PREVISÃO DE ATRASOS DE VOOS POR MACHINE LEARNING: REVISÃO BIBLIOMÉTRICA E METANALISE

Hélio da Silva Queiroz Júnior*, Viviane Adriano Falcão
Federal University of Pernambuco - UFPE, Technology and Geosciences Center - CTG

*** Corresponding author e-mail address**: helio.junior@ufpe.br

PAPER ID: SIT125

RESUMO

Um problema comum em aeroportos ao redor do mundo são os atrasos em voos comerciais. A crescente demanda por modal aéreo faz com que esses atrasos sejam cada vez mais recorrentes, causando custos adicionais e exigindo ajustes constantes na gestão de voos. Definir o método mais eficaz na previsão da ocorrência desses atrasos é um tema recorrente nas buscas de operações de trânsito. Nesses estudos, a especificidade da área analisada (seja uma companhia aérea, um aeroporto ou toda a operação de um país), a complexidade da produção buscada (previsão por regressão ou classificação) ou o tamanho da base de dados utilizada requerem métodos de análise mais robustos, e o aprendizado de máquina é uma alternativa comum ao uso de métodos estatísticos clássicos. No entanto, o número de métodos de previsão que usam o princípio da aprendizagem de máquina é vasto. A divergência entre os cenários estudados atribui diferentes verificações de precisão entre as respostas obtidas nos estudos realizados. Por isso, este projeto tem como objetivo definir os métodos mais adequados de *machine learning* para estimar a ocorrência de atrasos nos aeroportos. Com base nos métodos de revisão sistemática da literatura, grupos de publicações periódicas são levantados nas principais bases de dados internacionais, a partir de grupos de palavras-chave definidas. Assim, as análises indicam que a rede de crenças profundas, a floresta aleatória, o gradiente impulsionando árvores e redes neurais recorrentes e complicadas têm uma melhor eficácia nas respostas de previsão, e o cenário ideal de análise é aqueles específicos para aeroportos ou companhias aéreas. Portanto, o tamanho dos dados, a complexidade dos cenários e as previsões buscadas tendem a obter resultados mais adequados quando são aplicados métodos mais complexos de previsão.

Palavras-chave: Atraso de voo, Previsão, Meta-análise, Machine Learning, DEA.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico- Brasil (CNPq).

1. INTRODUÇÃO

Uma problemática comum nos aeroportos em todo o mundo são os atrasos de voos comerciais (WANG & WANG, 2019). A crescente demanda pelo modal aéreo faz com que estes atrasos (*delays*) sejam cada vez mais recorrentes, ocasionando custos adicionais e exigindo ajustes constantes no gerenciamento dos voos (LIU, YIN & HANSEN, 2019).

A aquisição de dados facilitada pela disponibilidade de novas e acessíveis tecnologias favorece a obtenção de informações úteis à aplicação em estudos de aviação em diversas áreas críticas, como em casos de análise de *big data*, gerenciamento de rede de transporte aéreo ou em previsão dos possíveis atrasos de voos utilizando para isso os métodos de aprendizado de máquina- *machine learning* (CHU et al, 2020).

Contudo, a complexidade dos dados do setor aéreo, aliada a diversidade dos modelos de *machine learning* para previsão formam a principal barreira na definição do cenário com maior acurácia na previsão.

Com isso, algumas barreiras são encontradas ao buscar estabelecer padrões que sirvam como base a aplicabilidade destes métodos de previsão no mercado. A área de análise, os parâmetros de influência nos atrasos, uma faixa de acurácia aceitável ou o modelo de previsão mais eficaz, compõem as esferas na determinação dos cenários adequados de estudos de previsão.

Diante disto, esta pesquisa objetiva analisar os distintos modelos de metodologias e métodos de previsão de atrasos e definir o cenário padrão de previsibilidade necessária para um modelo ser julgado razoável ou bom para tomada de decisão. Para isso, uma proposta metodológica foi desenvolvida, na qual foi realizada uma revisão bibliométrica para definir o comportamento das pesquisas sobre o tema, afim de que se levantasse as principais técnicas aplicadas, em seguida foi realizada uma metanalise para definir um valor padrão de acurácia a ser seguido pelo mercado na previsão de atrasos, para que então, fossem definidos os parâmetros e/ou métodos de *machine learning* mais eficientes para a realização de previsões de atrasos de voos.

2. METODOLOGIA

Afim de definir o cenário com a configuração mais adequada a realização de previsões utilizando *machine learning*, a presente pesquisa foi dividida em quatro etapas.

