



UNIVERSIDAD DE GRANADA

Programa de Doctorado en Física y Matemáticas

Tesis Doctoral

**UTILIZACIÓN DE ALGORITMOS DE MACHINE LEARNING /
DEEP LEARNING (ML/DL) PARA LA PREDICCIÓN DE
PARÁMETROS DE GESTIÓN DEL TRANSPORTE AÉREO**

Autor: Juan Gerardo Muros Anguita

Director: Oscar Díaz Olariaga

Granada, 15 de enero 2024

Editor: Universidad de Granada. Tesis Doctorales
Autor: Juan Gerardo Muros Anguita
ISBN: 978-84-1195-350-4
URI: <https://hdl.handle.net/10481/92738>

El doctorando Juan Gerardo Muros Anguita y el director de la tesis: Oscar Díaz Olariaga

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor and, as far as our knowledge reaches, the rights of the authors to be cited (when their results or publications have been used) have been respected in the performance of this work.

Lugar y fecha: Granada a 15 de enero 2024

El candidato a Doctor:

El Director Tesis:

Firmado Juan Gerardo Muros Anguita

Firmado Oscar Díaz Olariaga

Agradecimientos

Esta tesis hunde sus raíces más profundas en el amor y apoyo sin límites que siempre obtuve de mis padres José y Aurora. Quiero agradecer a mi esposa Laurence y a mi hijo Guillermo su paciencia.

Especial mención requiere mi director de tesis el Dr. Oscar Díaz sin cuya dirección y colaboración esta tesis no habría podido ser escrita. Igualmente deseo agradecer el continuo entusiasmo y soporte por parte de mi tutor y amigo el Dr. Eduardo Battaner.

A Antonio Navarro le agradezco su apoyo para la edición del documento.

- Sin curiosidad, no hay movimiento -

0. Resumen

Predecir los parámetros de gestión del transporte aéreo en los próximos años es crítico para los principales agentes del transporte aéreo como son: las líneas aéreas, autoridades aeroportuarias, proveedores del servicio de navegación aérea, además de la propia industria de fabricación de aviones, etc. Dado que la industria aeronáutica genera casi 87.7 millones de empleos (directos e inducidos) y 3.4 billones de US\$ lo que representa un 4.1% del PIB mundial. El uso de los algoritmos de Machine Learning /Deep Learning (ML/DL), donde recientemente ha irrumpido el ChatGPT para extender el uso masivo de la Inteligencia Artificial (IA), vienen desplazando a otros métodos estadísticos clásicos como ARIMA, etc. en el pronóstico de las series temporales.

Esta tesis muestra una doble viabilidad, la de desarrollar y entrenar modelos de ML/DL con reducidos medios propios, y por otro lado demostrar su aplicabilidad a la gestión del transporte aéreo, lo cual ha sido posible gracias sobre todo a la existencia de librerías como Keras, Tensorflow y lenguajes de programación como Python, de código abierto, orientados al aprendizaje profundo (DL). En particular, se abordan tres casos de estudios: el pronóstico de la demanda a corto y medio plazo de pasajeros (nacionales e internacionales), el pronóstico de la carga aérea (doméstica e internacional) y el de la predicción de la demora de salida de aviones, todos ellos aplicados usando los datos del caso de estudio de Colombia.

Los modelos singulares de ML/DL desarrollados para los dos primeros casos son de aprendizaje profundo con arquitecturas de codificadores-decodificadores realizados con redes neuronales artificiales híbridas que mezclan redes convolucionales con redes recurrentes tipo LSTM, denominadas “*ConvLSTM2D*” las cuales mejoran el aprendizaje de los patrones temporales de las series históricas multivariable, donde se mezclan las variables aeronáuticas objeto de estudio con otras variables socioeconómicas de contexto. Mientras que, para el tercer caso, relativo a la predicción de las demoras en las salidas de aviones, aplicada con los datos de los aeropuertos de Colombia, la investigación se ha orientado en realizar la selección del mejor del modelo, a través de la comparativa

en las prestaciones de varios modelos clásicos de ML y un modelo simple de DL (MLP), tomando como métrica comparativa el error cuadrático medio (RMSE).

Un esfuerzo importante de esta investigación se ha centrado en proporcionar robustez a las valoraciones obtenidas por los modelos, aplicando las técnicas de “*k-fold cross validation*”, dada la naturaleza estocástica del problema (modelo y datos). Mientras que otra parte importante de esta investigación se ha orientado a explicar los resultados obtenidos por los modelos ML/DL, ya que estos algoritmos se comportan como cajas negras al no explicitar los patrones aprendidos, lo que se ha resuelto mediante el uso de técnicas de “*permutation importance*”, así como aplicando análisis variacional al configurar múltiples escenarios obtenidos al combinar diferentes features, para seleccionar los más relevantes del problema.

El dataset usado para las demostraciones procede de la Autoridad Aeronáutica colombiana Aerocivil, el cual ha sido preprocesado con técnicas estadísticas de escalado, normalizado, categorizado, etc. para homogeneizar los datos que faciliten el aprendizaje. Es de subrayar que en el dataset usado para el pronóstico de las diferentes demandas no ha sido necesario realizar el proceso de “extracción de los *features*” o características de las variables de entrada, como por el contrario si se requiere en física mediante p.e. el uso de números adimensionales para categorizar un régimen, etc., ya que los modelos DL, como es el *ConvLSTM2D* no requieren de dicho preprocesado por parte del ingeniero (tarea que se conoce “*feature engineering*”), ya que esta extracción de features la realizan de forma automática los modelos de aprendizaje profundo.

Esta tesis ha sido posible al haber podido acceder a un gran y fiable volumen de datos aeronáuticos y socioeconómicos para el entrenamiento del modelo. Donde Colombia ha sido el caso de estudio elegido, porque representa el tercer país latinoamericano y el quinto de América en la industria aeronáutica. Por último, hay que señalar que se han entrenado los modelos ML/DL bajo la modalidad de aprendizaje supervisado (SL), por lo que el dataset contenía no solo las variables de entrada sino también las variables de salida (objetivas) a predecir.

Palabras Clave: aeropuertos, aprendizaje profundo, previsión demanda pasajeros aéreos, previsión demanda carga aérea, predicción demoras salida aviones, redes neuronales convolucionales, redes neuronales recurrentes.

0.1 Abstract

Predicting the management parameters of air transport in the coming years is critical for the main air transport agents such as: airlines, airport authorities, air navigation service providers, as well as for the aircraft manufacturing industry itself, etc. given that the aeronautical industry generates almost 87.7 million jobs (direct and induced) and 3.4 trillion US\$, which represents 4.1% of global GDP. The use of Machine Learning/Deep Learning (ML/DL) algorithms, where ChatGPT has recently emerged to extend the massive use of Artificial Intelligence (AI), are displacing other classic statistical methods such as ARIMA, etc. in time series forecasting. This thesis shows a double viability, that of developing and training ML/DL models with its own means, and on the other hand demonstrating its applicability to air transport management, thanks above all to the existence of libraries such as Keras and programming languages such as Python, open source, both aimed at deep learning (DL).

In particular, three case studies are addressed: the forecast of short and medium-term demand for passengers (national and international), that of air cargo (domestic and international) and that of the prediction of aircraft departure delays, all demonstrated using data from the Colombia application case. The unique ML/DL models developed for the first two cases are deep learning with encoder-decoder architectures made with hybrid artificial neural networks that mix convolutional networks with recurrent LSTM-type networks, called “*ConvLSTM2D*” which improve the learning of the temporal patterns of multivariable historical series, where the aeronautical variables under study are mixed with socioeconomic variables. While, for the third case, related to the prediction of delays in aircraft departures, demonstrated with data from Colombian airports, the research has been aimed at selecting the best model, through comparison in the performances of several classical ML models and a simple DL model (MLP), based on the measurement of the root mean square error (RMSE) metric. An important effort in this research has

been focused on providing robustness to the evaluations obtained by the models, applying “*k-fold cross validation*” techniques, given the stochastic nature of the problem (model and data). While the other important part of this research has been oriented on interpreting the results obtained by the ML/DL models, since these algorithms behave like black boxes by not making the learned patterns explicit, through the use of “*permutation importance*” techniques. As well as variational analysis when configuring scenarios obtained by combining different features, to select the most relevant to the problem.

The dataset used for the demonstrations comes from the Colombian Aeronautical Authority Aerocivil, which has been preprocessed with statistical techniques for scaling, normalized, categorized, etc. to facilitate the homogenization of data that facilitates learning. It is worth highlighting that in the dataset used for the demand forecast it has not been necessary to carry out the process called “feature extraction” or characteristics of the input variables such as it is necessary to implement e.g. in dimensionless numbers analysis in physics to categorize a regime, etc., since DL models, such as *ConvLSTM2D*, do not require such preprocessing by the engineer (a task known as “*feature engineering*”), since this feature extraction is carried out implicitly by deep learning models. In short, this thesis has also been possible by having been able to access a large and reliable volume of aeronautical and socioeconomic data for model training. Colombia has been the chosen case, because it represents the third Latin American country and the fifth in America in the aeronautical industry. Finally, it should be noted that the ML/DL models have been trained under the supervised learning (SL) modality, so the dataset contained not only the input variables but also the objective or output variables to be predicted.

Keywords: air cargo demand forecasting, air passenger demand forecasting, airport, convolutional neural networks, deep learning, flight delay prediction, long-short term memory.

ÍNDICE

1.	INTRODUCCIÓN	1
2.	JUSTIFICACIÓN.....	15
3.	OBJETIVOS.....	16
3.1	GENERALES	16
3.2	ESPECÍFICOS	17
4.	REVISIÓN BIBLIOGRÁFICA.....	20
4.1	DEMANDA DE LA CARGA AÉREA	20
4.2	DEMANDA DE PASAJEROS AÉREOS	23
4.3	DEMORA DE SALIDA AVIONES COMERCIALES	25
5.	CASO APLICACIÓN DE COLOMBIA.....	30
6.	METODOLOGÍA Y DATOS	35
6.1	DOS PRIMEROS CASOS DE ESTUDIO DE CARGA AÉREA Y PASAJEROS.....	35
6.1.1	<i>Dataset.....</i>	35
6.1.2	<i>Preparación Dataset.....</i>	36
6.1.3	<i>Arquitectura Modelo “ConvLSTM2D”</i>	39
6.1.4	<i>Procesos y Entrenamiento Modelo</i>	49
6.1.5	<i>Hiperparametrización y Métricas Evaluación Modelo</i>	53
6.2	TERCER CASO DE ESTUDIO DEMORA SALIDAS AVIONES COMERCIALES	56
6.2.1	<i>Dataset.....</i>	56
6.2.2	<i>Tipología de Modelos ML/DL</i>	56
6.2.3	<i>Entrenamiento Modelos</i>	60
6.2.4	<i>Tests Estadístico de Features.....</i>	62
6.2.5	<i>Eliminación Outliers Dataset</i>	63
6.2.6	<i>Determinación Importancia Features</i>	63
7.	RESULTADOS.....	65
7.1	PRONÓSTICO DEMANDA DE LA CARGA AÉREA.....	65
7.1.1	<i>Pronóstico Carga Aérea internacional.....</i>	65
7.1.2	<i>Pronóstico Carga Aérea Doméstica.....</i>	70
7.2	PRONÓSTICO DEMANDA DE PASAJEROS	75
7.3	TERCER CASO DE ESTUDIO CORRESPONDIENTE A LA PREDICCIÓN DEL RETRASO DE LA SALIDA DE AVIONES.	85
8.	CONCLUSIONES	97
8.1	DEMANDA CARGA AÉREA	97

8.2	DEMANDA PASAJEROS AÉREOS	98
8.3	DEMORA SALIDA AVIONES	100
9.	BIBLIOGRAFÍA	104
10.	LISTADO DE PUBLICACIONES	128
10.1	ANEXO I: AIR CARGO TRANSPORT DEMAND FORECASTING USING CONVLSTM2D, AND ARTIFICIAL NEURAL NETWORK ARCHITECTURE APPROACH	129
10.2	ANEXO II: AIR PASSENGER DEMAND FORECAST THROUGH THE USE OF ARTIFICIAL NEURAL NETWORK ALGORITHMS	171
10.3	ANEXO III: PREDICTION OF DEPARTURE FLIGHT DELAYS THROUGH THE USE OF PREDICTIVE TOOLS BASED ON MACHINE LEARNING/DEEP LEARNING ALGORITHMS.	205

1. Introducción

Esta tesis sobre las predicciones de la evolución futura de algunos parámetros de la gestión del transporte aéreo (como son la demanda de pasajeros y de la carga aérea), así como de los retrasos de las salidas de aviones, se ha centrado en el impacto de los mismos en los sistemas aeroportuarios y no tanto en las líneas aéreas o de la gestión del tráfico aéreo (ATM). Los planes para el desarrollo de los distintos componentes del sistema aeroportuario dependen en gran medida de los niveles de actividad que se prevén para el futuro. Para planificar las instalaciones e infraestructuras de un aeropuerto, red o conjunto de aeropuertos, y para poder satisfacer las necesidades futuras, es esencial predecir el nivel y la distribución de la demanda en los diversos componentes del sistema aeroportuario TRB ^[1] (2002). Pronosticar la demanda en una industria tan dinámica y sensible a factores exógenos, como es la aviación, es una tarea extremadamente compleja. No obstante, es necesario realizar estimaciones de tráfico aéreo como paso previo a la planificación y diseño de las instalaciones aeroportuarias, ya sea la correspondiente a un aeropuerto o a un sistema de red aeroportuaria según Horonjeff R. et al.^[2] (2010) y Wells A. & Young S.^[3] (2004).

La comprensión de los patrones de la demanda futura permite al planificador aeroportuario evaluar el desempeño futuro del aeropuerto y con ello recomendar programas de desarrollo consistentes con la estimación de los costes asociados a esos planes de desarrollo, así como proyectar los tipos y niveles de ingresos que respalden las inversiones y el retorno de la inversión (ROI) según Ashford N. et al.^[4] (2011). El pronóstico de la demanda es un requisito básico para desarrollar un “*plan director*”, el cual busca en primer lugar la eficiencia y rentabilidad, ya sea de un aeropuerto, o de una red aeroportuaria a nivel regional o nacional, de acuerdo con Janic M. ^[5] (2021), Janic M. ^[6] (2009) e ICAO ^[7] (1987).

Para evaluar la planificación de las actividades aeroportuarias, se hace necesario analizar y evaluar la demanda futura de las demandas de pasajeros y carga aérea que las originan. Existen numerosos factores que afectan la demanda, y los planificadores encargados de estimar la demanda, suelen utilizar para ello no

solo los propios datos históricos aeronáuticos (de tráfico aéreo), sino también los datos (de series históricas) socioeconómicas locales, que la conforman como son: la riqueza nacional (PIB), la paridad del poder adquisitivo (PPA) de los habitantes, la población, el índice de producción industrial (IPI), el índice de precios al consumo (IPC), la tasa de cambio (de la moneda local con respecto al dólar USA), exportaciones e importaciones, etc. Ya que estos índices tienen una gran influencia en el comportamiento de la demanda de tráfico aéreo según Rodríguez Y. et al. ^[8] (2020), García Cruzado M. ^[9] (2013), Horonjeff R. et al. ^[2] (2010) e ICAO ^[10] (2006).

La experiencia práctica demuestra que los pronósticos de tráfico aéreo no suelen ser precisos cuando las previsiones se realizan a largo (15-20 años) o muy largo plazo (25-30 años), pero en cambio las previsiones a corto y medio plazo (entre 5 y 15 años) son necesariamente críticas para el planificador y tomadores de decisiones, ya que comprenden un periodo habitual de planificación aeroportuaria (sistema o red aeroportuaria) de acuerdo con ACI ^[11] (2016), Kazda A. & Caves R. ^[12] (2015), de Neufville R. & Odoni A. ^[13] (2013).

La demanda de transporte de carga aérea ha venido experimentando un ininterrumpido crecimiento en los últimos 35 años. Más precisamente, en las últimas dos décadas (hasta el año 2019) a nivel mundial, el promedio anual de la tasa de crecimiento del indicador “Freight Tonne-Kilometer” (FTK) fue del 4.1%, mientras que el del “Mail Tonne-Kilometer” (MTK) fue del 4.05 %. En el año 2019 el FTK alcanzó la cifra de $225 \cdot 10^9$, mientras que el “Revenue Tonne-Kilometer” (RTK) fue de $1043 \cdot 10^9$ de acuerdo a ICAO ^[14] (2020), IATA 2021a ^[15] e IATA 2021b ^[16]. La aparición de la pandemia del Covid-19 produjo una caída del FTK en el año 2020 del 18.5% a nivel mundial con respecto a 2019 según el World Bank ^[17] (2022), IATA 2021a ^[15] e IATA 2021b ^[16]. Asimismo, varios estudios pronostican que la demanda de carga aérea (a nivel global), se recuperará, a índices de crecimiento pre-pandémicos a muy corto plazo, incluso antes que la demanda de pasajeros aéreos de acuerdo con Gudmundsson S. et al. ^[18] (2021), JADC ^[19] (2021) y Boeing ^[20] (2020).

En lo concerniente a las metodologías utilizadas para la realización de pronósticos, los académicos vienen presentando estudios e investigaciones desde

hace tiempo. En este tiempo se ha desarrollado una diversidad de modelos para pronosticar la demanda de carga/pasajeros, los cuales se pueden pre-clasificar en dos grandes grupos: métodos económicos y modelos de series temporales según Dantas T. et al. ^[21] (2017). Los métodos económicos se centran en la correlación entre la demanda (de pasajeros y/o carga aérea) y múltiples variables, como la influencia de los cambios en el entorno económico y el sistema de tráfico aéreo, para posteriormente establecer mediante una serie de ecuaciones los modelos de pronóstico. Los modelos comúnmente utilizados incluyen análisis de regresión según Abed S.Y. et al. ^[22] (2001), prueba de causalidad según Fernandes E. & Pacheco R. ^[23] (2010), modelo “logit” según Garrow L.A. & Koppelman F.S. ^[24] (2004), y modelo de gravedad según Grosche T. et al. ^[25] (2007). Los métodos de series temporales se basan principalmente en datos históricos para predecir mediante la extracción de la relación intrínseca entre los datos actuales y las series de observaciones pasadas. Los diversos modelos, a veces llamados también estimadores o filtros de series temporales que se han venido utilizando para pronosticar la demanda de carga/pasajeros han sido fundamentalmente: las técnicas de suavizado según Samagaio A. & Wolters M. ^[26] (2010), el modelo de Markov adaptado por Chin A.T.H. & Tay J.H. ^[27] (2001), el modelo ARIMA (autoregresivo integrado de media móvil)/SARIMA (estacional autoregresivo integrado de media móvil) de Tsui W.H.K. et al. ^[28] (2014), el método de ajuste estacional usando filtros de Kalman y simulaciones de Monte Carlo (aplicables según que las distribuciones sean gaussianas o no, respectivamente) de acuerdo con Aston J.A.D. & Koopman S.J. ^[29] (2006).

Ahora bien, debido a la característica no lineal de la demanda de pasajeros/carga aérea, los enfoques económicos y de series temporales han sido criticados duramente debido a una capacidad limitada de pronóstico según Tsui W.H.K. et al. ^[28] (2014). Por ello, algunos investigadores implementan otras metodologías para el pronóstico, como es el modelo DLM (Dynamic Linear Models) de acuerdo con Rodríguez Y. et al. ^[8] (2020); el cual presenta las siguientes ventajas respecto a las metodología habituales de estimación de predicciones: detecta tendencias estocásticas que se encuentran ocultas en las series temporales según West M. & Harrison J. ^[30] (2006), e identifica cambios estructurales que

permiten estimar el efecto en el tiempo de factores exógenos, sin aumentar el número de parámetros de acuerdo a Honjo K. et al. ^[31] (2018). Adicionalmente la estructura de independencia condicional sobre la que se basa (la dinámica de estados) permite considerar predicciones considerando un algoritmo recursivo de acuerdo con Petris G. et al. ^[32] (2009). Otro planteamiento que se ha utilizado para el pronóstico de demanda de pasajeros/carga aérea es la dinámica de sistemas de Tascón D. & Díaz O. ^[33] (2021), la de Suryani E. et al. ^[34] (2012) ^[35] (2010). La dinámica de sistemas incorpora un conjunto de herramientas que permiten analizar los entornos complejos, utilizando herramientas como diagramas de causalidad, fomentando el pensamiento sistémico entre los gestores de procesos, lo que captura la complejidad dinámica de un sistema dado y proporciona la gran ventaja de simular el modelo para evaluar los resultados de las interacciones entre las variables, anticipando los posibles efectos (colaterales) de las políticas del transporte aéreo antes de su implementación.

Desde hace algunos años, se han comenzado a utilizar metodologías basadas en redes neuronales artificiales (ANN), gracias a las difusiones, éxito y disponibilidad de estas técnicas. Las características diferenciales de las ANN que las hacen más relevantes para el cálculo del pronóstico según Zhang G. et al. ^[36] (1998), Srisaeng P. et al. ^[37] (2015), Mostafaeipour A. et al. ^[38] (2018) son: (a) a diferencia de los métodos tradicionales las ANN y particularmente las técnicas de Deep Learning (DL) son métodos “auto-adaptativos” dirigidos por los datos, que no necesitan ser pre-procesados en el sentido de que no se necesita extraer previamente los “features” que conforman el conocimiento previo del problema y sus variables (lo que se conoce como “*feature engineering*”). Ya que el modelo aprende de los ejemplos de datos de entreno, capturando las relaciones funcionales (incluso las más sutiles), aunque dichas relaciones subyacentes sean desconocidas o difíciles de describir. Por tanto, las ANN son adecuadas para problemas cuyas soluciones requieren conocimientos que son difíciles de especificar, pero para los que existen datasets o un corpus de observaciones suficientes. En este sentido, pueden tratarse como uno de los métodos estadísticos multivariable no lineales y no paramétricos. (b) Las ANN pueden generalizar bastante bien, salvo que se presenten los problemas de sobreajustes, los cuales suelen aparecer al entrenar

con pocos y ruidosos datos, lo que lleva al modelo a aprenderse incluso el propio ruido y adoptar el sesgo de los datos, lo cual le hace perder su capacidad para generalizar según Millstein F. ^[39] (2018) y Sewak M. et al. ^[40] (2018). Es decir, después de conocer los datos que se les presentan, las ANN a menudo pueden inferir correctamente la parte invisible de los patrones presentes en la población, incluso si los datos de la muestra contienen información ruidosa. (c) Las ANN son “aproximadores funcionales universales”; habiéndose demostrado que pueden aproximar cualquier función, con la precisión deseada, lo que se conoce como “teorema del aproximador universal” de acuerdo con G. Cybenko ^[41] (1989), quien uso para su demostración la función *sigmoide* como función de activación de las neuronas. Esta característica le permite a las ANN adoptar funciones más generales, dada su flexibilidad de adaptación. (d) Las ANN no presuponen si los procesos son o no lineales, mientras que por el contrario las predicciones que se han venido realizando con métodos estadísticos estaban basadas en hipótesis de linealidad de la función pronóstico. Así, los enfoques tradicionales de la predicción de series temporales, como el método de Box-Jenkins o ARIMA, asumen que las series temporales en estudio se generan a partir de procesos lineales, lo que no es razonable suponer para la generación de las series temporales asociadas a las demandas en el transporte aéreo. Por ello, las redes neuronales artificiales ANN, son capaces de establecer patrones no lineales, sin que exista ninguna asunción previa sobre el tipo de relaciones (lineales o no) entre las variables de entrada y salida. En resumen, se puede afirmar que las ANN son una herramienta de modelado más genérica y flexible para el pronóstico del transporte aéreo según Gupta V. et al. ^[42] (2019), Dingari M. et al. ^[43] (2019), Mostafaeipour A. et al. ^[38] (2018), Srisaeng P. et al. ^[37] (2015) y Bao Y. et al. ^[44] (2012).

Esta tesis se plantea pues como una investigación que pretende dar continuidad y profundizar sobre estos precedentes académicos de pronósticos del transporte aéreo, pero centrándose ahora en una mayor disponibilidad y uso de herramientas de ML que incorporan modernas técnicas del aprendizaje profundo (DL), para resolver el pronóstico a corto y medio plazo de demanda de pasajeros/carga aérea. Para ello se ha usado como caso de estudio, los datos de transporte aéreo correspondientes a Colombia y su red completa de aeropuertos

nacionales e internacionales. Ya que actualmente Colombia representa el tercer mercado aéreo latinoamericano y el quinto de las Américas, por volumen de tráfico gestionado de acuerdo con ACI ^[45] (2021) e IATA 2021a ^[15]. Además se ha hecho especial énfasis en diferenciar el antes y después de la crisis mundial del Covid-19, para ver cómo estos modelos son capaces de capturar los patrones de esta crisis, al incluir los datos del primer año del Covid-19 del 2020. Este primer año del Covid-19 recoge ya un impacto severo de la pandemia, por lo que se espera que a partir de la aparición de algún punto de inflexión futuro, se pueda determinar el periodo de recuperación, tanto del volumen de la demanda como de la tendencia de crecimiento, durante el periodo post-pandémico (más allá del 2021). Para ello, y como predictor, se ha propuesto una arquitectura de ML/DL, de desarrollo propio, que usa redes neuronales artificiales del tipo “ConvLSTM2D” según Muros JG & Diaz O. ^[46] (2023) y Muros JG & Diaz ^[47] (2022). Donde el prefijo “Conv” nombra la característica Convolutacional y el sufijo “LSTM2D” denomina el atributo de la memoria a corto-largo plazo en 2 dimensiones de la redes neuronales recurrentes (RNN). Este tipo de arquitectura usa un híbrido o mezcla entre las redes convolucionales neuronales (CNN), muy útiles para la extracción de los patrones invariantes espaciales de las variables, y las redes neuronales recurrentes (RNN) que sirven para extraer los patrones temporales de las variables implicadas según Millstein F. ^[39] (2018), Sewak M. et al. ^[40] (2018), Yang J. et al. ^[48] (2015), Malhotra et al. ^[49] (2015) y Hermans M. & Schrauwen B. ^[50] (2013). Estas técnicas de predicción basadas en Machine Learning/Deep Learning (ML/DL), pueden incorporar más elementos de análisis y por tanto ser potencialmente más eficaces de acuerdo con Ketkar N. & Moolayil J. ^[51] (2021). Otras de las ventajas de las redes “ConvLSTM2D” con respecto a los métodos clásicos de tipo autoregresivo, es que admiten un tratamiento multivariable (con muchos más “features” de entrada) según Pedrycz W. & Chen S. ^[52] (2020) y Calin O. ^[53] (2020). La aplicación a los dos primeros casos de estudio de la presente investigación con este tipo de modelo de red neuronal DL, viene a representar así, una suerte de modelo de conocimiento abstracto, inferido a partir del aprendizaje de los patrones históricos de las series temporales de tráfico aéreo interrelacionadas con las socioeconómicas, a partir de las cuales se proyecta su evolución futura.

Por último, se aborda el tercer caso de estudio relativo a la predicción de los retrasos en las salidas de los vuelos comerciales según Muros JG & Díaz O. ^[54] (2023) y Muros JG & Díaz O, lo cual tiene un gran impacto en la planificación y el desarrollo de las infraestructuras aeroportuarias. ^[55] (2023). Las líneas aéreas, de acuerdo al gran crecimiento de la demanda del transporte aéreo, necesitan cada vez mayor disponibilidad de infraestructuras aeroportuarias, para ofertar nuevas rutas y frecuencias. Efectivamente, la industria aeroportuaria mundial ha venido respondiendo a esta demanda, con una permanente ampliación de sus infraestructuras y con la construcción de nuevos aeropuertos, a través de un ingente volumen de inversión, tanto público como privado de acuerdo con Graham A. & Morrell P.S ^[56] (2017). Pero la adecuación de la oferta a la demanda no ha corrido siempre a la par, por lo que en muchos aeropuertos del mundo la capacidad ofertada ha estado por debajo de la demandada, lo que ha generado la congestión del aeropuerto, y como consecuencia de ello ha aparecido un incremento de las demoras de salidas de aviones comerciales de acuerdo a Neufville R. and Odoni A. ^[13] (2013), Horonjeff R. et al. ^[2] (2010), Ashford N. et al. ^[4] (2011). Muchos estudios, tanto técnicos como académicos, han concluido que la demora es un concepto que está íntimamente vinculado a la capacidad del aeropuerto según ACRP ^[57] (2014), Gelhausen M. et al. ^[58] (2020) y Janic M. ^[6] (2009). Ahora bien, ¿qué se entiende exactamente por demora y cómo se mide? Antes de dar respuesta a dicho interrogante, conviene mencionar que el sistema de transporte aéreo tiene varios actores, siendo los tres más relevantes desde el punto de vista operacional: los aeropuertos, el sistema de gestión y control del tráfico aéreo (ATM), y las líneas aéreas. Para mayor complejidad del problema, las relaciones entre estos tres agentes son difíciles de coordinar, pues cada uno de ellos dispone de su propia regulación, políticas y estrategias operativas según Kearns S. ^[59] (2018), Budd L. & Ison S. ^[60] (2017). Por eso, no se puede imputar el problema de la demora de salidas de vuelos comerciales solo y exclusivamente a la operación aeroportuaria (o a la gestión de su capacidad). Es más, tampoco se puede imputar el problema de la demora solo a la falta de operatividad y sincronización de estos tres actores, ya que existen otros factores externos que influyen y mucho en la aparición o agravamiento de las demoras, como es el de la adversidad meteorológica según Borsky S. &

Unterberger C. ^[61] (2019), Choi et al. ^[62] (2016), y Fernandes N. et al. ^[63] (2020). Por otro lado, cada país del mundo tiene su propia política regulatoria (de control y seguimiento de la actividad) a partir de la autoridad de su sistema de aviación civil, así como políticas propias de gestión de sus infraestructuras de transporte aéreo. Todo esto hace aún más difícil, que exista un estándar o norma universal única que defina y regule el concepto de demora y que sea aplicable por igual en todo el mundo de acuerdo con ACI ^[64] (2012). Asimismo, cada actor del sistema de transporte aéreo (y en cada país) gestiona el problema de las demoras según sus propios recursos y estrategias operativas. Lo que ha conducido a una búsqueda de consenso a nivel mundial siguiendo ciertos ejemplos o iniciativas, para adoptar un criterio común sobre cómo gestionar la demora, y que dicho criterio pueda ser extensible a todos los aeropuertos, tanto a nivel nacional como internacional. Además, los entes reguladores nacionales de la aviación comercial (u organismos de control y seguimiento) vigilan cada vez más la calidad del servicio aéreo, y es justamente donde el concepto de puntualidad entra en escena, que al fin y al cabo, se podría interpretar como un indicador del desempeño o prestaciones de acuerdo a Carvalho L. et al. ^[65] (2021). Por lo que las autoridades aeroportuarias en muchos países, están adoptando las recomendaciones o sugerencias de organizaciones supranacionales que son referencia de la industria aeronáutica, implementando el criterio de que un vuelo comercial (programado) es puntual si el mismo se inicia (o termina) dentro de un margen preciso establecido de 15 minutos (positivo o negativo) con respecto a la hora programada, de lo contrario, se concluye que ese vuelo sale o llega con demora de acuerdo a la FAA ^[66] (2021), EUROCONTROL ^[67] (2020) y OAG ^[68] (2017). La aplicación de este estándar puede ser ligeramente diferente en algunos países en lo que se refiere a la interpretación exacta de los 15 minutos, es decir, en algunos casos la puntualidad se acepta hasta los 15 minutos 59 segundos, y en otros la puntualidad se acota a los 14 minutos 59 segundos. Las salidas anticipadas (a veces llamadas demoras negativas) no pueden ser consideradas como demoras, pero dicha anticipación puede causar ciertos problemas operacionales en los aeropuertos, ya que las salidas inesperadas pueden interrumpir la secuencia programadas de vuelos de salida y las llegadas

anticipadas pueden influir también en la asignación de puestos de estacionamiento en plataforma según Pamplona D. & Alves C.J. ^[69] (2020).

Para fijar a partir de qué suceso u operación se cuenta específicamente el tiempo que conduce a que un vuelo se considere puntual o demorado, existen varios criterios a nivel mundial sobre dónde empezar a contar el tiempo, por lo que cada organismo regulador nacional, e incluso cada autoridad aeroportuaria define su propia política al respecto. Un primer estándar muy utilizado es que para los salidas de vuelos la hora programada de referencia sea comparada con la hora exacta y real del remolque o retroceso (*"push-back"*), la cual se produce en el momento en que el tractor o remolque del avión inicia el retiro de la aeronave de su puesto de estacionamiento en la plataforma. Aunque en un segundo estándar, la hora programada de salida se compara con el denominado tiempo real de fuera-calzas *"actual off-block time"* (AOBT), con el retraso de la llegada definido como la diferencia entre la hora programada de llegada y la hora real de puestas o bloqueo de calzas *"actual in-block time"* (AIBT) en el puesto de estacionamiento de la plataforma de acuerdo con ICAO ^[70] (2018), el European Observatory on Airport Capacity & Quality ^[71] (2015, y EUROCONTROL ^[72] (2007). Como se mencionó, las demoras pueden tener orígenes diversos, por lo que para los planificadores y gestores aeroportuarios, así como para los organismos reguladores nacionales del servicio de transporte aéreo, es muy importante conocer los motivos de las demoras. Este conocimiento sirve a los gestores aeroportuarios a generar estrategias operacionales para tratar de mitigar el problema. Por otro lado, para los agentes reguladores o de control del servicio del transporte aéreo, un problema adicional es identificar a los responsables de las demoras y a los actores perjudicados, para si procede, aplicar las oportunas penalizaciones u obligatoriedad de compensaciones económicas a terceros. Para ello, es común en muchos países adoptar el *"Standard IATA Delay Codes-AHM 730"* del IATA 2021b ^[16], en donde se especifican y codifican todos los posibles motivos que generan una demora. En ese estándar se encuentran tabulados (codificados) las posibles causas a las que se puede asociar una demora, según estén asociadas a las operaciones de la línea aérea, del aeropuerto, de la gestión de la flujo del tráfico aéreo (ATFM), sean

causas meteorológicas, y otras causas diversas como pueden ser de seguridad, inmigración, aduana, directrices gubernamentales, etc.

El volumen de las demoras ha ido creciendo en las últimas dos décadas en su magnitud, y por consiguiente en su negativo impacto económico, contabilizado en términos de desviaciones de costes de acuerdo con Borsky S. & Unterberger C. ^[61] (2019), University of Westminster ^[73] (2015), Peterson E. et al. ^[78] (2013), NEXTOR ^[79] (2010), y Cook A. et al. ^[80] (2009). Así, en el sistema de aviación de los Estados Unidos en el año 2019, el 18,87% de los vuelos tuvieron una demora según la agencia de estadísticas BTS ^[81] (2021). En Europa, en el año 2019, el 46,4% de todos los despegues partieron con demora, y el promedio temporal de dichas demoras, para los vuelos que partieron con demora, fue de 28,4 minutos según EUROCONTROL ^[82] (2020b). En Colombia, país que constituye el caso de aplicación de esta tesis, en el año 2019 el 28,60% y el 22,54% de los vuelos domésticos e internacionales respectivamente salieron con demora. Y en lo que se refiere al impacto económico de las demoras, según estimaciones de la FAA ^[83] (2020), el coste generado por las demoras en el sistema de aviación de Estados Unidos en el año 2019 fue de 33 mil millones de dólares (incluidos todos los concepto). En Europa, el coste total de las demoras para el año 2012 se estimó en 11,2 miles de millones de euros de acuerdo a IATA ^[84] (2013). Para Colombia el coste de las demoras, solo para las líneas aéreas y en concepto de compensaciones a pasajeros, en el año 2019 fue de 6,5 millones de dólares según el apartado de las actividades aeronáuticas de Aerocivil ^[85] (2023).

