



Previsão de Convecção Profunda Usando Árvore de Decisão na Região Metropolitana do Rio de Janeiro- Deep Convection Forecasting Using Decision Tree in Rio de Janeiro Metropolitan Area

Juliana Hermsdorff Vellozo de Freitas; Gutemberg Borges França & Wallace Figueiredo Menezes

Universidade Federal do Rio de Janeiro, Departamento de Meteorologia, Laboratório de Meteorologia Aplicada (LMA),

Campus Cidade Universitária, 21941-916, Rio de Janeiro, Rio de Janeiro, Brasil

E-mails: julianahvf@gmail.com; gutemberg@lma.ufrj.br; wallace.menezes@gmail.com

Recebido em: 09/11/2018 Aprovado em: 21/01/2019

DOI: http://dx.doi.org/10.11137/2019_1_127_134

Resumo

Neste trabalho são estudadas e analisadas descargas atmosféricas, e árvore de decisão é usada como ferramenta de *nowcasting* de eventos convectivos intensos, baseando-se no comportamento do conjuntos dos índices de instabilidade atmosférica, na área entre as latitudes 22,7°S e 23,1°S e longitudes 43°W e 44°W, da Região Metropolitana do Rio de Janeiro (RMRJ). Para seleção de eventos convectivos, foi usado dados de descargas atmosféricas e posteriormente houve a seleção de seus respectivos índices termodinâmicos. Os eventos foram classificados como A, B, C e D de acordo com a taxa de descargas por tempestade observada na área de estudo. Os índices de instabilidade atmosférica foram calculados através do perfil atmosférico da radiossondagens diariamente lançados às 09h e 21h na Estação de ar superior Galeão, no aeroporto internacional do Galeão (SBGL) na cidade do Rio de Janeiro - RJ, no período entre Abril de 2000 a Dezembro de 2016. O total de descargas ocorrido soma 81.317 e estes filtrado dos possíveis ruídos resultaram em 60.145. Resultados da série de treinamentos executados até a obtenção da árvore são apresentados e analisados. O desempenho da árvore foi testada através dos indicadores: taxa de verdadeiro positivo, alarme falso, precisão e taxa verdadeiro negativo por classes, e seus valores são (classe): 0,995 (A), 0,015 (A), 0,888 (A), 0,981 (A); 1,000 (B), 0,042 (B), 0,828 (B), 0,949 (B); 0,824 (C), 0,135 (C), 0,632 (C), 0,858 (C), e 0,717(D), 0,056 (D), 0,928 (D), 0,942(D), respectivamente. Um estudo de caso é apresentado e sua análise mostra que a delimitação da área de estudo influenciou no resultado das estatísticas e, principalmente, daqueles relacionados com as classes D, caracterizando uma árvore de decisão robusta.

Palavras-chave: Descargas atmosféricas; Tempestades; Nowcasting; Árvore de decisão; Inteligência artificial

Abstract

This work studies lightning and uses decision trees as a tool to nowcasting intense convective events, based upon the behavior and data from sets atmospherical of instability indexes in the area between 22,7°S and 23,1°S and 43°W and 44°W of Rio de Janeiro metropolitan area. For the selection of convective events, data were used for lightning and subsequent selection of their respective thermodynamic indexes. The events were classified as A, B, C and D according to the storm lightning rate observed in the study area. The thermodynamic indexes was calculated from daily launched radiosounding balloons launched at 09:00h and 21:00h from Galeão upper-air station of Rio de Janeiro International Airport in the period between April 2000 thru December 2016. The total of lightning ammounts 81.317 and the total ammount reduced to 60.145 due to data inaccuracy. Results from the series of tests executed until obtaining the optimal decision tree are preented and analysed. The decision tree performance was tested through the indicators: true positive rate, false alarm, precision and true negative rate per class, the values are as folow: 0,995(A), 0,015(A), 0,888(A), 0,980(A); 1,000(B), 0,042(B), 0,828(B), 0,949(B); 0,824(C), 0,135(C), 0,632(C), 0,858(C) e 0,717(D), 0,056(D), 0,928(D), 0,942(D). A case study is presented and it's analysis shows that the delimitation of the study area influenced the staticis results and, mainly, those related to the D class, characterizing a robust decision tree.