A primeira etapa é a de revisão sistemática de literatura, afim de que os principais trabalhos publicados a nível nacional e internacional pudessem ser coletados e avaliados. Para a realização desta etapa, foram consultados um dos principais periódicos (Scopus) sendo selecionados os artigos publicados em revistas com alto impacto e recentes.

A etapa seguinte é a de definição da acurácia universal dos métodos de *machine learning* utilizados nos estudos coletados. Nesta etapa foram utilizados os métodos de análise sistêmica de proporção e meta-análise por meta-regressão, afim de que fossem apontados os modelos com maior acurácia nas previsões de atraso por classificação (definição binária de resposta, em que “0” significa voo sem atraso e “1” voo atrasado) e por regressão (na qual o atraso é classificado de acordo com o tempo esperado de duração deste).

A divisão em cenários permite que a revisão sistemática por proporção e a meta-regressão associem o número de ocorrências de voos em um aeroporto com a proporção positiva de acertos do método de previsão aplicado, indicar a eficiência universal dos modelos de *machine learning* para previsão da ocorrência.

Baseado nas definições de Borenstein et al (2010), essas meta-análises, consideradas “clássica”, utilizam as proporções (não) transformadas e os erros padrão correspondentes no método de variância inversa, tendo o logit definindo o método de argumento "Inverso".

As etapas três e quatro são responsáveis por levantar os fatores de influência na acurácia obtida em cada estudo, além de selecionar o método mais eficaz para os cenários definidos, respectivamente.

A complexidade dos dados necessários, as áreas a serem abordadas, (seja aeroporto ou linha aérea numa escala universal ou com enfoque climático, foram avaliadas nestas etapas, sendo empregada a Análise Envoltória de Dados, DEA, afim de caracterizar a eficiência de cada estudo levantado.

3. RESULTADOS E ANÁLISES

3.1. Revisão Sistemática de Literatura

A partir do levantamento realizado na base Scopus utilizando os grupos de palavras-chave e strings selecionados, os artigos coletados no Grupo 01 (“Delay” AND "Air transportation" AND "prediction" AND ("Departure" OR "Arrival") OR "Airplane" OR "Airport") foram distribuídos em nuvem de palavras, Figura . Desta forma, foi possível identificar os conjuntos de palavras-chave que obtiveram o maior número de ocorrências nos estudos de previsão de atrasos em voos, além dos métodos utilizados com maior frequência.



Figura 1: Núvem de palavras com as palavras-chave de artigos preliminares selecionados.

Desta forma, o resultado do número de pesquisas coletadas para cada grupo foi de 150, 99 e 77 artigos para os grupos de palavras-chave e *strings* 01, 02 (“Delay” AND ("Air Transportation" OR "Airport" OR "Airline" OR " Airplane") AND ("Prediction" OR "Prevision") AND "Machine Learning") e 03 (“Delay” AND ("Air Transportation" OR "Airport" OR "Airline" OR " Airplane") AND ("Forecast" OR "Predict" OR "Estimate") AND "Machine Learning"), respectivamente.

Uma análise bibliométrica das palavras-chave dos artigos levantados para o grupo 01, com recorrência simultânea de no mínimo cinco pesquisas, demonstram que, numa visualização sobreposta ao longo dos anos, uma maior recorrência no uso de *machine learning* como método de previsão de atrasos de voos, a partir de 2019.

Por outro lado, métodos clássicos estatísticos de regressão, como a regressão logística e os modelos estocásticos, vão perdendo a participação nas análises realizadas. Este comportamento é um possível indicativo de que o uso das técnicas estatísticas servia como modelo preditivo principal, porém, depois foram utilizadas apenas como modelo de validação para a entrada de outros métodos, como os de *machine learning*, e assim deixando de serem aplicados.

Para o grupo 2, ao serem verificados os autores co-citados com maior representação nas análises de previsão, mais de 10 citações recorrentes, os resultados indicam que os autores que aparecem como referência em no mínimo vinte trabalhos correspondem a apenas 20% do total de artigos levantados.

Os três grupos linhas de pesquisa formados pelo software caracterizam a correlação entre os autores presentes em cada grupo, sendo cada linha de pesquisa liderada por um autor com um trabalho de maior impacto.

O grupo liderado por Schultz et al. (2021), analisou o impacto das condições climáticas na pontualidade dos voos em aeroportos americanos. Os métodos mais empregados nesse grupo foram as redes neurais artificiais, principalmente as redes para reconhecimento de padrões de dados como as recorrentes e convolucionais.