En base a lo comentado, y a la cada vez más relevante problemática de las demoras de los vuelos, la academia viene aportando investigaciones valiosas al respecto, desde diferentes puntos de vista como son: la caracterización, distribución, propagación, predicción, costes, y estrategias para analizar las demoras, desde hace al menos veinte años. En lo que se refiere exclusivamente al estudio de la predicción de las demoras de vuelos comerciales programados, dicho parámetro de gestión del transporte aéreo se ha analizado con diferentes planteamientos y metodologías. Los planteamientos usados de acuerdo con las investigaciones académicas se pueden agrupar en al menos seis grupos fundamentales de acuerdo con Carvalho et al. ^[65] (2021), Kim M. & Bae J. ^[74] (2021),

Truong D. ^[75] (2021), Guo Z. et al. ^[76] (2021) y Lambelho M. et al ^[77] (2020). Los cuales son: 1) análisis estadístico mediante modelos de regresión, análisis de correlación, modelos econométricos, método de Monte-Carlo, pruebas paramétricas, pruebas no paramétricas y análisis multivariable (MVA), 2) modelos probabilísticos (los cuales abarcan, o incluyen, herramientas de análisis que estiman la probabilidad de un evento basándose en datos históricos), 3) análisis de red, principalmente teoría de grafos, aunque estudios relacionados con el análisis de redes bayesianas también se incluyen en esta categoría, 4) investigaciones operacionales, lo que comprende métodos analíticos avanzados como optimización, simulaciones y teoría de colas, 5) Machine Learning con predicción en regresión, donde los métodos más usados incluyen el “*k-Nearest Neighbor*” (k-NN), las redes neuronales Recurrentes (RNN), las redes NN en Cascada (donde cada red está conectada directamente no solo a la anterior sino a todas las precedentes), el Deep Learning mediante redes del tipo CNN, LSTM, y otros métodos de ML como el SVM, fuzzy logic, y Random Forests, y por último 6) el Machine Learning se ha usado también para predicción de clasificaciones binarias (donde el problema se reduce a determinar si el vuelo estará o no demorado) y de multiclases (p.e pronosticando diferentes niveles de demora) donde dos métodos de clasificación comúnmente utilizados son el Decision Trees (DT) y el Naive Bayes (NB), con la mayoría de los modelos como el Random Forest (RF) incluidos en esta categoría.

Para este tercer caso de estudio de la tesis, se han usado los datos correspondientes a la red aeroportuaria completa de Colombia (que conforma el país de aplicación con 58 aeropuertos), comprendiendo los datos 357.595 registros o de salidas (departure) de vuelos comerciales programados correspondientes al año 2018 completo. Los vuelos comprendidos son tanto para salidas nacionales como internacionales, y donde cada uno de los registros está constituido a su vez por más de una decena de atributos asociados al vuelo. Dado que el entrenamiento es supervisado, el dataset usado han consistido tanto en los inputs de atributos de vuelo como de las salidas en valores reales de las demoras ocurridas, por ser una predicción en regresión, contrariamente a la predicción más simple o binaria (donde solo se predice si el vuelo estará o no demorado).

Esta tesis aporta varias novedades, por un lado, se realiza una comparativa (“*benchmarking*”) de las prestaciones medidas en términos de media y varianza estadística, para una bancada de 10 modelos de ML/DL, lo que permite posteriormente seleccionar el modelo más óptimo. De otro lado, se determinan los coeficientes de importancia de cada uno de los “*features*” o atributos de vuelo (inputs) en la determinación de la demora, a través del método de ML conocido como “*Permutation Importance*” según Altmann A. et al. ^[86] (2010) y Breiman L. ^[87] (2001). Este método permite clasificar por orden de importancia, la contribución de cada uno de los atributos del vuelo en la predicción del tiempo de demora. El filtrado de los atributos de vuelo más relevantes, se confirma al volver a evaluar las prestaciones de cada modelo, pero esta vez entrenados con los atributos reducidos a los más importantes. Lo que aparte de simplificar el modelo, disminuye los datos necesarios para su entreno, así como los tiempos de entreno, facilitando sin duda la comprensión de los patrones de conocimiento obtenidos por el algoritmo en la predicción. En definitiva, estas reducciones de los *features* del modelo ayudan a explicitar las decisiones adoptadas por el algoritmo.

Antes de finalizar esta introducción, se realiza una breve descripción de la historia del ML/DL, pues junto a la aplicación al transporte aéreo, esta herramienta de inteligencia artificial constituye el segundo pilar de esta tesis sobre predicción de los parámetros de gestión del transporte aéreo usando algoritmos de ML/DL. Por simplificar el relato, su origen comienza con McCulloch y Pitts ^[89] que en 1943 propusieron crear redes neuronales “artificiales”, para imitar las redes neuronales biológicas, conectando nodos o neuronas entre sí, a las cuales se les asigna determinadas funciones matemáticas, técnicamente denominadas “funciones de activación”. Cada nodo recoge la suma de entrada de los inputs de los nodos anteriores provenientes de las salidas de los nodos de la capa anterior, para que ese nodo genere a su vez una salida que se conecta con los nodos de la capa siguiente, así hasta llegar a la última capa de salida. Con ello se trata de emular las conexiones de las neuronas con sus sinapsis, dendritas (entradas) y axones (salida) que existen en el cerebro biológico. Estas neuronas fueron descubiertas por el científico español más importante de todos los tiempos S. Ramón y Cajal en 1887 ^[90]. Este tipo de aprendizaje con redes neuronales artificiales (ANN) fue

denominado en 1952 por Samuel A. ^[91] (de IBM) como aprendizaje máquina (“*Machine Learning*”). En 1962 el psicólogo Rosenblatt F. ^[92] propuso el modelo de perceptrón multicapa (MLP). Tras estos prometedores inicios se pasó por largos periodos de sequía investigadora, denominados “de invierno” en el argot de la historia de la inteligencia artificial”, con algunos éxitos puntuales, pero sin lograr resolver los problemas que venían apareciendo, y ni mucho menos siendo capaces de difundir sus técnicas y medios a gran escala, como ocurre en la actualidad. El desarrollo del ML/DL quedó así “hibernado” por no poder automatizar los cálculos que los algoritmos demandaban, así como por no disponer de capacidad de computación suficiente para acometer la complejidad de los problemas que se estaban tratando. La primera explosión de estas técnicas llegaría años más tarde, en los inicios del s.XXI gracias a la introducción del hardware de procesado en paralelo denominado GPU (procesado por unidades gráficas) que incorpora de una forma más eficiente el procesado en paralelo necesario para llevar a cabo grandes y complejas operaciones con tensores (denominación genérica para matrices de más de dos dimensiones), como los que aparecen durante las operaciones de entreno del modelo de ML/DL. Aunque fue a finales del s.XX y principios s.XXI cuando realmente se comenzó a avanzar en la resolución de los problemas matemáticos que involucraban estas técnicas, mediante el desarrollo de algoritmos más eficientes en la obtención de los mínimos globales buscados (p.e. de las funciones convexas). De hecho, Yan LeCun Y. ^[20] demostró que la no convexidad de una función (p.e. donde se pueden presentar varios mínimos locales) no era un gran obstáculo para resolver el problema de la optimización en el ML, por lo que el foco se puso en los métodos y algoritmos que lograsen de forma rápida y eficiente una buena aproximación de dicho mínimo. El método para obtener el mínimo, a través del ajuste de los pesos (o conexiones entre las neuronas del modelo ANN), se le denominó de ‘gradiente descendente estocástico’ (SGD). Para explicarlo de forma intuitiva significa avanzar en el ajuste de una variable yendo por la pendiente (derivada de la función) hacia abajo (negativa) para generar la disminución de la función de error (también denominada función de coste). Aunque al hacerlo sobre una parte de los datos (para que el cálculo del mismo sea eficiente en grandes datasets) el gradiente descendente se hace “estocástico”, propuesta que fue

desarrollada por Robbins H. & Monro S. ^[93] en 1951. Este nuevo método del SGD, junto a la “reinención” del método de ‘backpropagation’ (BP) o propagación del error hacía atrás, que se realiza para distribuir el error “total” (de la función de coste) sobre cada uno de los pesos que conforman las conexiones entre los diferentes nodos del modelo ML/DL. El BP se considera que fue inventado ya por G. Leibniz en 1673, con “la regla de derivación en cadena” para obtener la derivación de las funciones que dependen a su vez de otras funciones, sin embargo, hubo de esperarse para su implementación de forma flexible en algoritmos rápidos y eficientes, a fechas tan tardías como la de 1995 por parte de Russel S. & Norvig P ^[94]. Y. LeCun ^[95] así mismo ayudó a introducir y difundir el uso de las redes neuronales convolucionales (CNN), que constituyen uno de los pilares sobre los que se asienta la implementación de las redes neuronales de aprendizaje profundo (DL). Por otro lado, Hochreiter S. ^[96] introdujo como parte de las Redes Neuronales Recurrentes (RNN) las del tipo LSTM (Long Short-Term Memory), abordando al mismo tiempo el problema clásico del “desvanecimiento del gradiente” que tantos problemas daba para el correcto funcionamiento de estas redes. J. Bengio ^[97] continuó con el desarrollo de algoritmos de aprendizaje profundo, aunque parece que fue G. Hinton ^[201] en 2006 quien acuñó la palabra “*Deep Learning*” (DL) por la semejanza de estos tipos de aprendizaje ML, compuestos cada vez por un mayor número de capas, con el del aprendizaje humano, al extraer (automáticamente) los modelos DL las características o ‘features’ del problema, lo que liberaba así de la tarea de extraer dichas características de los datos, por parte de los ingenieros expertos en el problema (lo que se conoce como “*feature engineering*”). En 2019 Y. LeCun, J. Bengio y G. Hinton fueron galardonados por sus contribuciones al aprendizaje profundo con el premio Turing y en 2022 con el premio princesa de Asturias a la investigación científica y técnica, junto a Demiss Hassabi.

2. Justificación

Las motivaciones personales para abordar esta tesis han sido varias. En primer lugar, mostrar la viabilidad de desarrollar algoritmos propios basados en las últimas técnicas de Machine Learning /Deep Learning, a través de uso de lenguajes y librerías de código abierto, implementando modelos de redes neuronales convolucionales (CNN) y de redes neuronales recurrentes (RNN). En particular dentro de estas últimas las denominadas LSTM, todas ellas con amplios y exitosos antecedentes, provenientes de otras áreas del conocimiento como son la del “computer vision (CV)”, procesamiento del lenguaje natural (NLP), series temporales, etc. Para ello adquirimos las capacidades necesarias para generar los algoritmos de ML/DL a través de los tutoriales en internet de Jason Brownlee ^[98] (2023). En segundo lugar, mostrar la viabilidad de aplicar este aprendizaje profundo al campo de la gestión del transporte aéreo, donde no se ha constatado el uso de estos modelos híbridos de CNN+LSTM en la literatura científica, como el que esta tesis aborda con la propuesta del modelo denominado “*ConvLSTM2D*”. En tercer lugar, para explorar el acceso y uso de grandes datasets, como el utilizado de Aerocivil ^[99] (2023), para el entrenamiento, validación y test de los algoritmos aquí desarrollados para su aplicación al transporte aéreo, lo que se consiguió a través de la colaboración con el director de mi tesis. En cuarto lugar, mostrar cómo se pueden alcanzar resultados aceptables sin necesidad de usar grandes recursos de computación. En concreto, aquí se ha usado un Mac mini personal con un procesador Intel i.7 a 3.2 GHz de 6 núcleos, con gráficos Intel UHD Graphics 630 de 1536 MB, con memoria SDRAM DDR4 de 16 GB a 2667 MHz de frecuencia y memoria SSD de 512MB. Y en último lugar, mostrar la viabilidad de realizar esta investigación con recursos exclusivamente personales y sin financiación.

3. Objetivos

3.1 Generales

Mostrar la viabilidad del uso de determinados algoritmos Machine Learning / Deep Learning (ML/DL), cuyas tendencias a futuro enmarca y contextualiza F. Chollet ^[100] (2021), para predecir de forma eficiente y precisa los parámetros de gestión del transporte aéreo, lo que proporciona ventajas frente a las técnicas estadísticas clásicas de acuerdo con Shuwen et. al ^[101] (2021). Desarrollando para ello modelos de redes neuronales artificiales (ANN) propios, escritos en Python como lenguaje de programación, ampliamente aceptado por la comunidad científica, con la “API” de Keras ^[102] creada por F. Chollet ^[103] (2023) para aplicaciones de aprendizaje profundo (DL) bajo la cual se ejecutan las potentes librerías de Tensorflow ^[104] (2019) como ‘*backend*’ o servidor de Keras. Estas librerías ejecutan los grafos que resuelven los complejos cálculos operacionales con ‘matrices multidimensionales’ de rango n (denominados ‘tensores’ por Tensorflow) con el índice ‘ n ’ indicando el número de subíndices (o dimensiones) necesarias para definir un elemento de dicho tensor). Estos tensores emergen de forma natural al entrenar simultáneamente grandes datasets multidimensionales, definidos por múltiples “*features*” o variables, instantes de tiempo, numero de muestras, numero de secuencias, etc. Además, se aplican métodos de variaciones estadísticas como “*k-fold cross-validation*”, para validar la calidad de los resultados obtenidos. Los resultados de los modelos se evaluarán mediante la aplicación de las métricas RMSE (“Root Mean Square Error”) y MAPE (“Mean Absolute Percentage Error”). Esta tesis también busca mejorar la interpretabilidad de los resultados obtenidos por los algoritmos, los cuales trabajan en modo “caja negra”, es decir que el algoritmo no explicita los patrones aprendidos, mediante la aplicación de nuevas técnicas como las de “*permutation importance*”.

3.2 Específicos

Se abordan tres objetivos específicos, los cuales se corresponden con tres casos de estudio, todos ellos fuertemente interrelacionados y asociados con las capacidades y planificación de la gestión aeroportuaria relativos a las demandas de pasajeros aéreos, carga aérea y a las predicciones de las demoras de las salidas de vuelos comerciales, todos los cuales se predicen con herramientas de ML/DL.

El primer objetivo de acuerdo con Muros JG. y Diaz O. ^[46] (2023), del que se adjunta su artículo en el Anexo I (del listado de publicaciones), consiste en la predicción de la demanda de la carga aérea (nacional e internacional) para el caso de aplicación de Colombia, a corto-medio plazo (para los siguientes 5 años post Covid-19) tomando en cuenta las series temporales de los parámetros de gestión del transporte aéreo, pero combinadas con un conjunto de datos socioeconómicos, elegidos de entre los que se tenían disponibles para Colombia, e incluyendo como último año de la serie histórica el año 2020 en el que comienza la crisis global de la pandemia del Covid-19. Dicho estudio de series temporales se realiza con modelos de redes ANN híbridas, donde se mezclan las redes convolucionales (CNN) con las redes recurrentes (RNN) del tipo LSTM ("Long-Short Term Memory) para mejorar el aprendizaje de los patrones no lineales espaciotemporales de multivariantes. Estas predicciones son críticas para la aplicación de las políticas públicas del transporte aéreo encaminadas a priorizar: las inversiones necesarias para acometer la expansión y modernización de las infraestructuras aeroportuarias, de la navegación aérea, de la política de tasas aéreas, regulatorias, etc. Estas predicciones son igualmente importantes para la definición del plan director de los aeropuertos, con vistas a proponer un desarrollo coherente de un aeropuerto o sistema aeroportuario que atienda las demandas futuras de servicios, minimizando el impacto en su entorno.

El segundo objetivo de acuerdo con Muros JG & Diaz O. ^[47] (2022), es similar al objetivo 1, pero reemplazando el parámetro de la demanda de la carga aérea por el de pasajeros, del que se adjunta su artículo en el Anexo II (del listado de publicaciones). Este segundo caso de estudio consiste en la predicción de la demanda de pasajeros (nacionales e internacionales) para el caso de aplicación de

Colombia a corto-medio plazo (en los siguientes 5 años post Covid-19). Para lo que se toma en cuenta las series temporales de los parámetros de gestión del transporte aéreo, combinadas con un conjunto de datos socioeconómicos, elegidos de entre los que se tenían disponibles para Colombia, incluyendo el primer año 2020 de la crisis severa del Covid-19. Dicho estudio de series temporales se realiza con modelos de redes ANN híbridas, donde se mezclan las redes convolucionales (CNN) con las redes recurrentes (RNN) del tipo LSTM ("Long-Short Term Memory) para aprender los patrones no lineales espaciotemporales correspondientes al análisis no lineal de multivariantes. Predecir la demanda de pasajeros en una industria aeronáutica tan sensible a los cambios es una tarea difícil, pues depende de muchas variables exógenas a la propia aviación según Horonjeff R. et al. ^[2] (2010). Para ello se han usado varias multivariantes de Colombia como el PIB, la Paridad de Poder Adquisitivo (PPA), la población, la inflación (IPC), la tasa de cambio peso (colombiano), etc. a fin de correlacionar los patrones de la demanda de pasajeros con la influencia de estas variables socioeconómicas. Conviene señalar que este segundo análisis de la demanda de pasajeros es diferente en naturaleza al de la carga aérea, por presentar unos patrones de comportamiento muy diferentes de acuerdo con Wan MJ ^[105] (2013), en aspectos tan variados como el establecimiento de mercados, configuración de rutas, precios, políticas de marketing, etc.

El tercer objetivo de acuerdo con Muros JG & Díaz O. ^[54] (2023) y Muros JG & Díaz O. ^[55] (2023), del que se adjuntan sus publicaciones en el Anexo III (del listado de publicaciones), consiste en predecir los retrasos en las salidas de los aviones comerciales, para el caso de aplicación de los aeropuertos de Colombia, tanto en vuelos nacionales como internacionales. Dada la importancia que tiene la predicción de este parámetro del transporte aéreo, para evitar que estos retrasos se propaguen en la red aeroportuaria, con los consiguientes perjuicios para los servicios de gestión aeroportuaria y de navegación aérea. En este tercer caso de estudio, la misión ha sido doble, de un lado comparar las prestaciones de varios modelos, para poder seleccionar el mejor, en términos de media y varianza estadística referidos al RMSE (raíz del error cuadrático medio), para lo que se han elegido 10 modelos de ML/DL. Y de otro lado, medir y avanzar en la interpretabilidad

de los resultados obtenidos por los modelos ML/DL, evaluando, con ayuda de las técnicas de “*permutation importance*”, cuales de los “inputs” (o atributos de vuelo) son los más importantes para determinar el retraso en la salida de los aviones, a través de la medición de los coeficientes de importancia de los “features” de vuelo.

4. Revisión Bibliográfica

4.1 Demanda de la Carga Aérea

A pesar de su gran importancia, el análisis o investigación de la previsión de la demanda de transporte de carga aérea no presenta históricamente un gran número de publicaciones, o al menos no tanto como el pronóstico de demanda de pasajeros aéreos según Gelhausen M. et al.^[106] (2019) y Baier F. et al.^[107] (2021). En la literatura científica se pueden encontrar diferentes planteamientos para el pronóstico de la demanda de carga aérea, siendo fundamentalmente de tres tipos: modelos clásicos de gravedad, modelos de series temporales, como los modelos autorregresivos integrados con corrección de errores por medias móviles denominados ARIMA (o SARIMA si se incluye también el análisis de las estacionalidades), y los modelos más modernos de Redes Neuronales Artificiales ANN de ML/DL.

A continuación, se ofrece un breve repaso a los distintos planteamientos existentes en la literatura para la predicción de la carga aérea. Alexander D. et al.^[108] (2021) proponen un modelo de gravedad (estructurado en volumen y distancia) para realizar el pronóstico de la demanda de carga aérea internacional utilizando como caso de aplicación el mercado de los Estados Unidos. De sus resultados los autores afirman que los modelos de gravedad desarrollados por ellos, son lo suficientemente robustos (tanto a nivel agregado como desagregado) para dar cuenta con precisión de ciertos impactos relevantes en el desarrollo histórico pasado. Baier F. et al.^[10i] (2021) presentan un enfoque cuantitativo para los pronósticos de carga aérea utilizando datos aeroportuarios globales (de varias regiones del mundo) en modelos gravitacionales con efectos fijos aeroportuarios generalizados y linealizados.

Chen S.^[109] et al.(2012) comenzaron a utilizar redes neuronales artificiales (ANN), para mejorar al mismo tiempo la precisión del pronóstico de la demanda de pasajeros y de la carga aérea, indicando los factores comunes y diferentes que caracterizan a cada una de las demandas. De los resultados, los autores afirman

que la utilización de más variables de entrada no garantiza pronósticos más precisos (algo que confirmamos en los resultados de esta tesis), y que a mayor número de años de entrenamiento tampoco se garantiza una previsión más precisa. Estos primeros trabajos pioneros en el ML en 2012 deben ser reinterpretados ahora, teniendo en cuenta el contexto de desarrollo básico en el que se encontraban los modelos de redes ANN en esas fechas y la no aparición aún del aprendizaje profundo (DL), a juzgar por los deficientes resultados obtenidos aplicando ML, y como se deduce por la vuelta posterior a la realización trabajos mediante los métodos clásicos estadísticos como ARIMA etc. Chou T. et al. ^[110] (2010) desarrollan un modelo de pronóstico de regresión difusa (FRFM) aplicando los métodos de la lógica difusa, para pronosticar la demanda de carga aérea internacional. Los autores argumentan que, al pronosticar la demanda de carga aérea, los factores de incertidumbre a menudo provocan desviaciones significativas en las estimaciones derivadas de la aplicación de la regresión lineal tradicional. Por ello, y con el objetivo de mejorar la precisión del pronóstico que minimice las desviaciones, los autores proponen integrar el análisis de regresión difusa con el análisis de regresión lineal para reducir el impacto de estos factores inciertos.

Recientemente, Gudmundsson S. et al. ^[18] (2021) proponen un modelo ARIMA para pronosticar la recuperación post-Covid-19 de los mercados globales de pasajeros y carga aérea. Para ello usan el PIB como variable explicativa en la configuración de ARIMA. En la metodología, los autores aplican modelos ARIMA de media móvil integrada autoregresiva con cambios en sus estructuras para obtener un pronóstico a partir del historial pasado. Para el pronóstico de la demanda de carga aérea, usan los datos históricos de tres aeropuertos de China. Li H. et al. ^[111] (2020) proponen un nuevo enfoque de conjunto con descomposición secundaria (SDE) más un algoritmo de búsqueda 'Cuckoo' (CSA). Específicamente, la serie de tiempo de carga aérea original se descompone en: el modo variacional (VMD), el de la muestra (SE) y el modo empírico (EMD). Posteriormente, mediante un proceso combinado de ARIMA con Redes Neuronales Elman (ENN) optimizadas por CSA, se pronostica la componente de tendencia y los cíclicos de baja frecuencia, durante los cuales se realiza la reconstrucción del espacio de fase (PSR) para determinar la estructura de entrada idónea para las redes neuronales.

Los resultados finales del pronóstico se obtienen integrando los valores predichos de cada componente, y según los autores, los resultados obtenidos muestran mejoras con respecto a otros modelos de referencia en términos de robustez y precisión.

Liu J. et al. ^[112] (2020) realizan una evaluación empírica entre dos técnicas estadísticas y tres modelos de Machine Learning para la predicción de la demanda de carga aérea: la regresión lineal múltiple (MLR), la media móvil integrada de auto regresión (ARIMA), la regresión por el método del vector de soporte (SVR), la red neuronal (NN) y el árbol de regresión de aumento de gradiente (GBRT). Los datos de la carga aérea a nivel mundial de un año, se utilizan para valorar los cinco modelos de pronóstico. Para realizar la comparación con rigor, los autores ejecutan los experimentos bajo diferentes horizontes y estrategias. Los resultados de la investigación muestran que el modelo ARIMA es bueno para modelar el pronóstico a corto plazo, mientras que el modelo SVR obtiene mejores resultados en pronósticos a largo plazo.

Madhavan M. et al. ^[113] (2020) utilizan también modelos de media móvil integrada autorregresiva (ARIMA) y series de tiempo estructural bayesiano (BSTS) para pronosticar la demanda de carga aérea y de pasajeros aéreos en la India, a partir de un conjunto de datos históricos de 10 años. Los resultados de esta investigación muestran que el modelo ARIMA es útil para capturar la naturaleza multiplicativa de la tendencia de los datos de pasajeros aéreos, sin embargo, fue menos preciso cuando se presentaban aspectos de incertidumbre o variacionales. Por otro lado, el modelo BSTS fue el más preciso para pronosticar la demanda de carga aérea a corto plazo. Por lo que los autores concluyen que el modelo de ecuación estructural bayesiana parece más transparente y adecuado para manejar directamente datos de series temporales no estacionarios.

Onder E. and Kuzu S. ^[114] (2014) investigan la aplicación de técnicas de suavizado para la estimación de la demanda de tráfico aéreo (incluida la de la carga aérea). Los autores exploran los efectos estacionales de estos parámetros, y como herramientas de análisis proponen los métodos clásicos de predicción de series temporales de promedios móviles, con suavizados exponenciales, suavizados

lineales de un solo parámetro de Brown, suavizados exponenciales de segundo orden de Brown, métodos de descomposición y suavizado exponencial lineal de los dos parámetros de Holt aplicados a la predicción del volumen de tráfico para el periodo 2007-2013 del mercado aéreo turco. Finalmente, Rodríguez Y. et al. ^[8] (2020) realizan un pronóstico a corto plazo del tráfico aéreo (incluida la carga aérea) utilizando DLM (Modelos Lineales Dinámicos). Los autores argumentan que la utilización de DLM presenta algunas ventajas frente a las metodologías habituales como la: detección estocástica que ocultan las series temporales, y la determinación del efecto variable que produce el impacto de las variables exógenas sin incrementar el número de parámetros. De los resultados obtenidos, los autores destacan que la aplicación del modelo DLM obtiene valores MAPE (Mean Absolute Percentage Error) por debajo del 1%, lo que garantiza predicciones de mayor precisión, y un modelo alternativo para desarrollar pronósticos fiables del transporte aéreo en el corto plazo.

4.2 Demanda de Pasajeros Aéreos

En relación con la predicción de las demandas de pasajeros del transporte aéreo, en la literatura científica se encuentran diversos planteamientos como se han visto ya, incluidos los de redes neuronales ANN. A continuación se citan algunos estudios en relación con la demanda de pasajeros: (a) modelo neuronal híbrido con preprocesamiento de datos por descomposición de variables, lo que permite entresacar los “*features*” o atributos más relevantes del problema, método denominado de “*feature engineering*” o creación de características o propiedades, a partir de los datos “crudos” (sin procesar), lo que mejora el rendimiento de la red y la consiguiente optimización de los resultados de la previsión de la demanda según Alekseev K.P.G. & Seixas J.M. ^[115] (2009). El aprendizaje profundo, aparecido posteriormente ha convertido estas técnicas en irrelevantes, al extraer, como parte del proceso del propio modelo DL dichas propiedades, (b) Bao Y. et al. ^[44] (2012) proponen la descomposición de modo empírico del conjunto (EEMD) basado en los modelos de máquinas de vectores de soporte (SVM), las cuales buscan mediante hiperplanos, dentro de las múltiples dimensiones de las

características que constituyen el conjunto de datos, maximizar la separación para los diferentes valores o clases, dentro de un marco de modelado que incorpora métodos basados en la derivada para restringir el indeseado problema que ocurre durante el proceso de cambio de EEMD, por lo que se denomina modelo EEMD-Slope-SVMs, (c) ya se han citado los trabajos de Chen S. et al. ^[109] (2012) que proponen la utilización de redes neuronales ANN implementando el método automático de retro-propagación (BP) según Rumelhart et al. ^[116] (1986), el cual consiste en distribuir desde la salida hasta la entrada (por eso se habla hacia atrás), el error de aprendizaje entre las diferentes neuronas del modelo, en función de su sensibilidad o gradiente de cada nodo, para ajustar los pesos y minimizar el error total, en este caso se usó el aprendizaje supervisado (es decir, donde se conocen previamente las etiquetas o resultados de salida de los datos a entrenar). El BP fue ulteriormente popularizado gracias a los trabajos de Y. Lecun, J. Bengio y G. Hinton ^[117] (2015) al aplicarlo junto con las técnicas de aprendizaje profundo (DL). Estos tres autores recibieron el premio príncipe de Asturias 2022, (d) Jin F. et al. ^[118] (2020) presentan un enfoque híbrido VMD-ARMA / KELM-KELM que consiste en la descomposición en modo variacional (VMD) de un modelo de media móvil autorregresiva (ARMA) y con máquinas de aprendizaje extremo del Kernel (KELM), lo que significa que primero se adopta VMD para descomponer los datos originales en varias funciones, a fin de reducir su complejidad, y luego se utilizan los modelos ARMA y KELM para predecir los componentes estacionarios y no estacionarios, obteniéndose el resultado final como integración de un nuevo modelo KELM que incorpora los resultados de la previsión de todos los componentes, (e) Plakandaras V. et al. ^[119] (2019) y Sulistyowati R. et al. ^[120] (2018) emplean el Machine Learning con modelos simples como modelos SVM en regresión, también denominados regresión por vectores de soporte (SVR), (f) Sun S. et al. ^[121] (2019) plantean un modelo de auto regresión vectorial para problemas no lineales (NVARNN) basado en MIV, este enfoque consiste en: (1) utilización de un método de valor de impacto medio (MIV) para identificar y extraer las variables de entrada, y (2) usando el modelo NVARNN para capturar las irregularidades y volatilidades que presentan las series temporales, (g) Xiao Y. et al. ^[122] (2015) modelan la predicción de la demanda como un conjunto de modelos (“*ensemble*”) basado en el análisis del

espectro singular (SSA), lo que implica que la serie temporal original se descompone en sus tres componentes: tendencia, oscilaciones estacionales y la componente irregular. La tendencia se predice mediante una red neuronal de regresión generalizada (GRNN), mientras que las oscilaciones estacionales se predicen mediante redes de función de base radial (RNFN), (h) Dingari M. et al. ^[43] (2019) utilizan una arquitectura simple de redes ANN como es la de perceptrón multicapa (MLP), llevando implícito el conocido algoritmo de propagación retroalimentada BP, (i) Mostafaeipour A. et al. ^[38] (2018) en cambio usan las redes ANN pero incorporando mejoras con la implementación de algoritmos metaheurísticos evolutivos para la búsqueda de las configuraciones óptimas, (j) Blinova T.O. ^[123] (2007) propone la utilización de las redes neuronales de retroalimentación retardada en el tiempo, (k) y finalmente Gupta V. et al. ^[42] (2019) utilizan las conocidas redes neuronales recurrentes (RNN) basadas en memoria a corto-largo plazo (LSTM), las cuales es de señalar vienen siendo reemplazadas estos dos últimos años por los denominados “mecanismos de atención” de Bahdanau D. ^[124] (2014) que usan el método “transformer” creado por Ashish Vaswani ^[125] (2017) para focalizar el aprendizaje de forma flexible en los valores de mayor peso según el contexto, amplificando o disminuyendo la importancia de estos dentro de la secuencia evaluada. Estos mecanismos de atención, especialmente los de transformer (de autoatención) se corresponde con una forma de aprender patrones más rápida, precisando de menos datos, y permitiendo el procesado en paralelo. Lo cual ha constituido un éxito en el procesado de lenguaje natural (NLP) con grandes modelos de lenguaje LLM (“*Large Language Model*”), y también se han llevado al campo de la visión artificial”, “audio”, optimizando la extracción de patrones en secuencias más largas y complejas.

4.3 Demora de Salida Aviones Comerciales

Las referencias bibliográficas en relación con la demora de los vuelos comerciales (las relativas a los vuelos programados) es estudiada desde varios enfoques, siendo las dos más habituales (por recoger mayor cantidad de publicaciones): la de predicción de las demoras y la de propagación de las demoras.

Sobre esta segundo enfoque se puede encontrar una extensa literatura, entre las que cabe citar: Tan X. et al. ^[126] (2021), Kim M. & Park S. ^[127] (2021), Kafle N. & Zou B. ^[128] (2016), Fleurquin P. et al. ^[129] (2013), Cai Q. et al. ^[130] (2021), Zhang M. et al. ^[131] (2019), Campanelli B. et al. ^[132] (2016), Wang P. et al. ^[133] (2003), Markovic D. et al. ^[134] (2008), Xu N. et al. ^[135] (2008), Zhang J. et al. ^[136] (2010), Hao L. et al. ^[137] (2014), Boswell S. et al. ^[138] (1997), Wong J. & Tsai S. ^[139] (2012), Xu N. et al. ^[140] (2005), Wu W. et al. ^[141] (2016), Baspinar B. et al. ^[142] (2016), Pyrgiotis N. et al. ^[143] (2013), y Hansen M. ^[144] (2002).

Otras líneas de investigación ponen el foco en analizar la relación de las demoras con otros aspectos de la gestión del transporte aéreo, como son: a) la eficiencia del vuelo, sus aspectos económicos (p.e. costes) y logísticos según Peterson E. et al. ^[78] (2013), Cook A. et al. ^[80] (2009), Morrison S. et al. ^[145] (2008), Zou B. & Hansen M. ^[146] (2012), Qin Q. & Yu H. ^[147] (2014), Pathomsiri S. et al. ^[148] (2008), Kotegawa T. et al. ^[149] (2011), Pfeil D. & Balakrishnan H. ^[150] (2012), Abdelghany K. et al. ^[151] (2004), Ahmadbeygi S. et al. ^[152] (2008), Soomer M. & Franx G. ^[153] (2008), y Wu C. ^[154] (2005), b) las operaciones en tierra de las aeronaves de acuerdo con Mofokeng T. & Marnewick A. ^[155] (2017), c) la capacidad aeroportuaria según Zou B. & Hansen M. ^[156] (2014), Abdel-Aty M. et al. ^[157] (2007), y Zhong Z. et al. ^[158] (2017), d) los aspectos relacionados con los clientes según Ferrer J. et al. ^[159] (2012), y Lubbe B. & Victor C. ^[160] (2012), e) y finalmente con la generación, caracterización, comportamiento y estructura de los retrasos según Li Q. & Jing R. ^[161] (2021), Kim M. & Bae J. ^[74] (2021), Mohammadian I. et al. ^[162] (2019), Du W. et al. ^[163] (2018), Lambelho M. et al. ^[77] (2020), Chen Y. & Lin J. ^[164] (2021), Tu Y. et al. ^[165] (2008), Mueller E. & Chatterji G. ^[166] (2002), Pamplona D. & Alves C.J. ^[69] (2020), Fernandes N. et al. ^[63] (2020), Schaefer L. & Millner D. ^[167] (2001), Hunter G. et al. ^[168] (2007), Qin Q. & Yu H. ^[147] (2014), Bakhshandeh R. et al. ^[169] (2013), Borsky S. & Unterberger C. ^[61] (2019), Santos G. & Robin M. ^[170] (2010), Sternberg et al. ^[171] (2016), Wang C. & Wang X. ^[172] (2019), Wang Y. et al. ^[173] (2019), Wesonga R. et al. ^[174] (2012), y Zhang M. et al. ^[175] (2021).