Keywords: Lightinig; Thunderstorms; Nowcasting; Thermodynamic indices; Decision tree; Artificial intelligence

1 Introdução

O Brasil é um dos países de maior ocorrência de descargas atmosféricas do planeta, devido à sua grande extensão territorial e ao fato de estar localizado em uma região predominantemente tropical do globo terrestre. Estima-se que cerca de 100 milhões de descargas nuvem-solo ocorram no Brasil por ano (Pinto Jr. & Pinto, 2000).

A ocorrência de descargas atmosféricas (DA) pode gerar transtornos para a sociedade, provocando colapsos ou desligamentos na rede de distribuição de energia elétrica e na rede de telecomunicações, incêndios, acidentes na aviação e em embarcações marítimas, podendo, inclusive, provocar perdas de vida (Lima & Gomes, 2009).

De acordo com Nascimento (2005), a detecção de DA é essencial para o nowcasting e monitoramento do tempo. Os dados de precipitação, vento e especialmente a informação da atividade elétrica, tem ganhado destaque nos ambientes operacionais de meteorologia, assim como na mídia, e a previsão destes fenômenos se torna cada vez mais importante.

Neste trabalho, foram escolhidas as Árvores de Decisão (AD) como técnica primária de Inteligência Artificial (IA), para classificação e predição de tempestades convectivas profundas. Esta técnica está entre as mais poderosas ferramentas de Machine Learning, subdivisão da IA. Das técnicas de Machine Learning, as árvores de decisão ou matriz ou regra de decisão (Gagne II, 2009 apud Quinlan, 1986) foram escolhidas, pois tem se mostrado uma das técnicas mais bem-sucedidas para o aprendizado de classificação supervisionada, devido à sua utilização prática e intuitiva, e à sua portabilidade, onde é possível utilizá-la em diferentes sistemas operacionais.

Para a criação de uma AD que traga informações de severidade, primeiramente selecionou-se eventos com tempestades convectivas, através do número de DA. Após isto, índices termodinâmicos foram utilizados para “ensinar” e nortear a AD na classificação de atmosferas propícias à formação destas tempestades. Segundo Wetzel & Marin (2000), a duração de uma tempestade pode ser avaliada através das condições da severidade atmosférica, calculadas pelos índices

de instabilidade, também chamados de “indicadores de tempo severo” ou “ingredientes atmosféricos”. Os indicadores de tempo severo usados por exemplo por Silva Dias (2000), são baseados em perfis verticais de temperatura, umidade e vento, e sintetizam algumas características termodinâmicas e de cisalhamento vertical do vento, típico de situações convectivas. A utilização de forma antecedente e a análise de variáveis meteorológicas presentes nesta estrutura dinâmica e termodinâmica das tempestades são ferramentas úteis na identificação da susceptibilidade da atmosfera para a formação de sistemas convectivos mais intensos que dão origem à alta incidência de descargas atmosféricas.

Deste modo, neste trabalho se busca a definição de regras que as árvores de decisão geram, construindo um caminho de probabilidade, através da potencialidade deste conjunto de índices de instabilidade.

2 Dados Meteorológicos e Área de Estudo

Neste trabalho serão utilizados 2 tipos de dados entre o período de Abril de 2000 a Dezembro de 2106: dados de descargas atmosféricas e dados sondagens atmosféricas.

- Dados de Descargas Atmosféricas : Os dados de DA consistem em latitude, longitude, data, hora e minuto de descargas localizadas pelos sensores da Rede Integrada Nacional de Detecção de Descargas Atmosféricas (RINDAT). A RINDAT é uma rede de sensores e centrais que permite detectar em tempo real as descargas atmosféricas nuvem-solo, isto é, a maior parte das descargas que atingem o solo, em parte do território brasileiro.

