como *input* ou variáveis preditoras, da ocorrência ou não de EMC, caracterizados pela ocorrência de DA entre às 15 e 22h (local). O índice CAPE não foi utilizado como dado de entrada no treinamento dos algoritmos de AM das 00 UTC porque seus valores não variaram significativamente nesse horário e, portanto, não foram determinantes no processo de aprendizagem dos algoritmos, assim como em De Castro *et al.* (2022).

Organizou-se as DA em dois grupos, 00 e 12 UTC, cada grupo foi dividido em 20 planilhas de acordo com o número de DA (definidos no passo vi) com os resultados estatísticos do treinamento e teste apresentados na Tabela 12 e 13.

#### 5.4.1 Treinamento e teste com os dados das radiossondagens das 00 UTC

Os IIT (K, SH, SW e TT) obtidos através das sondagens atmosféricas das 00 UTC foram treinados e testados, conforme passo vi e vii do método, com o objetivo de avaliar a capacidade desses índices em fazer previsões de 16 horas de ocorrência ou não de EMC, assim como classificar se o evento é severo. O índice CAPE foi excluído desse treinamento em virtude de seus valores não terem variação significativa nesse horário.

A Tabela 12 apresenta as estatísticas dos melhores resultados, utilizando os dados das sondagens das 00 UTC, de previsões de EMC produzidos com os algoritmos de AM (coluna 3), também chamados de modelos nesse trabalho, para vinte intervalos de DA, considerando “SIM” como evento com  $DA > \text{limite da coluna 2}$ , caso contrário, “NÃO”. Nota-se que os modelos referentes aos experimentos de número 2, 3, 4, 5, 6 e 7 atendem ao critério estatístico para determinação do modelo ótimo descrito no passo viii do método e, assim, foram definidos como modelos ótimos para a classificação da ocorrência ou não de EMC com mais de 10 DA com previsão de 16 horas. Para a classificação da severidade do evento, os resultados não foram satisfatórios.

**Tabela 12:** Análise estatística dos melhores algoritmos de AM para previsão de EMC, na TMA-BR, para o PP, por intervalo de DA e entrada com os dados dos IIT extraídos das radiossondagens 00 UTC.

Experimento	Quantidade de DA	Algoritmo-número do experimento	POD	1-FAR	F-Measure	KAPPA	BIAS
1	0	RandomForest-1	0,92	0,63	0,91	0,58	0,84
2	>10	NaiveBayes-2	0,95	0,96	0,95	0,88	1,10
3	>20	MultilayerPerceptron-3	0,97	0,96	0,97	0,92	1,03
4	>30	MultilayerPerceptron-4	0,97	0,97	0,97	0,94	1,00
5	>40	MultilayerPerceptron-5	0,97	0,96	0,97	0,93	0,98
6	>50	NaiveBayes-6	0,96	0,95	0,96	0,92	0,97
7	>100	Logistic-7	0,93	0,91	0,93	0,85	0,89
8	>200	NaiveBayes-8	0,89	0,88	0,87	0,78	0,79
9	>300	NaiveBayes-9	0,85	0,86	0,85	0,70	0,75
10	>400	SimpleLogistic-10	0,84	0,85	0,84	0,68	0,90
11	>500	NaiveBayes-11	0,81	0,85	0,81	0,63	0,71
12	>600	NaiveBayes-12	0,78	0,84	0,78	0,57	0,68
13	>700	NaiveBayes-13	0,76	0,84	0,76	0,54	0,67
14	>800	NaiveBayes-14	0,75	0,84	0,75	0,51	0,66
15	>900	MultilayerPerceptron-15	0,76	0,76	0,77	0,49	0,90
16	>1000	MultilayerPerceptron-16	0,76	0,74	0,76	0,46	0,94
17	>2000	NaiveBayes-17	0,65	0,74	0,68	0,32	0,59
18	>3000	NaiveBayes-18	0,60	0,74	0,66	0,23	0,57
19	>4000	NaiveBayes-19	0,59	0,74	0,66	0,18	0,56
20	>5000	NaiveBayes-20	0,60	0,83	0,68	0,15	0,59

#### 5.4.2 Treinamento e teste com os dados das radiossondagens das 12 UTC

Os IIT obtidos através das sondagens atmosféricas das 12 UTC foram treinados e testados com o objetivo de avaliar a capacidade desses índices em fazer previsões de 04 horas de ocorrência ou não de EMC, assim como classificar a sua severidade.

A Tabela 13 apresenta as estatísticas dos melhores resultados, utilizando os dados das sondagens das 12 UTC, de previsões de EMC produzidos com os modelos (coluna 3) para vinte intervalos de DA, considerando “SIM” como evento com DA > limite da coluna 2, caso contrário, “NÃO”. Verifica-se que os modelos referentes aos experimentos de número 22, 23, 24, 25, 26, 27, 28 e 29, assim como o experimento de número 36 (>1000 DA) atendem ao critério estatístico definido no passo viii do método para detecção de EMC e a classificação da sua severidade. Assim, esses modelos foram definidos como ótimos para a classificação da ocorrência ou não de EMC com mais de 20 DA com previsão de 4 horas. Considerando que para

classificação EMC severo deve-se ter mais de mil descargas atmosféricas, o modelo *MultilayerPerceptron-36* detecta DA acima de mil e, portanto, esse foi o modelo ótimo selecionado para previsão de 4 horas de EMC severos.

A Tabela 13 mostra os melhores resultados estatísticos de acordo com vinte experimentos.

**Tabela 13:** Análise estatística dos melhores algoritmos de AM para previsão de EMC, na TMA-BR, para o PP, por intervalo de DA e entrada com os dados dos IIT extraídos das radiossondagens 12 UTC.

Experimento	Quantidade de DA	Algoritmo-número do experimento	POD	1-FAR	F-M	KAPPA	BIAS
21	0	RandomForest-21	0,90	0,50	0,89	0,45	0,89
22	>10	RandomForest-22	0,92	0,92	0,92	0,82	0,96
23	>20	MultilayerPerceptron-23	0,94	0,95	0,94	0,88	0,95
24	>30	MultilayerPerceptron-24	0,97	0,97	0,97	0,93	0,98
25	>40	MultilayerPerceptron-25	0,98	0,98	0,98	0,95	0,99
26	>50	RandomForest-26	0,97	0,97	0,97	0,94	1,01
27	>100	RandomForest-27	0,95	0,94	0,95	0,89	1,07
28	>200	RandomForest-28	0,92	0,93	0,92	0,85	1,08
29	>300	Logistic-29	0,91	0,91	0,91	0,82	1,08
30	>400	Logistic-30	0,88	0,89	0,88	0,76	1,11
31	>500	AdaBoostSimpleLogistic-31	0,87	0,88	0,87	0,74	0,94
32	>600	AdaBoostSimpleLogistic-32	0,87	0,88	0,87	0,74	0,96
33	>700	AdaBoostRandomForest-33	0,85	0,85	0,85	0,68	0,94
34	>800	AdaBoostRandomForest-34	0,84	0,82	0,84	0,65	0,96
35	>900	MultilayerPerceptron-35	0,82	0,78	0,82	0,60	0,97
36	>1000	MultilayerPerceptron-36	0,82	0,78	0,82	0,59	0,97
37	>2000	MultilayerPerceptron-37	0,82	0,60	0,81	0,45	0,66
38	>3000	Logistic-38	0,84	0,48	0,82	0,38	0,59
39	>4000	MultilayerPerceptron-39	0,87	0,40	0,85	0,33	0,34
40	>5000	MultilayerPerceptron-40	0,88	0,28	0,86	0,21	0,03

## 5.5 Análise estatística dos modelos ótimos testados com os dados de IIT modelados do WRF

Utilizando dados de DA do período de 2012 a 2017 e mensagens METAR de SBBR, foram selecionados dez dias para estudo de caso (Tabela 11) para obtenção dos IIT, calculados através de modelagem numérica, utilizando o modelo WRF. Desses dez dias, durante o PP, em sete houve detecção de mais de 1.000 DA na área de estudo e ocorrência de EMC em SBBR (presença de TS no código METAR). Nos três dias restantes, não houve detecção de DA na área de estudo. Os dados do modelo foram

disponibilizados pelo ICEA em duas simulações, com as configurações de parametrização conforme a Tabela 7.

Os IIT modelados por hora foram usados como dados de entrada nos modelos ótimos das 12 UTC (definidos na Seção 5.4), que possuem os melhores resultados estatísticos, com o objetivo de prever 30 horas de EMC dentro do PP, que corresponde a um período de 7 horas entre 15 e 22h (local) . Esses modelos são relativos aos experimentos de número 22, 23, 24, 25, 26, 27, 28 e 29 (detecção do evento).