O segundo grupo, liderado por Kim, Choi, Brinceno e Mavris (2016) definem estudos técnicos de aplicação de métodos de aprendizado para bancos de dados densos, as chamadas *deep learning*, no tráfego aéreo americano. Entre os métodos aplicados, são destaques do grupo as redes neurais recorrentes (*Recurrent Neural Networks - RNN*, em inglês) e o K-ésimo vizinho mais próximo (*k-nearest neighbors algorithm*, em inglês).

Por fim, o grupo de maior impacto de co-ocorrências, com 39 citações, e maior densidade de correlações (medidas por conexões), 982 conexões existentes, é o grupo liderado por Hansen. O terceiro grupo tem enfoque em análises de eficiência de rota (Liu, Hansen, Ball, & Lovell, 2021), na sazonalidade e nas condições climáticas (Gurkaran & Hansen, 2013).

Desta forma, caracterizar a eficiência média dos modelos de *machine learning* em realizar previsões de atrasos dos voos, além de definir uma situação adequada de dados e zonas

analisadas, permitirão servir como base a estudos futuros e a aproximar as respostas obtidas a situação real de um aeródromo ou linha aérea.

3.2. Acurácia universal dos métodos de *machine learning*

Ao analisar um cenário em que as regressões por predição classificatória binária (0 = voo *on-time*, 1 = voo atrasado), denominado de cenário A, 14 artigos que continham as informações de número total de voos aplicados na pesquisa e a acurácia do sistema para esse tipo de previsão foram utilizados.

O resultado do modelo revisão sistêmica randômica para o cenário A, Figura 3, indica que foram consideradas as dimensões das amostras aplicadas nos estudos, visto os mesmos pesos serem atribuídos aos artigos estudados. Esta análise apresenta como resultado para os métodos de *machine learning* no geral 82,59% de acurácia média, com variação de 87,37 % e 77,22% para mais e para menos, respectivamente.

Da mesma maneira, um modelo de meta análise de proporção com 15 artigos selecionados, os quais possuíam os mesmos parâmetros aplicados na análise para o cenário A, com distinção as acurácias das previsões analisadas, sendo do tipo regressão, ou seja, o modelo previa e atribuía os atrasos classes definidas pelos autores, de acordo com o tempo de atraso, denominado de cenário B. O resultado indica uma melhora na acurácia média dos métodos aplicados, com 85,96% de precisão dos resultados, numa faixa de erro variante em 4,56% e 5,28% para mais e para menos, respectivamente, Figura 4.

Contudo, com a alta heterogeneidade entre os artigos analisados no cenário A e B, é necessário realizar a distribuição dos estudos pela assimetria de gráfico de funil, com base no método de regressão linear.

Para isso foi aplicado um teste estatístico baseado no método FPV de Macaskill et al. (2001), o qual utiliza uma regressão linear ponderada do efeito do tratamento no tamanho total da amostra com pesos recíprocos à variância da probabilidade média do evento.

A alta similaridade entre os artigos levantados para os dois cenários fazem com que os resultados sejam iguais as duas análises. Desta forma, com uma distribuição t modular de 1,05 e p-valor de 0,3123, a estatística define o descarte da hipótese nula de existência de vieses de publicação entre as pesquisas analisadas.

Contudo, ao se verificar a distribuição do erro padrão pela prevalência normalizada pelo método da proporção transformada em arco duplo de Freeman-Turkey no gráfico de túnel, Figura 2, é possível tomar a conclusão contrária do teste estatístico, já que, apenas um trabalho se enquadra dentro do intervalo de confiança de 95% representado pelo funil.

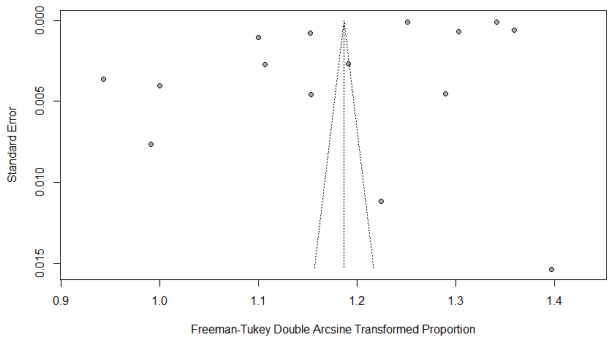


Figura 2 Gráfico de túnel, método de Freeman-Turkey.

Portanto, pode se concluir que, apesar da apresentação dispersa dos estudos no gráfico de funil, as análises estatísticas revelam que não há uma tendência viesada dos estudos, sendo este um comportamento característico de estudos de prevalência, os quais não obedecem a uma distribuição normal, podendo ser adotadas as acurácias universais obtidas.