- Dados de sondagens atmosféricas: Os dados da radiosondagens da atmosfera utilizados neste trabalho, foram extraídos do site da National Oceanic and Atmospheric Administration (NOAA), gerados pela REDEMET (Rede de Meteorologia do Comando da Aeronáutica). A estação utilizada é o aeródromo localizado no aeroporto internacional do Galeão (SBGL) na cidade do Rio de Janeiro - RJ, na latitude de 22° 48' 36" S e longitude 43° 15' 00" W.

A região de estudo consiste no quadrilátero formado entre as latitudes 22,7°S e 23,1°S e longi-

tudes 43°W e 44°W, com área aproximada de 4.573 km². Ele cobre totalmente a área das cidades: Rio de Janeiro, Itaguaí, Mesquita, Nilópolis e São João de Meriti e parcialmente o território das cidades: Mangaratiba, Rio Claro, Piraí, Seropédica, Queimados, Nova Iguaçu, Belford Roxo, Duque de Caxias, Magé, Guapimirim, Itaboraí, São Gonçalo, Maricá e Niterói; da Baía de Guanabara e Baía de Sepetiba e parte da região, conforme apresenta a Figura 1, onde o quadrilátero represta esta área.

3 Metodologia

A seguir serão apresentadas a ferramenta computacional e os passos metodológicos.

3.1 Ferramenta Computacional

A técnica de modelagem de IA escolhida foi a indução de árvores de decisão, para a realização da classificação destes eventos intensos. O classificador da árvore de decisão usado foi o J48, versão adaptada do algoritmo C4.5 de Quinlan (Salzberg, 1994), embutido no software livre Weka (Waikato Enviroment for Knowledge Analysis) desenvolvido pela Universidade de Waikato na Austrália (Witten & Frank, 2005).

3.2 Passos Metodológicos

Um fluxograma, observado na Figura 2, esquemmatiza o processo usada para a indução e criação desta árvore.

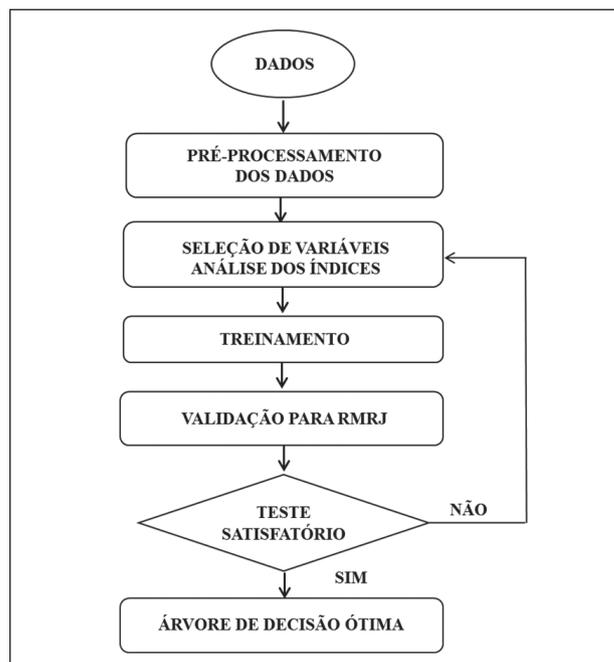
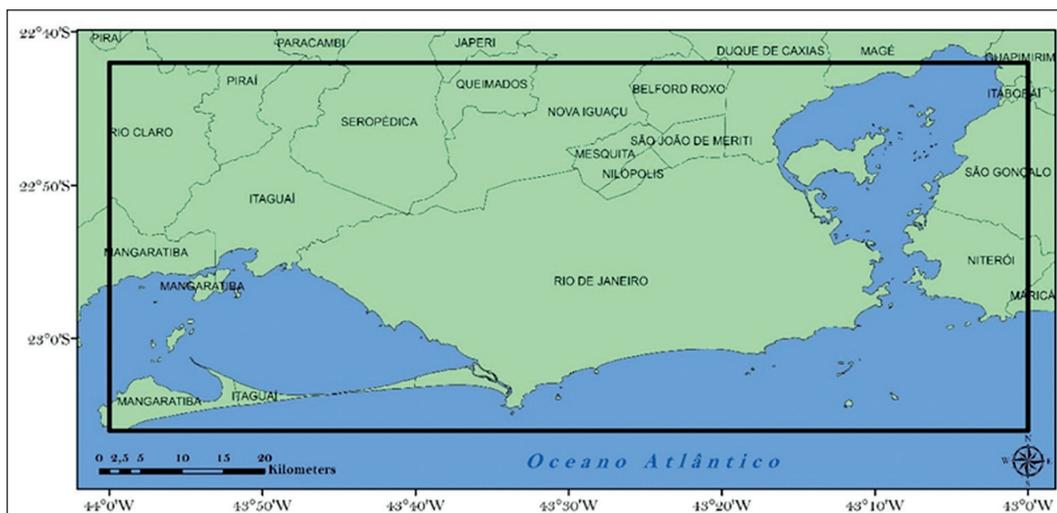