A Tabela 14 mostra a previsão horária para o PP, usando os IIT calculados com o modelo WRF, aplicados nos oito modelos ótimos das 12 UTC em duas simulações com esquemas de parametrização distintos (Tabela 7). Na Tabela 14, "A" e "E" significa Acerto e Erro respectivamente. Em geral, para todos os três dias sem DA no PP, pelo menos um modelo previu todo o período corretamente, independentemente da simulação utilizada. Para os sete dias com DA durante o PP, cinco dias tiveram o EMC corretamente previsto por pelo menos quatro modelos, e nos outros dois dias, em 04/12/2016 e 12/01/2017, todos os modelos detectaram o PP parcialmente corretamente e identificou-se que erros apareceram quando EMC encontrava-se em processo de dissipação, representado por uma queda na tarifa horária de DA. Este último ponto sugere que podemos gerar um alerta de 30 horas para o surgimento de EMC para TMA-Brasília através da abordagem sugerida.

Levando-se em consideração os resultados da Tabela 14, as métricas estatísticas relativas aos dados horários de previsão, conforme PP, da Simulação I e Simulação II para os oito modelos ótimos aplicados aos IIT modelados são apresentadas na Tabela 15. A previsão de EMC de 30 horas para o PP realizada com os modelos *Random Forest* dos experimentos 26 e 27 da Simulação II produziu os melhores resultados e obedecem ao critério estatístico definido no passo viii do método para detecção de EMC.



**Tabela 15:** Valores das estatísticas para os oito modelos da Tabela 14 .

Modelo-experimento	Simulação I					Simulação II				
	POD	1-FAR	F-M	KAPPA	BIAS	POD	1-FAR	F-M	KAPPA	BIAS
RandomForest-22	1.00	0.81	0.81	0.61	1.24	1.00	0.82	0.82	0.65	1.22
MultilayerPerceptron-23	1.00	0.79	0.79	0.58	1.27	1.00	0.83	0.83	0.70	1.20
MultilayerPerceptron-24	1.00	0.80	0.80	0.62	1.24	1.00	0.88	0.88	0.80	1.13
MultilayerPerceptron-25	1.00	0.80	0.80	0.62	1.24	1.00	0.88	0.88	0.80	1.13
RandomForest-26	1.00	0.80	0.80	0.63	1.25	1.00	0.90	0.90	0.84	1.11
RandomForest-27	1.00	0.78	0.78	0.61	1.28	1.00	0.90	0.90	0.84	1.12
RandomForest-28	0.65	0.85	0.85	0.44	0.77	0.56	0.86	0.86	0.37	0.65
Logistic-29	0.64	0.71	0.71	0.31	0.90	0.64	0.83	0.83	0.47	0.77

## 5.6 Estudo de Casos

Nesta seção são apresentados um *hindcast* de trinta dias (vinte dias aleatórios e os dez dias usados no teste dos dados modelados dos IIT extraídos do modelo WRF), que não fazem parte do conjunto de treinamento, retirados do conjunto de dados no período de 2012 a 2017 e são testados os modelos ótimos definidos na Seção 5.3 para previsão de ocorrência de EMC entre às 15 e 22h (local) e, caso ocorra EMC, se é severo.

De acordo com a Tabela 16, as previsões EMC de 16 horas e 4 horas são apresentadas aqui, com 6 e 8 modelos ótimos, respectivamente, definidos para 30 dias independentes que não foram usados no processo de treinamento dos algoritmos de AM descrito no item "b" da Seção 4.3. A previsão de severidade de EMC é destacada em amarelo e os resultados das previsões de EMC (e severidade) são representados por "A" e "E" que significam, respectivamente, acerto e erro. Em geral, o conjunto de modelos com tempos de entrada de 16 horas e 4 horas foram capazes de detectar EMC (e gravidade entre parênteses), no período de previsão definido na Seção 5.1, com alta precisão e com taxas de probabilidade de acerto variando de 0,87 a 0,90 e 0,93 a 1,00 (0,97), respectivamente.

**Tabela 16:** Resultado do estudo de caso com 30 dias de previsão de EMC (e severidade) para período entre 15 e 22h (local), usando modelos ótimos com os IIT (dados de entrada) das 00 UTC (modelos dos experimentos de números 2 a 7) e 12 UTC (modelos dos experimentos de números 22 a 29 e 36 para detecção de severidade).

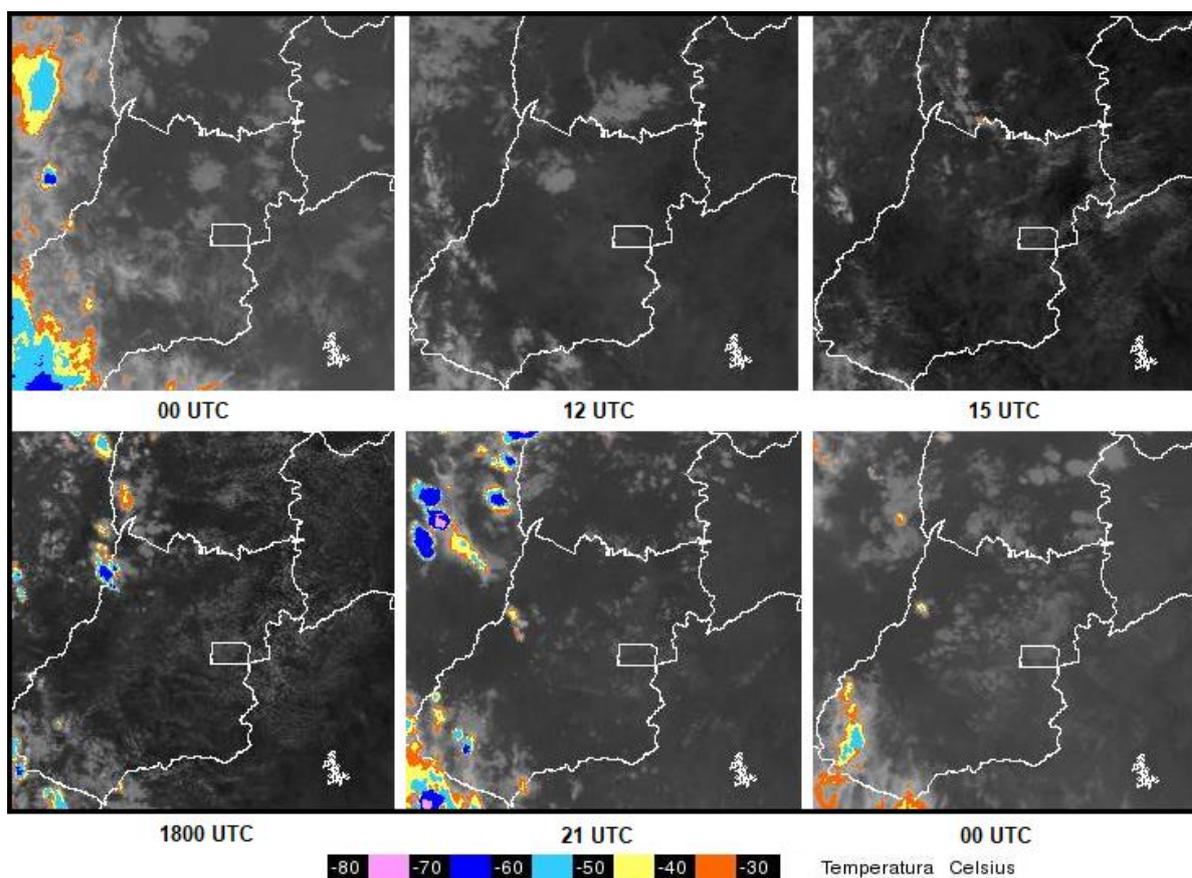
Dia	12/04/2012	23/10/2012	24/02/2013	12/05/2013	07/10/2013	25/10/2013	26/10/2013	06/03/2014	17/03/2014	10/04/2014	13/05/2014	18/11/2014	29/11/2014	03/01/2015	25/02/2015
Qtd AD	0	2038	6300	0	22215	83	60	3138	12029	6473	0	281	185	15925	358
NaiveBayes-2	E	A	A	A	A	E	E	A	A	A	A	A	A	A	A
MultilayerPerceptron-3	A	A	A	A	A	E	E	A	A	A	A	A	A	A	A
MultilayerPerceptron-4	A	A	A	A	A	E	E	A	A	A	A	A	A	A	A
MultilayerPerceptron-5	A	A	A	A	A	E	E	A	A	A	A	A	A	A	A
NaiveBayes-6	A	A	A	A	A	E	E	A	A	A	A	A	E	A	A
Logistic-7	A	A	A	A	A	A	A	A	A	A	A	A	E	A	A
RandomForest-22	E	A	A	A	A	E	E	A	A	A	A	A	A	A	A
MultilayerPerceptron-23	E	A	A	A	A	E	E	A	A	A	A	A	A	A	A
MultilayerPerceptron-24	E	A	A	A	A	E	E	A	A	A	A	A	A	A	A
MultilayerPerceptron-25	E	A	A	A	A	E	E	A	A	A	A	A	A	A	A
RandomForest-26	E	A	A	A	A	E	E	A	A	A	A	A	A	A	A
RandomForest-27	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
RandomForest-28	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Logistic-29	A	A	A	A	A	A	A	A	A	A	A	E	E	A	A
MultilayerPerceptron-36	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Dia	07/03/2015	09/03/2015	11/03/2015	02/05/2015	09/04/2016	26/06/2016	05/10/2016	02/11/2016	18/11/2016	04/12/2016	06/12/2016	12/01/2017	28/10/2017	09/11/2017	19/12/2017
Qtd AD	6044	6465	9711	206	21	0	304	420	1038	5535	6982	4884	597	102	390
NaiveBayes-2	A	A	A	A	E	A	A	A	A	A	A	A	A	A	A
MultilayerPerceptron-3	A	A	A	A	E	A	A	A	A	A	A	A	A	A	A
MultilayerPerceptron-4	A	A	A	E	A	A	A	A	A	A	A	A	A	A	A
MultilayerPerceptron-5	A	A	A	E	A	A	A	A	A	A	A	A	A	A	A
NaiveBayes-6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Logistic-7	A	A	A	E	A	A	A	E	A	A	A	A	A	A	A
RandomForest-22	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
MultilayerPerceptron-23	A	A	A	A	E	A	A	A	A	A	A	A	A	A	A
MultilayerPerceptron-24	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
MultilayerPerceptron-25	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
RandomForest-26	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
RandomForest-27	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
RandomForest-28	A	A	A	A	A	A	A	A	A	A	A	A	A	E	A
Logistic-29	A	A	A	A	A	A	A	E	A	A	A	A	A	E	A
MultilayerPerceptron-36	A	A	A	A	A	A	A	A	A	A	A	A	E	A	A