La predicción de las demoras de salidas de vuelos programados es un parámetro crítico tanto para el servicio a los pasajeros como para la gestión de aeropuertos y líneas aéreas. Alonso H. & Loureiro A. ^[176] (2015) desarrollan una

predicción de la demora en vuelos de salida con un planteamiento basado en un modelo unimodal. El modelo unimodal se puede implementar utilizando machine learning (destinado a problemas de clasificación supervisados donde las clases son conocidas para los datos de entreno), para ello los autores utilizan redes neuronales artificiales ANN. La idea principal detrás de este tipo de modelo es que la clase de variable aleatoria asociada con una consulta determinada debe seguir una distribución unimodal. Por otro lado, estos autores miden la importancia de las variables predictoras aplicando un análisis de sensibilidad, seleccionando las más significativas para la predicción del retraso de salida, para lo que utilizan árboles de decisión. Belcastro L. et al. ^[177] (2016) desarrollan un predictor del retraso, de varios días de antelación, para la llegada de un vuelo programado como función de las condiciones meteorológicas. El retraso de llegada previsto tiene en cuenta tanto la información del vuelo (aeropuerto de origen, aeropuerto de destino, hora prevista de salida y llegada) así como las condiciones meteorológicas en el aeropuerto de origen y el aeropuerto de destino de acuerdo con el horario de vuelo. Los autores estudiaron diferentes técnicas de machine learning para evaluar su idoneidad, y finalmente seleccionaron el algoritmo Random Forest (RF) como el más óptimo en cuanto a precisión, lo cual coincide con nuestros resultados. Los resultados muestran una alta precisión en la predicción de retrasos por encima de un umbral determinado. Por ejemplo, con un umbral de retraso de 15 minutos, los autores lograron una precisión del 74,2%, y con un umbral de 60 minutos la precisión fue del 85,8%. Choi S. et al. ^[62] (2016) propusieron un modelo para predecir las demoras de vuelos causadas por las inclemencias del tiempo, para ello los autores recurrieron a la utilización de Data Mining y algoritmos supervisados de ML. Para superar los efectos de los datos de entrenamiento desbalanceados (para aquellos que presentan más datos para unos perfiles de comportamiento que para otros), los autores aplicaron técnicas de re-muestreo y los implementaron en los siguientes modelos de ML: Decision Trees (DT), RF, "AdaBoost" y "*k-Nearest-Neighbors*" (k-NN). Posteriormente, compararon la precisión de predicción de cada uno de los algoritmos y la curva de característica operativa del receptor ROC (Receiver Operating Characteristic) para el caso de las predicciones de clasificación binaria (p.e. predecir si el vuelo estará o no demorado), usando los datos de vuelo y

meteorológicos. Guo Z. et al. ^[76] (2021) propone un método híbrido entre Random Forest Regression RFR y Maximal Information Coefficient (RFR-MIC) para la predicción del retraso en la salida de un vuelo. La RFR y el MIC están inherentemente fusionados en términos de coherencia de la información. El modelo RFR-MIC propuesto por los autores presenta un buen rendimiento en comparación con la regresión lineal (LR), "*k-Nearest Neighbors*" (k-NN), la Red Neuronal Artificial (ANN) y el standard Random Forest Regression (RFR). Por otro lado, los resultados de RFR-MIC muestran una mayor precisión en el coeficiente de correlación (R), el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE). Henriques R. & Feiteira I. ^[178] (2018) proponen predecir la ocurrencia de demoras en vuelos de llegada planteando la metodología Knowledge Discovery Database (KDD), donde aplican varias técnicas de Data Mining. Para entrenar el modelo los autores recopilaban datos históricos de información de los vuelos y del clima. En su análisis utilizaron modelos de Decision Trees (DT), Random Forest (RF) y Multilayer Perceptron (MLP), con estos modelos evaluaron sus actuaciones, y el mejor modelo resultó ser el MLP con un 85% de precisión. Kalliguddi A. & Leboulluec A. ^[179] (2017) proponen el desarrollo de modelos con técnicas de machine learning y modelos estadísticos para identificar y predecir las demoras, como son el DT, RF, y regresiones lineales múltiples (MLR). Pérez-Rodríguez J. et al. ^[180] (2017) desarrollaron un modelo de probabilidad logit asimétrico para estimar y predecir las probabilidades diarias de retraso en las llegadas de aeronaves. Los principales resultados del estudio muestran que, según los métodos logit frecuentista y estándar bayesiano, el retraso en la salida, el tamaño de la línea aérea, el tamaño del aeropuerto y el día del vuelo son los factores estadísticamente significativos para explicar la probabilidad de retraso. Adicionalmente, el modelo bayesiano asimétrico desarrollado identifica un nuevo factor de retraso importante con respecto a los modelos bayesianos frecuentistas y simétricos, cual es la distancia entre aeropuertos (estadísticamente significativa al 1%). Rebollo J. and Balakrishnan H. ^[181] (2014) presentan una nueva clase de modelos para predecir retrasos en el tráfico aéreo. Los modelos propuestos por los autores consideran los estados de retraso temporal y espacial (es decir de red) como variables explicativas, y utilizan algoritmos RF para predecir retrasos de salida de 2 a 24 h.

Los autores analizan dos tipos de mecanismos de predicción: los de clasificación, donde la salida es una predicción binaria de si hay o no demora de salida, a partir de un valor que sea mayor o menor que un cierto umbral predefinido, y de regresión, donde el retraso se calcula como un valor continuo. Finalmente, los autores también analizan el rendimiento de los modelos de predicción propuestos tanto para clasificación como para la regresión de los retrasos en las salidas de los vuelos. Los resultados muestran que, para un horizonte de pronóstico de 2h, el error de prueba promedio es del 19% cuando se clasifican retrasos por encima o por debajo de 60 min. Truong D. ^[75] (2021) utiliza los algoritmos denominados como Causal Machine Learning (CML) dirigidos por USELEI (Understanding, Sampling, Exploring, Learning, Evaluating, and Inferring) junto al Causal Data Mining para predecir la probabilidad de retrasos en los vuelos. Los procesos USELEI permiten construir una red causal completa válida y precisa, donde la red se utiliza para predecir la probabilidad de un incidente de retraso y para identificar tanto los predictores más significativos como los efectos sobre la variable a predecir (p.e. el riesgo de una demora en la llegada). Los resultados de la investigación indicaron efectos importantes en la capacidad, eficiencia y volumen de tráfico en los aeropuertos de origen y destino, como consecuencia del riesgo de retrasos en los vuelos. Los resultados muestran una alta precisión de predicción, ya que el modelo logró una precisión del 91,97%. Yu B. et al. ^[182] (2019) desarrollaron un modelo de predicción de demoras de vuelos, para ello, y siguiendo un enfoque factorial, propusieron el método Deep Belief Network (DBN) para extraer los patrones internos y profundos de los retrasos en los vuelos (utilizando factores micro-influyentes específicos, como p.e. el grado de saturación de la ruta aérea y del aeropuerto, a todos los cuales afecta directamente el retraso de los vuelos). El modelo Support Vector Regression (SVR) se encuentra embebido en la capa superior de la arquitectura desarrollada para realizar un último ajuste fino. Utilizando los datos operativos del aeropuerto de Colombia los autores entrenaron y probaron el método propuesto y evaluaron su rendimiento en función de la precisión, la robustez y el ajuste de parámetros. El método propuesto por los autores demostró ser altamente capaz para gestionar grandes conjuntos de datos y capturar los factores clave que más influyen en los retrasos.

5. Caso Aplicación de Colombia

Colombia ha sido el país elegido para particularizar la aplicación de nuestros modelos predictivos ML/DI bajo la modalidad de SL. Por lo que se va a describir brevemente el contexto aeronáutico y socioeconómico para proporcionar el soporte mínimo a la interpretabilidad de los casos de estudios abordados en esta tesis. El dataset usado para el entreno, validación y test de los modelos ML/DL se ha extraído de las propias bases de datos de Colombia; país que representa el tercer mercado aéreo del subcontinente latinoamericano, y el quinto de las Américas, por volumen de tráfico gestionado ACI ^[45] (2021) e IATA 2021a ^[15]. En Colombia se liberalizó la industria del transporte aéreo a principios de la década de 1990. Esto trajo consigo reformas estructurales en el sector aeroportuario, lo que condujo a la privatización de los principales aeropuertos de la red según Díaz O. & Pulido L. ^[183] (2019) y Díaz O. & Ávila J. ^[184] (2015), produciendo 19 concesiones aeroportuarias, incluyendo los aeropuertos más grandes e importantes de Colombia. Las líneas aéreas también fueron privatizadas como parte de una ininterrumpida batería de políticas públicas (aún hoy en marcha) que incluye no solo aspectos normativos y regulatorios sino también agresivos programas de inversión pública y privada en infraestructura y tecnología de acuerdo con Díaz O. ^[185] (2021). En Colombia el mercado de operadores aéreos privados se realizó primero con un modelo de negocio de servicio completo, o FSC (Full-Service Carrier), lo que se produjo después de la liberalización del sector (principios-mediados de la década de 1990), para posteriormente (desde 2012) dar entrada a las líneas aéreas de bajo coste o LCC (Low-Cost Carrier). Por otro lado, las tarifas aéreas se liberalizaron totalmente desde el año 2012 de acuerdo con Díaz O. & Zea J.F. ^[186] (2018). Como resultado de estas políticas públicas aeronáuticas, desde el inicio de la liberalización de la industria (1991) y hasta el año 2019, el transporte de pasajeros (totales) creció casi un 800% (liderado por el transporte de pasajeros domésticos), mientras que la demanda de carga aérea (total) en el mismo periodo creció casi un 200% (liderado por el transporte de carga aérea internacional) de acuerdo con los boletines de operaciones de Aerocivil ^[187] (2023).

El transporte aéreo en Colombia prácticamente no tiene competencia a nivel doméstico con otros medios de transporte, sobre todo para distancias medias y largas, debido a dos situaciones determinantes: la primera, por la compleja geografía del país atravesado de suroeste a noreste por tres cadenas montañosas de la cordillera de los Andes, y la segunda, por la deficiencia en infraestructura, cobertura, capacidad y tecnología de los sistemas de comunicación terrestre (vial y ferroviario) según Díaz O. & Carvajal A.F. ^[188] (2020).

En la figura 1 se presenta la ubicación geográfica de la mayoría de los aeropuertos de la red aeroportuaria colombiana (participantes en el presente estudio) y abiertos al tráfico comercial. En la figura 2 se presenta el desarrollo del tráfico aéreo de pasajeros en las últimas cuatro décadas.

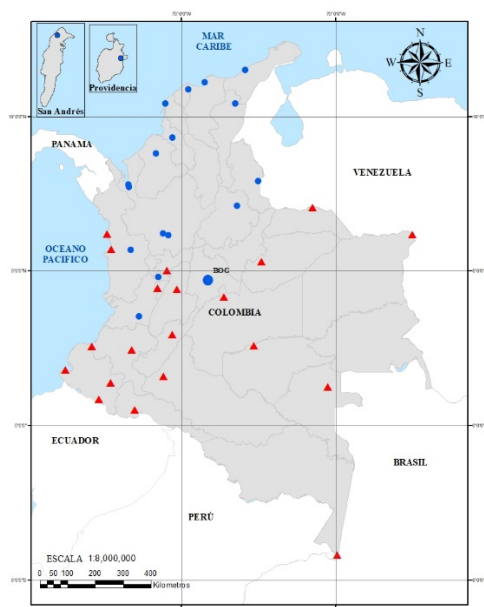


Fig. 1. Ubicación geográfica de la red aeroportuaria colombiana. *Descripción:* círculo azul (aeropuertos de gobernanza privada), triángulo rojo (aeropuertos de gobernanza pública). BOG es el código IATA del principal aeropuerto del país Bogotá-El Dorado, situado en la capital de Colombia. Fuente: Aerocivil ^[187].

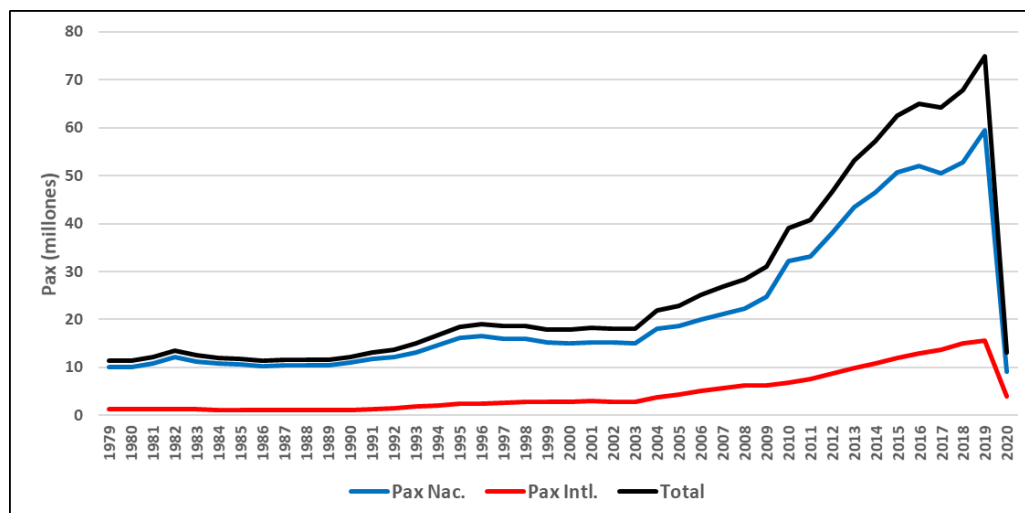


Fig. 2. Desarrollo del tráfico aéreo de pasajeros de Colombia (red completa aeroportuaria), periodo 1979-2020. Fuente: Aerocivil ^[187].

En el ámbito socioeconómico, hay que destacar que el crecimiento del PIB de Colombia, durante el periodo 1979-2020 fue del 850% de acuerdo con el Banco de la República de Colombia BRC ^[189] (2023). Se considera crítico aportar este y los siguientes datos socioeconómicos que se van a referir, ya que ha quedado bien establecida la fuerte y estrecha relación entre la demanda del transporte de pasajeros y carga aérea con la evolución del PIB en el país o región del estudio, de acuerdo con Airbus ^[190] (2016), Boeing ^[191] (2016), Hakim M.M. & Merkert R. ^[192] (2016), Morrell P.S. ^[193] (2011), Ashford N. et al. ^[194] (2013), y Halpern & Graham ^[195] (2018).

En relación con el impacto de la pandemia del Covid-19 en Colombia, dicha circunstancia generó (en el peor año de la pandemia 2020), una fuerte caída de la demanda de pasajeros aéreos (ver figura 2), debida exclusivamente a la completa suspensión del tráfico aéreo comercial de pasajeros entre mediados de marzo y principios de agosto de 2020, derivada de la situación de emergencia sanitaria ordenada por las autoridades sanitarias ante la pandemia del Covid-19. La reapertura y reactivación del sector aerocomercial fue muy lenta, y por etapas y aeropuertos, la cual comenzó a desplegarse en el último cuatrimestre del 2020. Este impacto en los vuelos consistió en una reducción del 85% de pasajeros domésticos y un 75% en los pasajeros internacionales (ambas tasas con respecto

al año 2019). Sin embargo, la demanda de carga aérea (total) solo cayó un 16% con respecto al 2019, por lo que el impacto de la pandemia fue mucho menor de acuerdo con Díaz O. & Alonso ^[196] (2021). En lo que se refiere al impacto económico causado por la pandemia en el país, en el año 2020 la riqueza nacional (PIB) cayó un 5% y el poder adquisitivo de los habitantes (PIB/cápita) disminuyó un 6% (ambos indicadores con respecto al año 2019). En cuanto al mercado laboral, la tasa de desempleo en Colombia se situó en el 15,9% en 2020, lo que significó un aumento de 5,4 puntos porcentuales más frente al 10,5 % de 2019.

En lo que se refiere al tercer caso de estudio sobre el retraso de la salida de vuelos en Colombia, este efecto se puede considerar una de las consecuencias directas del proceso de liberalización de la industria de la aviación comentado para Colombia, lo que causó una concentración de vuelos en unos pocos aeropuertos centrales (o hubs) según señalan Díaz O. & Zea J.F. ^[186] (2018); trabajo que está en línea con lo que señalan otros autores de que un grado de concentración excesiva de vuelos en un aeropuerto tipo “hub” o central puede resultar en demoras por congestión de acuerdo con Wensveen J.G. ^[197] (2016). Por ello, en la red aeroportuaria colombiana durante el año 2019, el 28,60% y el 22,54% de los vuelos domésticos e internacionales respectivamente salieron demorados. La Autoridad Aeroportuaria de Colombia adoptó el Standard IATA Delay Codes-AHM 730 de acuerdo con IATA 2021b ^[16] para codificar y registrar las posibles causas de todas las demoras que se producen en las salidas (departure) de los vuelos programados de la red nacional de aeropuertos abiertos al tráfico comercial. Esta información, precisa y detallada referida a los 357.595 registros de salidas de vuelos (nacionales e internacionales) del año 2018 de los 58 aeropuertos comerciales colombianos, es con la que se ha trabajado en esta tesis. En Colombia, el estándar utilizado por la Autoridad Aeroportuaria Nacional Aerocivil para medir las demoras es el siguiente: la puntualidad de la operación de salida (o llegada) se acepta hasta los 15 minutos 59 segundos (en positivo o negativo) con respecto a la hora programada, contados desde el tiempo real de la salida, medido este en el momento en que se inicia el remolque (“push-back”) del avión desde su puesto de estacionamiento en plataforma.

Los datos históricos usados para el entrenamiento, validación y test de la demanda de carga aérea/pasajeros (doméstica e internacional) corresponden al periodo de 42 años 1979-2020 (ambos inclusive), así como de los 357.595 registros de salidas de vuelos en Colombia del año 2018, los cuales se han obtenido del sistema estadístico de transporte aéreo de la Autoridad Aeronáutica de Colombia Aerocivil ^[187] (2023); mientras que los datos de las series históricas con los indicadores socioeconómicos más relevantes de Colombia usados en el estudio, se han obtenido de las instituciones oficiales colombianas, del Banco de la República de Colombia BRC ^[189] (2023) y del Departamento de Estadística de Colombia DANE ^[198] (2023).

Es importante reseñar que los datasets de Colombia usados en esta tesis son de carácter público y su acceso es libre, sin coste alguno.

6. Metodología y Datos

Dado que tanto el modelo “*ConvLSTM2D*”, como la metodología y datos (salvo en una de las series históricas) usados, son comunes a los procedimientos de la predicción de la demanda de pasajeros y de la carga aérea, se ha considerado oportuno agrupar sus análisis en un mismo apartado, el 6.1 relativo a los dos primeros casos de estudio sobre el pronóstico a corto-medio plazo de dichas demandas de tráfico aéreo (pasajeros y carga). Dejando para el apartado 6.2 la metodología y análisis de datos correspondientes al tercer caso de estudio sobre la predicción de la demora en la salida de vuelos comerciales, por ser diferente el proceso, datos y modelos usados.

6.1 Dos Primeros Casos de Estudio de Carga Aérea y Pasajeros

6.1.1 Dataset

De acuerdo con la metodología implementada, antes del desarrollo del modelo, se ha procedido a la preparación de los datos (o *dataset*) relativos a las series temporales históricas anuales de seis variables: siendo 5 los “features” o variables de entrada socioeconómicos elegidos que acompañan a la sexta variable, que es la propia serie histórica de la demanda de transporte (de carga o de pasajeros) que se pronostica. El análisis multi-variable está pues constituido por la propia variable aeronáutica a predecir: que será o bien la carga aérea doméstica o internacional (para el primer caso de estudio), o bien los pasajeros nacionales o internacionales (para el segundo caso de estudio); mientras que el resto de variables usadas como entrada son las correspondientes a la situación socioeconómica de Colombia, las cuales coinciden con las que recomiendan los trabajos académicos de Rodríguez Y. et al. ^[8] (2020), Díaz O. & Girón E. ^[199] (2020) para el estudio de estas demandas. Específicamente son: el PIB (Producto Interior Bruto), la Población, el IPI (Índice de Producción Industrial), el IPC (Índice de Precios al Consumidor), y la TRM (Tasa de Cambio entre dólar americano y el peso colombiano). Los datos históricos cubren el periodo continuado de 42 años entre 1979 y 2020, y fueron obtenidos del sistema estadístico de la Autoridad Aeronáutica

de Colombia Aerocivil ^[187] (2021), mientras que los datos socioeconómicos se obtuvieron de las fuentes oficiales competentes del Banco de la República de Colombia BRC ^[189] (2023), y del Departamento de Estadística Nacional de Colombia DANE ^[198] (2023), como ya se ha mencionado.

Aunque los datos del año 2020, tanto los de tráfico aéreo como los socioeconómicos son muy diferentes respecto a los años anteriores, debido a la comentada crisis global y local del Covid-19 en Colombia, se decidió incluir como último año de las series históricas el 2020, para obligar al modelo a aprender parte del patrón socioeconómico y aeronáutico asociado a la pandemia.

6.1.2 Preparación Dataset

Una parte importante del esfuerzo en la preparación del dataset ha consistido en extraer y dividir las secuencias temporales de entrada al modelo, en las subsecuencias relativas al número de pasos previos “*lags*” (o pasos anteriores) de cada variable de entrada, necesarios para realizar las predicciones “*outs*” (o pasos futuros), propios del modelo “*ConvLSTM2D*”. Aunque el algoritmo recorre durante el entrenamiento la serie temporal histórica completa de los 42 años, los recorre a través de los grupos de secuencias temporales, en las que se dividen los pasos de los años previos, definidos con el hiperparámetro “*n_lags*”, antes de considerar los pasos de predicción futura, definido con el hiperparámetro “*n_outs*”. Estas variables además se han “hiperparametrizado” en el algoritmo, para analizar su sensibilidad a la precisión en la predicción, a fin de elegir y posteriormente usar, las longitudes de años “*n_lags*” y “*n_outs*” que optimizan el aprendizaje del modelo. Como se ha señalado, se extraen también del dataset la propia serie histórica correspondiente a la salida a pronosticar, o bien la carga aérea, nacional o internacional, asociadas con el primer caso de estudio que contempla estos dos tipos de predicciones, o bien los pasajeros, nacionales o internacionales asociados con el segundo caso de estudio que contempla estos dos tipos de predicciones, pero siempre una sola a la vez (nacional o internacional). Así para el primer caso de estudio de la carga aérea se usa la serie de la carga domestica (o la internacional), puesto que durante el entrenamiento supervisado del modelo se le debe enseñar al algoritmo los resultados reales de las propia salida a predecir, para que pueda estimar el error

del modelo al comparar la salida estimada con la real, error que se define a través de la función de coste elegida. Para simplificar el estudio de “hiperparametrización”, se ha considerado el mismo número de pasos de años previos que el de pasos futuros a predecir; es decir $n_lags = n_outs$.

Para el primer caso de estudio, se parte inicialmente del dataset completo, antes de filtrarlo o reducirlo con ayuda de los modelos, compuesto por todos los features disponibles del dataset (que son 11 en total), donde dos son aeronáuticos correspondientes a la carga aérea (denominados DOM e INT) y los otros nueve son los socioeconómicos de GDP, PCG, POP, IPI, CPI, BMR, FDI, IMP, tabulados en series temporales desde el año 1979 hasta el año 2020 (ambos inclusive). Estas once variables se definen en la tabla 1 para una mejor comprensión. Así pues, estas variables de entrada no solo recogen la propia evolución de la demanda de la carga aérea nacional (DOM) o de la carga aérea internacional (INT), sino que también recogen la evolución de esas otras nueve variables socioeconómicas del contexto particular de Colombia, que sirven para definir el marco donde buscar los porqués de la evolución de la demanda de transporte de la carga aérea. Estas variables de contexto se han elegido en función de su relevancia apriorística y de la disponibilidad estadística de las mismas. Se habla de “a priori” porque luego, con ayuda del modelo se fijará la importancia de los inputs como veremos más adelante, con el análisis de los features que definen los diferentes escenarios o bien con la determinación de los coeficientes de importancia.

ACRONYM	DESCRIPTION	UNIT
GDP	Gross Domestic Product	US\$
PCG	Per Capita GDP	US\$
POP	Population	No unit
IPI	Industrial Production Index	Dimensionless index
CPI	Consumer Price Index	Dimensionless index
BMR	Benchmark (or Currency) Market Rate	COP (local currency)
FDI	Foreign Direct Investment	US\$
EXP	Exports	US\$
IMP	Imports	US\$
DOM	Domestic Air Cargo	Tonne
INT	International Air Cargo	Tonne

Tabla 1. Acrónimos de las variables aeronáuticas y socioeconómicas y sus unidades correspondientes al primer caso de estudio utilizadas en esta tesis. Fuente Autor.

Los datos de las series históricas se han normalizado, es decir, se han escalado las variables o *features* en una escala de valores de 0 (valor mínimo) a 1 (valor máximo) a fin de que todas las variables sean comparables en magnitud, para que el modelo “*ConvLST2D*” no las sobrevalore (o infravalore) por el valor de su magnitud inicial. Y sobre esa transformación se le aplicó una estandarización adicional, consistente en asignarle una distribución estándar a la serie, con media 0 y desviación típica 1. Con estas dos transformaciones (normalizado y estandarización) en las distribuciones de las series temporales se trata de facilitar el aprendizaje del modelo ponderando por igual todas las variables inicialmente. La elección específica de dichas transformaciones del dataset forma parte de los procesos heurísticos de búsqueda dentro de los espacios potenciales de posibles transformaciones, que mejor resultados ofrecen.

A continuación se separa el dataset en dos tensores: el de Inputs $\mathbf{X}$ y el de Outputs $\mathbf{Y}$. Dado que el modelo *ConvLSTM2D* es del tipo supervisado (se requiere entrenarlo con las salidas u objetivos reales, también denominados etiquetas). Para ello se precisa agrupar por bloques el dataset histórico; en este caso en tensores de 5 dimensiones para los inputs $\mathbf{X}$ y de 3 dimensiones para las salidas $\mathbf{Y}$, donde las dimensiones de esos tensores son las correspondientes a las requeridas por la

arquitectura de capas definidas para el modelo ConvLSTM2D, como se ve en la figura 6.

6.1.3 Arquitectura Modelo “ConvLSTM2D”

El modelo desarrollado en esta tesis, es uno derivado de las redes neuronales artificiales ANN de aprendizaje profundo DL, denominado “ConvLSTM2D” (“Conv” de Convolutacional y “LSTM2D” de memoria a corto-largo plazo), el cual es una aportación singular de esta tesis. Pues si bien ha sido usado o como red convolutacional (CNN) o como red recurrente RNN (LSTM) Millstein F. ^[39] (2018), Sewak M. et al. ^[40] (2018), Yang J. et al. ^[48] (2015), Malhotra P. et al. ^[49] (2015), y Hermans B. & Schrauwen S. ^[50] (2013), no se ha visto utilizado en la literatura científica como modelo hibrido para aplicaciones en el sector del transporte aéreo. Es decir, con esas dos características superpuestas de convolutacional y recurrente para el sector del transporte aéreo. El modelo “Conv2LSTM2D una vez entrenado o ajustado en las conexiones (o pesos) que hay entre las neuronas de las diferentes capas del modelo neuronal, es representado matemáticamente mediante la ecuación (1), donde su funcionalidad es sintetizada mediante la función $f_{model\ forecast}$. De esta forma el modelo predicen las variables de salida (denotadas genéricamente por el tensor $\hat{Y}$) con sus n_outs o pasos de secuencia temporales correspondientes a cada uno de los i “features” representados a su vez mediante la variable $\hat{y}_{it+n_outs}$, contados a partir del tiempo “ t ” y estimados a partir de las variables de entrada al modelo (denotada genéricamente por el tensor X) con sus n_lags de observaciones previas asociadas al instante de tiempo “ t ” considerado, para cada uno de los i “features” de entrada, representados mediante la variable x_{it-n_lags} , como se aprecia en la ecuación siguiente.

$$\hat{y}_{it+n_outs}, \hat{y}_{it+n_outs-1}, \dots, \hat{y}_{it} = f_{model\ forecast}(x_{it-1}, x_{it-2}, \dots, x_{it-n_lags}) \quad (1)$$

Para entender las ecuaciones que están detrás del modelo “ConvLSTM2D” se comienza por esquematizar la arquitectura general de las redes neuronales artificiales ANN, sobre las que se fundamenta el modelo de DL de “ConvLSTM2D”.

Estas redes se denominan redes de perceptrón de multicapa (MLP) según Vang-Mata R. ^[200] (2020), las cuales están constituidas por varias capas de neuronas, con una primera capa correspondiente a la entrada (asociada al tensor $\mathbf{X}$), y una última capa de salida (la asociada al tensor $\mathbf{Y}$), entre las que se intercalan otras capas ocultas de neuronas. Cada capa contiene un número variable de neuronas (o nodos), que son definidas mediante las correspondientes funciones de activación, las cuales pueden ser de muchos tipos de acuerdo con Hastie T. et al. ^[201] (2013). Por claridad se muestra en la figura 3 el flujo de conexiones entre la entrada y salida, para una sola neurona.

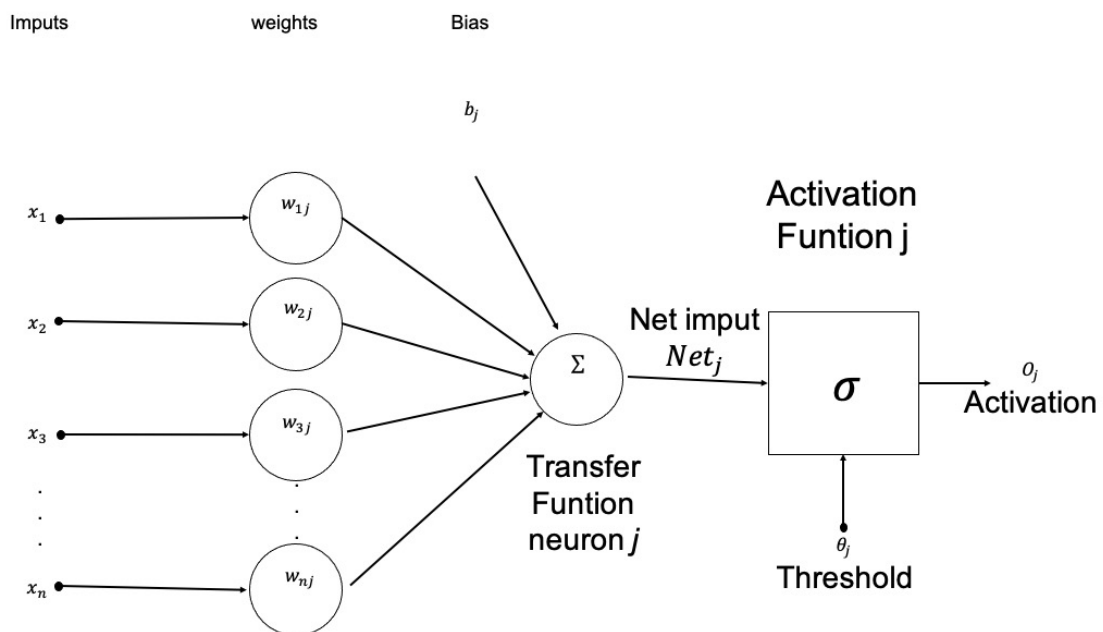


Fig. 3. Flujo de las variables asociadas a las conexiones de neuronas en redes MLP. Fuente: Kingma D. & Ba J. ^[218] (2014).

En la figura 3 las x_i representan los inputs correspondiente al "feature" i de la capa de entrada, con w_{ij} representando el peso o ponderación del "feature" i de la capa de entrada en la conexión con la neurona j de la capa posterior, Σ representa la función de transferencia, o suma de todas las entradas a la neurona j provenientes de todas las salidas de las neuronas de la capa anterior y corregidas

con el sesgo (o *bias*) b_j de la neurona (lo que en las neuronas naturales se denominan dendritas), produciendo la salida neta Net_j asociada a la neurona j . A continuación, a cada neurona se le aplica la función de activación σ elegida (p.e. *tanh* o *relu* o *sigmoide*), la cual produce finalmente la salida O_j de la neurona j (lo que en las neuronas naturales se denomina axón). La misión de la función de activación de una neurona es activar (o desactivar) la neurona, objetivo muy importante para desacoplar neuronas que no intervengan en la comprensión del problema. Al mismo tiempo, la función de activación permite introducir efectos no lineales, para que el modelo pueda aprender patrones más complejos (no simples relaciones lineales), pero siempre se exige que las funciones de activación sean derivables. Dicha característica de derivabilidad permite aplicar luego los métodos del “gradiente descendente”, que implican la derivación en cadena o Backpropagation (BP) en cada nodo, al objeto de repartir el error yendo hacia atrás, entre todas las neuronas del modelo, según sus sensibilidades o contribuciones al error. A partir de aquí se establecen las sucesivas conexiones, desde la salida de cada neurona a las neuronas de la capa siguiente, hasta llegar sucesivamente a la última capa de salida, donde se generan todas las salidas de las neuronas de la última capa asociadas al tensor de salida $\mathbf{Y}$. Este flujo de entrada-salida para estimar un valor, se denomina de propagación “hacia delante” desde la capa de entrada, pasando por todas las neuronas de las capas intermedias u ocultas, hasta la capa de salida final. Siendo θ_j el valor umbral de la neurona j , para activar (o no) su función de activación según Dorffner G. ^[202] (1996). Es importante subrayar que el poder desactivar las neuronas del modelo (gracias a la elección de la función de activación) es crítico, como ya se ha mencionado, a fin de poder desvincular o desacoplar durante el aprendizaje la contribución de determinadas neuronas/capas, que no contribuyen a la solución del problema, de manera que se genere un patrón de aprendizaje más simple y robusto entre exclusivamente las neuronas que contribuyen a la solución.

Del conjunto de datos (dataset) se extraen las x_{it} o “inputs” de entrada, donde los subíndices i y t representan el número del “*feature*” e instante del tiempo “ t ” de la observación (de la serie histórica). Estas x_{it} se agrupan en secuencias temporales, como se ha explicado anteriormente, mediante un número de tiempos

o pasos previos al instante t asociado al de comienzo de la predicción futura, y que se han denotado con el hiperparámetro n_lags . De igual forma, del dataset se extrae la salida y_{it} con la secuencia de pasos futuros de tiempo a predecir para el feature i , que se ha denominado n_outs . Estos dos tensores, de entrada $\mathbf{X}$ y salida $\mathbf{Y}$, son multidimensionales según Ketkar N. & Moolayil J.^[51] (2021). En efecto, el tensor $\mathbf{Y}$ de salida tiene tres dimensiones, las cuales se corresponden: con el número de muestras, el número de pasos de tiempo a predecir n_outs , y el número de “features” a predecir. En cambio, el tensor $\mathbf{X}$ de entrada posee cinco dimensiones correspondientes al: número de muestras, el número de subsecuencias en la que se dividen los n_lags para su análisis, el número 1 por ser las series temporales unidimensionales, el número de años comprendidos en cada subsecuencia para construir los n_lags desde los pasos previos, y finalmente el número de “features” de entrada i .

El modelo “ConvLSTM2D” posee una estructura codificador-decodificador esencial, donde los inputs $\mathbf{X}$ son primeramente codificados, con las capas híbridas que realizan las operaciones convolucionales (CNN) y recurrentes RNN (LSTM) según Aggarwal C.^[203] (2018), Bianchi F. et al.^[204] (2017), Blokdyk F.^[205] (2017), y Mandic D. & Chambers J.^[206] (2001). En cambio, en nuestro modelo desarrollado para esta tesis, la decodificación se realiza solo con las capas recurrentes de memoria de corto-largo plazo (LSTM), porque ha dado mejores resultados. La codificación que aportan las capas convolucionales facilita la representación de las invariancias de las secuencias de los “features” de entrada, mientras que la aportación de las células LSTM en la codificación permite memorizar el contexto temporal o histórico de esos *inputs*. La parte descodificadora del modelo, realizada con las capas LSTM, permite extraer e identificar los patrones entre los “features” y el tiempo, como son la tendencia y estacionalidad de las series históricas de acuerdo a Pedrycz W. & Chen S.^[52] (2020). Finalmente, la salida de las redes LSTM, dentro de la arquitectura del modelo según se aprecia en la figura 6, es “aplanada” y entregada a la última parte del modelo, cuya configuración consiste en capas densas (todas las neuronas conectadas con todas), del tipo MLP (ya descrito) cuya misión es realizar el aprendizaje de los patrones extraídos automáticamente por el bloque codificador-decodificador, anteriormente descrito.