3.3. Fatores de influência a acurácia

Para definir os ambientes mais adequados a atingir melhores resultados de acurácia dos modelos preditivos, foram definidos cenários de artigos em que a área de análise (aeroporto ou linha aérea) e de estudo (região analisada, neste caso o país) fossem conhecidos, cenários C e D, respectivamente.

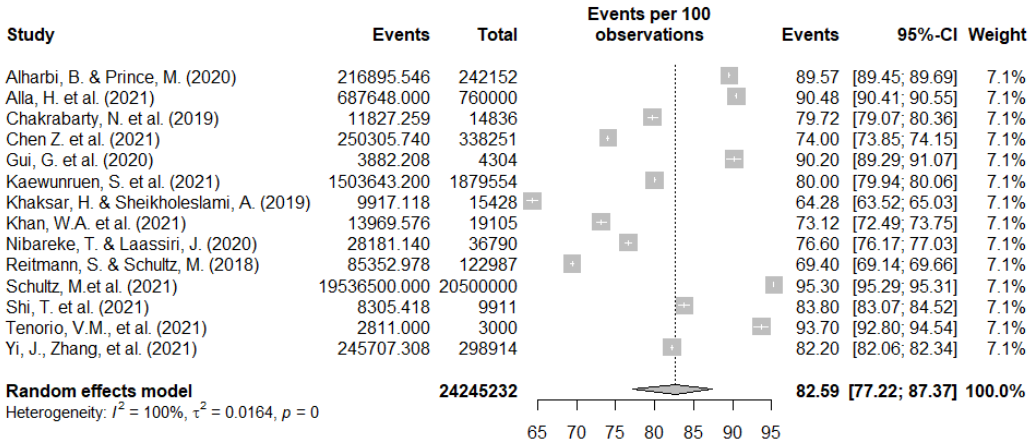


Figura 3: Cenário A: cenário de regressões por predição classificatória binária (0 = voo on-time, 1 = voo atrasado).

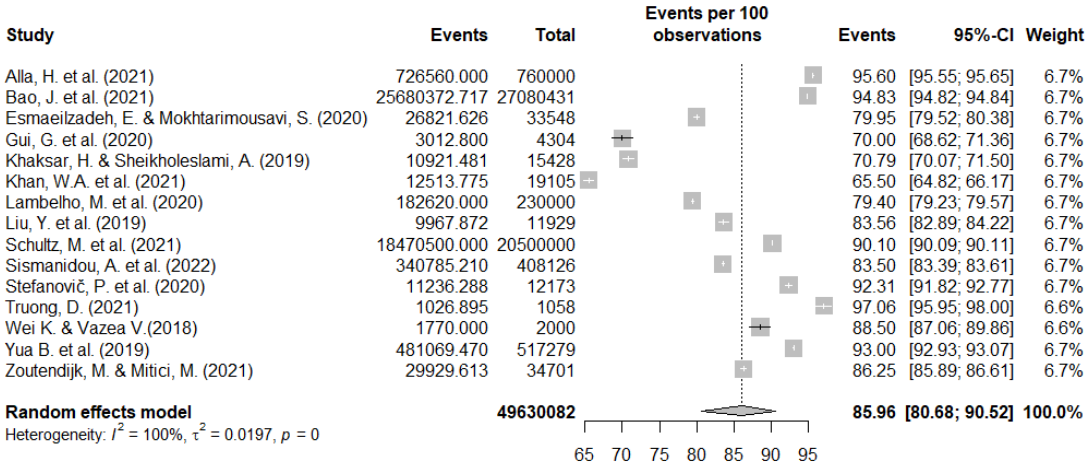


Figura 4: Cenário B: cenário de regressão dos atrasos por classes definidas.

O cenário C utiliza a acurácia dos modelos de previsão por classificação binária para caracterizar os estudos, assim como o cenário A. Os efeitos das áreas de análise, Tabela 1, demonstram que áreas mais amplas, como prever atrasos para todo um aeroporto, possuem menor capacidade assertiva.

Tabela 1 - Cenário C acurácia dos modelos de previsão por classificação binária para caracterizar os estudos.

Área de análise	*Nº de artigos*	*Acurácia média*	*Intervalo de variação positiva*	*Intervalo de variação negativa*
Linhas Aéreas	7	85,47%	6,12%	7,44%
Aeroportos	7	79,50%	6,77%	7,72%

Da mesma forma, foram verificadas as proporções tendo como subgrupos os países estudados pelos artigos do cenário C, Figura 5. Os resultados apresentam indicativos de que, as análises mais específicas, como a de Alharbi & Prince (2020) que utilizaram dados de companhias aéreas específicas da Arábia Saudita e uma maior quantidade de variáveis conhecidas, obtiveram uma maior eficácia na previsão dos atrasos.