A seguir é apresentada uma discussão de alguns aspectos sinóticos para área de estudo, nos dias 25/10/2013 e 26/10/2013, em que a maioria dos modelos das 12 UTC errou no processo de previsão de EMC, assim como no 28/10/2017, em que o modelo *MultilayerPerceptron-36* previu a ocorrência de EMC severo (nesse estudo definido quando temos a ocorrência de mais de 1000 DA durante o PP), mas foi detectado pelo sistema da STARNET o número de 597 (quinhentos e noventa e sete)

descargas atmosféricas, o que caracteriza a não ocorrência severidade e, em consequência, o modelo errou.

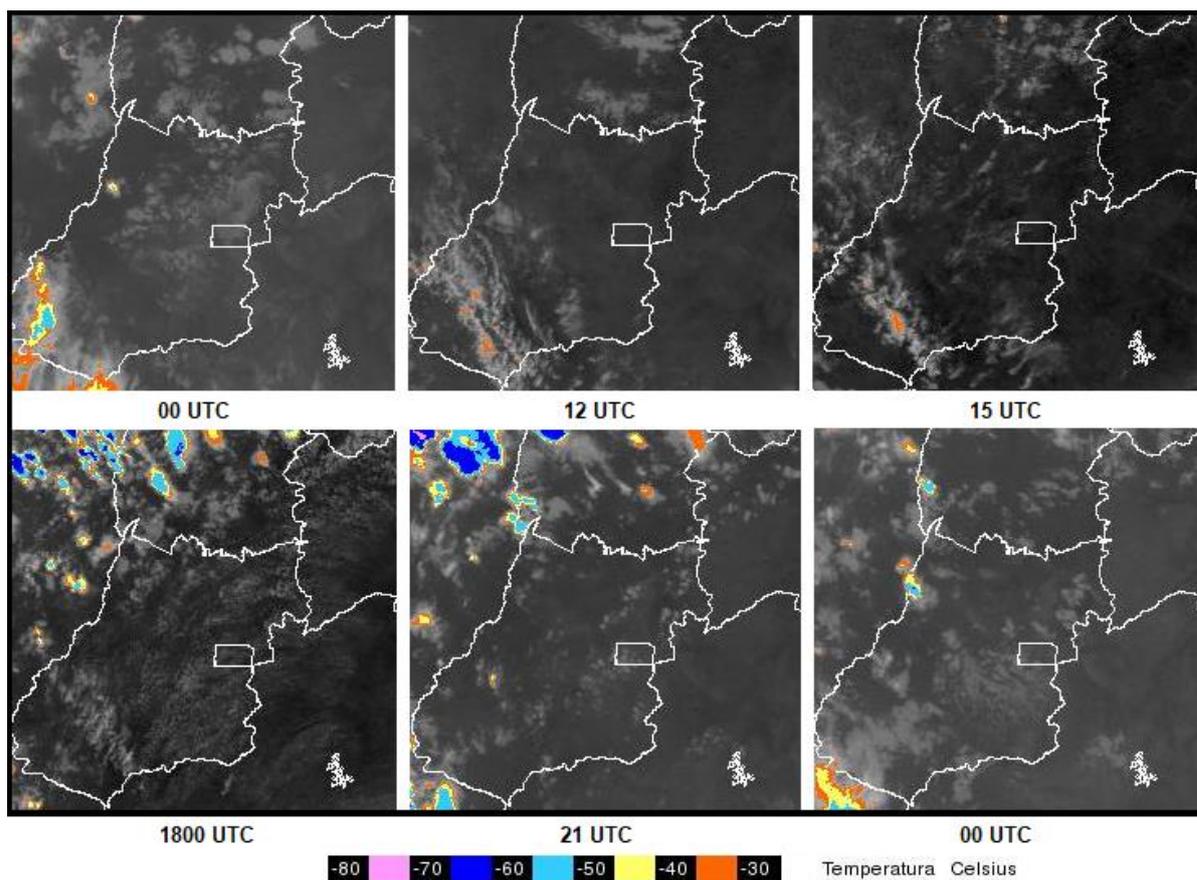
#### 5.6.1 Análise sinótica dos dias 25 de outubro de 2013 e 26 de outubro de 2013

Nos dias 25 de outubro de 2013 e 26 de outubro de 2013, a atmosfera apresentava-se estável durante todo o período de 24 horas, sem ocorrência de DA, de acordo, respectivamente, com as Figuras 14 (imagens de satélite a partir das 00 UTC do dia 25 de outubro de 2013 até às 00 UTC do dia 26 de outubro de 2013) e 15 (imagens de satélite a partir das 00 UTC do dia 26 de outubro de 2013 até às 00 UTC do dia 27 de outubro de 2013). Os IIT caracterizaram estabilidade conforme previsto pela maioria dos modelos da 00 UTC e 12 UTC. Nesses dias, supostamente, houve uma falha no sistema de detecção de descargas atmosféricas da STARNET ou a formação e dissipação de um EMC de baixo impacto em um intervalo de tempo muito curto.



**Figura 14:** Imagens de satélite GOES-13, temperatura realçada, do dia 25 de outubro de 2013.

Fonte: Adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>.