En resumen, la aplicación de la codificación-descodificación con capas convolucionales CNN y recurrentes LSTM en el modelo “*ConvLSTM2D*” equivale a producir un doble aprendizaje profundo de acuerdo con Brownlee J. ^[207] (2018), Trifa A. et al. ^[208] (2017), y Yang J. ^[48] et al. (2015).

La figura 4 representa los mapas de features que producen las capas de convolución, al aplicar la operación de convolución a los datos de la capa precedente, según sean los valores de los diferentes filtros (kernel) y los tamaños de las ventanas deslizantes definidos. Estos valores de la configuración de las capas Convolucionales (número de filtros y tamaño ventanas) se hiperparametrizan para su posterior optimización. La Convolución consiste en la proyección en espacios vectoriales de más dimensiones (por el número de filtros), donde comienzan a emerger los patrones de correlación de los diferentes *features*. Las capas CNN se pueden apilar sucesivamente en varios bloques, tras realizar el submuestreo detrás de cada operación de convolución de acuerdo a Yan W.Q. ^[209] (2020). Con la operación del submuestreo (denominada “pooling”), que viene a continuación, se pretende reducir los problemas clásicos del “sobreajuste”. Este problema, suelen aparecer en el aprendizaje de los modelos ML/DL, cuando se sobre-entrena un modelo que p.e. tiene muchos parámetros a ajustar, pero en cambio se ha usado un dataset pequeño, por lo que se le impide al modelo la capacidad de generalizar, al acabar por “memorizar” todos los sesgos y ruidos del dataset específico usado durante el entreno gracias a sus muchos parámetros de ajuste que le proporciona su configuración de arquitectura profunda. Otras técnicas que se han usado en esta tesis para evitar el sobreajuste son las técnicas de “*k-fold cross validation*” que más adelante se explican, junto a otras conocidas operaciones de regularización o suavización, como son las operaciones de “dropout”, en las que se desconectan aleatoriamente algunas neuronas durante el entrenamiento, para forzar así a que el resto de las neuronas aprendan o generalicen mejor. Dicho suavizado o diluciones, se puede conseguir también limitando los valores máximos de los pesos de las neuronas, de acuerdo con Millstein F. ^[39] (2018) y Sewak M. et al. ^[40] (2018). El mapa de “*features*” es generado automáticamente por el propio modelo, proceso que se denomina de “*extracción de features*”, lo que es una propiedad de los modelos DL. Ya que cuando no se usan

modelo DL, se hace necesario realizar dicho proceso previo de extracción “manual” de features mediante la intervención de un experto en el problema, operación que se conoce como “*feature engineering*” o bien requiere de la implementación adicional de otro modelo o algoritmo específico que realice dicha tarea.

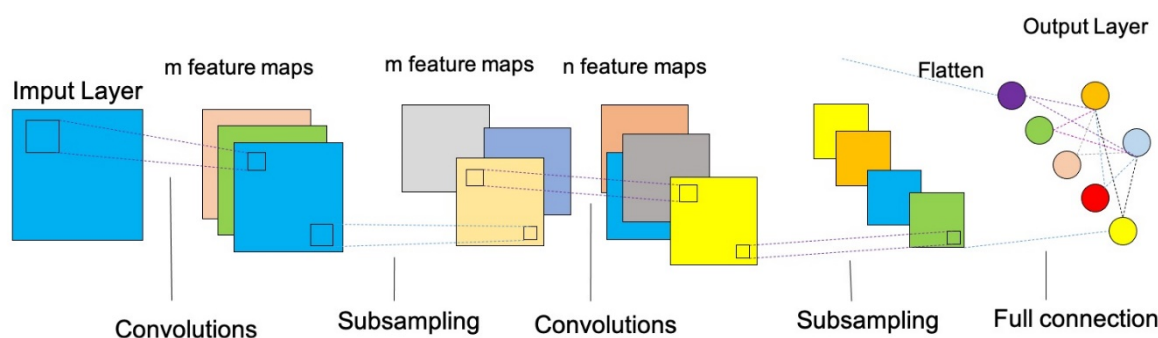


Fig. 4. Esquema general de la producción de mapas de “*features*” realizados por capas convolucionales. Fuente: Hu W. et al.^[210] (2019).

A fin de poder describir las ecuaciones del modelo *ConvLSTM2D*, se muestra en la figura 5 el esquema de una célula de memoria a corto-largo plazo LSTM, donde se observa la presencia de los tres tipos de puertas empleadas para su gestión; con (*i*) denotando la entrada, (*f*) el olvido, (*o*) la salida y (*c*) los estados de la célula. Las redes recurrentes (RNN) se caracterizan por realimentar las entradas de una neurona con las salidas de las mismas, es decir que son de tipo “cíclicas” (o bucles), contrariamente a la mayoría de las redes ML/DL, incluidas las convolucionales, que son de “propagación directa hacia adelante”, sin realimentación. Esta característica permite a la red guardar memoria de contextos secuenciales pasados (como los usados en el lenguaje) o de contextos temporales (como los de las series históricas de esta tesis). Las redes LSTM creadas por Hochreiter S. & Schmidhuber J. ^[212] (1997) se desarrollaron para mejorar los problemas que aparecían en las redes RNN, al aumentar la capacidad de memoria de las dependencias a largo plazo (o de secuencias muy largas), a través de una arquitectura en la que se valora en cada momento, que información significativa hay que guardar y cual olvidar, criterio que depende de valorar el *feature* dentro de

cada secuencia o contexto. En concreto, estas redes LSTM combaten mejor el problema clásico de las RNN de “desvanecimiento del gradiente”. Problema que aparece cuando los pesos no se actualizan porque los gradientes llegan a ser casi nulos en las neuronas de las capas iniciales, como consecuencia de ir multiplicado muchas veces entre si valores muy pequeños del gradiente.

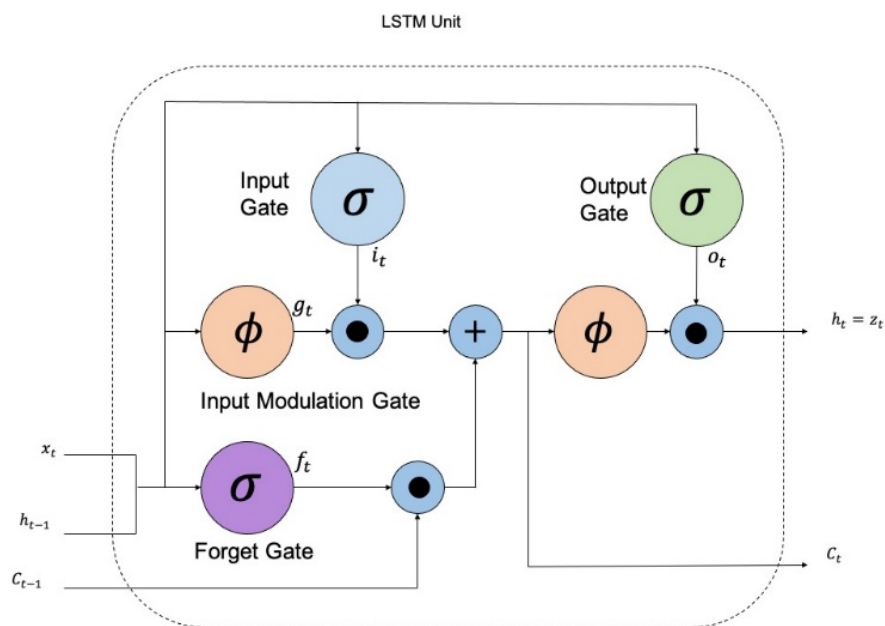


Fig. 5. Esquema de una célula o unidad de memoria de corto-largo plazo LSTM. Fuente: Donahue J. et al. ^[211] (2015).

A continuación se describe el grupo de ecuaciones (2) que definen las capas LSTM, pero con las operaciones de convolución dentro de ellas mismas, para definir el modelo ConvLSTM2D, de acuerdo con el esquema representado en las Figuras 3, 4 y 5. La superioridad en las prestaciones de la arquitectura de capas “ConvLSTM2D” radica por un lado, en la mencionada capacidad de gestionar la memoria temporal a corto y largo plazo que le proporciona la arquitectura LSTM, y de otro lado la proveniente de la incorporación de las capas convolucionales, que le proporcionan los patrones espaciales o de correlación entre las variables, a la vez que disminuye la dimensionalidad al problema, a través de los procesos de submuestreo (“pooling”) a los que se les somete tras la obtención de los mapas de

features. Características propias del modelo híbrido de “ConvLSTM2D” de acuerdo a Hu W. et al. ^[210], Trifa A. et al. ^[208] y Shi X. et al. ^[213] (2015). Las ecuaciones del modelo son:

$$i_t = \sigma(W_{xi} * x_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \quad (2a)$$

$$f_t = \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \quad (2b)$$

$$o_t = \sigma(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{co} \circ c_{t-1} + b_o) \quad (2c)$$

$$g_t = \tanh(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c) \quad (2d)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t \quad (2e)$$

$$h_t = O_t \circ \tanh(c_t) \quad (2f)$$

Donde: i_t , f_t , o_t representan respectivamente la puerta de entrada, la puerta del olvido y la puerta de salida de la célula ConvLSTM2D en el instante “ t ”, x_t son los *inputs* temporales de la célula para ese instante “ t ”, h_{t-1} y c_{t-1} son la salida y el estado respectivamente de la célula en el instante anterior “ $t-1$ ”, siendo “ $*$ ” el operador convolución, y W el filtro o kernel de la convolución, con $W_{\cdot i}$ y b_i representando a los pesos y sesgos de la puerta de entrada, $W_{\cdot f}$ y b_f los pesos y sesgos de la puerta del olvido, $W_{\cdot o}$ y b_o los de la puerta de salida y $W_{\cdot c}$ y b_c los pesos y sesgos del estado de la célula. Las dimensiones y procesamiento de las capas ConvLSTM2D, aun siendo análogos a los de la célula LSTM, son diferentes, dado que la célula híbrida ConvLSTM2D realiza la operación convolución de acuerdo con Hu W. et al. ^[210] (2019). Finalmente, el operador “ $\circ$ ” denota el producto de Hadamard y σ la función de activación. Cabe mencionar que en el algoritmo desarrollado en la presente tesis la función de activación σ elegida ha sido la de la función de unidad lineal rectificada (ReLU) con mejores prestaciones que la indicada en la fórmula (2d) de “ $\tanh$ ”, por converger más rápido cuando se usa el gradiente estocástico SGD y por desactivar las neuronas en el aprendizaje (cuando la entrada es negativa), lo que hace el cálculo del modelo más simple y rápido de acuerdo con Nair & Hinton ^[214] (2010).

En la figura 6 se muestra la arquitectura completa y detallada del modelo ConvLSTM2D desarrollado en esta tesis, para una de las configuraciones de los hiperparámetros utilizadas, definida al usar solo dos “*features*” de entrada (inputs) y un solo *feature* de salida, y para el caso de entrenar el modelo con seis pasos temporales previos n_lags , subdivididos a su vez en dos dimensiones nuevas asociadas a 3 subsecuencias de 2 elementos cada una, para predecir otros seis pasos futuros n_outs , para cada año. En dicha figura 6, se pueden observar las cinco dimensiones de los inputs del tensor de entrada $\mathbf{X}$ y las tres dimensiones del tensor de salida $\mathbf{Y}$. El signo de interrogación ‘?’ (a veces denominado “*none*”) de la primera dimensión, denota el grado de libertad en la elección de la cantidad de muestras a usar. También, se observa la definición de la arquitectura específica que se ha hecho en esta tesis, con un codificador resultante de la concatenación de tres bloques compuestos cada uno de ellos de una capa ConvLSTM2D seguida de una capa MaxPooling3D (finalmente solo se ha usado una capa de MaxPooling3D en el primer bloque por dar mejores resultados en los dos casos de estudio). Esta capa de “*pooling*” es la responsable de sub-muestrear las salidas de la capa anterior para evitar el sobreajuste en el aprendizaje. La salida del codificador de “*features*” se aplanan antes de ser entregada al decodificador de la capa LSTM, como ya se ha mencionado. Entremedias, están las capas denominadas *repeat_vector* y *time_distributed*, las cuales simplemente sirven para acoplar o adaptar las dimensiones de entrada/salida de las capas LSTM según Calin O. ^[53] (2020).

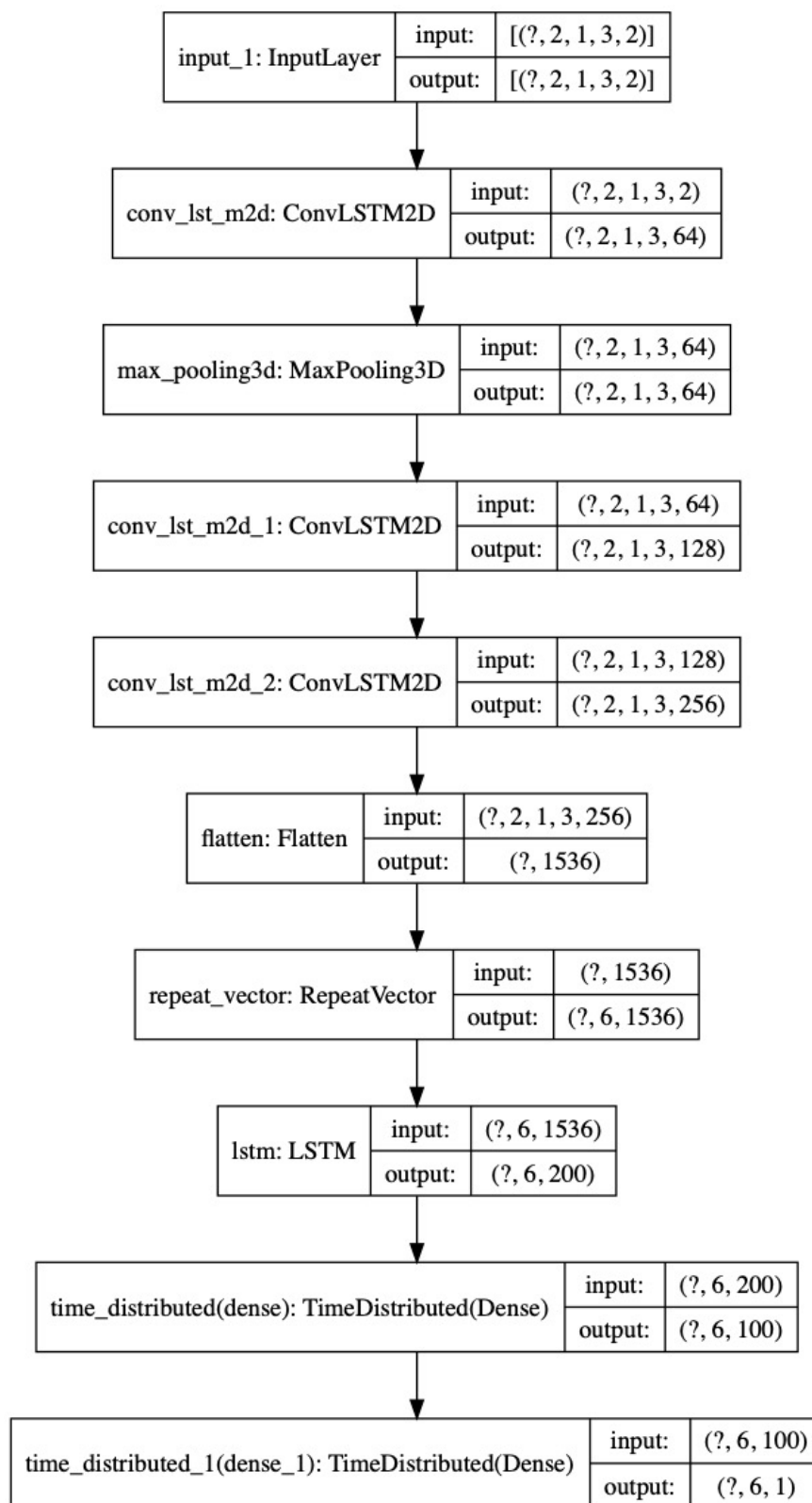


Fig. 6. Modelo completo de ConvLSTM2D usando tres bloques codificadores apilados de ConvLSTM2D y un decodificador LSTM. Fuente: Autor.

6.1.4 Procesos y Entrenamiento Modelo

Una vez definida la arquitectura del modelo *ConvLSTM2D* de la fig.6, y antes de realizar el entrenamiento del modelo, se implementa el proceso de compilación del algoritmo, en el cual se definen otros hiperparámetros del modelo como son el tipo de optimizador del gradiente (SGD, Adam, RMSprop, etc.), la función de coste o error durante el aprendizaje (MSE, etc.), y la métrica para la evaluación final del modelo una vez entrenado (MAPE, RMSE, etc.). El siguiente paso en la metodología comprende el entrenamiento del modelo *ConvLSTM2D*. Antes de entrenar el modelo *ConvLSTM2D* en aprendizaje supervisado, hay que elegir la función de coste a usar, la cual será la responsable de comparar el valor de las salidas reales (objetivas) de la serie temporales esperables (tensor $\mathbf{Y}$) con las predichas o estimadas (tensor $\hat{\mathbf{Y}}$) obtenidas por el modelo *ConvLSTM2D* a partir del tensor de entrada $\mathbf{X}$. La ecuación (3) siguiente es la forma compacta equivalente a la ecuación (1).

$$\hat{\mathbf{Y}} = f_{model\ forecast}(\mathbf{X}) \quad (3)$$

En este trabajo, se ha elegido para la función de coste el error cuadrático medio (MSE). La ecuación (4) define dicho error con la norma 2 $\| \cdot \|$, donde $\mathbf{E}$ es el error (función de coste) y n el número de muestras de la media.

$$E = \frac{1}{n} \sum_x \| (y(x) - \hat{y}(x)) \|^2 \quad (4)$$

En ML, a partir de la función de coste, se implementa el proceso fundamental para el ajuste automático de los pesos del modelo, denominado *backpropagation* (BP) de acuerdo a Goodfellow et al. ^[216] (2016). El cual calcula el gradiente del error total o función de coste (en este caso a través de la métrica MSE definida la ecuación 4) con respecto a cada uno de los pesos o conexiones de las neuronas

de cada capa del modelo. Para lo que aplica la regla de la cadena en la derivación, a fin de propagar dicho error hacia atrás del modelo (desde la capa de salida hasta la entrada), para repartir el error, según la sensibilidad (o derivada) del mismo a cada conexión, entre todos los pesos del modelo. O dicho de otra forma para obtener cuanto contribuye cada conexión o peso de modelo al error obtenido por la función de coste, para ajustar así el peso en cada iteración (o “epoch”) durante el entrenamiento. Con ello se busca que el nuevo ajuste de pesos del modelo produzca un error total (o función de coste) menor en cada iteración realizada con el lote de muestras del tensor **X** (datos de entrada). Si el problema converge, dicho error se va haciendo cada vez menor o asintótico, por lo que se podrá parar el entrenamiento cuando se haya obtenido un error aceptable, como se muestra en las figuras 7 y 8. El grupo de ecuaciones (5) muestra dicha regla de la cadena de derivación en las redes ANN según Aggarwal C. ^[203] (2018), Hassoun M. ^[217] (1995):

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \quad (5a)$$

$$\frac{\partial net_j}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \left(\sum_{k=1}^n w_{kj} o_k \right) = \frac{\partial}{\partial w_{ij}} w_{ij} o_i = o_i \quad (5b)$$

La búsqueda del mínimo de la función de coste se implementan a través de diferentes optimizadores o gradientes, como son el gradiente estocástico descendente (SGD), según Kingma D. & Ba J. ^[218] (2014). Estos optimizadores tratan, entre otras cosas, de evitar que la solución quede “atrapada” en alguno de los mínimos locales que presenta la función de coste cuando esta no es una función convexa. Para lo que se han implementado diferentes estrategias, al igual que con las diferentes funciones de coste propuestas, para buscar la mejor solución en términos de eficacia y eficiencia. En la presente tesis se ha usado uno de los optimizadores más eficientes, en el sentido de alta velocidad de convergencia y estabilidad robusta, conocido como estimación del momento adaptivo “ADAM” según Kingma D. & Ba J. ^[218] (2014). El cual es una versión mejorada del “RMSProp”, al implementar una velocidad (o paso de aprendizaje) variable o

adaptativa y un momento o inercia suficiente para evitar quedar atrapado en los mínimos locales de la función de coste por los que pueda pasar. El grupo de ecuaciones (6) que describen la actualización de los pesos con el optimizador ADAM según Kingma D. & Ba J. ^[218] (2014) son:

$$m_w^{(t+1)} \leftarrow \beta_1 m_w + (1 - \beta_1) \nabla_w l^{(t)} \quad (6a)$$

$$v_w^{(t+1)} \leftarrow \beta_2 v_w^{(t)} + (1 - \beta_2) (\nabla_w l^{(t)})^2 \quad (6b)$$

$$\hat{m}_w = \frac{m_w^{(t+1)}}{1 + \beta_1^{(t+1)}} \quad (6c)$$

$$\hat{v}_w = \frac{v_w^{(t+1)}}{1 + \beta_2^{(t+1)}} \quad (6d)$$

$$w^{(t+1)} \leftarrow w^t - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_w + \epsilon}} \quad (6e)$$

Donde $w^{(t+1)}$ representan los pesos en la iteración $t+1$, η la velocidad o paso de aprendizaje, $\nabla_w l^{(t)}$ representa el gradiente de la función de coste l con respecto a los pesos w , ϵ es un valor mínimo para evitar dividir por 0 (en las librerías de *keras* se usa el valor de 10^{-7}), β_1 y β_2 son los factores de olvido para el gradiente o primer momento y el segundo momento respectivamente, $m_w^{(t+1)}$ es la media móvil de los pesos o primeros momentos en la iteración $t+1$, $v_w^{(t+1)}$ es la varianza móvil (o segundo momento) de los pesos en la iteración $t+1$. Hay que señalar que en la notación de las ecuaciones (6) los superíndices con la variable t corresponden a las iteraciones (o ciclos de aprendizaje), mientras que la notación con los subíndices t del resto de las ecuaciones mostradas hasta ahora aquí, corresponden a los diferentes instantes de tiempo de la serie temporal.

Cuando se acaban los lotes de parejas de tensores de entreno $\mathbf{X}$ input e $\mathbf{Y}$ output se repite el ciclo un número determinado de veces (para el primer caso de estudio de la demanda de carga aérea se usaron 400 ciclos completos del dataset), hasta que se obtiene un error relativo pequeño correspondiente un nivel de ajuste de pesos del modelo aceptable. En las Figuras 7 y 8 se muestra a modo de ejemplo,

tomado del primer caso de estudio, la evolución del parámetro MSE relativo a la función de coste durante las sucesivas iteraciones de entreno de todos los lotes, tanto para las parejas de entreno como para las de test y tanto para la función de coste MSE como para la métrica de evaluación RMSE respectivamente. Nótese que las muestras fueron normalizadas y estandarizadas por lo que sus valores no están referidos a la escala real de las muestras. Observar cómo cada vez que finaliza un ciclo o iteración de todos los lotes se calcula tanto el error de las muestras de entreno como las de validación, de manera que se obtiene una métrica de validación del espacio muestral que es independiente del espacio muestral de las parejas de entreno.

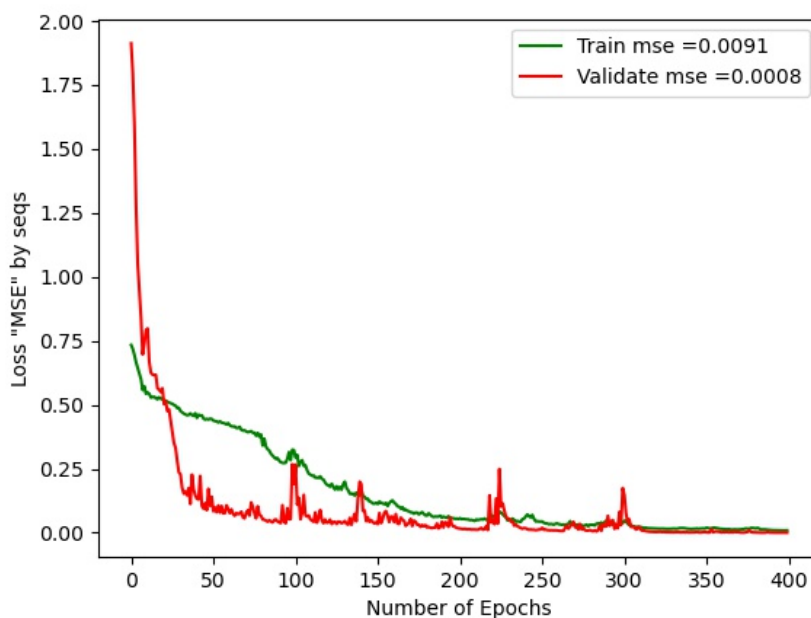


Fig. 7. Evolución de la función de coste MSE durante el entreno o número de ciclos (epoch) tanto para las muestras de entreno como las de validación. Fuente: Autor.

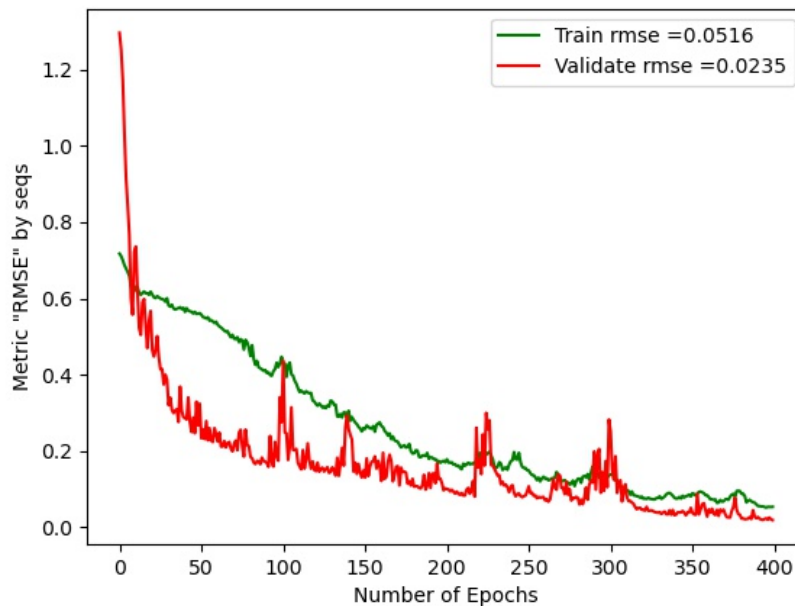


Fig. 8. Evolución de la métrica RMSE durante el entreno o repetición de los ciclos tanto para las muestras de entreno como las de validación. Fuente: Autor.

6.1.5 Hiperparametrización y Métricas Evaluación Modelo

En esta tesis se ha hiperparametrizado tanto el número de iteraciones como el tamaño del lote en el aprendizaje, para como se ha venido repitiendo obtener una optimización de los resultados. Un lote es una unidad del conjunto de datos a partir del cual se estima el error de la función de coste, y por tanto se ajustan los pesos. Estos gradientes se denominan estocásticos, porque el error se calcula a partir de los lotes del dataset, sin esperar a entrenar el dataset completo en cada “epoch”, debido a los elevados tiempos de ejecución y memoria de computación que dicho proceso total requeriría. Lo que convierte al proceso de aprendizaje en sensible al tamaño del lote elegido, por lo que fue hiperparametrizado este valor en el algoritmo para determinar su tamaño óptimo de acuerdo con Pedrycz W. & Chen S. ^[52] (2020), Calin O. ^[53] (2020), y Yi Z. & Tan K. ^[215] (2004).

El siguiente paso en la metodología es la evaluación del modelo, mediante el cual se pretende evaluar la bondad del entrenamiento por comparación de los resultados bajo los diferentes escenarios o configuraciones de los hiperparámetros

descritos. Por lo que el mejor resultado en la predicción del modelo no solo depende de la arquitectura del modelo elegido y de la calidad y cantidad del dataset (factores ampliamente comentados aquí), sino que también sus resultados son sensibles a la configuración de los hiperparámetros empleados en su entrenamiento. En el campo del ML/DL la búsqueda de la mejor solución, se convierte en un problema heurístico, en el sentido de que se proponen y prueban diferentes soluciones de arquitecturas, configuraciones y procesos de entreno, extraídos del espacio infinito de soluciones posibles, para, a continuación, medir sus actuaciones y poder seleccionar aquellos modelos y configuraciones más apropiados según Aggarwal C. ^[203] (2018), Patterson J. & Gibson A. ^[219] (2017). Por este motivo, en esta fase de evaluación se requiere, de un lado definir el espacio de los hiperparámetros con los que configurar el modelo y su entrenamiento, y por otro lado definir las métricas con las que se evalúa y compara finalmente la bondad de los resultados del modelo. La métrica de evaluación elegida para estos problemas de regresión es la del error porcentual absoluto medio (MAPE), muy usada en este tipo de problemas, el cual mide el error de predicción como un porcentaje de acuerdo con Rodríguez Y. et al. ^[8] (2020), Kim S. & Kim H. ^[220] (2016), y Ren L. & Glasure Y. ^[221] (2009). Lo cual constituye una manera intuitiva de interpretar y evaluar el error del modelo, pues cuanto más pequeño sea el MAPE mejor será el pronóstico. Adicionalmente también se implementó la raíz del error cuadrático medio (RMSE). Para ello, cada vez que se entrena o itera el modelo durante el aprendizaje, es decir cada vez que quedan fijados los pesos de todas las conexiones de las neuronas/capas del modelo, se procede a predecir de nuevo los valores de salida **Y** del conjunto de datos, a partir de los valores de entrada **X**. Estos gráficos están recogidos en las figuras 7 y 8 y ya han sido comentados.

Los hiperparámetros manejados para optimizar la arquitectura del modelo ConvLSTM2D han sido, entre otros: el número de filtros por capa convolucional-recurrente, el tamaño de las ventanas deslizantes de la convolución, el número de veces que se repite el codificador de capas ConvLSTM2D, la presencia (o no) de capas específicas del tipo *dropouts* y *batchnormalization* (que ayudan a evitar los sobreajustes y entrenar de forma robusta), y finalmente el número de neuronas de

la capa LSTM según Donahue J. et al. ^[211] (2015). Con todo ello se ha obtenido, finalmente, un modelo relativamente robusto en sus prestaciones.

Los hiperparámetros relativos al proceso de entreno y partición de las secuencias previas n_lags para predecir secuencias futuras n_outs , tras varias pruebas con diferentes valores, se han centrado en dos escenarios posibles: el primero de 6 años (corto plazo) y el segundo a 12 años (medio plazo), con los dos hiperparámetros iguales entre sí ($n_lags = n_outs$). Al implementar las técnicas “callbacks” de Keras como “Earlystopping” y “ModelCheckpoint”, relativas a detener el aprendizaje cuando se alcanza el error umbral y guardar los pesos del modelo óptimo hallados por el algoritmo, estas técnicas han sido el método elegido de búsqueda del mejor resultado de entreno, y no el del número de iteraciones usadas, hiperparámetro denominado “epoch”. Finalmente, los hiperparámetros que si se han tenido en cuenta han sido el del número de subsecuencias (dentro de n_lags), el número de “features” usados como inputs, para predecir los pasajeros nacionales o los pasajeros internacionales para el segundo caso de estudio, y de la carga domestica o internacional para el primer caso de estudio, junto a los hiperparámetros ya mencionados de pasos previos n_lags y pasos futuros a predecir n_outs .

Finalmente, y como última etapa de la metodología, se desarrolla la predicción del modelo, que consiste en realizar, con el modelo ya ajustado, las predicciones de las series futuras (n_outs) a partir de la serie de los años previos (n_lags) elegidos y de los “features” elegidos. Para lo que se consideró conveniente aplicar dicha predicción a la serie histórica (real) Y , y poder así observar la desviación con las salidas estimadas $\hat{Y}$ o predichas por el modelo *ConvLSTM2D*, dada la escasez de datos. Este proceso de predicción es en definitiva el objetivo final de esta metodología, dado que lo que se busca con este desarrollo del algoritmo es realizar la predicción de la demanda de pasajeros aéreos o de carga aérea para años futuros. Aunque el objetivo se ha centrado en dos de los “features”, pasajeros nacionales y pasajeros internacional (o de carga aérea nacional o internacional), el modelo podría haber sido usado para predecir en los años futuros, cualquiera de los otros “features” adicionales incluidos en el conjunto de datos histórico (p.e. PIB,

IPI, IPC, TRM, población, etc.). Hay que señalar que los pesos y funciones representativas del modelo, para cada una de las posibles combinaciones entre cualquier número de los features (los socioeconómicos + los dos aeronáuticos) posibles de salida en función de los “features” de entrada, sería diferente, puesto que el entrenamiento $Y = f_{model\ forecast}(X)$, o ajuste de pesos del modelo ConvLSTM2D, se hace con los “features” de los tensores apropiados X de entrada e Y de salida extraídos del conjunto de datos históricos a predecir.

6.2 Tercer Caso de estudio Demora Salidas Aviones Comerciales

6.2.1 Dataset

Para el tercer caso de estudio se ha dispuesto de una base datos de vuelos de salidas (“departure”) comerciales programadas, domésticos e internacionales, de la red aeroportuaria colombiana completa (58 aeropuertos) para todo el año 2018, lo que constituye un conjunto de 357.595 vuelos con un total de 5.721.520 datos, al contener cada registro de vuelo 16 “features” o atributos asociados. Cada vuelo de salida contiene los siguientes *features*: a) tipo de tráfico (doméstico o internacional), b) identificación aerolínea, c) aeropuerto de origen (siempre nacional), d) aeropuerto de destino (nacional o internacional), e) número de vuelo (vinculado a la aerolínea, ruta y frecuencia de dicho vuelo), f) fecha programada de salida, g) hora programada de salida, h) fecha de remolque (push-back), i) hora de remolque (push-back), j) la demora horaria, k) la demora expresada en minutos, l) la situación del vuelo, m) el estado de la salida (según estándar de la autoridad aeroportuaria del país), n) el código de la demora (según el Standard IATA Delay Codes-AHM 730 de IATA ^[16] 2021b), donde cada código (de los 89 posibles) se asocia con un motivo que puede provocar la demora, o) motivo de la demora, p) observaciones realizadas por la autoridad aeroportuaria asociadas al vuelo y/o su demora.

6.2.2 Tipología de Modelos ML/DL

Como se ha comentado los modelos de ML/DL representan matemáticamente unos “aproximadores universales” de acuerdo con Guss W. & Salakhutdinov R. ^[222] (2019), quienes abordaron la dimensionalidad necesaria (infinita o finita en su

aproximación) del mapa de la función f que cumple $Y = f(X)$. Para este tercer caso de estudio, se analizan diez modelos ML/DL que mapean la salida Y (tiempo de la demora) con los inputs X (atributos del vuelo que pueden ser hasta nueve en su versión más extensa del dataset una vez filtrados de los 16 iniciales), al objeto de comparar las prestaciones de los modelos. Dentro de estos diez modelos solo hay un modelo DL, el denominado Multi Layer Perceptrón (MLP) de acuerdo con Vang-Mata R. ^[200] (2020).