Com artigos de previsão de atrasos por regressão, foram realizadas metanálises de proporção por subgrupos com a área de análise e estudo para o cenário D, Tabela 2.

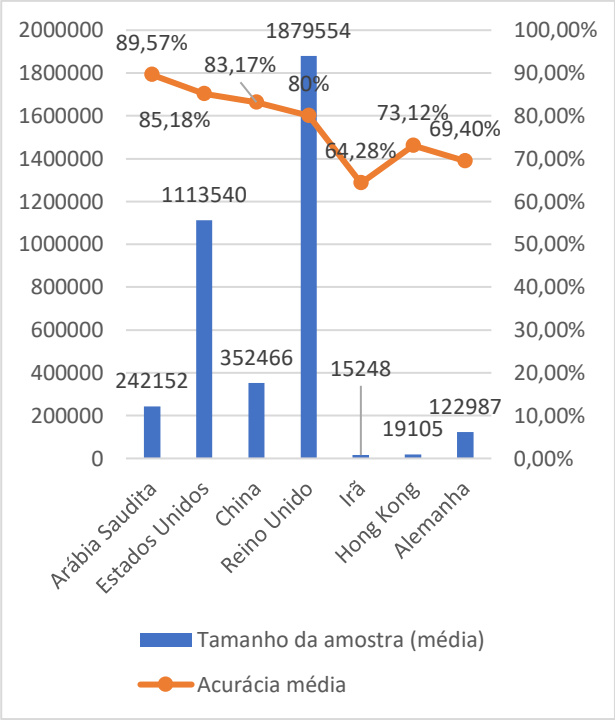


Figura 5: Proporções de acurácia e tamanho da amostra tendo como subgrupos os países estudados pelos artigos do cenário C.

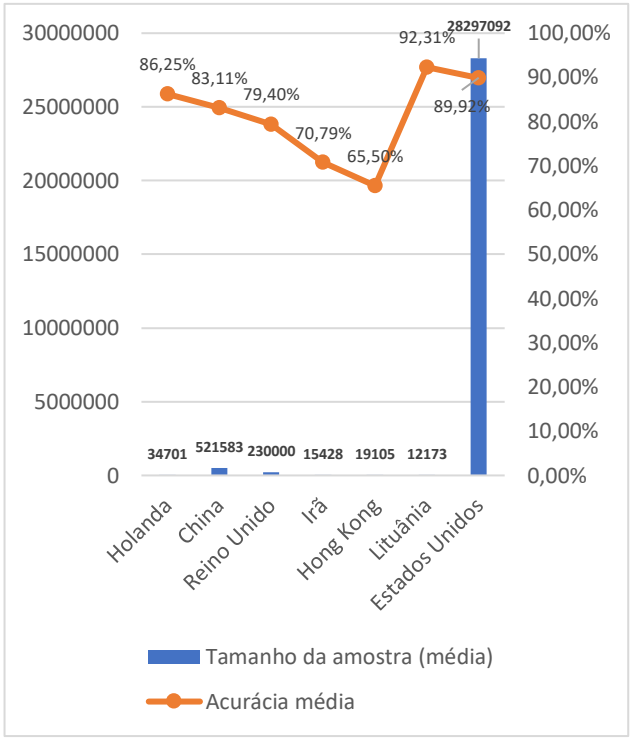


Figura 6: Proporções de acurácia e tamanho da amostra tendo como subgrupos os países estudados pelos artigos do cenário D.

O estudo em linhas aéreas apresentou uma maior margem de erro para a previsão em classes de atrasos. Por outro lado, as análises de aeroportos obtiveram uma maior acurácia média e um erro variante tolerável. Uma hipótese para este comportamento é a de que: quanto maior a especificidade das informações do voo para a obtenção de uma previsão em classes de atrasos, melhor a capacidade de generalização da resposta pelo modelo.

Tabela 2 - Cenário D análise de subgrupo com a área de estudo.

Área de análise	*Nº de artigos*	*Acurácia média*	*Intervalo de variação positiva*	*Intervalo de variação negativa*
Linhas Aéreas	3	80,77%	14,64%	21,47%
Aeroportos	12	85,96%	4,32%	5,05%

Assim como observado para o cenário C na Figura 6, para a análise de subgrupo com a área de estudo, o cenário D apresenta um comportamento em que, as especificidades do país analisado são mais importantes do que a quantidade de dados.