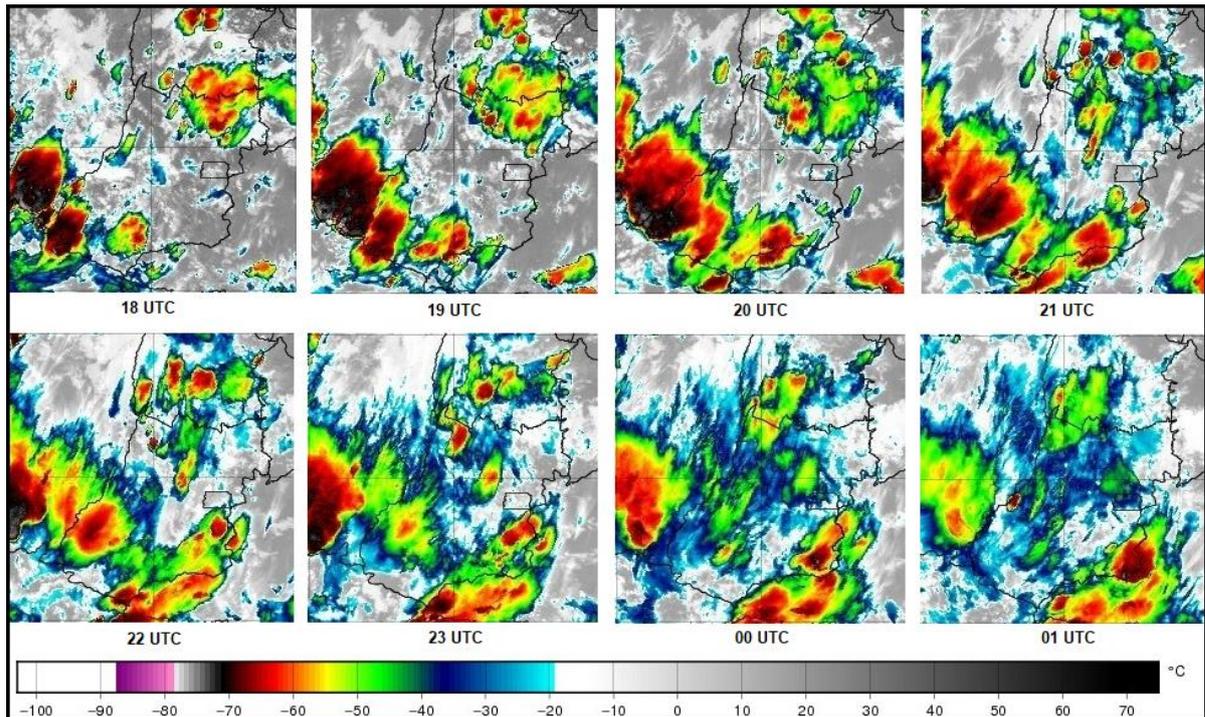


**Figura 15:** Imagens de satélite GOES-13, temperatura realçada, do dia 26 de outubro de 2013.

Fonte: Adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>.

### 5.6.2 Análise sinótica do dia 28 de outubro de 2017

Nesse dia, a formação de EMC começou no início da tarde e prolongou-se até o fim da noite, porém fora da área de estudo. Os IIT apontavam para uma atmosfera instável favorecendo a ocorrência de eventos meteorológicos severos. Analisando-se as imagens do satélite GOES-16, canal 13, da Figura 16, a partir das 18 UTC do dia 28 de outubro de 2017 até às 01 UTC do dia 29 de outubro de 2017, verifica-se que há formação de nuvens convectivas em vários setores da região centro-oeste do Brasil, o que atenua o erro do modelo multilayerperceptron-36 na classificação da severidade, pois a maioria das DA estavam ocorrendo fora da área de estudo.



**Figura 16:** Imagens de satélite GOES-16, canal Infravermelho (CH13), do dia 28 de outubro de 2017. Fonte: Adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>.

No geral, avaliando-se todo o conjunto de modelos para um determinado dia, verifica-se que eles são eficazes em prever os eventos meteorológicos convectivos e sua severidade. Os resultados também demonstram que quando o evento meteorológico convectivo severo acontece próximo à área de interesse, o modelo tem tendência a apontar para severidade.

## 6. CONCLUSÕES

Este trabalho desenvolveu modelos de previsão, utilizando-se aprendizado de máquina, visando a detecção de EMC e sua severidade para a TMA-BR, utilizando IIT obtidos de sondagens atmosféricas da EMA-BR, além de avaliar o desempenho dos modelos, usando os IIT modelados obtidos do modelo WRF, na detecção desses eventos meteorológicos.

Observou-se que as descargas atmosféricas ocorrem em maior frequência no verão quando a massa equatorial continental, cuja origem está na planície amazônica, tende a avançar para a região centro-oeste, onde provoca aumento das temperaturas, umidade e precipitações.

Foi identificado o período entre 18 e 19h (local) como o de maior ocorrência de atividade convectiva e os setores oeste, noroeste e norte da TMA-BR, os locais de maior densidade de descargas atmosféricas, assim como ficou definido o período de detecção de EMC (e sua severidade) entre às 15 e 22h (local).

Foram desenvolvidos modelos baseados em técnicas de aprendizado de máquina com resultados estatísticos satisfatórios, baseados nos dados de IIT das radiossondagens das 00 e 12 UTC com capacidade de prognosticar o EMC com antecedência, respectivamente, de 16 horas e 4 horas. O modelo ótimo desenvolvido para classificar a severidade de EMC (*MultilayerPerceptron-36*), com até 4 horas de antecedência, também teve índices estatísticos aceitáveis com POD de 0,82 e FAR de 0,22.

Na tentativa de estender a previsão para 30 horas dos modelos de AM, foi testada a aplicabilidade dos IIT, modelados numericamente com o modelo WRF, nos modelos ótimos desenvolvidos com os IIT das 12 UTC, sendo os resultados apresentados na detecção horária de EMC, durante o PP, animadores em 10 dias (estudo de caso). Verificou-se que os modelos *RandomForest-26* e *27* da Simulação II (teste de sensibilidade com as parametrizações cumulus e microfísica de DA SILVA, 2021) estão de acordo com critério estatístico definido neste trabalho para a classificação de modelos ótimos na detecção de EMC para o período de previsão.

Os resultados de *hindcasts* (30 dias) para previsão de EMC, na área de estudo, indicaram que os modelos apresentaram desempenhos satisfatórios para identificar a ocorrência, ou não, desse fenômeno meteorológico, com 16h e 4h de antecedência e classificar sua severidade com antecedência de 4h.

Esta pesquisa demonstra os benefícios e o quão promissora é a aplicação de técnicas de AM na previsão de EMC que se formam localmente e são difíceis de serem previstos somente com a utilização de modelagem numérica, extrapolação de imagens de satélite e radar, além do uso individual de IIT. A operacionalização dessa ferramenta no CIMAER e CGNA ajudaria o previsor a prestar uma informação mais rápida, permitindo que o aeronavegante e o controle de tráfego aéreo tenham mais tempo para a tomada de decisão, além de melhorar significativamente a assertividade da previsão de EMC, aumentando a confiabilidade das informações prestadas.

Como trabalhos futuros, sugere-se o ampliamiento do conjunto de dados e áreas de interesse a serem modeladas numericamente, a fim de definir configurações de parametrização regionalmente específicas para o modelo WRF, com o intuito de serem calculados IIT mais precisos e, em consequência, melhorar a aplicação operacional desses dados na previsão de EMC em diversas regiões do Brasil.

Sugere-se a inclusão de outras variáveis meteorológicas (*input*) no desenvolvimento de modelos baseados em IA, a fim de aperfeiçoar a previsão de EMC utilizando técnicas de AM, como, por exemplo, a umidade em baixos níveis, SBCAPE, SBLI (*Surface Based Lifted Index* – Índice de Levantamento calculado a partir da parcela de ar de superfície), estimativa do valor de CAPE usando a temperatura máxima prevista, umidade específica e razão de mistura em superfície.

## 7. REFERÊNCIAS

ALEXANDER, G. D., WEINMAN, J. A., KARYAMPUDI, V. M., *et al.* "The Effect of Assimilating Rain Rates Derived from Satellites and Lightning on Forecasts of the 1993 Superstorm", **Monthly Weather Review**, v. 127, n. 7, p. 1433–1457, jul. 1999. DOI: 10.1175/1520-0493(1999)127<1433:TEOARR>2.0.CO;2.

ALMEIDA, V. A. de, FRANÇA, G. B., CAMPOS VELHO, H. F. de. "Short-range forecasting system for meteorological convective events in Rio de Janeiro using remote sensing of atmospheric discharges", **International Journal of Remote Sensing**, v. 41, n. 11, p. 4372–4388, 2 jun. 2020. DOI: 10.1080/01431161.2020.1717669.

ALMEIDA, V. A. **Sistema de Previsão de Eventos Meteorológicos Convectivos para a Área Terminal do Rio de Janeiro**. Tese (Doutorado) – Programa de Pós-graduação em Engenharia Civil, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2020.

American Meteorological Society (AMS). **Lightning**. Disponível em: <<https://glossary.ametsoc.org/wiki/Lightning>>. Acesso em: 28 nov 2021>.

American Meteorological Society (AMS). **Stability Index**. Disponível em: <[https://glossary.ametsoc.org/wiki/Stability\\_index](https://glossary.ametsoc.org/wiki/Stability_index)>. Acesso em: 28 nov 2021.

ANQUETIN, S., YATES, E., DUCROCQ, V., *et al.* "The 8 and 9 September 2002 flash flood event in France: a model intercomparison", **Natural Hazards and Earth System Sciences**, v. 5, n. 5, p. 741–754, 6 out. 2005. DOI: 10.5194/nhess-5-741-2005.

BANTA, R. M., "The Role of Mountain Flows in Making Clouds". In: BANTA, R. M., BERRI, G., BLUMEN, W., *et al.*, **Atmospheric Processes over Complex Terrain**, Boston, MA, American Meteorological Society, 1990. p. 229–283. DOI: 10.1007/978-1-935704-25-6\_9. Disponível em: [http://link.springer.com/10.1007/978-1-935704-25-6\\_9](http://link.springer.com/10.1007/978-1-935704-25-6_9). Acesso em: 20 mar. 2022.