El modelo MLP tipo DL, cuya arquitectura se muestra en la figura 9, ha sido generado usando las librerías keras/tensorflow, y consiste en dos capas densas (conexiones de todas las neuronas de la capa anterior con todas las neuronas capa posterior) de 9 neuronas (obsérvese que este número debe coincidir con el número de *features* de los inputs) en la primera capa y 36 neuronas en la segunda, usando la función de activación ReLU (Rectified Linear Unit) de acuerdo a Agarap A. ^[223] (2019), Arora S. et al. ^[224] (2018), Hara et al. ^[225] (2015), y Arora S. & Barak ^[226] (2009). De esta forma la función de activación incorpora las no linealidades que pueda contener el mapeo entre la salida y la entrada al modelo, ya que las conexiones entre las neuronas son siempre de tipo lineal (peso más sesgo), como se aprecia en la figura 3. La función ReLU se define como: $y = \max(0, x)$, donde “ y ” representa la salida de la neurona y “ x ” la entrada de la neurona. La función ReLU es una de las más usadas actualmente al presentar múltiples ventajas: su computación es fácil y por tanto los tiempos de entreno son cortos, converge rápidamente sin presentar saturaciones (problema que presentan otras funciones de activación como la sigmoide y la tangente hiperbólica) por tener una salida lineal, además ReLU no presentan los problemas de desvanecimiento del gradiente (que hace que no se actualicen los pesos del modelo por ser cuasi-nulas las derivadas de sensibilidad del error a los pesos sobre todo en las primeras capas del modelo), estas funciones se activan de forma escasa (“sparsity”) lo que simplifica las conectividades o acoplamientos existentes entre todas neuronas, y finalmente, permite que el modelo ML/DL aprenda las relaciones no lineales entre la entrada x y la salida y .

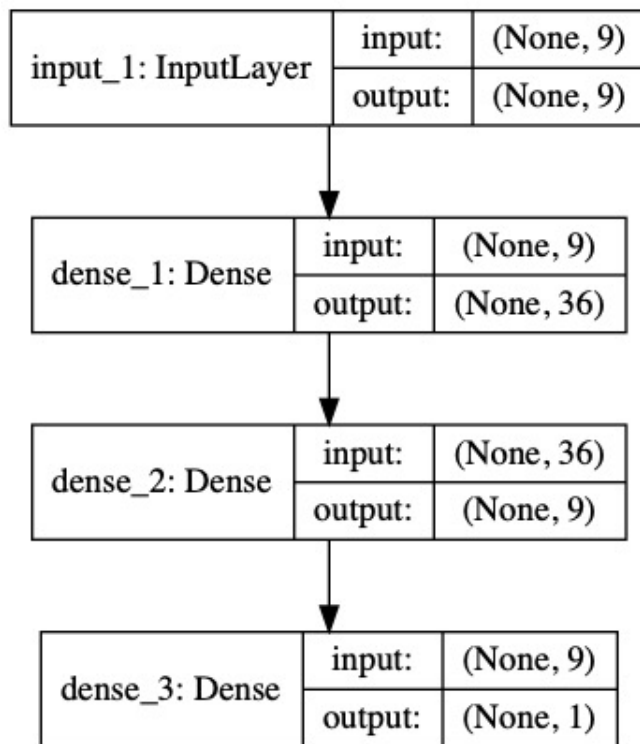


Fig. 9. Arquitectura tipo de capas y neuronas del modelo MLP desarrollado. Fuente: Autor.

El resto de los 9 modelos configurados en el presente trabajo son modelos clásicos que no son del tipo DL, siendo los siguientes: el modelo Support Vector Machine (SVR) que es muy robusto tanto en predicciones lineales como en no lineales, lo que se consigue aumentando el espacio dimensional de los “*features*” según Bao Y. et al.^[44] (2012), el modelo Random Forest (RFR) que son modelos tipo “*ensemble*” que mejoran los resultados de los algoritmos simples que los constituyen al ponderar o combinar las salidas de estos según Schonlau M & Zou R.^[227] (2020), y Breiman L.^[87] (2001), el modelo de Regresión Lineal Simple (LIR) muy sencillo de usar según Montgomery D. et al.^[228] (2012), el modelo Bagging Gradient (BGR) que mejora las predicciones produciendo multigrupos por combinaciones y repeticiones de acuerdo con Rokach L.^[229] (2019), Ganjisaffar Y. et al.^[230], 2011, y Friedman J.^[231] (2000), el modelo Extreme Gradient Boosting (XGR), el cual a partir de modelos más “débiles” generan modelos más fuertes usando el gradiente descendiente como método optimizador de los mismos, a la

vez que presenta una respuesta robusta cuando se entrena con datasets desbalanceados (que son aquellos que presentan una frecuencias más altas para determinados valores) de acuerdo a Hanif I. ^[232] (2019), y Chen T. & Guestrin C. ^[233] (2016), el modelo Huber Regressor (HBR) que es muy robusto entrenando los datos con “outliers” (o valores atípicos o extremos) al penalizarlos a través de funciones de coste *específicas* de acuerdo con Wang L. et al. ^[234] (2022), y Sun Q et al. ^[235], (2018), el modelo Extra Trees (ETR) que implementa más árboles de decisión para ajustar mejor el aprendizaje de acuerdo a Camana M. et al. ^[236] (2020) y Geurts et al. ^[237] (2006), el modelo de “*Decision Trees Regressor*” (DTR) que es el modelo simple de árboles decisión que se ha considerado conveniente retener por la capacidad auto-explicativa de los patrones aprendidos y para disponer de un modelo básico de referencia con el que comparar las prestaciones del resto según Izza Y. et al. ^[238] (2020), el modelo *k- Nearest Neighbors* (KNR) que predice en función de las predicciones que realizan sus “*k*” vecinos más próximos de acuerdo a Azadkia M. ^[239] (2020). Se hace notar que la “R” al final del nombre de todos los modelos, denota que los modelos se han configurado para predicciones en Regression (donde la salida es un valor numérico continuo).

En los diferentes escenarios implementados para el tercer caso de estudio, según diferentes selecciones de paquetes de *feature*, el modelo que presenta mejores prestaciones en términos de menor RMSE (Root-Mean-Square Error), que es la métrica elegida para comparar las prestaciones, es el Random Forest; estos modelos presentan un mejor rendimiento que, por ejemplo, los modelos MLP de redes neuronales artificiales tipo DL (profundas) debido a que se estima que o bien no se ha llegado a optimizar la arquitectura del modelo MLP aquí implementada (en términos de cantidad y tipo de capas de neuronas, así como del número de neuronas presentes en cada capa), y porque los modelos densos y profundos como los MLP, son más propensos a presentar sobreajustes (al tener más parámetros capaces de memorizar el ruido y sesgo específico del pequeño dataset con el que se entrena), y además son más proclives a presentar los problemas de desvanecimiento del gradiente (por ser más profundos). Otra explicación de por qué pueden dar peores prestaciones puede radicar en que los otros modelos que no son Deep Learning simplifican el análisis, lo que los hace más robustos y propensos

a generalizar para este tipo de problemas. Los modelos más simples generalizan mejor al ser menos sensibles al ruido de los inputs, matices que no tratan de captar, por su naturaleza simple, lo que sirve para suavizar la función de mapeo buscada, frente a los modelos más complejos que por tratar de retener más detalles e información son más propensos a perder la capacidad de generalización, lo que a su vez convierte a los modelos complejos en fuertemente dependientes de los datasets elegidos para el entrenamiento. En otras palabras, los modelos más complejos suelen tener más precisión, pero son más sensibles a la varianza del conjunto de datos elegido para su entrenamiento. En cambio, los modelos más simples suelen presentar peor precisión, pero al generalizar mejor presentan menor varianza (sensibilidad) a la elección del dataset de entrenamiento. En resumen, un modelo complejo como el MLP que es del tipo DL, puede estar requiriendo una modificación en su arquitectura, añadiendo capas convolucionales, o capas recurrentes tipo LSTM, o bien necesitar un ajuste en el número y tipo de capas y neuronas que lo conforman, junto al empleo de dataset más grandes y complejos, en el sentido de que haya más registros de vuelo y estos tengan más atributos para poder desarrollar toda su potencialidad.

6.2.3 Entrenamiento Modelos

El entrenamiento de los modelos forma parte del proceso central del ML. En esta tesis se han entrenado separadamente cada uno de los 10 modelos ML/DL con el mismo dataset y bajo dos versiones o escenarios diferentes; el primer escenario usando 10 *features* o inputs y el segundo bajo un escenario reducido con tan solo 6 *features* (con el objetivo de valorar si se pueden simplificar los atributos de los vuelos para ganar en simplicidad, velocidad ejecución y de paso seleccionar los atributos más sensibles para la predicción). Otra aportación reseñable de esta investigación, es haber realizado los entrenos estadísticamente, mediante la implementación de las técnicas de “*k-fold cross-validation*”, para asegurar que no existen sesgos en el entrenamiento por haber elegido un tipo particular de partición muestral de unos datos para entrenamiento y otros para validación/test. El entrenamiento se ha implementado dividiendo el dataset en lotes de 128 muestras cada uno, por lo que al entrenar el modelo ML/DL con cada uno de los lotes, procesando los datos hacia adelante (desde los inputs hasta la capa de salida) se predice o estiman los

valores de salida, momento en el que se vuelve hacia atrás (desde la capa salida hasta la capa de entrada) para actualizar los pesos (conexiones entre neuronas) de todas las neuronas de todas las capas del modelo. Este proceso de actualización hacia atrás es el comentado método de “Backpropagation” (BP), que ha sido suficientemente descrito ya en el apartado 6.1, en el que se han analizado los dos primeros casos de estudio correspondientes al pronóstico de la demanda de carga aérea y de pasajeros para Colombia.

Indicar que para este tercer caso de estudio la función de coste elegida ha sido la de la métrica RMSE (en vez del MSE de los dos casos de estudio anteriores). La ecuación (7) define dicho error RMSE con la norma 2 $\| \cdot \|$, donde E es el error y n el número de muestras del dataset, siendo “ $y(x)$ ” la salida real (proporcionada por el dataset en el aprendizaje supervisado) mientras que “ $\hat{y}(x)$ ” es la salida estimada por el modelo.

$$E = \sqrt{\frac{1}{n} \sum_x \|y(x) - \hat{y}(x)\|^2} \quad (7)$$

El resto de los procesos aplicados al único modelo MLP han sido suficientemente descritos anteriormente para los dos primeros casos de estudio, por lo que no se repiten aquí. En cambio, conviene señalar que los modelos basados en arboles de decisión el ajuste se hace, principalmente, por métodos de inducción y poda (“*pruning*”) de los árboles decisión según Jiang X. et al. ^[240] (2017), y Kulkarni V. & Sinha P. ^[241] (2012). El proceso de aprendizaje del dataset, se ha repetido 150 veces (iteraciones) de todo el dataset, para el modelo MLP, lo que se ha considerado como suficiente para alcanzar un valor estable o asintótico del RMSE. En el caso del resto de los modelos no DL el proceso de entreno se detiene cuando se obtiene un error aceptable por debajo del umbral marcado. Estos otros modelos ML, pero no DL se caracterizan por estimar la predicción mediante propiedades del conjunto, por acelerador, apilamiento, etc. que en inglés se denominan tipo “*ensemble*” o “*boosting*” o “*stacking*” o “*bagging*”. Es decir, dichos

modelos ponderan y mejoran las predicciones del conjunto, mediante las apropiadas combinaciones de las predicciones de los modelos más simples que los componen de acuerdo con Rokach L. ^[229] (2019).

Para darle más robustez (u objetividad) al valor de las métricas obtenidas con el RMSE, como ya se ha comentado, se han empleado técnicas “*k-folds cross-validation*” de acuerdo con Pedrycz W. & Chen S. ^[52] (2020), y Chang C. & Lin C. ^[88] (2011). En concreto el dataset se ha dividido en tres grupos, o tres formas diferentes de separar los datos de entreno y valida/test. Los datos de entrenamiento, como su nombre indica, se usan para entrenar el modelo y los de valida/test para evaluar las prestaciones del aprendizaje del modelo, evitando en todo momento que el modelo vea los datos de valida/test durante el aprendizaje, para que la evaluación sea objetiva. Además, con las técnicas de “*cross-validation*” se repiten estos entrenos un número variable de veces. Con todo ello se caracteriza el RMSE obtenido por cada modelo ML/DL, mediante una media y una desviación típica, constituyendo una forma robusta de evaluar las prestaciones de los modelos, frente a los otros casos cuando se usa una sola partición en grupos para entreno y test. Hay que enfatizar que todos estos métodos son de naturaleza estocástica, tanto por el uso del gradiente descendente estocástico SGD implementado en el aprendizaje, como por la propia naturaleza estocástica de las particiones del dataset en grupos de entreno y valuación, por lo que sus resultados no son determinísticos, sino que presentan una distribución de probabilidad (caracterizadas por su media y desviación típica).

6.2.4 Tests Estadístico de *Features*

Antes del entreno de los datos, para este tercer caso de estudio, se aplicaron diferentes controles estadísticos previos, pero solo al *feature* que es de tipo numérico, es decir solo al atributo del tiempo de la demora del vuelo. Para ello se ha usado una batería de métodos para la caracterización estadística del *feature* demora, al objeto de determinar si la distribución de sus valores es de tipo gaussiano; lo que se ha implementado a través de los tests de Shapiro-Wilk, D'Agostino y de Anderson-Darling, o si son distribuciones independientes que no están correlacionadas con la salida; lo que se realiza con los tests de Pearson lineal,

Spearman Rank, Kendall Rank, y Chi-Cuadrado. También se analizó si no presentan estacionalidad con el test ADFuller, y si presentan la misma distribución que la de salida con ayuda de los test de Student, Paired Student, Anova, Mann-Whitney U, Kruskal-Wallis H, y Friedman. Estos tests de caracterización estadística de los features nos indican, entre otras cosas, si se será fácil la extracción de patrones por parte del modelo por el tipo de distribución que siguen los features del dataset con el que se entrena. Adicionalmente, esta información estadística de los features permite seleccionar los métodos de pre-procesado más apropiados del dataset antes del entreno, mediante la aplicación de las apropiadas transformaciones, escalados, etc. que faciliten el entreno del modelo.

6.2.5 Eliminación *Outliers* Dataset

Igualmente, antes de entrenar los 10 modelos ML/DL implementados, se procedió a la eliminación de los datos “*outliers*” (o valores atípicos o extremos) del dataset, puesto que en caso de retenerlos los modelos pueden estar aprendiendo patrones marginales a partir de dichos atributos, lo que los predispone a tener una mala capacidad de generalización en la predicción. En esta tesis, se han utilizado cuatro métodos para identificar y eliminar los outliers, dos de los cuales son en sí métodos ML, pero no supervisados, como son los métodos “*Local Outlier Factor*” y “*One Class SVM*”, los cuales trabajan agrupando en clúster los datos normales y marcando como datos atípicos aquellos que quedan fuera de esos perímetros de valores normales. El tercer método es el de “*Isolation Forest*” que aplica teoría de árboles de decisión binarios que no requiere pre-agrupar en perfiles de normalidad, por lo que es más rápido y eficiente en la detección de los outliers. El cuarto método es el de “*Elliptic Envelope*” de acuerdo a Rousseeuw P. & Van Driessen K. ^[242] (1999), el cual calcula los outliers por el método del determinante de la covarianza mínima, el cual requiere para su aplicación que el número de muestras (o vuelos) sea mayor que el cuadrado de los features, lo cual se cumple en este caso.

6.2.6 Determinación Importancia *Features*

Otra aportación de esta tesis en el estudio del tercer caso relativo a la predicción de las demoras de las salidas de los vuelos comerciales, ha consistido en la determinación de la importancia que tienen los atributos del vuelo en la

determinación del tiempo de demora de cada vuelo. Para ello se ha implementado el método de “*Permutation Importance*” (PI) según Altmann A. et al. ^[86] (2010), y Breiman L. ^[87] (2001), consistente en aplicar técnicas de inspección de los features del problema para estimadores que sean entrenables. Donde se observa que cada uno de los 10 modelos propuestos ofrece su propia selección de los features más significativos de los vuelos para fijar el tiempo de la demora, como se muestra en los resultados del apartado 7. Lo que lleva a pensar que para la selección de los atributos más relevantes, se podría realizar una ponderación de los features predichos por cada modelo, análoga a las técnicas de “*bagging*”, “*boosting*”, “*stacking*”, o “*ensemble*”, donde unos ponderan por igual a todos los modelos para obtener la media, otros priman más a los modelo por su precisión, etc. al objeto de alcanzar el objetivo buscado (p.e. una menor varianza, o un menor sesgo o una mayor precisión). El método PI consiste básicamente en permutar aleatoriamente los feature entre sí y medir para cada caso cómo varía la salida (tiempo de demora) a través del entrenamiento del nuevo dataset permutado, para cada modelo considerado, lo cual permite establecer el coeficiente de importancia o sensibilidad del output (tiempo de demora) como función de cada feature j , denominado coeficiente c_j , de acuerdo con la ecuación (8).

$$c_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j} \quad (8)$$

Donde “ s ” es el resultado del entreno del modelo ML cuando se entrena con el dataset D sin permutar. A continuación, cada feature “ j ” asociado a una columna del dataset “ D ” se permuta aleatoriamente con las otros features para producir una versión permutada del dataset $\tilde{D}(j, k)$ donde el subíndice “ k ” representa las repeticiones con las que se han gestionado los features, siendo “ K ” el número máximo de permutaciones, y “ $s_{k,j}$ ” el nuevo resultado del entreno del modelo para cada feature “ j ” repetido “ k ” veces con el dataset permutado $\tilde{D}(j, k)$.

7. Resultados

A continuación, se detallan los resultados obtenidos por los algoritmos para los 3 casos de estudio relativos a la gestión de los parámetros del transporte aéreo aplicados al caso de Colombia.

7.1 Pronóstico Demanda de la Carga Aérea

Con el objetivo de encontrar los valores óptimos en la longitud de años de la secuencia a predecir y para valorar los “*features*” más importantes que definen el mejor pronóstico de las dos variables de carga aérea: doméstica (denominada DOM) y de carga aérea internacional (denominada INT), se realiza un estudio variacional de los hiperparámetros de configuración del modelo a fin de elegir y pronosticar la demanda para el mejor escenario posible.

7.1.1 Pronóstico Carga Aérea internacional

El primer análisis variacional implementado consiste en buscar en el espacio de longitudes de secuencias la correspondiente al número de años n_outs a predecir que sea el óptimo, en el sentido de que proporcionen el menor MAPE (métrica elegida para valorar el pronóstico) de la variable de carga aérea internacional (INT). Midiendo este MAPE en el último ciclo de la secuencia n_outs de la serie histórica. Para simplificar el problema se ha elegido n_outs (años futuros pronosticados) igual a n_lags (años previos), bajo un escenario de solo dos inputs (GDP y el propio INT), pero se han hiperparametrizado las subsecuencias anuales en las que se puede dividir cada secuencia de n_lags . La elección de los hiperparámetros de los ciclos y subciclos en los que se divide la serie histórica, es relevante como se verá, a la hora de facilitar al modelo la extracción de los patrones de comportamiento de las series históricas que configuran la predicción de INT. Como el proceso de aprendizaje de los modelos ML/DL es estocástico, se ha repetido 10 veces el entrenamiento bajo cada configuración hiperparamétrica del modelo, por lo que el resultado viene caracterizado en términos de media y desviación típica de los MAPE obtenido. En la tabla 2 se muestran los resultados de los MAPE como función de los diferentes n_outs y del número de subsecuencias

y número de años por cada subciclo. Hay que destacar que dichos valores MAPE están referidos a un dataset cuyos valores han sido escalados, normalizados y estandarizados y por tanto no se corresponden con la escala original, como se ha explicado en el apartado 6.1.2

n_lags years	n_Subseqs	subseq years	mean MAPE	std MAPE	Ranking
12	4	3	90.47	30.28	6th best
12	3	4	83.79	47.08	4th best
10	2	5	102.33	38.64	7th best
8	2	4	87.15	53.91	5th best
6	3	2	14.87	4.66	2nd best
6	2	3	8.35	2.51	1st best
4	2	2	16.49	21.36	3rd best

Tabla 2. Análisis variacional para encontrar las longitudes óptimas de las secuencias y subsecuencias que dan el menor error MAPE en la predicción de la carga aérea internacional. Fuente: Autor.

Un primer resultado a destacar (ver la tabla 2), es que el número n_lags más óptimo (igual al de n_lags), es de 6 años, dentro del cual se toman 2 subsecuencias de 3 años cada una. Este escenario mejora levemente al de 6 años cuando se toman 3 subsecuencias de 2 años cada una, lo que muestra la sensibilidad de la precisión del modelo *ConvLSTM2D* a la elección de los hiperparámetros relativos a la longitud de la secuencia (n_lags) y de la subsecuencia o subciclo ($n_subseqs$). Una explicación plausible de por qué resulta el MAPE mayor cuando se toman menos años, p.e. 4 años y 2 subciclos de 2 años cada uno lo que debería tener un menor error, puede radicar en que con estas longitudes menores de años puede que no sean lo suficientemente extensas para recoger los patrones temporales que aparecen cuando se toman ciclos más largos como el de 6 años (con 2 subciclos de 3 años cada uno). Finalmente, cuando estos ciclos son de mayor longitud como 8 o 12 años, los errores son mayores, porque por un lado puede haber solapamiento de ciclos de patrones, lo que dificulta su detección y porque obviamente a mayor cantidad de años a predecir mayor es el error. Este comportamiento recogido en la información que contienen los tensores de 5D de los inputs $\mathbf{X}$ y de 3D en las salida

Y, es un comportamiento singular, a destacar como aportación de esta tesis, si se compara con los métodos estadísticos, donde el error al ser acumulativo siempre es mayor para una mayor cantidad de años a predecir, dado que no segmentan sus análisis en secuencias y subsecuencias, como si hace el modelo *ConvLSTM2D*.

El segundo análisis variacional implementado, denominado de sensibilidad a la importancia de los features, es determinar el número y tipo de features que más ayudan a predecir la demanda de carga aérea internacional a partir de los restantes 10 features, 9 de ellos sacados del contexto socioeconómico y el décimo tomado de la propia carga aérea doméstica (DOM) a predecir. Para ello, y dada la gran multiplicidad de combinaciones posibles de *features* para generar escenarios posibles, se han realizado solo algunas de ellas. P.e. se han tomado todas las posibles combinaciones de 2 features, además de considerar el escenario más simple de un solo feature. Las siguientes configuraciones proceden de implementar el pronóstico para; un escenario con 3 features, 2 escenarios con 5 features, un escenario de 10 features y un último escenario (o configuración) completo con los 11 features. Se insiste en la complejidad de búsqueda en el espacio de todas las posibles combinaciones de features, por lo que se han considerados solo unos pocos, y cuando se han considerado combinaciones de varios features se han elegido los que mejor pronóstico daban para el caso de referencia de combinaciones de 2 features. La mejor media del MAPE resultante al ejecutar cada escenario 10 veces, es el criterio elegido para seleccionar la combinación de features más óptima. Para ello se han fijado los n_lags y n_outs igual a 6 con 2 subsecuencias de 3 años cada una, obtenidos del análisis variacional previo. En la tabla 3 se muestran los resultados.

Input feature scenario	Number of features	Mean MAPE	std MAPE	Ranking
INT	1	13.24	3.84	5th best
GDP, INT	2	8.35	2.51	1st best
PCG, INT	2	9.72	7.18	2nd best
POP, INT	2	22.99	1.5	11th best
IPI, INT	2	23.94	1.76	12th best
CPI, INT	2	26.3	1.6	13th best

BMR, INT	2	44.72	4.43	15th best
FDI, INT	2	12.17	2.51	4th best
EXP, INT	2	20.96	12.52	9th best
IMP, INT	2	11.79	4.6	3rd best
DOM, INT	2	31.14	15.5	14th best
GDP, DOM, INT	3	51.66	17.58	16th best
GDP, EXP, IMP, DOM, INT	5	18,13	2,4	7th best
GDP, PCG, FDI, IMP, INT	5	15,47	1,82	6th best
GDP, PCG, POP, IPI, CPI, BMR, EXP, IMP, DOM, INT	10	21,5	6,96	10th best
GDP, PCG, POP, IPI, CPI, BMR, FDI, EXP, IMP, DOM, INT	11	18,25	3,99	8th best

Tabla 3. Análisis variacional para encontrar los mejores features, en términos de la mejor predicción medida con la métrica MAPE y sus combinaciones, para predecir la demanda de carga aérea internacional. Fuente: Autor.

La tabla 3 muestra unos resultados relevantes; como cuando solo se usa un feature (el mismo INT que se desea predecir) su MAPE (o error) es superior a cuando se emplean el feature INT junto a otros features. Esto quiere decir que, los otros features socioeconómicos de Colombia como cabía esperar influyen y moldean la serie aeronáutica a predecir. También muestra que cuando se utilizan todos los features disponibles (11 en total), la precisión de la predicción, (medida con la métrica MAPE) empeora, debido a que presumiblemente algunas de esas variables además de no influir en la variable de carga aérea internacional (INT), introducen ruido y complejidad no manejable, en los patrones de aprendizaje o de las correlaciones entre los features, lo que supone un empeoramiento de los resultados. Como se aprecia en la tabla 3, los mejores resultados predictivos los proporcionan las combinaciones de dos variables (incluido el que se desea predecir), destacando como pareja ganadora el feature GDP (PIB) con el que dicha variable aeronáutica INT se encuentra más fuertemente acoplado. Coincidiendo así el algoritmo, con la identificación de la importancia del GDP en el INT obtenida en varias investigaciones previas según Hakim M. & Merkert R. ^[192] (2016), Morrell P. ^[193] (2011), Ashford et al. ^[194] (2013), y Halpern N. & Graham A. ^[195] (2018). El GDP es seguido en importancia por el PCG (Per Capita GDP). El siguiente feature

en importancia para determinar la predicción de la variable INT son las importaciones (IMP). El feature FDI (Foreign Direct Investment) es la siguiente variable en importancia. Vuelve a llamar la atención, que cuando se combinan los 5 mejores features la predicción está por detrás de las mejores predicciones hechas por cada uno por separado. Lo cual se puede asociar a la complejidad que implica extraer los patrones cuando se correlacionan varios features a la vez, lo que estaría indicando los límites de predicción de los modelos *ConvLSTM2D* aquí implementados.

Para concluir este análisis, en la figura 10 se muestra la predicción temporal (a corto plazo) de la demanda de carga aérea internacional (INT en color naranja), en contraste con las curvas de las series históricas (reales) de demanda de la carga aérea internacional (INT) y de GDP (en color azul) como feature más relevante para la determinación de la demanda de carga aérea internacional. En color rojo se identifica la línea predictiva que hace el modelo *ConvLSTM2D* de la serie histórica INT completa, donde se aprecia la manera de cómo va mejorando el aprendizaje, ajustándose mejor en los últimos años de la serie histórica. En la curva de predicción se observa una recuperación importante, pero que no alcanza a recuperar el nivel existente en el periodo pre-pandémico, junto a la aparición de un nuevo ciclo que hará que vuelva a caer en el futuro inmediato la demanda de carga aérea internacional. Este escenario óptimo es el resultante del análisis variacional anterior, donde n_lags es igual a n_outs e igual a 6 años considerando 2 subsecuencias de 3 años cada una y para el caso de 2 features donde el feature más relevante para la carga aérea internacional es el GDP.

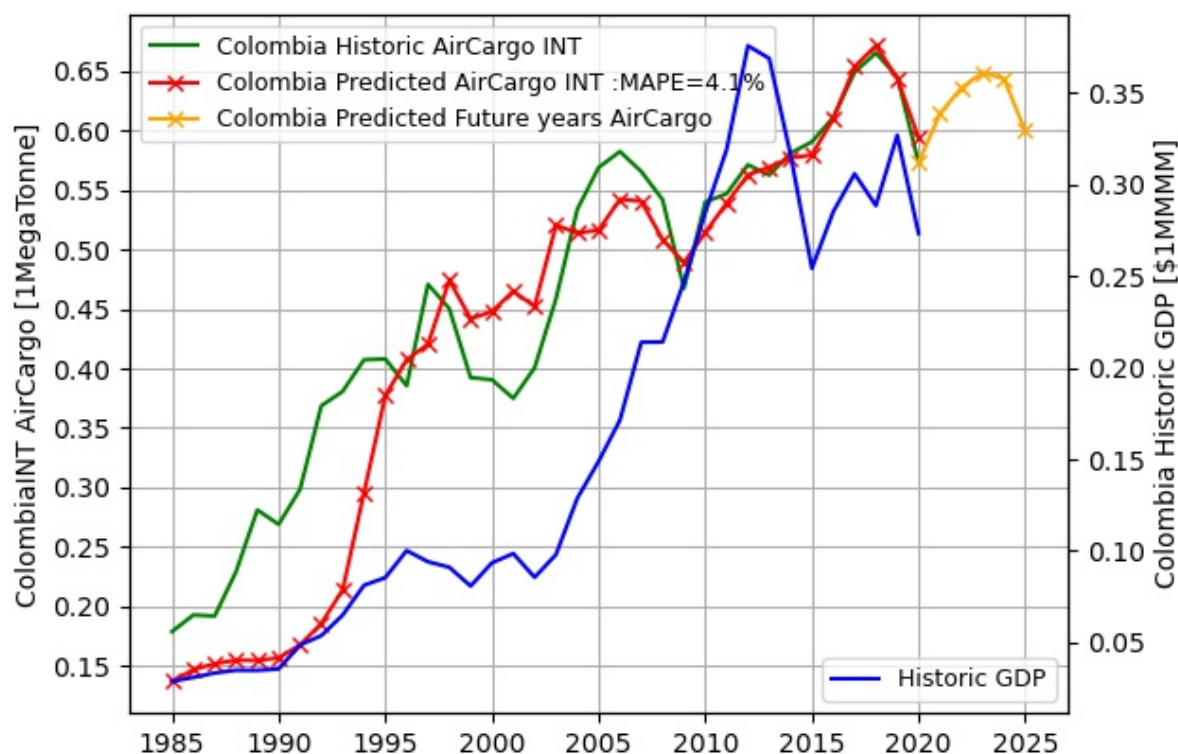


Fig. 10. Pronóstico de la demanda de carga aérea internacional (INT) en millones de toneladas, como función de la propia historia de dicha variable y de la evolución de la serie histórica del GDP (en US\$ $\times 10^{12}$). Fuente: Autor

7.1.2 Pronóstico Carga Aérea Doméstica

El proceso de análisis variacional implementado para el caso de la carga aérea doméstica (DOM) es equivalente al mostrado para la carga aérea internacional (INT), consistente en buscar primero en el espacio de las longitudes de secuencias la correspondiente al número de años n_{outs} a predecir que sean óptimo, en el sentido de que proporcionen el menor MAPE, medido este en el último ciclo de n_{outs} de la serie histórica. Para simplificar el problema se ha elegido n_{outs} (años futuros a predecir) igual a n_{lags} (años previos), y un escenario con solo dos inputs PCG y la propia demanda de carga aérea doméstica DOM, y se han hiperparametrizado las subsecuencias o subciclos en los que se puede dividir cada secuencia de n_{lags} . En la tabla 4 se muestran los resultados de los MAPE como

función de los diferentes n_outs y del número de subsecuencias. Hay que destacar que dichos valores MAPE están referidos a un dataset cuyos valores han sido escalados, normalizados y estandarizados y por tanto no se corresponden con la escala original.

n_lags years	n_Subseqs	subseq years	mean MAPE	std MAPE	Ranking
4	2	2	68.68	1.73	3rd best
8	2	4	65.83	3.46	2nd best
12	3	4	187.88	62.25	3rd best
12	4	3	209.31	55.05	4th best
6	2	3	42.05	2.14	1st best

Tabla 4. Análisis variacional para encontrar las longitudes óptimas de las secuencias y subsecuencias que dan el menor MAPE en la predicción de la demanda de carga aérea doméstica. Fuente: Autor

Un primer resultado relevante en las predicciones de los escenarios del DOM (mostrado en la tabla 4) es que los MAPE presentan una fuerte varianza (es cinco veces superior el valor mayor respecto al menor), esto puede ser consecuencia de la posible alta variabilidad o volatilidad intrínseca de la serie histórica de la variable carga aérea doméstica DOM. Lo que implicaría que el aprendizaje de los patrones de comportamiento en la carga aérea doméstica DOM puede ser un problema más complejo que el de la demanda de la carga aérea INT. O dicho de otra forma, la volatilidad o variaciones de los patrones colombianos intrínsecos (domésticos) son mucho mayores que los internacionales (externos), debido a que los primeros se encuentran más ligados a un entorno nacional colombiano más variable o volátil y los segundos al entorno internacional (aparentemente más estable o robusto que el entorno nacional colombiano). En la tabla 4 se puede apreciar que el número de n_outs , que es igual al de n_lags más óptimo, es también de 6 años, los cuales se toman en 2 subsecuencias de 3 años cada una (nótese que es igual al resultado obtenido para la INT). La explicación de por qué ha resultado un MAPE mayor cuando se toman menos años, p.e. 4 años y 2 subciclos de 2 años, parece ser la misma que la comentada para la carga INT. E igualmente parece aplicable el mismo argumentario encontrado para la carga INT cuando se pronostican más años (p.e.

8 o 12) a partir de secuencias y subsecuencias de más años de la serie histórica. Lo que es un comportamiento singular hallado por este algoritmo predictivo.

El segundo análisis variacional implementado, también denominado de sensibilidad a la importancia de los features, es determinar el número y tipo de features que mejor ayudan a predecir la demanda de carga aérea doméstica a partir de los restantes 10 features, 9 de ellos socioeconómicos recogidos de la serie histórica. Para ello y dada la gran multiplicidad de combinaciones posibles de features, se han analizado unos pocos escenarios, p.e. todas las posibles combinaciones de 2 features, además de considerar el escenario más simple de 1 feature. Adicionalmente se han implementado un escenario con 3 features, otro escenario con 4 features, y un último escenario con todos los features (11). El MAPE medio (con su desviación típica) obtenido al ejecutar cada escenario 10 veces ha sido el criterio elegido para seleccionar la combinación de features más óptima. Para ello se han fijado los n_lags y n_outs en 6 con 2 subsecuencias de 3 años, obtenidos del análisis variacional previo. En la tabla 5 se muestran los resultados.

Input feature scenario	Number of features	mean MAPE	std MAPE	Ranking
DOM	1	66.41	10.31	5th best
GDP, DOM	2	51.18	3.70	3rd best
PCG, DOM	2	42.05	2.14	1st best
POP, DOM	2	133.34	17.19	12th best
IPI, DOM	2	135.37	12.77	13th best
CPI, DOM	2	138.62	11.52	14th best
BMR, DOM	2	95.45	10.50	10th best
FDI, DOM	2	69.00	7.32	7th best
EXP, DOM	2	42.14	4.87	2nd best
IMP, DOM	2	71.47	16.55	8th best
INT, DOM	2	121.80	11.00	11th best
PCG, EXP, DOM	3	74.11	14.76	9th best
PCG, EXP, FDI, DOM	4	54.56	2.66	4th best

GDP, PCG, POP, IPI, CPI, BMR, FDI; EXP, IMP, INT, DOM	11	67.96	4.46	6th best
---	----	-------	------	----------

Tabla 5. Análisis variacional para encontrar los mejores features y sus combinaciones para predecir la demanda de carga aérea doméstica. Fuente: Autor.

Los resultados, mostrados en la tabla 5, muestran que cuando se usa un solo feature (el mismo DOM que se desea predecir) su MAPE es superior (quedando en quinto lugar del ranking) a cuando se emplean el feature DOM junto a otros features. Los resultados también muestran que cuando se utilizan todos los features disponibles (11), la predicción MAPE no es muy buena, debido a que algunas de esas variables además de no influir en la demanda de carga aérea doméstica, pueden estar introduciendo ruido/complejidad en las correlaciones de los features, que el modelo *ConvLSTM2D* no es capaz de manejar. Como se aprecia en la tabla 5, los mejores resultados predictivos lo proporcionan las combinaciones de dos variables (incluida la que se desea predecir). Siendo ganador el *feature* de la renta per cápita (PCG), con el que se encuentra más fuertemente vinculado la demanda de carga aérea doméstica, seguido del feature EXP (exportaciones). Lo que está indicando que las exportaciones en Colombia dependen en gran medida del transporte aéreo nacional de la carga. El siguiente feature en importancia, para determinar la predicción de la demanda de carga aérea doméstica, es el GDP. La cuarta configuración en importancia es la combinación de los 4 features PCG, EXP, FDI y DOM, lo que significa que el modelo encuentra patrones relevantes en la combinación de estos 4 features. El quinto feature en importancia es considerar solo a la serie (histórica) temporal de la demanda de carga aérea doméstica, lo que evidencia la aportación positiva de combinarlos con otros *features* frente a considerar solo la serie aislada DOM. Finalmente, hay que destacar que la demanda de carga aérea doméstica es sensible a incorporar más cantidad de features para predecir el desarrollo de la misma, frente al acoplamiento a otros features aislados como FDI (inversión exterior directa), IMP, IPI, CPI (o IPC en español), etc., lo cual resulta otro hallazgo diferencial de esta investigación relativa a la demanda de la carga aérea doméstica DOM respecto a las predicciones de la demanda de carga aérea internacional INT.