Figura 2 Fluxograma do processo de criação de uma árvore de decisão.

Desta forma, primeiramente fez-se uma seleção dos dados de DA para a contagem de tempestades e suas horas com atividade elétrica. Após isto, fez-se o pré-processamento, com o intuito de remover dados espúrios. Em sequência, selecionou-se os casos de tempestades e seus respectivos índices termodinâmicos, extraídos das sondagens atmosféricas. Estes dados oriundos das sondagens foram inseridos no Weka para treinamento e validação da AD, e após teste, selecionou-se a melhor AD, baseando-se em índices estatísticos.

Figura 1 Mapa da região de estudo, com quadrilátero formado entre as latitudes 22,7°S e 23,1°S e longitudes 43°W e 44°W representando a área em que foram filtrados os dados da RINDAT. Fonte: Google Maps.



Passo 1: Pré-Processamento dos dados

No trabalho de Lima & Gomes (2009) sobre a rede de descargas da rede RINDAT, foi utilizado um filtro de -15 kA a $+15$ kAm (kiloampere). Outros estudos (Biagi *et al.*, 2007; Orville, 2002; Cummins *et al.*, 1998) também utilizaram limites de pico de corrente, pois pequenas descargas com correntes pequenas, tornam difícil a classificação. Assim, os dados utilizados nas análises, estatísticas e resultados foram filtrados por intensidade de DA, e houve a remoção de dados espúrios em relação ao seu pico de intensidade.

Com isso, os dados deste trabalho contemplam apenas descargas com pico de corrente igual ou superior a 10 kA em módulo. Inicialmente o total de raios era de 81317, e após o filtro seu valor apresentou um total de 60145, reduzindo uma média de 26% dos dados de raios, utilizando esta metodologia.

Passo 2: Seleção de Variáveis e Análises dos Índices Termodinâmicos

Índices de instabilidade termodinâmica são designados especificamente para avaliar o potencial de tempestades convectivas, segundo Blanchard (1998). Assim, neste trabalho se busca a definição de regras que as árvores de decisão geram, que construam um caminho de probabilidade, através da potencialidade deste conjunto de índices.

Desta forma, foi verificado através também dos treinamentos da AD, que a inclusão de todos os índices termodinâmicos apresentados na metodologia, corroboraria na construção desta árvore, que indicará uma atmosfera intável e favorável às tempestades. Foram usados os seguintes índices termodinâmicos descritos na Tabela 1.

Índices de Instabilidade	Nomeclatura
K	Índice K
TT	Índice Total Totals
LI	<i>Lifted Index</i> ou Índice de Levantamento
SWEAT	<i>Severe Weather Threat</i> ou Índice Sweat
SHOWALTER	Índice Showalter
CAPE	Energia Potencial Convectiva
CINE	Inibição Convectiva

Tabela 1 Índices atmosféricos utilizados.