BARROS, J. R. **A chuva no Distrito Federal: o regime e as excepcionalidades do ritmo**. Dissertação (mestrado) – Programa de Pós-Graduação em Geografia. Universidade Estadual Paulista, São Paulo, 2003.

BARROS, J. R. A. e BALERO, J. C. S. **A influência do Clima e do tempo do Centro-Oeste do Brasil nas Condições de Voo na Região**. Elisee: Revista de Geografia da UEG. v. 1, n. 2, p. 25-50, 2012.

BARTHLOTT, C., CORSMEIER, U., MEISSNER, C., *et al.* "The influence of mesoscale circulation systems on triggering convective cells over complex terrain", **Atmospheric Research**, v. 81, n. 2, p. 150–175, ago. 2006. DOI: 10.1016/j.atmosres.2005.11.010.

BATTAN, L. J. "Radar observation of the atmosphere". **Quarterly Journal of the Royal Meteorological Society**, v. 99, n. 422, p. 793–793, out. 1973. DOI: 10.1002/qj.49709942229.

BHUIYAN, M. A. E, B., BEGUM, F., ILHAM, S. J., *et al.* "Advanced wind speed prediction using convective weather variables through machine learning application", **Applied Computing and Geosciences**, v. 1, p. 100002, out. 2019. DOI: 10.1016/j.acags.2019.100002.

BHUIYAN, M. A. E., ANAGNOSTOU, E. N., KIRSTETTER, P.-E. "A Nonparametric Statistical Technique for Modeling Overland TMI (2A12) Rainfall Retrieval Error", **IEEE Geoscience and Remote Sensing Letters**, v. 14, n. 11, p. 1898–1902, nov. 2017. DOI: 10.1109/LGRS.2017.2728658.

BREIMAN, L. "Random Forests", **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. DOI: 10.1023/A:1010933404324.

BRITANNICA ENCYCLOPEDIA. **Machine Learning: Artificial Intelligence**. Disponível em: <<https://www.britannica.com/technology/machine-learning>>. Acesso em: 10 dez 2021.

BROWN, R. D., ROBINSON, D. A. "Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty", **The Cryosphere**, v. 5, n. 1, p. 219–229, 16 mar. 2011. DOI: 10.5194/tc-5-219-2011.

BROWNLEE, J. Master Machine Learning Algorithms. **Discover How They Work and Implement Them from Scratch**; Machine Learning Mastery: Vermont, VIC, Australia, 2016.

BYERS, H. R., "Thunderstorms". In: BYERS, H. R., LANDSBERG, H. E., WEXLER, H., *et al.*, **Compendium of Meteorology**, Boston, MA, American Meteorological Society, 1951. p. 681–693. DOI: 10.1007/978-1-940033-70-9\_56.

CHAMEIDES, W.L. "The role of lightning in the chemistry of the atmosphere", in **The Earth's electrical Environment**, National Academy Press, Washington, D.C., 70, 1986. Disponível em: [https://books.google.com.br/books?hl=pt-BR&lr=&id=D9OKQ\\_KOzwMC&oi=fnd&pg=PA70&dq=Chameides,+W.L.,+The+role+of+lightning&ots=JjUkyNPhjl&sig=a\\_5-hgYDIRbQ-O5TbICVM2Y5u1U#v=onepage&q=Chameides%2C%20W.L.%2C%20The%20role%20of%20lightning&f=false](https://books.google.com.br/books?hl=pt-BR&lr=&id=D9OKQ_KOzwMC&oi=fnd&pg=PA70&dq=Chameides,+W.L.,+The+role+of+lightning&ots=JjUkyNPhjl&sig=a_5-hgYDIRbQ-O5TbICVM2Y5u1U#v=onepage&q=Chameides%2C%20W.L.%2C%20The%20role%20of%20lightning&f=false). Acesso em: 23 out. 2022.

CHOU, M. D., and SUAREZ, M. J., 1999: A solar radiation parameterization for atmospheric studies. **NASA Tech. Memo.** 104606, 15, 40 pp.

CHRONIS, T. G., ANAGNOSTOU, E. N. "Error analysis for a long-range lightning monitoring network of ground-based receivers in Europe: LONG-RANGE LIGHTNING RETRIEVAL ERROR", **Journal of Geophysical Research: Atmospheres**, v. 108, n. D24, p. n/a-n/a, 27 dez. 2003. DOI: 10.1029/2003JD003776.

CRAVEN, J. P., JEWELL, R. E., BROOKS, H. E. "Comparison between Observed Convective Cloud-Base Heights and Lifting Condensation Level for Two Different Lifted Parcels", **Weather and Forecasting**, v. 17, n. 4, p. 885–890, ago. 2002. DOI: 10.1175/1520-0434(2002)017<0885:CBOCCB>2.0.CO;2.

DA SILVA, Y. U. **Previsão de Eventos Convectivos severos com WRF e Técnicas de Aprendizado de Máquinas**. Dissertação (Mestrado) – Programa de Pós-Graduação em Meteorologia do Instituto de Geociências do Centro de Ciências Matemáticas e da Natureza, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

DA SILVA, Y. U., FRANÇA, G. B., RUIVO, H. M., et al. "Forecast of Convective Events Via Hybrid Model: Wrf and Machine Learning Algorithms", **SSRN Electronic Journal**, 2022. DOI: 10.2139/ssrn.4113508.

DE CASTRO, J. N., FRANÇA, G. B., DE ALMEIDA, V. A., *et al.* "Severe Convective Weather Forecast Using Machine Learning Models", **Pure and Applied Geophysics**, v. 179, n. 8, p. 2945–2955, ago. 2022. DOI: 10.1007/s00024-022-03088-8.

DECEA. **Anuário Estatístico de Tráfego Aéreo de 2020**. Disponível em: <[https://www.decea.mil.br/?i=midia-e-informacao&p=pg\\_noticia&materia=cgna-divulga-o-anuario-estatistico-2020](https://www.decea.mil.br/?i=midia-e-informacao&p=pg_noticia&materia=cgna-divulga-o-anuario-estatistico-2020)>. Acesso em: 30 jan. 2022.

DECEA. **Instrução do Comando da Aeronáutica (ICA 100-12)**. Regras do ar, 2016. Disponível em: <<https://publicacoes.decea.mil.br/publicacao/ica-100-12>>. Acesso em: 23 out. 2022.

DECEA. **Instrução do Comando da Aeronáutica (ICA 105-17)**. Centros Meteorológicos, 2020. Disponível em: <<https://publicacoes.decea.mil.br/publicacao/ica-105-17>>. Acesso em: 23 out. 2021.

DECEA. **Manual de Estações Meteorológicas de Altitude (MCA 105-9)**. Centros Meteorológicos, 2015. Disponível em: <<https://publicacoes.decea.mil.br/publicacao/mca-105-9>>. Acesso em: 23 out. 2022.

DIAS, M. A. F. S., 2000. Índices de instabilidade para previsão de chuva e tempestades severas. **Universidade de São Paulo**, SP. Domingues, P.

DIXON, M., WIENER, G. "TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-based Methodology", **Journal of Atmospheric and Oceanic Technology**, v. 10, n. 6, p. 785–797, dez. 1993. DOI: 10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2.

DOSWELL, C. A. "The Distinction between Large-Scale and Mesoscale Contribution to Severe Convection: A Case Study Example", **Weather and Forecasting**, v. 2, n. 1, p. 3–16, mar. 1987. DOI: 10.1175/1520-0434(1987)002<0003:TDBLSA>2.0.CO;2.

FACELI, K., LORENA, A. C., GAMA, J., *et al.* **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. 2ª ed. Rio de Janeiro: LTC, 2021.

FRANÇA, G. B., DE ALMEIDA, M. V., ROSETTE, A. C. "An automated nowcasting model of significant instability events in the flight terminal area of Rio de Janeiro,

Brazil", **Atmospheric Measurement Techniques**, v. 9, n. 5, p. 2335–2344, 26 maio 2016. DOI: 10.5194/amt-9-2335-2016.

FRANCO, C. R. **Inteligência Artificial**. 1ª ed. Londrina: Editora e Distribuidora Educacional, 2014.

FREITAS, J. H. V., FRANÇA, G. B., MENEZES, W. F. "Deep Convection Forecasting Using Decision Tree in Rio de Janeiro Metropolitan Area", **Anuário do Instituto de Geociências - UFRJ**, v. 42, n. 1, p. 127–134, 25 mar. 2019. DOI: 10.11137/2019\_1\_127\_134.

FUELBERG, H. E., BIGGAR, D. G. "The Preconvective Environment of Summer Thunderstorms over the Florida Panhandle", **Weather and Forecasting**, v. 9, n. 3, p. 316–326, set. 1994. DOI: 10.1175/1520-434(1994)009<0316:TPEOST>2.0.CO;2.