Finalmente, en la figura 11 se muestra la predicción temporal de la demanda de carga aérea doméstica a corto plazo (en color naranja), presentando una fuerte caída y recuperación, para volver a caer, en contraste con las curvas de las series históricas (reales) de la carga aérea doméstica (en color verde) y de PCG (en color azul) como *features* más relevantes. En color rojo se identifica la curva predictiva que hace el modelo *ConvLSTM2D* de la serie histórica DOM (donde se observa como esta curva va ajustándose mejor, conforme aumentan los años de entrenamiento).

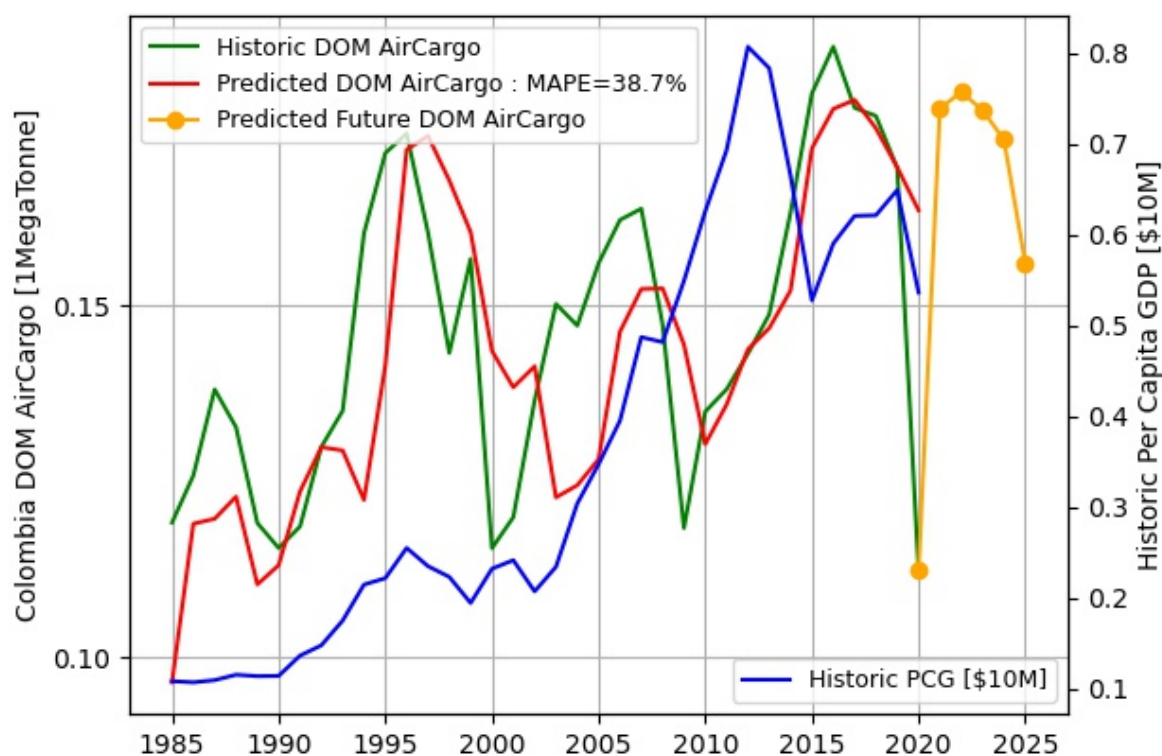


Fig. 11. Pronóstico de la demanda de carga aérea doméstica (DOM) (en millones de toneladas) como función de la serie histórica de la propia variable y de la evolución histórica del PCG (en US\$ $\times 10^4$). Fuente: Autor

7.2 Pronóstico Demanda de Pasajeros

A continuación, se presentan, interpretan y discuten los resultados obtenidos con el algoritmo definido, entrenado y evaluado, para la predicción de la demanda (a corto y medio plazo) de pasajeros nacionales (NAT) e internacionales (INT) registrados en el sistema aeroportuario colombiano. Los dos hiperparámetros considerados para definir los posibles escenarios de predicción son: de un lado, los relativos a los años previos y a los años de predicción, respectivamente denominados n_lags y n_outs , como en el caso de la carga aérea, y de otro lado los n_lags y n_outs se han subdividido a su vez en subsecuencias. Al igual que en la predicción de la carga aérea, los resultados del modelo se muestran muy sensibles a dichos hiperparámetros, dado que definen el contexto de ventanas de temporalidad desde las que observar y extraer los patrones de las variables con el tiempo. Esta sensibilidad es propia del análisis de las capas “ConvLSTM2D” y justifica las múltiples dimensiones del tensor de datos de entrada $\mathbf{X}$ (5D) y el de salida $\mathbf{Y}$ (3D). La salida $\mathbf{Y}$ tiene dos *features* como variables dependientes pero solo se usará una al mismo tiempo; pasajeros nacionales (NAT) o internacionales (INT), mientras que la entrada $\mathbf{X}$ contiene cinco *features* o dimensiones como variables independientes (PIB, Población, IPI, IPC, TRM), además de uno de los dos *features* a predecir, ya que los escenarios considerados siempre incorporan como input la propia historia temporal de la variable a predecir, con lo que se consigue darle mayor robustez y predictibilidad al modelo.

Antes de pasar a describir los hiperparámetros y resultados alcanzados en cada uno de los escenarios elegidos como más relevantes, se muestra la figura 12 correspondiente a la curva de aprendizaje (ajuste de los pesos) del modelo según el número de iteraciones o ciclos de aprendizaje (donde se observan la relativa robustez en la convergencia hacia valores asintóticos del MAPE con escasas fluctuaciones y en no muchas iteraciones).

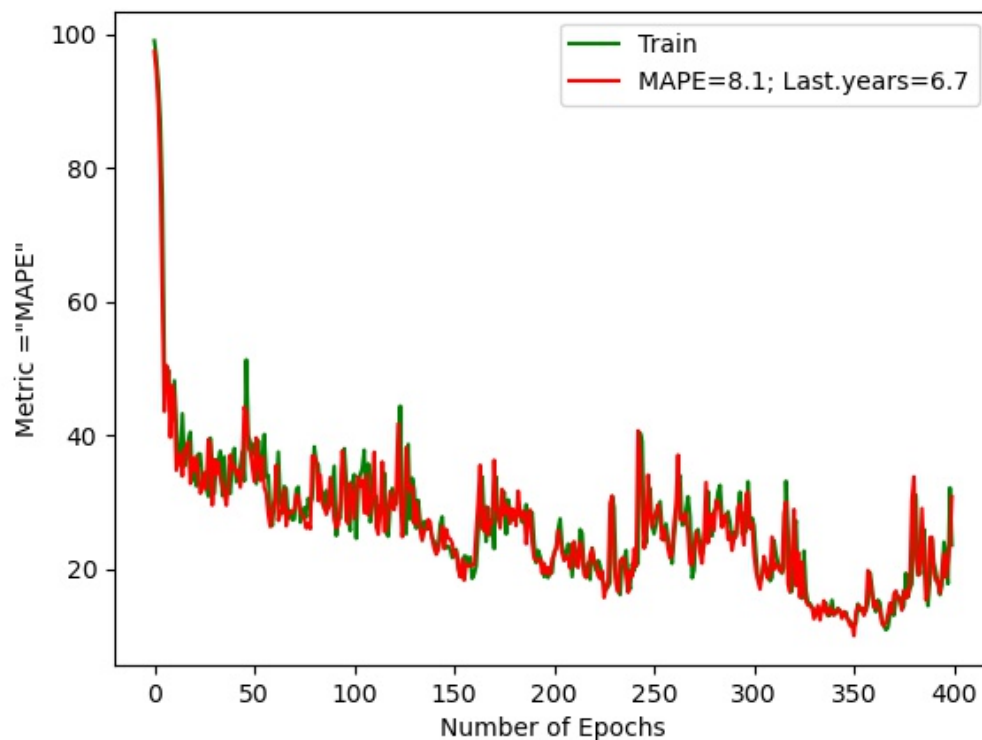


Fig. 12. Evolución del MAPE según la curva de aprendizaje del modelo entrenado descrito en el **escenario 1**. Fuente: Autor.

En la figura 13 se representa la predicción de la serie histórica con el modelo ya entrenado, donde se observa como la disminución del tráfico de pasajeros del año 2020 (en plena situación pandémica) no es predicha por el modelo cuando reproduce la serie histórica (otro aspecto es la predicción futura que hace razonablemente bien en la figura 14). Ambas curvas han sido tomadas como ejemplos del **escenario 1**, que se describe más adelante.

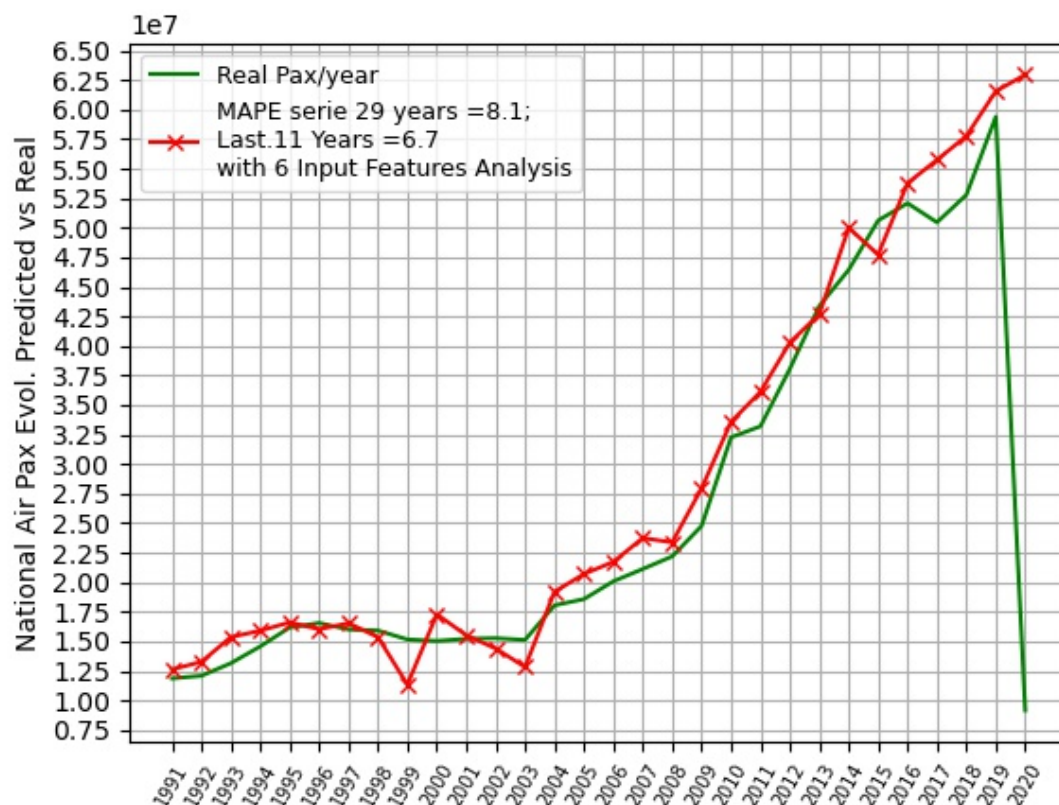


Fig. 13. Evolución de la predicción de los pasajeros de la serie histórica (en rojo) vs. la serie histórica real (verde), que debe iniciarse n_lags años después de 1979 del modelo ya entrenado y descrito en el **escenario 1**. Fuente: Autor.

Escenario 1. Se presenta una primera predicción considerando secuencias de 12 años previos al año t junto a los 12 años siguientes, sacados de la serie histórica total, considerando las series temporales de seis *features* elegidos como variables independientes (pasajeros nacionales, PIB, población, IPI, IPC y TRM) y con pasajeros nacionales como única variable de salida a predecir. Los 12 años previos n_lags se dividen en tres subsecuencias de 4 años cada uno. La figura 14 recoge dicha predicción a medio plazo con un valor del MAPE de 6.7%, medido en los últimos 11 años y de 8.1% para la serie histórica de 29 años (señalar que se comienza a contabilizar 12 años después de 1979, para permitir disponer de los 12 años previos de n_lags , a partir de los cuales predecir los 12 años siguientes n_outs). En la figura 14 se puede observar que para principios del año 2024 el

tráfico (o demanda) de pasajeros nacionales, recuperaría el nivel existente en época prepandémica (2019); con una recuperación de la tendencia en la demanda ligeramente superior al del año 2019 pero presentando oscilaciones que podrían estar indicando cierta inestabilidad en la predicción de los pasajeros nacionales.

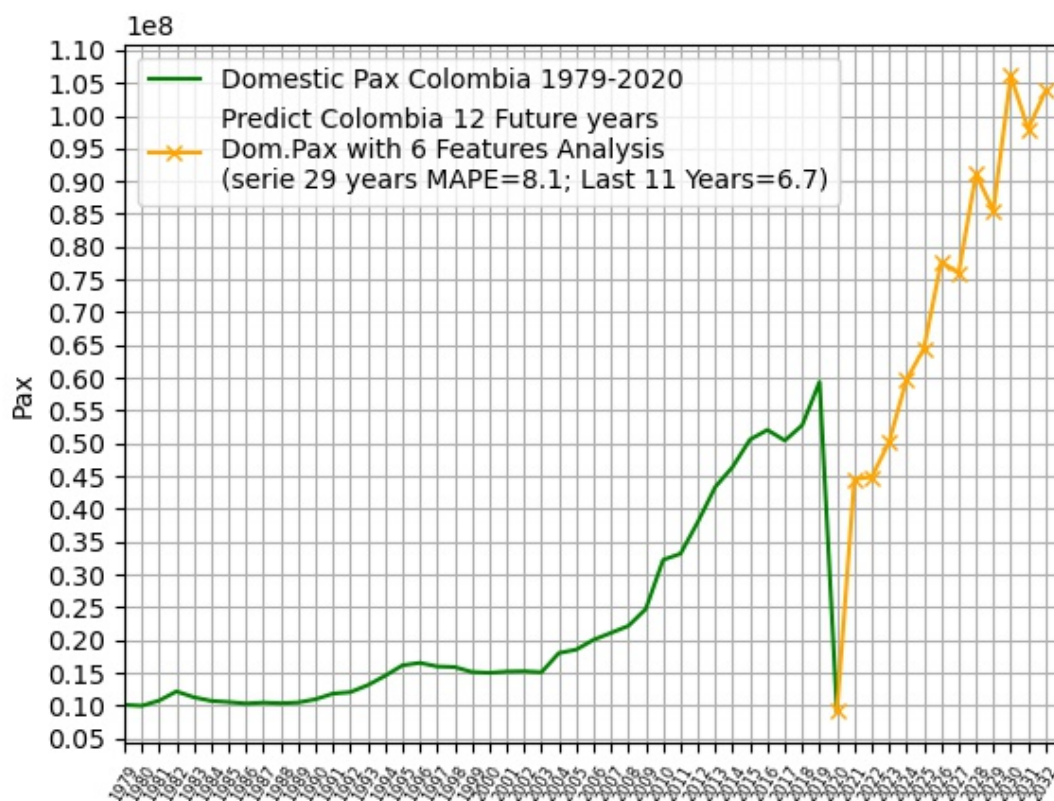


Fig.14. Pronóstico de la demanda de pasajeros nacionales de Colombia (naranja frente al verde de la serie histórica), para el periodo 2021-2032, correspondiente al **escenario 1**. Fuente: Autor.

Escenario 2. Se presenta una segunda predicción considerando secuencias de 12 años previos y de los 12 años siguientes, tomadas de la serie histórica total, pero considerando ahora las series temporales de solo dos *features* elegidos como variables independientes (pasajeros internacionales, PIB) y con pasajeros internacionales INT como única variable de salida a predecir. Agrupando los 12 años previos n_lags en tres subsecuencias (o subciclos posibles) de 4 años cada uno. La figura 15 muestra dicha predicción a medio plazo con un valor del MAPE de 9.3% para los últimos 11 años y de 9.7% para la serie histórica de 29 años

(recuérdese que se contabilizan 12 años después de 1979 como pasos previos n_lags para predecir los 12 años siguientes n_outs). En la figura 15 se puede observar que para principio del año 2024 el tráfico (o demanda) de pasajeros internacionales de Colombia se recuperaría al nivel existente en época prepandémica (2019), con una tendencia en la demanda muy similar a la del año 2019 y sin oscilaciones. La no presencia de fluctuaciones en la demanda de los pasajeros internacionales recuerda a la robustez encontrada en las tendencias de la demanda de carga aérea internacional frente a la nacional, lo que vuelve a indicar la presencia de mayores volatilidades, frente a una mayor robustez y estabilidad de las demandas de internacionales (pasajeros y carga aérea), lo que podría estar indicando mayor volatilidad de la economía nacional colombiana frente a los mercados internacionales. Esto es una aportación singular de este modelo *ConvLSTM2D*.

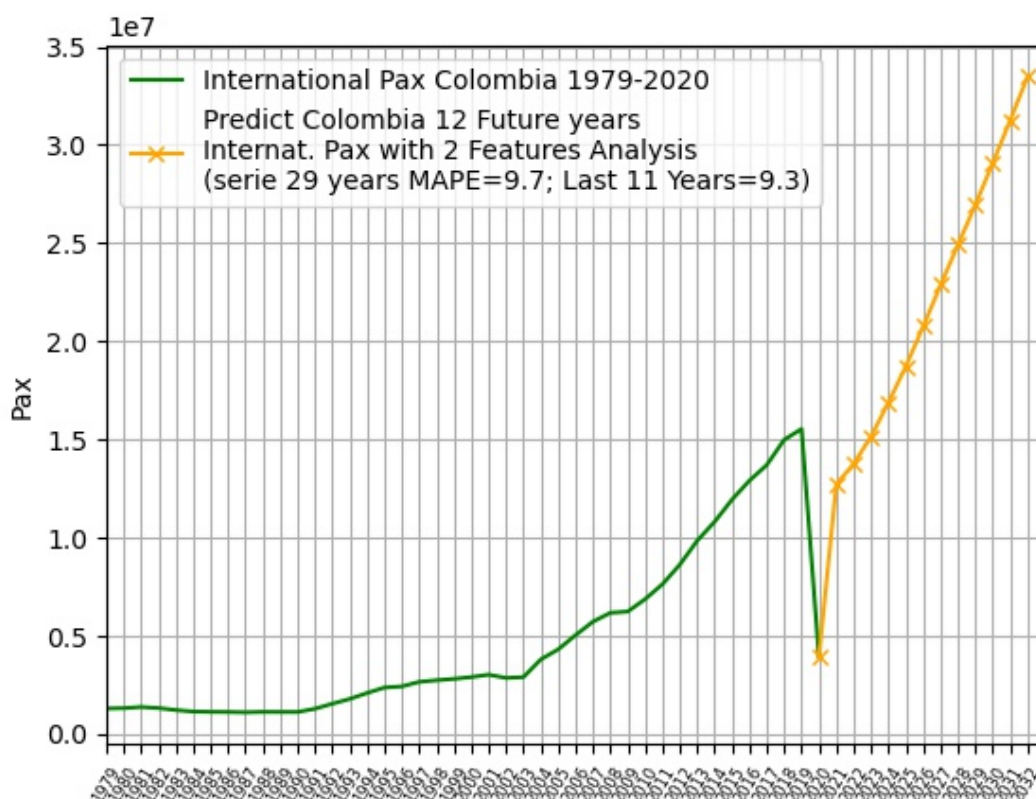


Fig. 15. Pronóstico de la demanda de pasajeros internacionales de Colombia (en naranja frente al verde serie histórica) para el periodo 2021-2032 correspondiente al **escenario 2**. Fuente: Autor.

Escenario 3. Se presenta un tercer escenario de predicción considerando secuencias de seis años previos y de los seis años siguientes a predecir, de entre la serie histórica total, reteniendo las series temporales de solo dos *features* elegidos como variables independientes (pasajeros nacionales, PIB), junto a la de pasajeros nacionales como única variable de salida a predecir, y agrupando los seis años previos n_lags en dos subsecuencias, por lo que tienen tres años cada una. En la figura 16 se representa dicha predicción a corto plazo con un valor del MAPE de 4.5% para los últimos cinco años (se excluye el año 2020) y de 6.0% para la serie histórica de 35 años (al empezar seis años después de 1979, año histórico de inicio de la serie, para predecir los n_outs años siguientes). Los resultados muestran que para finales de 2022 o principios del 2023, el tráfico (o volumen de demanda) de pasajeros nacionales, a nivel país, habría recuperado el nivel prepandémico (2019) y sin presentar oscilaciones.

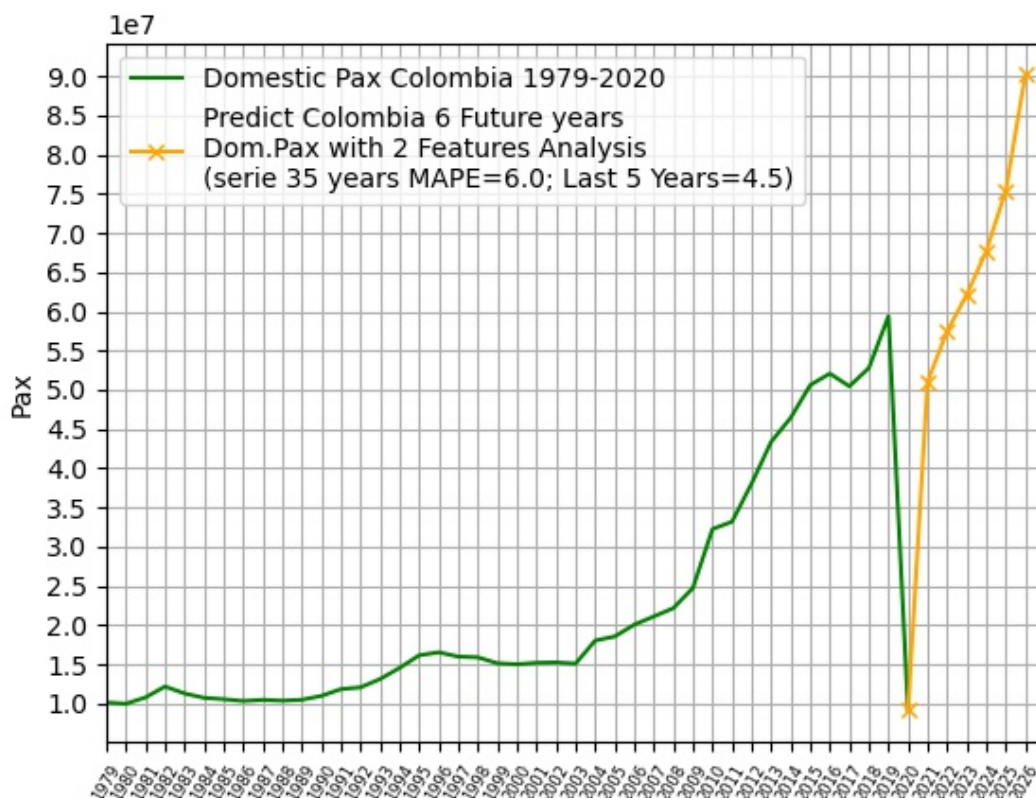


Fig.16. Pronóstico de la demanda de pasajeros nacionales de Colombia (en color naranja frente al verde real) para el periodo 2021-2026; del **escenario 3**. Fuente: Autor.

Escenario 4. Se presenta una cuarta predicción, calculada con las tendencias de las series temporales de pasajeros nacionales e internacionales usadas como *features* de entrada, para analizar su impacto, dejando en tres la cantidad total de *features* usados como *inputs*, ya que se añade el PIB como feature de entrada. Se han considerado secuencias de seis años previos y de los seis años siguientes, de entre la serie histórica total. Con las series temporales de solo 3 *features* elegidos como variables independientes (pasajeros internacionales, pasajeros nacionales, PIB) y con pasajeros internacionales como variable a predecir. Agrupando los seis años previos n_lags también en dos subsecuencias, por tanto de tres años cada una. Los resultados de este cálculo se muestran en la figura 17; donde los valores de MAPE resultantes son de 7.6% para los últimos cinco años (se excluye el año 2020) y de 5.8% para la serie histórica de 35 años (al empezar seis años después

de 1979, año de inicio de los datos históricos disponibles, para predecir los n_outs años siguientes). Los resultados indican que para mediados del año 2022 la demanda de pasajeros internacionales, a nivel país, recuperaría el volumen de demanda prepandémica (año 2019) e incluso con una tendencia superior (en el crecimiento de la demanda). Tampoco se observan oscilaciones en la tendencia predicha.

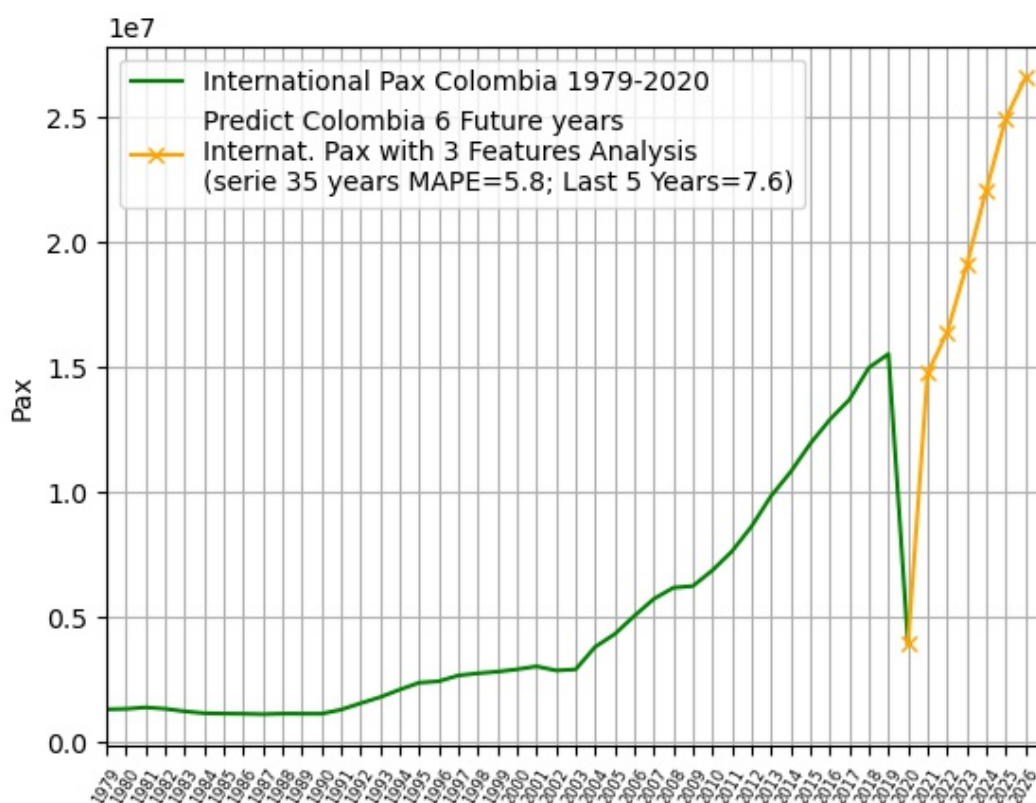


Fig.17. Pronóstico de la demanda de pasajeros internacionales de Colombia (en color naranja frente al verde histórico) para el periodo 2021-2026; del **escenario 4**. Fuente: Autor.

Escenario 5. Finalmente, se presenta una quinta predicción con secuencias de seis años previos y de los seis años siguientes, de entre la serie histórica total, tomando solo dos *features* como variables independiente (pasajeros internacionales y PIB) y con pasajeros internacionales como única variable de salida a predecir. Agrupando los seis años previos n_lags también en dos subsecuencias (o subciclos posibles), por tanto, de tres años cada uno. Los

resultados se muestran en la figura 18, donde los valores del MAPE son de 3% para los últimos cinco años (se excluye el año 2020) y de 4.8% para la serie histórica de 35 años (al empezar seis años después de 1979, año de inicio del histórico de datos disponible, y para predecir los n_outs años siguientes). En este escenario, los resultados indican que para mediados del año 2022 la demanda de pasajeros internacionales de Colombia se habría recuperado al nivel de la demanda prepandémica (año 2019). Se puede observar cómo este mejor modelo (en términos de MAPE), pasados los años de la recuperación de la crisis del Covid-19, la tendencia futura reproduce fielmente la forma y vigor de la subida que tenía el patrón pre pandémico del Covid-19.

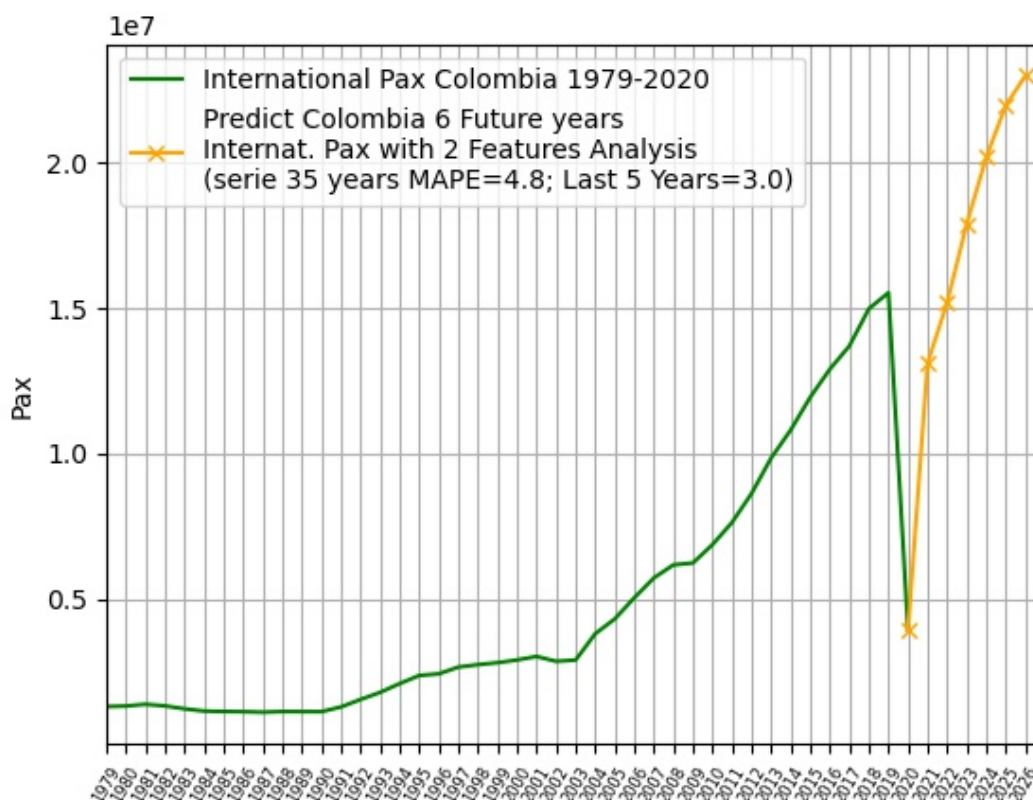


Fig.18. Pronóstico de la demanda de pasajeros internacionales de Colombia, del periodo 2021-2026; **escenario 5**. Fuente: Autor.

Como conclusión del análisis de los resultados se desea hacer tres observaciones. En primer lugar, el algoritmo aquí desarrollado tiene una

complejidad básica-media, caracterizada por la presencia de 2 millones (exactamente 1.920.009) de parámetros de conexiones neuronales (pesos) a aprender, los cuales deben de ser ajustados como incógnitas en cada iteración durante el proceso de aprendizaje. Particularizando dicho valor para el **escenario 5** (con dos *features* y seis años de previsión), estas incógnitas o pesos crecen hasta los 3.149.833 conexiones neuronales cuando se manejan los seis *features* y 12 años de previsión del **escenario 1** (que aun así obtiene ligeramente peores resultados a pesar de ser un modelo más complejo al manejar más *features* y por consiguiente más parámetros de conexiones neuronales).

En segundo lugar, como se ha mostrado exhaustivamente en la descripción del modelo de ML/DL del ConvLSTM2D, se trata de procesos estocásticos, originados implícitamente en el mismo proceso del cálculo del gradiente estocástico descendente (SGD), que en el presente desarrollo se ha mejorado con la selección del optimizador ADAM de acuerdo con Kingma D. & Ba J.^[218] (2014) y por tanto con resultados sujetos a variaciones. Una forma de valorar las incertidumbres asociadas a los procesos estocásticos de los modelos de predicción de redes neuronales presentados en esta tesis, es mediante la repetición de la ejecución del algoritmo varias veces, con los mismos hiperparámetros, lo que se ha hecho caracterizando los resultados mediante la media y la desviación típica, lo que se conocen como técnicas de validación cruzada “*K-fold cross-validation*” de acuerdo con Pedrycz W. & Chen S.^[52] (2020), y Chang C. & Lin C.^[88] (2011).

Finalmente, cabe mencionar que los resultados aquí obtenidos están en la línea con los pronósticos realizados en un estudio reciente de Gudmundsson S. et al.^[18] (2021) sobre la recuperación del transporte aéreo a nivel mundial. El cual predice que la demanda de pasajeros aéreos se recuperará, a niveles de prepandemia a finales del 2022 (escenario optimista). Y los resultados también coinciden con un estudio de IATA (2020) el cual prevé la recuperación del tráfico de pasajeros aéreos del 2019, pero en este caso a nivel mundial, para a finales de 2022 (pasajeros domésticos) y principios de 2024 (pasajeros internacionales).

7.3 Tercer Caso de Estudio correspondiente a la Predicción del retraso de la Salida de Aviones.

La evaluación del aprendizaje de los 10 modelos ML/DL considerados para evaluar las prestaciones de los mismos en cuanto a la predicción del retraso de salida de vuelos, se ha implementado variacionalmente a través de varios escenarios, según que atributos de vuelo, de los disponibles, se consideren como inputs o variables independientes, a fin de determinar la sensibilidad del aprendizaje en la predicción de los retrasos. En la tabla 6 se recoge la lista general de los 16 atributos de vuelo disponibles del dataset, con la definición en inglés de los acrónimos de uso.

Feature	Description
TRF	Type of traffic (domestic / international)
AEL	Airline
ORI	Origin airport (departure)
DES	Destination airport
NVU	Flight number
FPS	Scheduled departure date
HPS	Scheduled departure time
FRM	Push-back date
HRM	Push-back time
DEA	Time Delay (according to airport authority standard)
DEM	Actual time Delay (minute)
EVU	Flight status (On time or Delayed)
CDE	Delay code (according to Standard IATA Delay Codes-AHM 730)
MDE	Reason for delay (according to Standard IATA Delay Codes-AHM 730)
OBS	Observations (from the airport authority)
EAC	Flight status (on time, delayed, early, canceled, according to airport authority standard)

Tabla 6. Lista de atributos de cada vuelo. Fuente: Autor

Los resultados comparativos de las prestaciones, medidos en RMSE para el **escenario 1**, definido este al seleccionar los 10 features (TRF, AEL, ORI, DES,

NVU, CDE, MDE, OBS, EAC y DEM), donde los 9 primeros son inputs de entrada y el último es el output a predecir (DEM o tiempo de demora expresado en minutos), para cada uno de los 10 modelos, se muestra en la figura 19. Donde se observa que las prestaciones medidas con la métrica RMSE son muy parecidas para todos los modelos, estando su valor por debajo del 0.5, salvo el comportamiento de dos modelos más dispersos (MLR y SVR) y quizás también el HBR. El modelo RFR es el de mayores prestaciones con un RMSE de 0.449 y una desviación típica de 0.0013, lo que da una idea de que se trata de un modelo muy robusto (o de baja varianza).