Um indicativo observado em ambas as situações é o de que: em países com movimentação aérea inferior a americana, como o caso da Holanda e da Arábia Saudita, o nível de generalização das respostas dos modelos de previsão aplicados é mais eficaz em relação ao tamanho amostral analisado.

Logo, caracterizar o cenário quanto a área estudada está mais relacionada a disponibilização e capacidade dos dados obtidos em descrever o fenômeno, no caso o atraso em um voo específico, do que ao tamanho da amostra ou da região de investigação.

3.4. Seleção do método mais eficaz

Não apenas especificar a área, mas o método utilizado é capaz de definir melhores resultados de acurácia nas previsões em ambas as possibilidades de análise, regressão ou classificação.

O método da Análise Envoltória de Dados (*Data Envelopment Analysis* – DEA, em inglês) foi aplicado, tendo como finalidade definir, entre os artigos levantados pelos modelos C e D, qual o método capaz de maximizar as eficiências obtidas a partir das amostras utilizadas, Tabela 3.

Tabela 3 Distribuição dos artigos pelo ano, tipo de análise e área de estudo. Os métodos mais acurados de *machine learning* aplicados nos estudos foram: [a] Árvore de decisão, [b] *deep belief network*, [c] *Ensemble learning*, [d] Floresta Randômica, [e] *Gradient Boosting Trees*, f Máquina de vetores de suporte, [g] Memória de curto prazo longa, [h] Rede Neural Artificial Profunda, [i] Rede Neural de Convolucional Gráfica, [j] Rede *Neural MultiLayer Perceptron* e [k] Redes Neurais Recorrentes.

Anos	*Previsão Climática*		*Previsão Universal*	
	Linha Aérea	*Aeroporto*	*Linha Aérea*	*Aeroporto*
2018		[g]Reitmann & Schultz		[e]Wei & Vazea
2019	[e]Etani		[e]Chakrabarty *et al.* [a]Khaksar & Sheikholeslami [h]Saadat & Moniruzzaman. [d]Liu et al.	[b]Yua *et al.*
2020			[j]Alharbi & Prince [g]Gui *et al.*	[f]Esmaeilzadeh & Mokhtarimousavi [a]Lambelho *et al.* [a]Nibareke & Laassiri [e]Stefanovič *et al.*
2021		[c]Chen *et al.* [d]Kaewunruen *et al.* [k]Schultz *et al.*	[j]Alla *et al.* [j]Shi *et al.* [e]Tenorio *et al.*	[i]Bao *et al.* [j]Cheevachaipimol *et al.* [e]Khan *et al.* [c]Yi *et al.* [d]Zoutendijk & Mitici
2022				[j]Sismanidou *et al.*

As variáveis selecionadas para o DEA definem a acurácia obtida de acordo com as condições adequadas para os cenários de estudo verificadas anteriormente, Tabela 4. Deste modo, são saídas do sistema tanto a acurácia obtida pelo modelo aplicado, quanto o fator de impacto da revista de publicação do artigo, considerando assim a confiabilidade da pesquisa realizada.

As entradas do modelo desenvolvido, os *inputs*, caracterizam cada técnica de *machine learning* aplicada pelo respectivo autor em seu estudo, garantindo a individualidade das Unidades Tomadoras de Decisão (*Decision Making Unit* – DMU, em inglês) exigidas pelo *software* empregado na análise, *MaxDEA*.

Tabela 4 Variáveis selecionadas para o DEA.

Variável aplicada	*Tipo*
Autores	DMU
Acurácia do modelo aplicado	Output
Fator de Impacto da revista	Output
Ano de publicação	Input
Tamanho da amostra	Input
Número de unidades previstas corretamente	Input

Assim, o tamanho da amostra e o número de voos previstos corretamente pelo sistema definem as DMU's pelas proporções analisadas anteriormente nas metaanálises desenvolvidas. O ano da publicação definem o

peso da originalidade e fator inédito da pesquisa.

Desta forma, foi utilizado o modelo com retornos variáveis de escala – BCC, orientado maximização das saídas do sistema, dadas as entradas definidas para cada DMU. Contudo, os resultados iniciais de eficiência foram muito semelhantes para o cenário C, fazendo necessária a aplicação de um critério de desempate entre os modelos.

O modelo de desempate utilizado foi o de Savage (1950), aplicado com coeficientes de otimismo (α) de 50% para ambas as fronteiras, clássica e invertida, Tabela 5. A proximidade entre os resultados, apesar dos critérios de desempate aplicados, demonstra que a eficiência dos métodos para previsão por classificação é relativamente próxima.

Tabela 5 Modelos de desempate das eficiências para o Cenário C.