Para a realização de contagem de um evento de tempestade profunda, baseou-se no critério do trabalho de Paulucci (2017), onde um evento de tempestade é definido como uma sequência de DA detectadas sobre a área de estudo, sendo que o intervalo de tempo entre duas descargas consecutivas desta sequência não exceda uma hora. Assim, para inserção na árvore de decisão, foi estabelecida a criação de 4 classes baseadas no percentil do número de DA por tempestade, segundo a Tabela 2. Desta forma, a árvore poderá distinguir as peculiaridades de cada classe e seu ambiente favorável ou não às ocorrências de tempestades profundas associadas à DA. A classe A possui indicativos fortes para essa associação. As classes B e C, intermediárias, possuem menor potencial respectivamente e a D, número de eventos com ausência de DA. Porém, um dado evento com DA pode ter ocorrido fora da área de estudo, mas não implica que não houve instabilidade. Este aspecto é analisado - através de estudo de caso - na seção de resultados, onde é possível também que eventos com quadro termodinâmico característicos da classe A possam ser observados como sendo B, C ou D, uma vez no teste a árvore considera apenas densidade de DA ocorrida na área limitada pela de estudo.

CLASSES	PERCENTIL	Acumulado de Descargas por Evento (RMRJ)
A	≥ 90	≥ 267
B	90-75	$56 \leq DA \leq 266$
C	≤ 75	$1 \leq DA \leq 55$
D	Ausência de descarga	0

Tabela 2 Classes e seus respectivos percentis e acumulados de descargas atmosféricas na RMRJ.

Após isto, a seleção da sondagens (00Z e 12Z) foi feita. Esta foi realizada em função da hora que os eventos com DA ocorreram, sempre considerando o ambiente meteorológico anterior à formação de uma tempestade, com a finalidade da AD identificar o possível potencial para convecções profundas. Assim, para uma tempestade que se iniciou às 10Z (07 horas local), foi considerada a sondagem das 00Z do dia (21 horas local do dia anterior); para uma tempestade que se iniciou às 22Z (19 horas local), se considerou a sondagem das 12Z do dia (09 horas lo-

cal). Foram selecionados eventos em diferentes horários do dia, seguindo o critério de Paulucci (2017).

Passo 3 :Treinamento

No treinamento procurou-se simular no Weka a variabilidade temporal de eventos mais intensos, com a finalidade deste software enxergar estas condições. Desta forma, as classes ficaram com as seguintes porcentagens de eventos: A – 10%, B - 15%, C – 20% e D – 55%. Assim, os casos mais intensos, são mais raros no conjunto total de dados.

Para a avaliação de desempenho, utilizou-se a verificação de instâncias classificadas para cada classe na matriz de confusão, conforme Tabela 3.

		CLASSE PREVISTA	
		Sim (+)	Não (-)
CLASSE REAL	Sim (+)	TP	FN
	Não (-)	FP	TN

Tabela 3 Matriz de Confusão para 2 classe.

Na Tabela 3 os significados das letras são: TP = (true positives) positivos verdadeiros; FP = (false positives) positivos falsos; TN = (true negatives) negativos verdadeiros; FN = (false negatives) negativos falsos. As medidas de análise utilizadas neste trabalho, se encontram na Tabela 4.

Medidas Estatísticas	Significado	Equação
Taxa de verdadeiros positivos (True Positive Rate, TPR)	$TPR = \frac{TP}{TP+FN}$ É a proporção de vezes em que o evento foi previsto dentro do número de vezes em que ele ocorreu.	$TPR = \frac{TP}{TP+FN}$
Taxa de falsos positivos (False Positive Rate, FPR)	Proporção de eventos previstos que acabaram classificados incorretamente, sobre o total de negativos.	$FPR = \frac{FP}{FP + TN}$
Precisão (Precision)	Proporção de eventos positivos previstos corretamente sobre o total de eventos classificados como positivos (corretos e incorretos)	$Precisão = \frac{TP}{TP + FP}$
Taxa de verdadeiros negativos (True Negative Rate, TNR)	Proporção de não-eventos previstos corretamente sobre o total de não eventos	$TNR = \frac{TN}{FP + TN}$

Tabela 4 Índices estatísticos derivados da matriz de confusão.