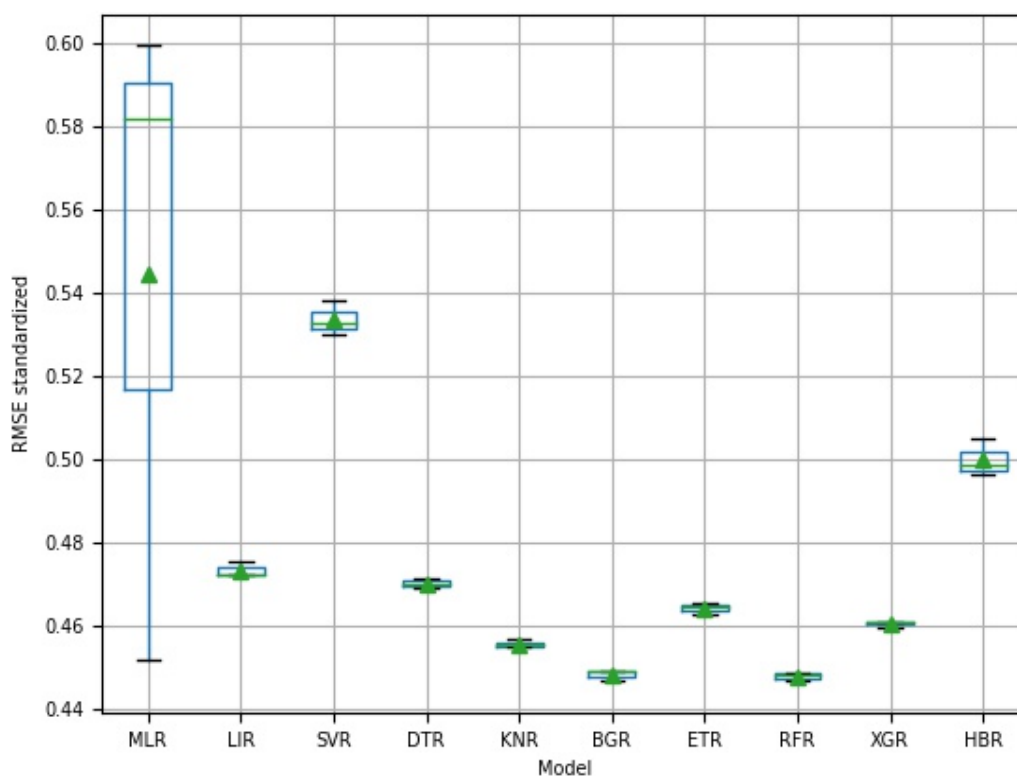


Fig. 19. Comparativa medida en RMSE, resultante del aprendizaje obtenido para cada uno de los 10 modelos, con el feature tiempo de demora transformado en una distribución estándar (de media 0 y desviación típica 1), referido al **escenario 1** resultante de elegir 10 features; 9 de inputs (TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC) más 1 de la salida (DEM). Fuente: Autor.

Dado que el “*feature*” DEM se encuentra estandarizado para facilitar el aprendizaje de los modelos ML/DL, en la figura 20 se muestra los resultados del RMSE reconstruyendo su escala original de tiempos de demora en minutos (lo que se hace invirtiendo el proceso de estandarización realizado en el DEM). En el que se aprecia que los errores de predicción (en su escala natural de minutos) son inferiores a 16 minutos (para el mejor modelo RFR). Es decir, que al introducir los 9 atributos mencionados de vuelo se predice un tiempo de demora que tiene un error inferior a 16 minutos, lo cual es un resultado notable para los features considerados.

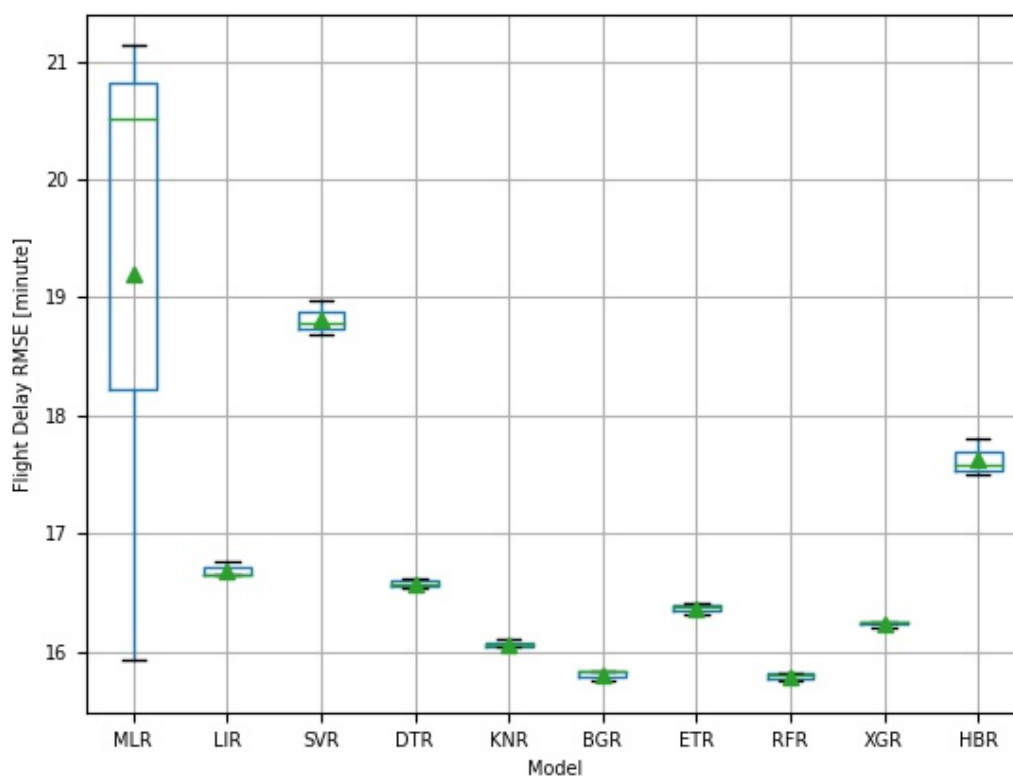


Fig. 20. Comparativa medida en RMSE, resultante del aprendizaje obtenido para cada uno de los 10 modelos, en su escala original de minutos de demora, referida al **escenario 1** configurado al elegir 10 features; 9 de entrada (TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC) más 1 de salida (DEM). Fuente: Autor.

A continuación, se genera **el escenario 2 reducido**, resultante de aplicar el método referido de “*Permutation Importance*” mediante el cual cada uno de los 10

modelos ML/DL considerados autoseleccionan sus 5 mejores features, los que más impactan en determinar el tiempo de demora, de entre los 9 features inputs iniciales del escenario. Es decir, cada modelo clasifica a los features inputs por su importancia en la determinación del tiempo de demora, asignándoles un coeficiente (relativo) de importancia, descrito en el apartado 6.2.6. En la figura 21 se observa el ranking de importancia asignada por el mejor modelo RFR a los 9 features de entrada, donde los 5 últimos son los más importantes (ORI, AEL, NVU, CDE, EAC), mostrados en orden creciente.

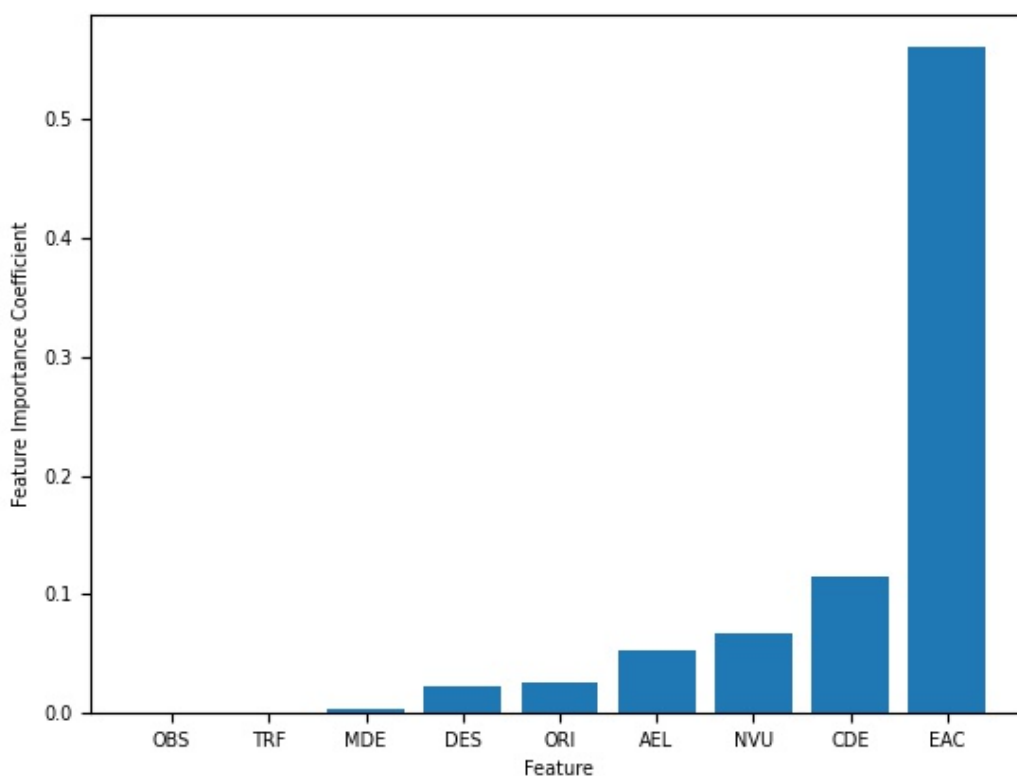


Fig. 21. Coeficientes de importancia relativos, por orden creciente, de los 9 features considerados como inputs del **escenario 1**, para la determinación del tiempo de demora de vuelo (DEM). Fuente: Autor.

Para valorar la calidad en la selección de los 5 features más importantes (EAC, CDE, NVU, AEL, ORI, ordenados ahora en orden decreciente) por parte de los

modelos, se les vuelve a entrenar bajo este nuevo **escenario 2 pero ahora reducido** a los 5 features de entrada mencionados y 1 de salida (DEM), a fin de comparar el aprendizaje de los modelos con los *features* reducidos con respecto al **escenario 1** donde estaban los 9 *features*. Los resultados de la precisión del error RMSE bajo este escenario reducido 2 se muestran en la figura 22. Donde se observa que son los mismos que los obtenidos al emplear los 10 features del **escenario 1**; de hecho, el RMSE mejora levemente para el **escenario 2**, lo que demuestra que los patrones de aprendizaje de los modelos se mejoran cuando se reduce la cantidad de features, siempre que se usen los features más significativos. Es decir que los modelos generalizan mejor cuando se simplifican los inputs a los features más significativos. Lo que es una aportación singular de esta tesis.

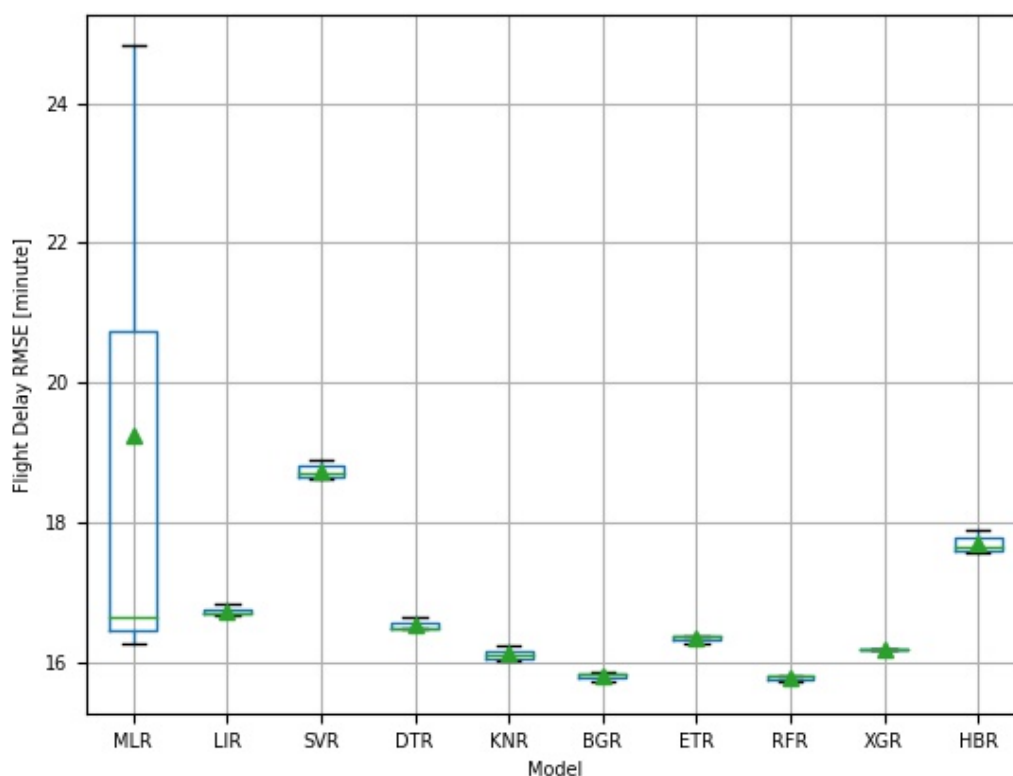


Fig. 22. Resultados comparativos medidos en RMSE obtenidos por los 10 modelos ML/DL cuando se entrenan solo con 5 features (los más significativos), correspondientes al **escenario 2** reducido del dataset. Se observa como los valores del aprendizaje de los modelos (en términos RMSE) son idénticos (o levemente mejores) a cuando se usa el **escenario 1**. Fuente: Autor.

Pero, dado que tanto en el **escenario 1 como el 2 (reducido)** intervienen una serie de atributos que son medidos a posteriori (cuando el vuelo salió en tiempo o se demoró) , y por tanto no pueden ser conocidos de antemano, como son el EAC (el estado del vuelo de acuerdo con los estándares de la autoridad aeroportuaria), CDE (El código del retraso de acuerdo al Standard IATA Delay Codes-AHM 730), MDE (Motivos del retraso según el estándar IATA Delay Codes-AHM 730), y OBS (observaciones del vuelo realizadas por la autoridad aeroportuaria), a fin de que estas predicciones sean útiles y lo más realistas posibles para predecir las demoras de los vuelos de salida, en esta investigación se decidió eliminar dichos 4 features del dataset. Por lo que el **escenario 3**, queda ahora definido únicamente por 6 features conocibles de antemano, 5 de los cuales son los inputs (TRF, AEL, ORI, DES, NVU) y 1 output (DEM). Tras su entreno se obtiene un RMSE de 33.8 minutos por parte del mejor modelo RFR (aproximadamente el doble que cuando se retenían los 4 features eliminados), resultado que se muestra en la figura 23.

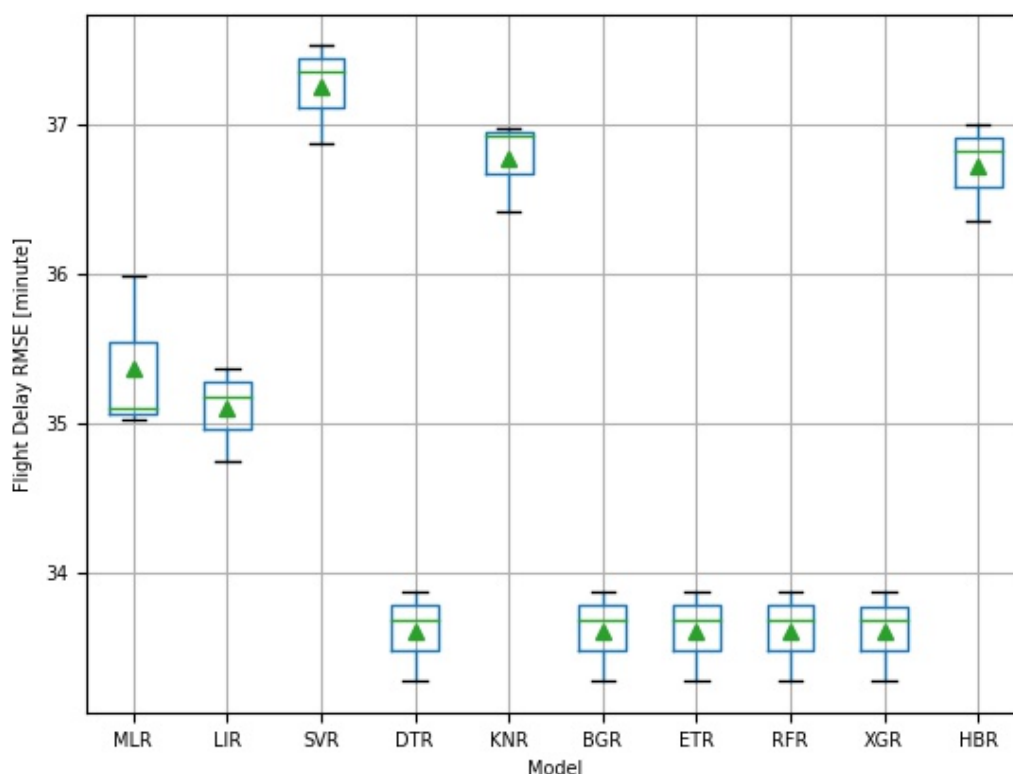


Fig. 23. Comparativa medida en RMSE de los 10 modelos cuando se usa el **escenario 3** del dataset; con 5 inputs (TRF, AEL, ORI, DES, NVU) y como output la demora a predecir (DEM). Donde el mejor modelo RFR obtiene un error RMSE en la predicción de la demora del vuelo de 33.8 minutos. Fuente: Autor.

La razón de esta pérdida notable de precisión al eliminar esos 4 features (EAC, CDE, MDE, OBS), además de que estos 4 features son conocidos a posteriori, habría que buscarla en que hay otras variables que influyen en la demora y que no han sido incluidas (por no estar disponibles) en el dataset considerado, como puedan ser el tipo de aeronave (principalmente su tamaño), el nivel de congestión del aeropuerto, las condiciones meteorológicas, condiciones ATM, etc.

Cuando a este **escenario 3** de solo 5 +1 features, se le aplica el método “*Permutation Importance*” para conocer la importancia (relativa) de los features en la determinación del tiempo demora del vuelo, se obtiene el ranking de significancia para el caso del mejor modelo RFR, que se muestra en la figura 24.

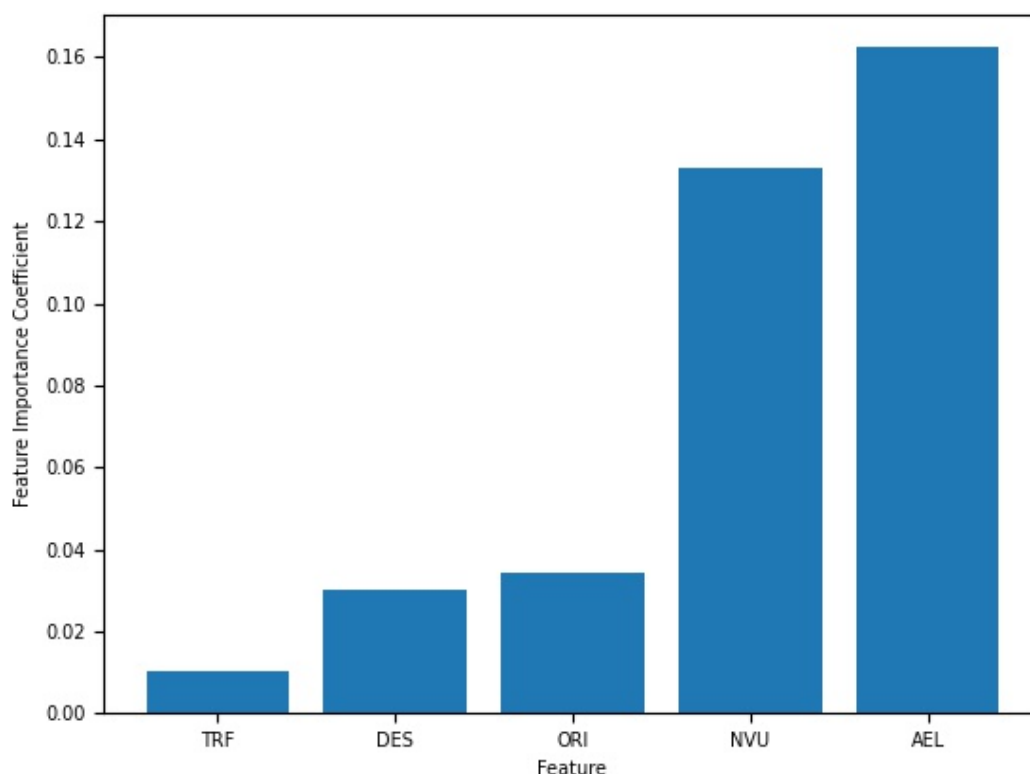


Fig. 24. Orden de importancia de los 5 inputs features considerados en el **escenario 3** (TRF, DES, ORI, NVU, AEL) por orden creciente. Fuente: Autor.

El escenario 4 (análogamente al **escenario 2**) se genera al aplicar otra reducción más de features al **escenario 3**, en este caso para reducirlo a los 3 mejores que se autoseleccionan por los propios modelos al aplicar el método “*Permutation Importance*”. En la figura 25 se observa la comparativa de la medida RMSE para los 10 modelos ML/DL, bajo el **escenario 4** de 4 features consistente en 3 inputs (ORI, NVU, AEL) más 1 de output el DEM. Los resultados muestran que a pesar de la reducción de features (a los 3 más importantes) se obtienen los mismos resultados que en el **escenario 3** cuando se usan 5 features de entrada. Lo que muestra que las auto-reducciones de features realizadas por el método “*Permutation Importance*” han sido exitosas y lo que es más importante, el dataset puede ser reducido o simplificado notablemente, consiguiendo con ello ahorro de tiempo y simplicidad de entrenamiento de los modelos, sin merma alguna en la

precisión de la predictibilidad del tiempo de demora DEM. Lo que constituye una aportación singular de esta tesis.

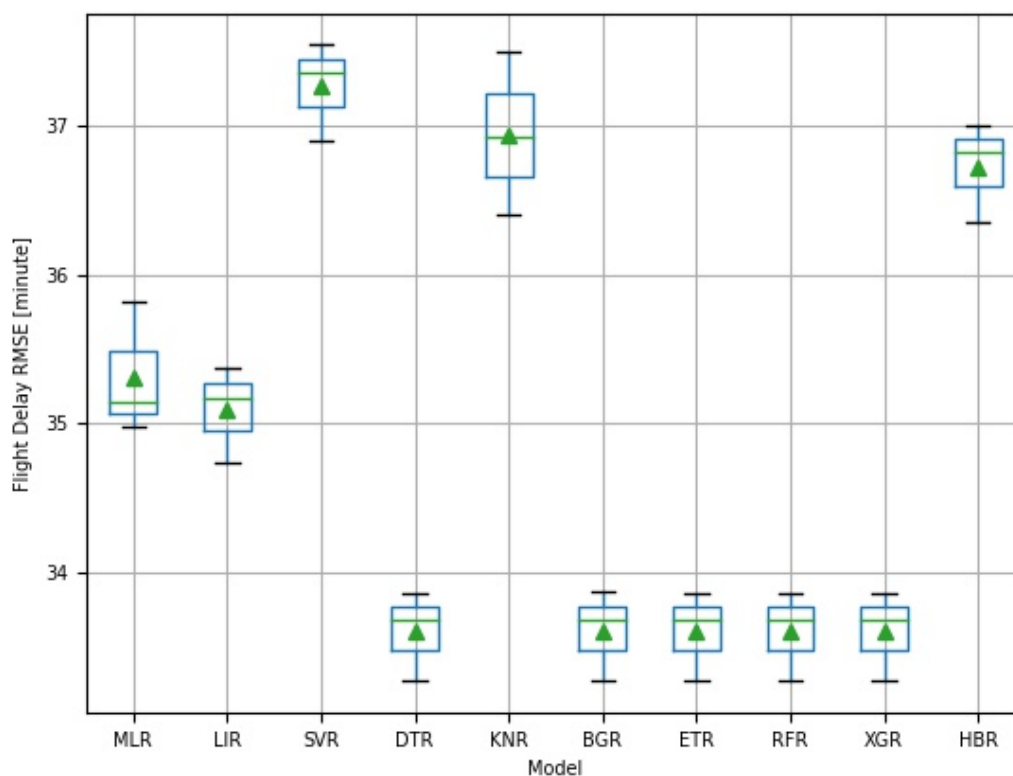


Fig. 25. Comparativa RMSE de modelos para el **escenario 4**, con 3 inputs solo (ORI, NVU, AEL) y la salida DEM como demora a predecir. Donde el mejor modelo RFR obtiene el mismo error RMSE de la predicción de la demora del vuelo de 33.8 minutos que el del **escenario 3**. Fuente: Autor.

Para caracterizar estadísticamente a los features se aplicaron 14 tests estadísticos, tal y como se describe en el apartado 6.2.4, al objeto de determinar el tipo de distribución del feature de demora de vuelo DEM, dado que es la única variable numérica a la que se pueden aplicar dichos tests, ya que el resto de atributos de vuelo del dataset son variables categóricas (TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC). Los resultados de estos test se muestran en la tabla 7. Donde se observa que el test ADFuller señala que el tiempo de demora de la salida (departure) del vuelo es una variable estacionaria. El resto de hallazgos de la distribución del tiempo de demora indican que no es una distribución Gaussiana,

motivo por el cual se forzó la transformación de dicha distribución en una distribución estándar (de media 0 y desviación típica 1), para facilitar el aprendizaje de los modelos ML/DL.

Statistical test	Flight Delay
Shapiro-Wilk (if feature is Gaussian)	False
D'Agostino K^2 (if feature is Gaussian)	False
Anderson-Darling (if feature is Gaussian)	False
Linear Pearson (if feature is independent of the output)	False
Spearman Rank (if feature is independent of the output)	False
Kendall Rank (if feature is independent of the output)	False
Chi-square Cat. (if feature is independent of the output)	False
Adfuller (if feature is not stationary)	False
Student's t-test (if distribution is equal to the output)	True
Paired Student's t-test, (if distribution is equal to output)	True
Anova, (if distribution is equal to output)	True
Mann-Whitney U, (if distribution is equal to output)	True
Kruskal-Wallis H, (if distribution is equal to output)	True
Friedman, (if distribution is equal to output)	True

Tabla 7. Resultados de aplicar 14 test estadísticos a la distribución del feature demora del vuelo (DEM).
Fuente: Autor.

En la figura 26 se muestra la escala o márgenes del tiempo de la demora del vuelo DEM en términos de mínimo (-15 minutos) y máximo (240 minutos) y de media (15 minutos).

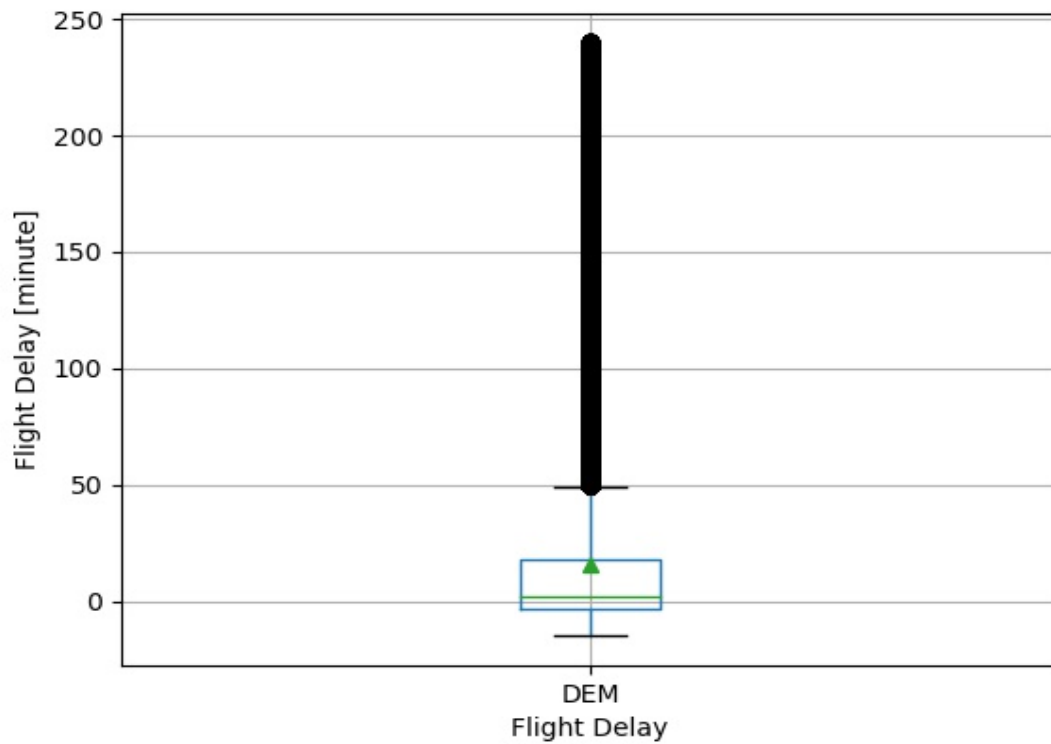


Fig. 26. Distribución del tiempo de la demora de la salida del vuelo. Fuente: Autor.

Finalmente, en la figura 27 se muestra como la función de distribución continua (Density) del tiempo de demora del vuelo es unimodal. Estando la media de la distribución en los 15 minutos.

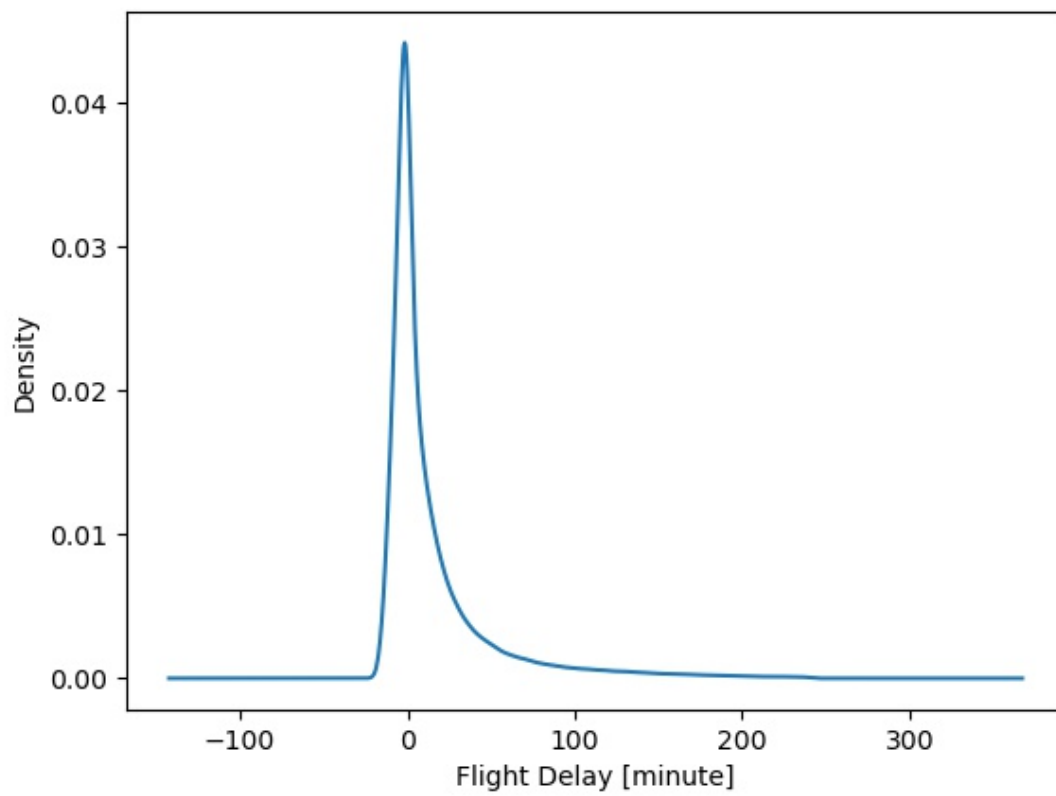


Fig. 27. Función de distribución (densidad) unimodal de la demora (en la salida) de vuelo. Fuente: Autor.

En resumen, se puede afirmar que con los atributos proporcionados para el entreno de los modelos con el dataset considerado, las predicciones del tiempo de demora realizadas por los modelos ML/DL son aceptables, ya que el error se mueve entre 16 y 33 minutos dependiendo del escenario considerado. Lo que suponen unas prestaciones aceptables y muy similares a las obtenidas por las investigaciones según Guo Z. et al. ^[76] (2021), Truong D. ^[75] (2021), Henriques R. & Feiteira I. ^[178] (2018), Belcastro L. et al. ^[177] (2016).

8. Conclusiones

La previsión de la demanda de transporte de carga aérea es extremadamente importante para los operadores aéreos y autoridades aeroportuarias. Hacer las previsiones lo más precisas posibles son fundamentales de cara a la planificación a corto y medio plazo (para el diseño de rutas, planificación de la flota, procesos operacionales, presupuestos, necesidades financieras, materiales y de personal, etc.). Para las autoridades aeronáuticas nacionales y gestores aeroportuarios (individuales o de red) disponer de pronósticos precisos es crítico para implementar sus planes directores aeroportuarios p.e., y en cualquier caso es un asistente vital para la ayuda a la toma de decisiones en temas tan variados como la política de inversiones (pública y privada) o para definir estrategias necesarias que puedan atender la demanda futura de pasajeros y carga aérea. Lo que implica atender debidamente a la expansión y modernización de los aeropuertos existentes y construcción de nuevos, así como para la propuesta de políticas tarifarias que incentiven la captación de nuevos operadores aéreos y la apertura de nuevas rutas. Evidentemente, acompañando a estas expansiones y actualizaciones de los sistemas aeroportuarios se tienen que definir las operativas de gestión para atender ese tráfico, donde la gestión de la demora de los vuelos será más crítica si cabe, cuando dichas ampliaciones o actualizaciones de equipos e infraestructuras no se hayan acometido a tiempo.

Este ha sido el objetivo general de la presente investigación; investigar con ayuda de las nuevas herramientas de aprendizaje profundo (DL) como anticipar las demandas futuras del transporte aéreo, para mejorar la gestión del mismo.

8.1 Demanda Carga Aérea

El primer objetivo del presente trabajo fue estimar el patrón de recuperación post-Covid19 de la demanda de carga aérea en Colombia, a partir del fuerte impacto que ocasionó la pandemia mundial del Covid-19 en el año 2020 (cuyo año fue incluido en la investigación). Los resultados muestran una rápida recuperación de la demanda de carga aérea, muy similar (aunque ligeramente más rápida) a la

que experimentará la demanda de pasajeros aéreos, de acuerdo con Muros JG & Díaz O.^{[46][47]} (2022 y 2023).

En cuanto a la metodología propuesta para el análisis de este pronóstico, la principal aportación de esta investigación a la literatura científica, ha sido mostrar la idoneidad del uso de métodos de ML/DL como el *ConvLSTM2D* en las predicciones de las series temporales aplicadas al transporte aéreo frente a otros métodos ML/DL y especialmente frente a los métodos estadísticos clásicos. Estas herramientas de aprendizaje profundo permiten incorporar análisis no lineales y multivariable para la extracción de los patrones de comportamiento de la demanda. Para ello se han tenido que segmentar (hiperparametrizando al mismo tiempo) en ciclos y subciclos las series históricas, como una característica diferencial de estos modelos de aprendizaje profundo frente a otros existentes. El análisis variacional añadido, a través de la definición de varios escenarios al configurar diferentes hiperparámetros, ha servido para determinar los hiperparámetros óptimos asociados con las longitudes temporales de estos ciclos y subciclos. El análisis de la variabilidad en cuanto a cantidad y selección de los “*features*” según diferentes escenarios, ha permitido igualmente la determinación de los features más significativos que obtienen las mejores predicciones, referidas a la métrica MAPE, tanto de las series temporales de carga aérea doméstica como internacional. La metodología ML/DL aquí aplicada permite filtrar no solo los features que más influyen en la demanda de transporte de carga aérea, sino también mostrar las sensibilidades de la demanda de la carga aérea con los *features* seleccionados.