GEORGE, J. J. **Weather Forecasting for Aeronautics**. Academic Press, p. 673, 1960.

GOODMAN, S.J. "Predicting thunderstorm evolution using ground based lightning detection networks", **NASA Tech. Memo**, TM-10352, 210 pp., 1990.

GRECU, M., ANAGNOSTOU, E. N., ADLER, R. F. "Assessment of the Use of Lightning Information in Satellite Infrared Rainfall Estimation", **Journal of Hydrometeorology**, v. 1, n. 3, p. 211–221, jun. 2000. DOI: 10.1175/1525-7541(2000)001<0211:AOTUOL>2.0.CO;2.

GRELL, G. A. and Freitas, S. R. A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, *Atmos. Chem. Phys.*, 14, 5233–5250, 2014. doi:10.5194/acp-14-5233-2014.

GRELL, G. A., FREITAS, S. R. "A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling", **Atmospheric Chemistry and Physics**, v. 14, n. 10, p. 5233–5250, 27 maio 2014. DOI: 10.5194/acp-14-5233-2014.

GRELL, G. A., and Devenyi, D. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. **Geophys. Res. Lett.**, 29(14), 2002.

GUIKEMA, S. D., NATEGHI, R., QUIRING, S. M., *et al.* "Predicting Hurricane Power Outages to Support Storm Response Planning", **IEEE Access**, v. 2, p. 1364–1373, 2014. DOI: 10.1109/ACCESS.2014.2365716.

GUIKEMA, S. D., QUIRING, S. M. "Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data", **Reliability Engineering & System Safety**, v. 99, p. 178–182, mar. 2012. DOI: 10.1016/j.ress.2011.10.012.

GUIKEMA, S. D., QUIRING, S. M., HAN, S.-R. "Prestorm Estimation of Hurricane Damage to Electric Power Distribution Systems: Prestorm Estimation of Hurricane Damage", **Risk Analysis**, v. 30, n. 12, p. 1744–1752, dez. 2010. DOI: 10.1111/j.1539-6924.2010.01510.x.

GULTEPE, I., SHARMAN, R., WILLIAMS, P. D., *et al.* "A Review of High Impact Weather for Aviation Meteorology", **Pure and Applied Geophysics**, v. 176, n. 5, p. 1869–1921, maio 2019. DOI: 10.1007/s00024-019-02168-6.

HAKLANDER, A. J., VAN DELDEN, A. "Thunderstorm predictors and their forecast skill for the Netherlands", **Atmospheric Research**, v. 67–68, p. 273–299, jul. 2003. DOI: 10.1016/S0169-8095(03)00056-5.

HONG, S.-Y., and LIM, J.-O. J., 2006: The WRF single-moment 6-class microphysics scheme (WSM6). **J. Korean Meteor. Soc.**, 42, 129–151.

HOUZE, R. A. **Cloud dynamics**. Second edition ed. Amsterdam, v. 104; New York, Academic Press is an imprint of Elsevier, 2014.

HUBBERT, J. C., WILSON, J. W., WECKWERTH, T. M., *et al.* "S-Pol's Polarimetric Data Reveal Detailed Storm Features (and Insect Behavior)", **Bulletin of the American Meteorological Society**, v. 99, n. 10, p. 2045–2060, out. 2018. DOI: 10.1175/BAMS-D-17-0317.1.

HUNTRIESER, H., SCHIESSER, H. H., SCHMID, W., *et al.* "Comparison of Traditional and Newly Developed Thunderstorm Indices for Switzerland", **Weather and Forecasting**, v. 12, n. 1, p. 108–125, mar. 1997. DOI: 10.1175/1520-0434(1997)012<0108:COTAND>2.0.CO;2.

ICAO. Global Air Navigation Plan. 6ª ed., 2016. 142 f., **Organização de Aviação Civil Internacional**, Montreal, Canadá.

ISAAC, G. A., BAILEY, M., BOUDALA, F. S., *et al.* "The Canadian Airport Nowcasting System (CAN-Now): The Canadian Airport Nowcasting System (CAN-Now)", **Meteorological Applications**, v. 21, n. 1, p. 30–49, jan. 2014. DOI: 10.1002/met.1342.

ISAAC, G. A., BAILEY, M., BOUDALA, F. S., *et al.* "The Canadian Airport Nowcasting System (CAN-Now): The Canadian Airport Nowcasting System (CAN-Now)", **Meteorological Applications**, v. 21, n. 1, p. 30–49, jan. 2014. DOI: 10.1002/met.1342.

ISAAC, G. A., BAILEY, M., BOUDALA, F., *et al.* "Decision Making Regarding Aircraft De-Icing and In-Flight Icing Using the Canadian Airport Nowcasting System (CAN-Now)". In: **SAE 2011 International Conference on Aircraft and Engine Icing and Ground Deicing**, 13 jun. 2011. **Anais [...]** [S.l.: s.n.], 13 jun. 2011. p. 2011-38–0029. DOI: 10.4271/2011-38-0029.

ISAAC, G., BAILEY, M., COBER, S., *et al.* "Airport Vicinity Icing and Snow Advisor (AVISA)". In: **44th AIAA Aerospace Sciences Meeting and Exhibit**, 9 jan. 2006. **Anais [...]** Reno, Nevada, American Institute of Aeronautics and Astronautics, 9 jan. 2006. DOI: 10.2514/6.2006-1219.

JIMÉNEZ, P. A., DUDHIA, J., GONZÁLEZ-ROUCO, J. F., *et al.* "A Revised Scheme for the WRF Surface Layer Formulation", **Monthly Weather Review**, v. 140, n. 3, p. 898–918, 1 mar. 2012. DOI: 10.1175/MWR-D-11-00056.1.

JOHNSON, J. T., MACKEN, P. L., WITT, A., *et al.* "The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm", **Weather and Forecasting**, v. 13, n. 2, p. 263–276, jun. 1998. DOI: 10.1175/1520-0434(1998)013<0263:TSCIAT>2.0.CO;2.

KANE, R. J. "Correlating Lightning to Severe Local Storms in the Northeastern United States", **Weather and Forecasting**, v. 6, n. 1, p. 3–12, mar. 1991. DOI: 10.1175/1520-0434(1991)006<0003:CLTSLS>2.0.CO;2.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **International joint Conference on artificial intelligence**. [S.l.: s.n.], 1995. v. 14, p. 1137–1145

KUNZ, M. "The skill of convective parameters and indices to predict isolated and severe thunderstorms", **Natural Hazards and Earth System Sciences**, v. 7, n. 2, p. 327–342, 3 maio 2007. DOI: 10.5194/nhess-7-327-2007.

LANDIS, J. R., KOCH, G. G. "The measurement of observer agreement for categorical data", **Biometrics**, v. 33, n. 1, p. 159–174, mar. 1977.

LEE, A. C. L. "An experimental study of the remote location of lightning flashes using a VLF arrival time difference technique", **Quarterly Journal of the Royal Meteorological Society**, v. 112, n. 471, p. 203–229, jan. 1986<sup>a</sup>. DOI: 10.1002/qj.49711247112.

LEE, A. C. L. "An Operational System for the Remote Location of Lightning Flashes Using a VLF Arrival Time Difference Technique", **Journal of Atmospheric and Oceanic Technology**, v. 3, n. 4, p. 630–642, dez. 1986<sup>b</sup>. DOI: 10.1175/1520-0426(1986)003<0630:AOSFTR>2.0.CO;2.

LEE, R. R., PASSNER, J. E. "The Development and Verification of TIPS: An Expert System to Forecast Thunderstorm Occurrence", **Weather and Forecasting**, v. 8, n. 2, p. 271–280, jun. 1993. DOI: 10.1175/1520-0434(1993)008<0271:TDAVOT>2.0.CO;2.

LIMA, I., PINHEIRO, C. A. M., SANTOS, F. A. O. **Inteligência Artificial**. 1<sup>a</sup> ed. Rio de Janeiro: Campus, 2014.

LIMA, K. C., GOMES, R. G. "Detecção de descargas elétricas atmosféricas em sistemas convectivos com dados do SIMEPAR", **Revista Brasileira de Geofísica**, v. 27, n. 1, p. 5–16, mar. 2009. DOI: 10.1590/S0102-261X2009000100001.

MACHINE LEARNING: ARTIFICIAL INTELLIGENCE. In: **Britannica Encyclopedia**. Disponível em: <<https://www.britannica.com/technology/machine-learning>>. Acesso em: 10 dez 2021.

MANZATO, A. "The Use of Sounding-Derived Indices for a Neural Network Short-Term Thunderstorm Forecast", **Weather and Forecasting**, v. 20, n. 6, p. 896–917, 1 dez. 2005. DOI: 10.1175/WAF898.1.