DMU	*Eficiência*		
	Clássica	*Invertida*	*Relativa*
Chakrabarty, N. et al. (2019)	1	0,9995	0,5002
Gui, G. et al. (2020)	1	0,9995	0,5002
Khan, W.A. et al. (2021)	1	1	0,5
Nibareke, T. & Laassiri, J. (2020)	1	1	0,5
Schultz, M.et al. (2021)	1	1	0,5
Tenorio, V.M., et al. (2021)	1	1	0,5
Khaksar, H. & Sheikholeslami, A. (2019)	1	1	0,5
Alharbi, B. & Prince, M. (2020)	0,9930	1	0,4965
Shi, T. et al. (2021)	0,8943	1	0,4471
Yi, J., Zhang, et al. (2021)	0,8833	1	0,4416
Kaewunruen, S. et al. (2021)	0,8574	1	0,4287
Chen Z. et al. (2021)	0,8174	1	0,4087

Entretanto, as respostas obtidas pelos métodos aplicados nos sete primeiros artigos: *gradient boosting trees*, memória de curto prazo longa, árvore de decisão e redes neurais recorrentes, respectivamente, são as que conseguiram atribuir uma melhor acurácia visto os cenários e dimensão dos dados aplicados.

Vale salientar ainda que o método de *gradient boosting trees* pode ser nomeado o de melhor performance, visto que foi aplicado em três dos sete artigos com melhor eficiência pelo DEA, Chakrabarty et al. (2019), Khan et al. (2021) e Tenorio et al. (2021).

Já para a definição do método de previsão por regressão, cenário D, o DEA BCC orientado a *output* foi utilizado definindo a fronteira clássica e a eficiência relativa pelo critério de Savage (1950). Foram novamente utilizados coeficientes de otimismo (α) igual a 50% para ambas as fronteiras, Tabela 6.

Tabela 6 Modelos de desempate das eficiências para o Cenário D.

DMU	*Eficiência*		
	Clássica	*Invertida*	*Relativa*
Wei K. & Vazea V.(2018)	1	0,9981	0,5009
Liu, Y. et al. (2019)	1	0,9985	0,5007
Yua B. et al. (2019)	1	0,9985	0,5007
Truong, D. (2021)	1	0,9995	0,5002
Gui, G. et al. (2020)	1	0,9998	0,5000
Schultz, M. et al. (2021)	1	0,999	0,5000
Bao, J. et al. (2021)	1	1	0,5
Khan, W.A. et al. (2021)	1	1	0,5
Stefanovič, P. et al. (2020)	0,9794	0,9991	0,4901
Sismanidou, A. et al. (2022)	0,8948	1	0,4474
Zoutendijk, M. & Mitici, M. (2021)	0,8886	1	0,4443
Lambelho, M. et al. (2020)	0,8575	0,999	0,4291

A distinção entre as DMU's é mais clara para a análise do cenário D. A complexidade dos dados e das previsões exigem dos modelos

aplicados uma maior capacidade de generalização.

Apesar dos oito primeiros artigos possuírem valores significativamente próximos de eficiência relativa, os métodos que se destacaram para os cenários aplicados foram: *deep belief network*, floresta randômica, *gradient boosting trees* e as redes neurais recorrentes e convulacionais.

4. CONCLUSÕES

A partir das discussões apresentadas, é clara a tendência em aplicar métodos mais robustos de análise e regressão de dados para a previsão de atrasos na aviação mundial.

Diante da infinidade de métodos classificados como *machine learning*, alguns se destacam frente a complexidade e disponibilidade das informações necessárias a realização destas previsões.

Para caracterizar um atraso, a dimensão buscada dos dados deve ser mais próxima as especificidades das etapas a serem realizadas pelos voos do que da quantidade de dados em si.

Portanto, cabe a acurácia dos métodos estar atrelada a definição específica da área a ser analisada e dos principais fatores que geram essas ocorrências.

CONCLUSÕES

Alharbi, B., & Prince, M. (2020). *A Hybrid Artificial Intelligence Approach to Predict Flight Delay*. International Journal of Engineering Research and Technology, 814-822.

Bao, J., Yang, Z., Zeng, W.,2019.*Graph to sequence learning with attention mechanism for network-wide multi-step-ahead flight delay prediction*.Transportation Research Part C: Emerging Technologies

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010*). A basic introduction to fixed-effect and random-effects models for meta-analysis*. Res Synth Methods, 97-111.