4 Resultados e Discussões

Após realizada a divisão de classes, se iniciou o treinamento e criação da AD. Para este trabalho, seguindo Almeida (2009), optou-se por realizar um grupo de testes e outro de validação. O Weka possui ferramentas que permitem a criação destes grupos silmultaneamente com a criação da AD, assim como a alteração da porcentagem de cada. Desta forma, foram realizados diversos treinamentos/validações e testes ao longo do processo de aprendizagem da AD que corresponde ao lopping na Figura 2, e foram avaliados os índices estatísticos para as classificações gerais e para a classe A, mostrados na Tabela 4.

Através da Tabela 5A, observa-se que as classes no geral (A, B, C, D,) foram classificadas corretamente para todas as árvores geradas e apresentaram valores maiores que 70%, indicando um bom acerto até nos conjuntos com poucos casos para treinamento/validação. À medida que o conjunto de treinamento aumenta, isto é, mais casos de todas as classes são inseridos na construção da AD, melhores são seus resultados de desempenho classificatório - que significa a previsão do evento. Porém, deve-se considerar que um grupo muito pequeno de validação também irá gerar maiores acertos.

Após análise das árvores de decisão geradas, observou-se que a número 6, com 90% de treina-

mento/validação e 10% teste, apresentou melhor resultado, considerando um conjunto de validação com representatividade. Através da Tabela 5A observa-se um total de 81,80% de proporção de acertos em geral (casos classificados corretamente) e com 18,2% de Alarme Falso (casos classificados supostamente incorretamente). Seu valor de kappa foi o segundo maior dos treinamentos (0.7379).

A árvore de decisão escolhida, significa que o conjunto de índices de instabilidade atmosférica indicou que existem condições atmosféricas para eventos convectivos significativos (ou seja, classe A). Na Tabela 6 são apresentados os valores das

estatísticas que mensuram o desempenho da AD escolhida, baseadas nos resultados. Pode-se observar que a TRP, FPR, Precisão e TNR são bons para as classes A e B, e estes sutilmente diminuem para as classes C e D. No entanto, é importante enfatizar que talvez a forma estabelecida para testar a AD, isto é, delimitando a representatividade dos índice em função da área, neste caso a RMRJ, tenha influência nos valores menores de TRP, FPR, Precisão e TNR para as classes C e D. Neste sentido, na seção 5, um estudo de caso será analisado quando a árvore classifica o evento com A e que no conjunto de dados é definido como D.

A											
Treinamentos Geral											
Porcentagem											
Número - Árvore de Descisão	1	2	3	...	4	5	6	7	8	9	10
Treinamento / Validação	50%	55%	60%	...	80%	85%	90%	95%	96%	97%	98%
Teste	50%	45%	40%	20%	15%	10%	5%	4%	3%	2%
Estatísticas											
TPR	0.722	0.778	0.789	...	0.794	0.806	0.818	0.811	0.822	0.808	0.791
FPR	0.115	0.111	0.1	...	0.09	0.075	0.065	0.073	0.066	0.062	0.074
Precisão	0.771	0.778	0.794	...	0.806	0.826	0.842	0.906	0.845	0.847	0.819
Kappa	0.6621	0.6718	0.6896	...	0.6998	0.7196	0.7379	0.7275	0.7428	0.7252	0.7004
Total Classificadas Corretamente	77.17%	77.81%	78.85%		79.39%	80.63%	81.80%	81.14%	82.17%	80.81%	79.08%
Total Classificadas Incorretamente	22.83%	22.19%	21.15%		20.61%	19.37%	18.20%	18.86%	17.83%	19.19%	20.92%
B											
Treinamentos Classe A											
Porcentagem											
Treinamento	50%	55%	60%	...	80%	85%	90%	95%	96%	97%	98%
Teste	50%	45%	40%	...	20%	15%	10%	5%	4%	3%	2%
Estatísticas											
TPR	1	1	1	...	0.995	0.993	0.995	1	1	1	1
FPR	0.024	0.02	0.017	...	0.017	0.019	0.015	0.013	0.01	0.011	0.023
Precisão	0.835	0.857	0.876	0.874	0.856	0.888	0.906	0.926	0.917	0.843

Tabela 5 Estatísticas de treinamento na criação de uma árvore de decisão. (A) Estatísticas para diferentes conjuntos de treinamento; (B) Estatísticas referentes à classe A, para diferentes conjuntos de treinamento. Os maiores valores estão destacados em vermelho.