MCCANN, D. W. "A Neural Network Short-Term Forecast of Significant Thunderstorms", **Weather and Forecasting**, v. 7, n. 3, p. 525–534, set. 1992. DOI: 10.1175/1520-0434(1992)007<0525:ANNSTF>2.0.CO;2.

MCCULLOCH, W. S., PITTS, W. "A logical calculus of the ideas immanent in nervous activity", **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, dez. 1943. DOI: 10.1007/BF02478259.

MEIBNER, C., KALTHOFF, N., KUNZ, M., *et al.* "Initiation of shallow convection in the Black Forest mountains", **Atmospheric Research**, v. 86, n. 1, p. 42–60, out. 2007. DOI: 10.1016/j.atmosres.2007.03.003.

MERRITT, L. P. "Comparison of Airborne and Ground Based Weather Radars", **Journal of Applied Meteorology**, v. 8, n. 6, p. 963–974, dez. 1969. DOI: 10.1175/1520-0450(1969)008<0963:COAAGB>2.0.CO;2.

MILLER, R. C. "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, 1972.

MOLLER, A. R. "Severe Local Storms Forecasting", **Meteorological Monographs**, v. 50, p. 433–480, 1 nov. 2001. DOI: 10.1175/0065-9401-28.50.433.

MONTEIRO, C. A. F. **Sobre a análise geográfica de sequências de cartas de tempo** (Pequeno ensaio metodológico sobre o estudo do clima no escopo da Geografia). *Revista Geográfica*, Rio de Janeiro, v. 32, n. 58, pp. 169-179, 1963.

MOOR, J. "The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years". **AI Magazine**, v. 27, No., 4, pp. 87-9, 2006.

MORALES, C. A., ANAGNOSTOU, E. N. "Extending the Capabilities of High-Frequency Rainfall Estimation from Geostationary-Based Satellite Infrared via a Network of Long-Range Lightning Observations", **Journal of Hydrometeorology**, v. 4, n. 2, p. 141–159, abr. 2003. DOI: 10.1175/1525-7541(2003)4<141:ETCOHR>2.0.CO;2.

MORALES, W. A. **Convergência do Fluxo de Umidade Verticalmente Integrada e os prognósticos do modelo WRF**. 101f. Dissertação (Mestrado) – Programa de Pós-Graduação em Meteorologia. Universidade Federal de Pelotas, Pelotas. 2013.

MORRISON, H., G. THOMPSON, V. TATARSKII. Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One– and Two–Moment Schemes. **Mon. Wea. Rev.**, 137, 991–1007, 2009.

MUELLER, C., SAXEN, T., ROBERTS, R., *et al.* "NCAR Auto-Nowcast System", **Weather and Forecasting**, v. 18, n. 4, p. 545–561, ago. 2003. DOI: 10.1175/1520-0434(2003)018<0545:NAS>2.0.CO;2.

NASCIMENTO, E. D. L. "Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: uma estratégia operacional adotável no Brasil?", **Revista Brasileira de Meteorologia**, v. 20, p. 20, 2005.

NILSSON, N. Artificial Intelligence: A New Synthesis. **China Machine Press**, p. 1-2, 1998.

ORVILLE, H. D. "A Numerical Study of the Initiation of Cumulus Clouds over Mountainous Terrain", **Journal of the Atmospheric Sciences**, v. 22, n. 6, p. 684–699, nov. 1965. DOI: 10.1175/1520-0469(1965)022<0684:ANSOTI>2.0.CO;2.

PAULUCCI, T. B., FRANÇA, G. B., LIBONATI, R., *et al.* "Long-Term Spatial–Temporal Characterization of Cloud-to-Ground Lightning in the Metropolitan Region of Rio de Janeiro", **Pure and Applied Geophysics**, v. 176, n. 11, p. 5161–5175, nov. 2019. DOI: 10.1007/s00024-019-02216-1.

POLIFKE, F. S. **Previsão de tempestades severas e de transbordamentos no rio Quitandinha – Petrópolis/RJ**. Tese (Doutorado) – Programa de Pós-Graduação em Engenharia Civil, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.

POLIFKE, F. S., SILVA, M. G. A. J., MENEZES, W. F., *et al.* "Atmospheric Indicators Evaluation Using WRF Numerical Model in Rainfall Events in the City of Rio de Janeiro", **Anuário do Instituto de Geociências - UFRJ**, v. 38, n. 2, p. 81, 16 fev. 2016. DOI: 10.11137/2015\_2\_81\_90.

POWERS, J. G., KLEMP, J. B., SKAMAROCK, W. C., *et al.* "The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions", **Bulletin of the American Meteorological Society**, v. 98, n. 8, p. 1717–1737, 1 ago. 2017. DOI: 10.1175/BAMS-D-15-00308.1.

RASMUSSEN, R., POLITOVICH, M., SAND, W., *et al.* "Winter Icing and Storms Project (WISP)", **Bulletin of the American Meteorological Society**, v. 73, n. 7, p. 951–974, jul. 1992. DOI: 10.1175/1520-0477(1992)073<0951:WIASP>2.0.CO;2.

RYZHKOV, A. V., ZRNIC, D. S., HUBBERT, J. C., *et al.* "Polarimetric Radar Observations and Interpretation of Co-Cross-Polar Correlation Coefficients", **Journal of Atmospheric and Oceanic Technology**, v. 19, n. 3, p. 340–354, 1 mar. 2002. DOI: 10.1175/1520-0426-19.3.340.

SASAKI, Y. "The Truth of the F-Measure". **School of Computer Science**, University of Manchester, Manchester, 2007.

SCHULTZ, P. "Relationships of Several Stability Indices to Convective Weather Events in Northeast Colorado", **Weather and Forecasting**, v. 4, n. 1, p. 73–80, mar. 1989. DOI: 10.1175/1520-0434(1989)004<0073:ROSSIT>2.0.CO;2.

SCHUMANN, W. O., KONIG, H. " **Über die Beobachtung von Atmosphericis bei geringsten, Naturwiss-senschaften**", v. 41, n. 8, p. 183–184, 1954. DOI: 10.1007/BF00638174.

SCHUUR, T. J., PARK, H.-S., RYZHKOV, A. V., *et al.* "Classification of Precipitation Types during Transitional Winter Weather Using the RUC Model and Polarimetric

Radar Retrievals", **Journal of Applied Meteorology and Climatology**, v. 51, n. 4, p. 763–779, abr. 2012. DOI: 10.1175/JAMC-D-11-091.1.

SHANNON, C. E. "XXII. Programming a computer for playing chess", **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 41, n. 314, p. 256–275, mar. 1950. DOI: 10.1080/14786445008521796.

SHOWALTER, A. K. "A stability index for forecasting thunderstorms". **Bulletin of the American Meteorological Society**, v. 34, p. 250-252, 1947.

SKAMAROCK, W. C., KLEMP, J. B., DUDHIA, J., et al. A Description of the Advanced Research WRF Model Version 4. [S.l.], UCAR/NCAR, 29 mar. 2019. Disponível em: <https://opensky.ucar.edu/islandora/object/opensky:2898>. Acesso em: 13 nov. 2021.

SMITH, T. M., LAKSHMANAN, V., STUMPF, G. J., et al. "Multi-Radar Multi-Sensor (MRMS) Severe Weather and Aviation Products: Initial Operating Capabilities", **Bulletin of the American Meteorological Society**, v. 97, n. 9, p. 1617–1630, 1 set. 2016. DOI: 10.1175/BAMS-D-14-00173.1.

SOARES, C. M. **Modelo de Previsão de Eventos Convectivos Baseado em Inteligência Artificial para a Rota Aérea Rio de Janeiro – São Paulo**. Dissertação (Mestrado) – Programa de Pós-Graduação em Meteorologia do Instituto de Geociências do Centro de Ciências Matemáticas e da Natureza, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

SOARES, C. M., FRANÇA, G. B., DE ALMEIDA, M. V., et al. "Study of GOES-R Thermodynamic Indices for Short-Term Forecasting of Convective Weather Events Using Machine Learning", **Pure and Applied Geophysics**, v. 178, n. 11, p. 4651–4663, nov. 2021. DOI: 10.1007/s00024-021-02889-7.

SZWARCFITER, J. L., MARKENZON, L. **Estruturas de dados e seus algoritmos**. Rio de Janeiro: LTC, ed. 3, v. 1, p. 320, 1994.

TEIXEIRA, M. S., SATYAMURTY, P. "Dynamical and Synoptic Characteristics of Heavy Rainfall Episodes in Southern Brazil", **Monthly Weather Review**, v. 135, n. 2, p. 598–617, 1 fev. 2007. DOI: 10.1175/MWR3302.1.