Chakrabarty, N., Kundu, T., Dandapat, S., Sarkar, A., Kole, D.K.,2019.*Flight arrival delay prediction using gradient boosting classifier*.Advances in Intelligent Systems and Computing

Cheevachaipimol, W., Teinwan, B., Chutima, P.,2021.*Flight delay prediction using a hybrid deep learning method*.Engineering Journal

Chung, S.-H., Ma, H.-L., Hansen, M., & Choi, T.-M. (2020). *Data science and analytics in aviation.* Transportation Research Part E, 134.

Esmaeilzadeh, E., Mokhtarimousavi, S.,2021.*Machine Learning Approach for Flight Departure Delay Prediction and Analysis*.Transportation Research Record

Gui, G., Liu, F., Sun, J., (...), Zhou, Z., Zhao, D.,2018.*Flight delay prediction based on aviation big data and machine learning*.IEEE Transactions on Vehicular Technology

Gurkaran, B., & Hansen, M. (2013). *Generating day-of-operation probabilistic capacity scenarios from weather forecasts*. Transportation Research Part C: Emerging Technologies, 153-166.

Kaewunruen, S., Sresakoolchai, J., Xiang, Y.,2021.*Identification of weather influences on flight punctuality using machine learning approach*.Climate

Khaksar, H., Sheikholeslami, A.,Ano.*Airline delay prediction by machine learning algorithms*.Scientia Iranica

Khan, W.A., Ma, H.-L., Chung, S.-H., Wen, X.,2020.*Hierarchical integrated machine learning model for predicting flight departure delays and duration in series.*Transportation Research Part C: Emerging Technologies

Kim, Y., Choi, S., Briceno , S., & Mavris, D. (2016). *A deep learning approach to flight delay prediction*. 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), (pp. 1-6).

Lambelho, M., Mitici, M., Pickup, S., Marsden, A.,2019.*Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions*.Journal of Air Transport Management

Liu, Y., Hansen, M., Ball, M. O., & Lovell, D. J. (2021). *Causal analysis of flight en route inefficiency*. Transportation Research Part B: Methodological, 91-115.

Liu, Y., Liu, Y., Hansen, M., Pozdnukhov, A., Zhang, D.,2021.*Using machine learning to analyze air traffic management actions: Ground delay program case study*.Transportation Research Part E: Logistics and Transportation Review

Liu, Y., Yin, M., & Hansen, M. (2019). *Economic costs of air cargo flight delays related to late package deliveries*. Transportation Research Part E, 125, 388-401.

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A *comparison of methods to detect publication bias in meta-analysi*s. Statistics in Medicine, 641-654.

Nibareke, T., Laassiri, J.,2020.*Using Big Data-machine learning models for diabetes prediction and flight delays analytics*.Journal of Big Data

Reitmann, S., Schultz, M.,2020.*Computation of air traffic flow management performance with long short-term memories considering weather impact*.Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

Saadat, M.N., Moniruzzaman, M.,2020.*Enhancing airlines delay prediction by implementing classification based deep learning algorithms*.Advances in Intelligent Systems and Computing

Savage, L. (1950). *The Foundations of Statistics*. Willey.

Schultz, M., Reitmann, S., & Alam, S. (2021). *Predictive classification and understanding of weather impact on airport performance through machine learning*. Transportation Research Part C: Emerging Technologies.

Shi, T., Lai, J., Gu, R., Wei, Z.,2021.*An Improved Artificial Neural Network Model for Flights Delay Prediction*.International Journal of Pattern Recognition and Artificial Intelligence

Stefanovič, P., Štrimaitis, R., Kurasova, O.,2019.*Prediction of Flight Time Deviation for Lithuanian Airports Using Supervised Machine Learning Mode*l.Computational Intelligence and Neuroscience

Tenorio, V.M., Marques, A.G., Cadarso, L.,2021.*Signal Processing and Machine Learning for Air Traffic Delay Prediction*.Transportation Research Procedia

Wang, C., & Wang, X. (2019). *Airport congestion delays and airline networks.* Transportation Research Part E, 122, 328–349.

Wei K., Vazea V.,2021.*Modeling Crew Itineraries and Delays in the National Air Transportation System*.Transportation Science

Yi, J., Zhang, H., Liu, H., Zhong, G., Li, G.,2021.Flight Delay *Classification Prediction Based on Stacking Algorithm*.Journal of Advanced Transportation

Yua B., Guoa Z., Asian S., Wangc H., Chend G.,2020.Flight *delay prediction for commercial air transport: A deep learning approach*.Transportation Research Part E

Zoutendijk, M., Mitici, M.,2021.*Probabilistic flight delay predictions using machine learning and applications to the flight-to-gate assignment problem*.Aerospace