Classes	Taxa de Verdadeiro Positivo (TPR)	Falso Alarme (FPR)	Precisão	Taxa de Verdadeiro Negativo (TNR)
A	0,995	0,015	0,888	0,981
B	1,000	0,042	0,828	0,949
C	0,824	0,135	0,632	0,858
D	0,717	0,056	0,928	0,942

Tabela 6 Valores da taxa de verdadeiro positivo, falso alarme, precisão e taxa de falso positivo para a árvore de decisão selecionada, nas suas classes A, B, C e D, respectivamente.

5 Estudo de Caso

- Evento de D classificado como A (ocorrido em 12/03/2015)

Neste caso, onde a classe D representa casos sem DA, os pluviômetros da cidade do Rio de Janeiro (RJ) não registram precipitação, mas houve a formação de áreas de instabilidades pontuais em São Paulo (SP), como pode ser visto na imagem de satélite na Figura 3A. Porém, próximo da divisa de SP com o estado do Rio de Janeiro e na Região Serrana do estado do RJ, observado na Figura 3B, houve grande número de DA e assim, indicando novamente instabilidades fora da RMRJ. Essa área também está incluída na representatividade dos índices, reforçando mais ainda o acerto da AD.

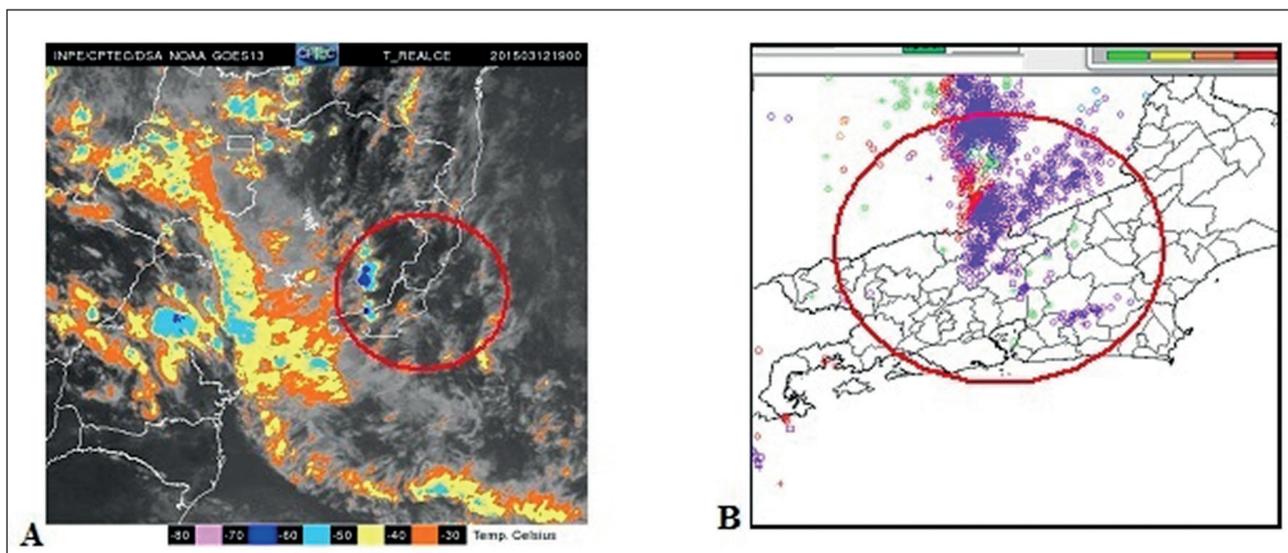


Figura 3 Imagens de ferramentas meteorológicas. (A) Imagem do satélite GOES de data 12/03/2015 às 16h00min (hora local) Fonte: CPTEC/INPE; (B) Acumulado de 24 horas de descargas atmosféricas na rede RINDAT.