TEWARI, M., CHEN, F., WANG, W., DUDHIA, J., LEMONE, M. A., MITCHELL, K., EK, M., GAYNO, G., WEGIEL, J., and CUENCA, R. H. Implementation and verification of the unified NOAA land surface model in the WRF model. **20th conference on weather analysis and forecasting/16th conference on numerical weather prediction**. 2004.

TURMAN, B. N., TETTELBACH, R. J. "Synoptic-Scale Satellite Lightning Observations in Conjunction with Tornadoes", **Monthly Weather Review**, v. 108, n. 11, p. 1878–1882, nov. 1980. DOI: 10.1175/1520-0493(1980)108<1878:SSSLOI>2.0.CO;2.

WEISMAN, M. L., KLEMP, J. B., "Characteristics of Isolated Convective Storms". In: RAY, P. S. (Org.), **Mesoscale Meteorology and Forecasting**, Boston, MA, American Meteorological Society, 1986. p. 331–358. DOI: 10.1007/978-1-935704-20-1\_15.

WILKS, D. S. **Statistical methods in the atmospheric sciences**. 2nd ed. Amsterdam; Boston, Academic Press, 2006. (International geophysics series, v. 91).

WILLIAMS, E. R. "The Schumann Resonance: A Global Tropical Thermometer", **Science**, v. 256, n. 5060, p. 1184–1187, 22 maio 1992. DOI: 10.1126/science.256.5060.1184.

WILLIAMS, E. R., WEBER, M. E., ORVILLE, R. E. "The relationship between lightning type and convective state of thunderclouds", **Journal of Geophysical Research**, v. 94, n. D11, p. 13213, 1989. DOI: 10.1029/JD094iD11p13213.

WILSON, J. W., CROOK, N. A., MUELLER, C. K., *et al.* "Nowcasting Thunderstorms: A Status Report", **Bulletin of the American Meteorological Society**, v. 79, n. 10, p. 2079–2099, out. 1998. DOI: 10.1175/1520-0477(1998)079<2079:NTASR>2.0.CO;2.

WILSON, J., and WILK, K., 1981, "Nowcasting applications of Doppler radar (invited paper)", v. 165 (05), pp. 123.

## ANEXO I

### LISTA DE ALGORITMOS UTILIZADOS NA PESQUISA<sup>4</sup>

ALGORÍTIMOS DO WEKA	DESCRIÇÃO DO ALGORÍTIMO
BayesNet	Algoritmo de classificação com a aprendizagem sendo realizada utilizando vários algoritmos de pesquisa e medidas de qualidade. Fornece estruturas de dados (estrutura de rede, distribuições de probabilidade condicional etc.) e recursos comuns aos algoritmos de aprendizagem como K2 e B.
NaiveBayes	Algoritmo de classificação utilizando classes de estimadores.
Logistic	Classe de construção e utilização de modelo de regressão logística multinominal com estimador de crista.
MultilayerPerceptron	Um classificador que usa retropropagação para aprender um perceptron multicamadas para classificar instâncias. A rede pode ser construída manualmente ou configurada usando uma heurística simples. Os parâmetros de rede também podem ser monitorados e modificados durante o tempo de treinamento. Os nós nesta rede são todos sigmóides (exceto quando a classe é numérica, caso em que os nós de saída se tornam unidades lineares sem limites)
SGD	Implementa descida de gradiente estocástico para aprender vários modelos lineares (classe binária SVM, classe binária de regressão logística, perda quadrada, perda de Huber e regressão linear de perda insensível a $\epsilon$ ). Substitui globalmente todos os valores ausentes e transforma os atributos nominais em binários. Ele também normaliza todos os atributos, portanto, os coeficientes na saída são baseados nos dados normalizados. Para atributos de classe numéricos, a função de perda ao quadrado, Huber ou $\epsilon$ -insensível deve ser usada. Não sensível a $\epsilon$ e perda de Huber podem exigir uma taxa de aprendizado muito maior.
SVM	Implementa o algoritmo de otimização sequencial mínima de John Platt para treinar um classificador de vetores de suporte. Essa implementação substitui globalmente todos os valores ausentes e transforma os atributos nominais em binários. Ele também normaliza todos os atributos por padrão. (Nesse caso, os coeficientes na saída são baseados nos dados normalizados, não nos dados originais - isso é importante para interpretar o classificador.)

<sup>4</sup> Dados obtidos do sítio de Internet do Manual do Software WEKA, disponível em <https://weka.sourceforge.io/doc.dev/allclasses-noframe.html>.

ALGORÍTIMOS DO WEKA	DESCRIÇÃO DO ALGORÍTIMO
Simple Logistic	Classificador para construção de modelos de regressão logística linear. LogitBoost com funções de regressão simples como aprendizes básicos é usado para ajustar os modelos logísticos. O número ideal de iterações LogitBoost a serem executadas é validado cruzado, o que leva à seleção automática de atributos.
IBk	Classificador de vizinhos K mais próximos. Pode-se selecionar o valor apropriado de K com base na validação cruzada. Também pode-se fazer ponderação de distâncias.
KStar	Classificador baseado em instância, ou seja, a classe de uma instância de teste é baseada na classe dessas instâncias de treinamento semelhantes a ela, conforme determinado por alguma função de similaridade. Ele difere de outras classes baseadas em instâncias porque usa uma função de distância baseada em entropia.
LWL	Usa um algoritmo baseado em instância para atribuir pesos de instância que são então usados por um <i>Weighted Instances Handler</i> especificado. Pode fazer classificação (por exemplo, usando Bayes ingênuo) ou regressão (por exemplo, usando regressão linear).
AdaBoost M1	Classe para impulsionar um classificador de classe nominal usando o método Adaboost M1. Apenas problemas de classe nominal podem ser resolvidos.
Attribute Selected Classifier	O dimensionamento dos dados de treinamento e teste é reduzido pela seleção de atributos antes de serem transmitidos a um classificador.
Bagging	Classe para empacotar um classificador para reduzir a variância. Pode fazer classificação e regressão dependendo da classe base.
LogitBoost	Classe para realização de regressão logística aditiva. Esta classe realiza a classificação usando um esquema de regressão como a classe base e pode lidar com problemas de várias classes.
RandomCommITee	Classe para construir um conjunto de classificadores de base aleatórios. Cada classificador base é construído usando um <i>seed</i> aleatório diferente (mas com base nos mesmos dados). A previsão final é uma média direta das previsões geradas pelos classificadores base individuais.
Decision Table	Classe para construir e usar uma tabela de decisão simples para classe majoritária.

ALGORÍTIMOS DO WEKA	DESCRIÇÃO DO ALGORÍTIMO
JRip	Esta classe implementa um aprendizado de regra proposicional, Poda incremental repetida para produzir redução de erros (RIPPER), que foi proposta por William W. Cohen como uma versão otimizada do IREP.
PART	Classe para gerar lista de decisão. A premissa é separar e conquistar. Constrói uma árvore de decisão parcial em cada interação e transforma a “melhor” folha em uma regra.
HoeffdingTree	Uma árvore Hoeffding é um algoritmo de indução de árvore de decisão incremental, a qualquer momento, que é capaz de aprender a partir de fluxos de dados massivos, assumindo que os exemplos de geração de distribuição não mudam com o tempo. As árvores Hoeffding exploram o fato de que uma pequena amostra pode muitas vezes ser suficiente para escolher um atributo de divisão ideal. Essa ideia é apoiada matematicamente pelo limite de Hoeffding, que quantifica o número de observações necessárias para estimar algumas estatísticas dentro de uma precisão prescrita.
DecisionStump	Classe de construção e uso de toco de decisão. Normalmente utilizado em conjunto com um algoritmo de reforço. Faz regressão (com base no erro quadrático médio) ou classificação (com base na entropia). Os dados ausentes são tratados como um valor separado.
J48	Classe para gerar uma árvore de decisão podada ou não.
LMT	Classificador para construir árvores modelo logístico, que são árvores de classificação com funções de regressão logística nas folhas. O algoritmo pode lidar com variáveis de destino binárias e multiclases, atributos numéricos e nominais e valores ausentes.
RandomForest	Classe para construir uma floresta de árvores aleatórias.
RandomTree	Classe para construção de uma árvore que considera K atributos escolhidos aleatoriamente em cada nó. Não executa poda. Também tem uma opção para permitir a estimativa de probabilidades de classe (ou média alvo no caso de regressão) com base em um conjunto de validação.

---

ALGORÍTIMOS DO WEKA	DESCRIÇÃO DO ALGORÍTIMO
REPTree	Constrói uma árvore de decisão ou de regressão usando ganho ou variância de informação e a poda usando erro reduzido. Classifica valores para atributos numéricos apenas uma vez. Os valores ausentes são tratados dividindo as instâncias correspondentes em partes .