6 Conclusões

Neste trabalho testa-se a potencialidade da IA de classificar e realizar previsão de eventos convectivos significativos na RMRJ, considerando o intervalo de 12 horas entre os lançamentos das radiosondagens. Assim, conclui-se que a árvore criada apresentou bons resultados ao realizar o nowcasting dos eventos convectivos significativos para todas as classes. Já o falso alarme obtido é função da área limitada pelo trabalho, para testar a árvore de decisão, conforme revela investigação do estudo de caso, onde verifica-se que representatividade espacial do conjunto de índices utilizado e, conseqüentemente, a previsão de árvore de decisão obtida é representativa além de 100 km dos limites da RMRJ.

Com isso, a metodologia utilizada neste trabalho indica uma regra de decisão forte, para ser usada como ferramenta à tomada de decisão em ambientes de meteorologia operacional.

Como futuros trabalhos, é importante expandir a área e classificar os casos de taxas de descargas atmosféricas elevadas por diferentes eventos meteorológicos, para melhor classificação da sua taxa e assim da árvore de decisão.

7 Referências

Almeida, M.V. 2009. *Aplicação de Técnicas de Redes Neurais na Previsão de Curtíssimo Prazo da Visibilidade e Teto para o Aeroporto de Guarulhos – SP*. Programa de Pós

- Graduação em Engenharia Civil, COPPE, Universidade Federal do Rio de Janeiro, Tese de Doutorado, 182p.
- Biagi, C.J.; Cummins, K.L.; Kehoe, K.E. & Krider, E.P. 2007. National lightning detection network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003–2004. *Journal of Geophysical Research*, 112, D05208.
- Blanchard, D.O. 1998. Assessing the vertical distribution of convective available potential energy. *Weather Forecasting*, 13: 870–877.
- Cummins, K.L.; Murphy, M.J.; Bardo, E.A.; Hiscox, W.L.; Pyle, R.B. & Pifer, A.E. 1998. A combined TOA/MDF technology upgrade of the US National Lightning Detection Network. *Journal of Geophysical Research*, 103(D8): 9035-9044.
- Gagne II, D.J.; McGovern, A. & Brotzge, J. 2009. Classification of convective areas using decision trees. *Journal of Atmospheric and Oceanic Technology*, 26(7): 1341-1353.
- Lima, K.C. & Gomes, R.G. 2009. Detecção de descargas elétricas atmosféricas em sistemas convectivos com dados do SIMEPAR. *Revista Brasileira de Geofísica*, 27(1): 5-16.
- Nascimento, E.L. 2005. Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: uma estratégia operacional adotável no Brasil?. *Revista Brasileira de Meteorologia*, 20(1): 121-140.
- Orville, R.E.; Huffines, G.R.; Burrows, W.R.; Holle, R.L. & Cummins, K.L. 2002. The North American lightning detection network (NALDN) - First results: 1998–2000. *Monthly Weather Review*, 130(8): 2098-2109.
- Paulucci, T.B. 2017. *Caracterização Espaço-Temporal de Descargas Atmosféricas e Tempestades Elétricas na Região Metropolitana do Rio de Janeiro entre 2001 e 2016*. Curso de Meteorologia, Universidade Federal do Rio de Janeiro, Monografia, 73p.
- Pinto Jr O. & Pinto Irca. 2000. *Tempestades e Relâmpagos no Brasil*. Instituto Nacional de Pesquisas Espaciais (INPE), São Paulo, 196 p.
- Salzberg, S.L. 1994. Book Review: C4.5: programs for Machine Learnings by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Machine Learning*, 16(3): 235–240.
- Silva Dias, M.A.F. 2000. Índices de Instabilidade para previsão de chuva e tempestades severas. Departamento de Ciências Atmosféricas, Instituto Astronômico Geofísico, Universidade de São Paulo (USP).
- Wetzell, S.W. & Martin, J.E. 2001. An Operational Ingredients – Based Methodology For Forecasting Midlatitude Winter Season Precipitation. *Weather Forecasting*, 16: 156-167.
- Witten, I.H. & Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques. second edition*. São Francisco, Morgan Kaufmann. 525 p.