8.2 Demanda Pasajeros Aéreos

El segundo objetivo ha consistido en estimar a corto y medio plazo la demanda de pasajeros aéreos a nivel nacional e internacional en Colombia, lo que aporta valiosa información para que los planificadores del transporte aéreo colombiano puedan diseñar, planificar e implementar, con suficiente antelación: (a) las estrategias de desarrollo relativas a las infraestructuras, instalaciones, equipos, modernización tecnológica, formación de personal técnico, etc., (b) una hoja de ruta de inversión de capital para hacer frente a los programas de desarrollos propuestos,

(c) y finalmente, políticas públicas para consolidar y reforzar el desarrollo de la industria del transporte aéreo local. Actualmente académicos, analistas, planificadores y autoridades, junto a consultores de la industria de la aviación civil internacional, vienen estudiando cómo mejorar estos pronósticos para saber cuándo y cómo se recuperará el tráfico aéreo (principalmente el volumen de la demanda de pasajeros aéreos) a niveles prepandémicos (año 2019), debido a la importancia de la industria de la aviación, no solo en lo que se refiere a ofertar la conectividad local/regional/mundial idónea, sino también por su importante contribución a la economía. La presente tesis tiene el objetivo de contribuir, desde el perfil académico, a estos dos puntos mencionados. Por otro lado, esta investigación aporta una novedad y singularidad de carácter geográfico, al ser el primer estudio de carácter académico (publicado), a nivel del subcontinente latinoamericano, sobre la previsión del tráfico aéreo para la época post-Covid-19.

En lo que se refiere a la metodología, se optó por el desarrollo de un modelo propio basado en una arquitectura de redes neuronales artificiales frente a otras posibles técnicas, al considerar que las técnicas de predicción de ML/DL pueden incorporar más elementos de análisis y por tanto ser potencialmente más eficaces. En particular, para el análisis de las series temporales, se han aplicado técnicas de redes con bloques codificadores-decodificadores del tipo *ConvLSTM2D*. El modelo *ConvLSTM2D* aquí desarrollado admite un tratamiento multivariable (no pasa lo mismo con los otros métodos clásicos estadísticos). Por otro lado, el modelo *ConvLSTM2D* desarrollado es del tipo aprendizaje supervisado, lo que significa que se entrena al modelo con las entradas y salidas (en forma de datos tensoriales multidimensionales) sacadas de las series históricas del dataset. La posibilidad de hibridación de las redes convolucionales CNN con las redes recurrentes de memoria a corto-largo plazo LSTM, ha permitido la implementación singular del modelo *ConvLSTM2D* en la presente investigación. Siendo importante la aportación de esta tesis, al demostrar la viabilidad de su aplicación en la obtención de resultados (aceptables) en la predicción de las series temporales del tráfico aéreo. Si bien el uso de los bloques codificadores-decodificadores de redes convolucionales-recurrentes híbridas ha sido una herramienta exitosa en los últimos años, para el tratamiento de series temporales, estas técnicas no se habían visto

aplicadas hasta ahora al campo del transporte aéreo. Finalmente, los resultados de la aplicación de modelo basado en ML/DL presenta valores MAPE muy aceptables (del orden del 3% al 9%, según el escenario), lo que convierte al modelo aquí desarrollado en una propuesta alternativa viable, para realizar pronósticos fiables de tráfico aéreo a corto y medio plazo.

Futuras líneas de investigación tanto para la demanda de carga aérea como para la de pasajeros podrían, por ejemplo, incorporar nuevas mejoras en la arquitectura de capas de neuronas aquí desarrollada que mejoren la codificación y extracción de *features*, o mediante la exploración de nuevos conceptos como los introducidos por la competición entre modelos que aportan las redes tipo GAN (*“Generative Adversarial Network”*), o la de incorporar el uso de códigos genéticos en la búsqueda de las soluciones óptimas. Lo que requerirá del empleo de métodos heurísticos, dentro del espacio de las soluciones posibles, analizando los resultados al ejecutar los modelos candidatos, según las diferentes configuraciones de los hiperparámetros de diseño del modelo y del proceso de entrenamiento.

8.3 Demora Salida Aviones

Como muestran los periódicos informes de las principales organizaciones de la aviación civil internacional (IATA, ACI, ICAO) se espera que el CAGR (o tasa de crecimiento anual compuesta por sus siglas en inglés) sea del 25.5% para la industria aeronáutica mundial en el periodo 2022-2027 ^[243], habiendo alcanzado la demanda de transporte aéreo un continuo y persistente sólido crecimiento desde hace casi tres décadas. Las autoridades aeroportuarias en todo el mundo intentan dar respuesta a dicha demanda con la ampliación de las infraestructuras aéreas, construcción de nuevos aeropuertos, optimización en el uso de la capacidad existente, mejoras en los procedimientos ATM (gestión del tráfico aéreo) , etc. pero así y todo muchos aeropuertos (principalmente los grandes e internacionales) siguen presentando una situación de congestión (muchas veces severa en ciertos periodos del día) lo que conduce irremediabilmente a la generación de demoras, situación que se complica aún más con situaciones meteorológicas adversas. Por lo que la predicción, lo más precisa posible, de las demoras se ha vuelto crítica, ya

que contribuye a formular estrategias operativas para un uso más eficiente de la capacidad, mejora de la gestión de las operaciones, optimización de procedimientos ATM, etc. que ayuden a mitigar la generación y propagación de las demoras a otros aeropuertos de la red.

El tercer objetivo de esta tesis ha consistido en un análisis comparativo de las prestaciones (medidas con la métrica RMSE) en la predicción de la demora de la salida de un vuelo comercial, de 10 modelos ML/DL bajo diferentes escenarios configurados a partir de los atributos conocidos de cada vuelo. Los modelos son del tipo ML pero en sus variantes más avanzadas (combinatoria de modelos) como son las técnicas de “*Ensemble*” con sus algoritmos de “*Voting*” para incentivar a los más precisos, “*Bagging*” (segmentando el dataset para obtener las ventajas de entrenar cada modelo con diferentes subconjuntos), “*Stacking*” (usando salidas de unos modelos para entrenar otros) o “*Boosting*” (penalizando específicamente los errores no deseados). Todas estas estrategias se enfocan en compensar los errores o desventajas de cada modelo de manera que el conjunto o “*Ensemble*” de los modelos generalice mejor que cualquiera de sus componentes. También se ha incluido en la comparativa el modelo MLP de DL, sin que haya resultado el modelo con mejores prestaciones, no solo por no estar ajustado en sus hiperparámetros de arquitectura y configuración, sino porque probablemente presenta problemas de desvanecimiento del gradiente por ser más profundo en cuanto a capas y conexiones, es decir que probablemente sea susceptible de mejora añadiendo nuevas capas del tipo CNN, RNN, etc. Los resultados comparativos medidos en RMSE para los tiempos (transformados) han estado entre 0.45 y de 0.96 (para el modelo ganador RFR) según el *escenario* considerado en el entreno, lo cual se traduce en errores (aceptables) de entre 16 y 33 minutos para la predicción del tiempo de la demora de las salidas (departure) de los vuelos comerciales.

Por otro lado, se han implementado métodos como el de “*Permutation Importance*” que ha permitido auto-clasificar la importancia de los *features* en la determinación de la demora por parte del propio modelo. Dicha bondad en la selección de features fue confirmada al reevaluar el modelo bajo los escenarios de versiones reducidas de los features (los más significativos), a fin de valorar la sensibilidad del error (RMSE) en el cribado de los features. Verificándose que

dichas auto-reducciones de “*features*” no afectan a la precisión (de la métrica RMSE) para la previsión de la demora, lo que valida la reducción y aligeramiento del tamaño del dataset, simplificación del modelo a la par que una reducción significativa de los tiempos ejecución y entreno del modelo. Aportación que consideramos singular de esta tesis, en línea también con la búsqueda de la interpretabilidad de los patrones aprendidos por el algoritmo.

Aunque el dataset empleado para la caracterización de la demora de la salida de vuelos ha sido muy extenso, correspondiente a un año completo de toda la red aeroportuaria colombiana y con muchos atributos vinculados para cada vuelo, hay que señalar que tiene limitaciones propias que han impedido una mejor predicción. Es decir, han faltado ciertos datos, como p.e. las condiciones meteorológicas previa al vuelo, el tamaño de la aeronave, el nivel de congestión del aeropuerto, el estado de la navegación aérea (ATM), la capacidad del sistema de pistas, el tiempo de rotación de la aeronave (“*aircraft turnaround time*”), etc., que podrían haber mejorado la predicción de la demora del vuelo. Por ello, una línea de investigación que se propone a futuro, sería aumentar el número de features relevantes, aunque esto suele representar un desafío importante, debido a que no siempre se dispone de tanta información para un aeropuerto (y mucho menos para un sistema aeroportuario). De otro lado, se podría profundizar en la investigación que afine los modelos de mayor potencial de aprendizaje, como son los del tipo DL implementado aquí con el modelo MLP, en sus arquitecturas de configuración de tipo y cantidad de capas y neuronas.

Obviamente, con la aparición del ChatGPT a finales 2022 y principios de 2023, se abre la posibilidad de investigar como los modelos LLM (“*large Language Model*”) se pueden implementar en la predicción de series temporales, al objeto de explorar como estos pueden mejorar la inferencia de los patrones históricos, al tiempo que explicar en lenguaje natural los resultados alcanzados, de acuerdo con la propuesta de López-Lira A. & Yuehua Tang Y ^[244] (2023), quienes han mostrado las ventajas de usar el ChatGPT para predicciones en el campo financiero. Así como explorar otras líneas de investigación, a través de la realización de simples códigos de prueba para mejorar la formulación de las preguntas (“*prompt engineering*”) al ChatGPT sobre predicciones de series temporales de acuerdo con

los trabajos de Krzysztof Joachimiak ^[245] (2023). En cualquier caso, actualmente existen importantes limitaciones de los “tokens” o extensiones del texto a entregar al ChatGPT, o en las condiciones de pago para las versiones “premium” de las aplicaciones que mejoran las prestaciones. Y lo que es más importante, de limitaciones para realizar el “*ajuste fino*”, ya que no se dispone del modelo original sobre el que poder realizar modificaciones de ajuste en la arquitectura y de re-entrenos específicos con el modelo para especializarse en un determinado sector del conocimiento (como puede ser el del transporte aéreo). Aunque todo parece indicar que pronto veremos otras aplicaciones LLM escritas *en código abierto* de acuerdo con A. McFarland (2023) ^[246] que permitan su uso en investigaciones como las desarrolladas en esta tesis. Las cuales habrán de acompañarse de los apropiados análisis y estudios comparativos que validen y valoren las ventajas e inconvenientes del uso de estas nuevas herramientas LLM.

En cambio, si que puede ser una línea de trabajo de rápida mejora, incorporar los “*mecanismos de atención*”, especialmente los denominados de “*transformer*”, para reemplazar las funciones de memoria que aportan las redes recurrentes RNN (LSTM) implementadas en este modelo *ConvLSTM2D*. Ya que estas nuevas técnicas “*transformer*”, consideradas como una aportación singular de los exitosos modelos LLM, pueden aportar mejoras muy relevantes en todo lo que se refiere a los tiempos de ejecución requeridos para entrenar grandes y complejos modelos, lo que les permiten que puedan acometer entrenos de dataset más grandes, y sobre todo admiten arquitecturas con un mayor número de parámetros, por tanto, más complejas y de mayor poder predictivo. Otra ventaja de los “*transformer*” sobre las redes RNN tipo LSTM aquí implementadas, es que capturan mejor las dependencias de la memoria a largo plazo, fundamental en el aprendizaje a partir de las series históricas, al evitar los problemas de desvanecimiento del gradiente que presentan las redes. Por todo lo cual es esperable una mejora sustancial en las predicciones usando codificadores-decodificadores *transformer*.

9. Bibliografía

- [1] TRB (2002). *Aviation Demand Forecast: Survey of Methodologies*. Washington DC: Transportation Research Board.
- [2] Horonjeff R., McKelvey F., Sproule W., Young S. (2010). *Planning and Design of Airports*. McGraw-Hill.
- [3] Wells A. & Young S. (2004). *Airport Planning & Management*. New York: McGraw-Hill.
- [4] Ashford N., Mumayiz S., Wright P. (2011). *Airport Engineering*. New Jersey: John Wiley & Sons.
- [5] Janic M. (2021). *System Analysis and Modelling in Air Transport*. Boca Raton: CRC Press
- [6] Janic M. (2009). *Airport analysis, planning and design: demand, capacity, and congestion*. New York: Nova Science Publishers.
- [7] ICAO (1987). *Master Planning. Part 1. Doc 9184*. Montreal: ICAO.
- [8] Rodríguez Y., Pineda W., Díaz Olariaga O. (2020). *Air traffic forecast in post-liberalization context: a Dynamic Linear Models approach*. *Aviation*, 24(1), 10-19. <https://doi.org/10.3846/aviation.2020.12273>
- [9] García Cruzado M. (2013). *Aeropuertos. Planificación, Diseño y Medio Ambiente*. Madrid: Ibergarceta Publicaciones.
- [10] ICAO (2006). *Manual on Air Traffic Forecasting*. Doc 8991. Montreal: ICAO.
- [11] ACI (2016). *Traffic Forecast*. Montreal: ACI (Airports Council International).
- [12] Kazda A., Caves R. (2015). *Airport design and operations*. Bingley: Emerald.
- [13] de Neufville R., Odoni A. (2013). *Airport Systems, Planning, Design, and Management*. New York: McGraw-Hill.
- [14] ICAO (2020). *Presentation of 2019 Air Transport Statistical Results*. Montreal: ICAO. https://www.icao.int/annual-report-2019/Documents/ARC_2019_Air%20Transport%20Statistics_es.pdf

- [15] IATA (2021a). *World Air Transport Statistics 2021*. Geneva: IATA.
- [16] IATA (2021b). *Air Cargo Market Analysis*. Geneva: IATA.
- [17] World Bank (2022). *Air transport*.
<https://data.worldbank.org/indicator/IS.AIR.GOOD.MT.K1>
- [18] Gudmundsson S.V., Cattaneo M., Redondi R. (2021). *Forecasting temporal world recovery in air transport markets in the presence of large economic shocks: The case of Covid-19*. *Journal of Air Transport Management*, 91, 102007. <https://doi.org/10.1016/j.jairtraman.2020.102007>
- [19] JADC (2021). *Worldwide Market Forecast 2020-2040*. Tokyo: Japan Aircraft Development Corporation.
- [20] Boeing (2020). *World Air Cargo Forecast 2020-2039*. Seattle: Boeing.
- [21] Dantas T., Oliveira F., Repolho H. (2017). *Air transportation demand forecast through Bagging Holt Winters methods*. *Journal of Air Transport Management*, 59, 116–123. <https://doi.org/10.1016/j.jairtraman.2016.12.006>
- [22] Abed S. Y., Ba-Fail A. O. & Jasimuddin S. M. (2001). *An econometric analysis of international air travel demand in Saudi Arabia*. *Journal of Air Transport Management*, 7, 143–148. [https://doi.org/10.1016/S0969-6997\(00\)00043-0](https://doi.org/10.1016/S0969-6997(00)00043-0)
- [23] Fernandes E. & Pacheco R. R. (2010). *The causal relationship between GDP and domestic air passenger traffic in Brazil*. *Transportation Planning and Technology*, 33, 569–581. <https://doi.org/10.1080/03081060.2010.512217>
- [24] Garrow L. A. & Koppelman F. S. (2004). *Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information*. *Journal of Air Transport Management*, 10, 401–411.
<https://doi.org/10.1016/j.jairtraman.2004.06.007>
- [25] Grosche T., Rothlauf F. & Heinzl A. (2007). *Gravity models for airline passenger volume estimation*. *Journal of Air Transport Management*, 13, 175–183. <https://doi.org/10.1016/j.jairtraman.2007.02.001>

- [26] Samagaio A. & Wolters M. (2010). *Comparative analysis of government forecasts for the Lisbon Airport*. Journal of Air Transport Management, 16, 213–217. <https://doi.org/10.1016/j.jairtraman.2009.09.002>
- [27] Chin A. T. H. & Tay J. H. (2001). *Developments in air transport: implications on investment decisions, profitability, and survival of Asian airlines*. Journal of Air Transport Management, 7, 319–330. [https://doi.org/10.1016/S0969-6997\(01\)00026-6](https://doi.org/10.1016/S0969-6997(01)00026-6)
- [28] Tsui W. H. K., Balli H., Gilbey A. & Gow H. (2014). *Forecasting of Hong Kong airport's passenger throughput*. Tourism Management, 42, 62–76. <https://doi.org/10.1016/j.tourman.2013.10.008>
- [29] Aston J. A. D. & Koopman S. J. (2006). *A non-Gaussian generalization of the airline model for robust seasonal adjustment*. Journal of Forecasting, 25, 325–349. <https://doi.org/10.1002/for.991>
- [30] West M. & Harrison J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.
- [31] Honjo K., Shiraki H. & Ashina S. (2018). *Dynamic linear modeling of monthly electricity demand in Japan: Time variation of electricity conservation effect*. PLOS ONE, 13(4), e0196331. <https://doi.org/10.1371/journal.pone.0196331>
- [32] Petris G., Petrone S. & Campagnoli P. (2009). *Dynamic linear models with R*. Springer. https://doi.org/10.1007/b135794_2
- [33] Tascón D. & Díaz O. (2021). *Air traffic forecast and its impact on runway capacity. A System Dynamics approach*. Journal of Air Transport Management, vol. 90, 1-10. <https://doi.org/10.1016/j.jairtraman.2020.101946>
- [34] Suryani E., Chou S. & Chen C. (2012). *Dynamic simulation model of air cargo demand forecast and terminal capacity planning*. Simulation Modelling Practice and Theory, 28, 27–41.
- [35] Suryani E., Chou S. & Chen, C. (2010). *Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework*. Expert Systems with Applications, 37, 2324–2339.

- [36] Zhang G., Patuwo B. & Hu M.Y. (1998). *Forecasting with artificial neural networks: The state of the art*. International Journal of Forecasting, 14, 35–62.
- [37] Srisaeng P., Baxter G. & Wild G. (2015). *Using an artificial neural network approach to forecast Australia's domestic passenger air travel demand*. World Review of Intermodal Transportation Research, 5(3), 281-313.
- [38] Mostafaeipour A., Goli A. & Qolipour M. (2018). *Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study*. The Journal of Supercomputing, 74(10), 5461-5484.
- [39] Millstein F. (2018). *Convolutional Neural Networks in Python*. Scotts Valley: CreateSpace Independent Publishing Platform.
- [40] Sewak M., Karim R. & Pujari P. (2018). *Practical Convolutional Neural Networks*. Birmingham: Packt Publishing.
- [41] Cybenko G. (1989). "Approximation by superpositions of a sigmoidal function". Mathematics of Control, Signals, and Systems. 2 (4): 303–314. <https://doi.org/10.1007/BF02551274>
- [42] Gupta V., Sharma K., Sangwan M. (2019). *Airlines passenger forecasting using LSTM based recurrent neural networks*. International Journal Information Theories and Applications, 26(2), 178-187.
- [43] Dingari M., Reddy M. & Sumalatha V. (2019). *Air Traffic Forecasting Using Artificial Neural Networks*. International Journal of Scientific & Technology Research, 8(10), 556-559. <https://doi.org/10.1007/s11227-018-2452-0>
- [44] Bao Y., Xiong T. & Hu, Z. (2012). *Forecasting Air Passenger Traffic by Support Vector Machines with Ensemble Empirical Mode Decomposition and Slope-Based Method*. Discrete Dynamics in Nature and Society, ID 431512, 1-12. <https://doi.org/10.1155/2012/431512>
- [45] ACI (2021). Annual World Airport Traffic Report. Montreal: ACI (Airports Council International).

- [46] Muros JG. & Díaz O. (2023). *Air cargo transport demand forecasting using ConvLSTM2D, an artificial neural network architecture approach*. Case Studies on Transport Policy, Volume 12, 2023, 101009, <https://doi.org/10.1016/j.cstp.2023.101009>
- [47] Muros JG. & Díaz O. (2022). *Air passenger demand forecast through the use of Artificial Neural Network algorithms*. International Journal of Aviation, Aeronautics, and Aerospace, 9(3). <https://doi.org/10.15394/ijaaa.2022.1726>
- [48] Yang J., Nguyen M., San P., Li X. & Krishnaswamy S. (2015). *Deep convolutional neural networks on multichannel time series for human activity recognition*. Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 3995–4001.
- [49] Malhotra P., Vig L., Shroff G. & Agarwal P. (2015). *Long-short term memory networks for anomaly detection in time series*. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 22-24 April 2014, Bruges.
- [50] Hermans M. & Schrauwen B. (2013). *Training and analyzing deep recurrent neural networks*. Working Paper, Ghent University.
- [51] Ketkar N. & Moolayil J. (2021). *Deep Learning with Python*. New York: Springer.
- [52] Pedrycz W. & Chen S. (2020). *Deep Learning: Concepts and Architectures*. Cham: Springer.
- [53] Calin O. (2020). *Deep Learning Architectures. A Mathematical Approach*. Cham: Springer.
- [54] Muros JG. & Díaz O. (2023). *Prediction of departure flight delays through the use of predictive tools based on machine learning/deep learning algorithms*. The Aeronautical Journal, 1-23. <https://doi.org/10.1017/aer.2023.41>
- [55] Muros JG., Díaz O. (2023). *Flight departure delay forecasting*. Journal of Airport Management, 17(2), 197-209. <https://www.henrystewartpublications.com/jam/v17>

- [56] Graham A. & Morrell P. S. (2017). *Airport finance and investment in the global economy*. London: Routledge.
- [57] ACRP (2014). *Defining and measuring aircraft delay and airport capacity thresholds*. ACRP Report 104. Washington, D.C.: Transportation Research Board.
- [58] Gelhausen M., Berster P., Wilken D. (2020). *Airport capacity constraints and strategies for mitigation*. London: Academic Press.
- [59] Kearns S. (2018). *Fundamentals of International Aviation*. London: Routledge.
- [60] Budd L. & Ison S. (2017). *Air Transport Management*. London: Routledge.
- [61] Borsky S. & Unterberger C. (2019). *Bad weather and flight delays: The impact of sudden and slow onset weather events*. *Economics of Transportation*, 18, 10–26. <https://doi.org/10.1016/j.ecotra.2019.02.002>
- [62] Choi S., Kim Y., Briceno S., Mavris D. (2016). *Prediction of Weather-induced Airline Delays Based on Machine Learning Algorithms*. IEEE/AIAA 35th Digital Avionics Systems Conference Proceeding. <https://doi.org/10.1109/DASC.2016.7777956>
- [63] Fernandes N., Moro S., Costa C., Aparício M. (2020). *Factors influencing charter flight departure delay*. *Research in Transportation Business & Management*, 34, 100413. <https://doi.org/10.1016/j.rtbm.2019.100413>
- [64] ACI (2012). *Guide to Airport Performance Measures*. Montreal: ACI World.
- [65] Carvalho L., Sternberg A., Goncalves L., Cruz A., Soares J., Brandão D., Carvalho G., Ogasawara E. (2021). *On the relevance of data science for flight delay research: a systematic review*. *Transport Reviews*, 41(4), 499–528. <https://doi.org/10.1080/01441647.2020.1861123>
- [66] FAA-Federal Aviation Administration (2021). *Definitions of variables* <https://aspmhelp.faa.gov/index.php/APM: Analysis: Definitions of Variables>

- [67] EUROCONTROL (2020). *Performance Review Report*. Brussels: EUROCONTROL.
- [68] OAG (2017). *OAG Punctuality League*. Chicago: OAG Aviation Worldwide.
- [69] Pamplona D.& Alves C.J. (2020). *An overview of air delay: A case study of the Brazilian scenario*. Transportation Research Interdisciplinary Perspectives, 7, 100189. <http://dx.doi.org/10.1016/j.trip.2020.100189>
- [70] ICAO (2018). *Manual on Collaborative Air Traffic Flow Management (ATFM)*. Doc 9971. Montreal: ICAO.
- [71] European Observatory on Airport Capacity & Quality (2015). *Delays to air transport in Europe: Methods of measuring, reporting, and analyzing*. Brussels: European Commission.
- [72] EUROCONTROL (2007). *A Matter of Time: Air Traffic Delay in Europe*. Brussels: EUROCONTROL.
- [73] University of Westminster (2015). *European airline delay cost reference values*. London: University of Westminster.
- [74] Kim M.& Bae J. (2021). *Modeling the flight departure delay using survival analysis in South Korea*. Journal of Air Transport Management, 91, 101996.
- [75] Truong D. (2021). *Using causal machine learning for predicting the risk of flight delays in air transportation*. Journal of Air Transport Management, 91, 101993. <https://doi.org/10.1016/j.jairtraman.2020.101993>
- [76] Guo Z., Yu B., Hao M., Wang W., Jiang Y., Zong F. (2021). *A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient*. Aerospace Science and Technology, 116, 106822. DOI: <https://doi.org/10.1016/j.ast.2021.106822>
- [77] Lambelho M., Mitici M., Pickup S., Marsden A. (2020). *Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions*. Journal of Air Transport Management, 82, 101737. <https://doi.org/10.1016/j.jairtraman.2019.101737>

- [78] Peterson E., Neels K., Barczi N., Graham T. (2013). *The economic cost of airline flight delay*. Journal of Transport Economics and Policy, 47(1), 107–121. <https://www.jstor.org/stable/24396355>
- [79] NEXTOR (2010). *Total Delay Impact Study*. Washington D.C.: National Center of Excellence for Aviation Operations Research (Federal Aviation Administration).
- [80] Cook A., Tanner G., Jovanović R., Lawes A. (2009). *The cost of delay to air transport in Europe - Quantification and management*. 13th Air Transport Research Society World Conference, June 27-30, 2009, Abu Dhabi.
- [81] BTS (2021). Bureau of Transportation Statistic (U.S. Department of Transportation). <https://www.transtats.bts.gov/HomeDrillChart.asp>
- [82] EUROCONTROL (2020b). *CODA Digest. All-causes delay and cancellations to air transport in Europe*. Brussels: EUROCONTROL.
- [83] FAA (2020). *Cost of Delay Estimates. FAA APO-100*. Washington D.C.: Federal Aviation Administration.
- [84] IATA (2013). *Inefficiency in European Airspace*. Geneva: IATA.
- [85] Actividades Aeronáuticas Aerocivil (2023). <https://www.aerocivil.gov.co/atencion/estadisticas-de-las-actividades-aeronauticas>
- [86] Altmann A., Tolosi L., Sander O., Lengauer T. (2010). *Permutation importance: a corrected feature importance measure*. Bioinformatics, 26(10), 1340–1347. DOI: <https://doi.org/10.1093/bioinformatics/btq134>
- [87] Breiman L. (2001). Random forests. Machine Learning, 45, 5-32.
- [88] Chang C. & Lin C. (2011). LIBSVM: *A library for support vector machines*. ACM Transactions on Intelligent Systems and Technology, 2(3), 1-27.
- [89] McCulloch W. & Pitts W. [A Logical Calculus of the ideas Imminent in Nervous Activity](#)
- [90] S. Ramón y Cajal. [Histología del sistema nervioso del hombre y de los vertebrados. Tomo I: Estudios sobre el plan estructural y composición](#)

[histológica de los centros nerviosos adicionados de consideraciones fisiológicas fundadas en los nuevos descubrimientos](#)

- [91] Samuel Arthur (1959). *"Some Studies in Machine Learning Using the Game of Checkers"*. IBM Journal of Research and Development.
<https://doi.org/doi:10.1147/rd.33.0210.S2CID2126705>
- [92] Rosenblatt, Frank (1962). *Principles of Neurodynamics*. Spartan, New York.
- [93] Robbins H. & Monro S. (1951). *"A Stochastic Approximation Method"*. The Annals of Mathematical Statistics. 22 (3): 400. <https://doi.org/10.1214/aoms/1177729586>.
- [94] Russel S. & Norvig P. (2003) [1995]. *Artificial Intelligence: A Modern Approach* (2nd ed.). Prentice Hall. ISBN 978-0137903955.
- [95] LeCun Y. et al. (1998). *"Gradient-based learning applied to document recognition"*. Proceedings of the IEEE. **86** (11): 2278–2324. <https://doi.org/10.1109/5.726791>
- [96] Hochreiter S., *"Untersuchungen zu dynamischen neuronalen Netzen"*. Archived 2015-03-06 at the Wayback Machine. *Diploma thesis. Institut f. Informatik, Technische Univ. Munich. Advisor: J. Schmidhuber, 1991.*
- [97] Bengio Yoshua (2012). *"Practical recommendations for gradient-based training of deep architectures"*. <https://arxiv.org/abs/1206.5533>
- [98] J. Brownlee. Machine Learning Magistery.
<https://machinelearningmastery.com/start-here/#getstarted>
- [99] Dataset Aerocivil (2023).
<https://www.aerocivil.gov.co/analitica/Paginas/DataSet.aspx>
- [100] Deep Learning Q&A: *What's Coming Next with Francois Chollet*
<https://www.forbes.com/sites/googlecloud/2021/07/27/deep-learning-qa-whats-coming-next-with-francois-chollet/?sh=47696e4560e7>

- [101] Shuwen C. Hong *Kong air traffic: explanation and prediction based on sparse seasonal ARIMA model*. School of Statistics Renmin University of China, August 13, 2021. <https://arxiv.org/pdf/2108.05817.pdf>
- [102] Keras API. <https://keras.io/>
- [103] François Chollet. *Deep learning with python*. ISBN 9781617294433. <https://tanthiamhuat.files.wordpress.com/2018/03/deeplearningwithpython.pdf>
- [104] tensorflow. <https://www.tensorflow.org/>
- [105] Wan MJ. 2013. <https://maktubaat.wordpress.com/2013/07/04/differences-between-the-air-passenger-and-air-freight-markets/>
- [106] Gelhausen M., Berster P, Wilken D (2019) *Airport capacity constraints and strategies for mitigation – a global perspective*. Academic Press, Elsevier, pp 1–338
- [107] Baier F., Berster P., Gelhausen M. (2021). Global cargo gravitation model: airports matter for forecasts. *International Economics and Economic Policy*. <https://doi.org/10.1007/s10368-021-00525-2>
- [108] Alexander D., Merkert R. (2021). Applications of gravity models to evaluate and forecast US international air freight markets post-GFC. *Transport Policy*, 104, 52–62. <https://doi.org/10.1016/j.tranpol.2020.04.004>
- [109] Chen S., Kuo S., Chang K., Wang Y. (2012). *Improving the forecasting accuracy of air passenger and air cargo demand: the application of back-propagation neural networks*. *Transportation Planning and Technology*, 35(3), 373–392. <https://doi.org/10.1080/03081060.2012.673272>
- [110] Chou T., Liang G., Han T. (2010). *Application of fuzzy regression on air cargo volume forecast*. *Quality and Quantity*, 45, 1539–1550. <https://doi.org/10.1007/s11135-010-9342-8>
- [111] Li H., Bai J., Cui X., Li Y., Sun S. (2020). *A new secondary decomposition-ensemble approach with cuckoo search optimization for air cargo*

- forecasting*. *Applied Soft Computing Journal*, 90, 106161.
<https://doi.org/10.1016/j.asoc.2020.106161>
- [112] Liu J., Ding L., Guan X., Gui J., Xu J. (2020). *Comparative analysis of forecasting for air cargo volume: Statistical techniques vs. machine learning*. *Journal of Data, Information and Management*, 2, 243–255.
<https://doi.org/10.1007/s42488-020-00031-1>
- [113] Madhavan M., Sharafuddin M., Piboonrungrroj P., Yang C. (2020). *Short-term Forecasting for Airline Industry: The Case of Indian Air Passenger and Air Cargo*. *Global Business Review*. <https://doi.org/10.1177/0972150920923316>
- [114] Onder E., Kuzu S. (2014). *Forecasting air traffic volumes using smoothing techniques*. *Journal of Aeronautics and Space Technologies*, 7(1), 65-85.
<https://doi.org/10.7603/s40690-014-0006-0>
- [115] Alekseev K.P.G. & Seixas J.M. (2009). *A multivariate neural forecasting modeling for air transport – Preprocessed by decomposition: A Brazilian application*. *Journal of Air Transport Management*, 15, 212–216.
- [116] Rumelhart, Hinton, Williams (1986). *"Learning representations by back-propagating errors"* (PDF). *Nature*. **323** (6088):
<https://doi.org/10.1038/323533a0>.
- [117] LeCun Yann, Bengio Yoshua, Hinton Geoffrey (2015). *"Deep learning"*. *Nature*. **521** (7553): <https://doi.org/10.1038/nature14539>
- [118] Jin F., Li Y., Sun S. & Li H. (2020). *Forecasting air passenger demand with a new hybrid ensemble approach*. *Journal of Air Transport Management*, 83, 101744. <https://doi.org/10.1016/j.jairtraman.2019.101744>
- [119] Plakandaras V., Papadimitriou T. & Gogas P. (2019). *Forecasting transportation demand for the U.S. market*. *Transportation Research Part A*, 126, 195–214. <https://doi.org/10.1016/j.tra.2019.06.008>
- [120] Sulistyowati R., Suhartono, Kuswanto H., Setiawan & Astuti, E. (2018). *Hybrid forecasting model to predict air passenger and cargo in Indonesia*.

- International Conference on Information and Communications Technology (ICOI/ACT), 6-7 March 2018. Yogyakarta, Indonesia.
- [121] Sun S., Lu H., Tsui K. & Wang S. (2019). *Nonlinear vector auto-regression neural network for forecasting air passenger flow*. Journal of Air Transport Management, 78, 54–62. <https://doi.org/10.1016/j.jairtraman.2019.04.005>
- [122] Xiao Y., Liu J., Xiao J., Hu Y., Bu H. & Wang S. (2015). *Application of multiscale analysis-based intelligent ensemble modeling on airport traffic forecast*. Transportation Letters: International Journal of Transportation Research, 7(2), 73-79. <https://doi.org/10.1179/1942787514Y.0000000035>
- [123] Blinova T.O. (2007). *Analysis of possibility of using neural network to forecast passenger traffic flows in Russia*. Aviation, 11(1), 28-34.
- [124] Bahdanau Dzmitry (2016-05-19). "Neural Machine Translation by Jointly Learning to Align and Translate". [arXiv:1409.0473](https://arxiv.org/abs/1409.0473) [cs.CL].
- [125] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. 2017. *Attention is all you need* <https://doi.org/10.48550/arXiv.1706.03762>
- [126] Tan X., Jia X., Yan J., Wang K., Bian L. (2021). *An Exploratory analysis of flight delay propagation in China*. Journal of Air Transport Management, 92, 102025. <https://doi.org/10.1016/j.jairtraman.2021.102025>
- [127] Kim M. & Park S. (2021). *Airport and route classification by modelling flight delay propagation*. Journal of Air Transport Management, 93, 102045. <https://doi.org/10.1016/j.jairtraman.2021.102045>
- [128] Kafle N., Zou B. (2016). *Modeling flight delay propagation: A new analytical-econometric approach*. Transportation Research Part B, 93, 520–542. <http://dx.doi.org/10.1016/j.trb.2016.08.012>
- [129] Fleurquin P., Ramasco J., Eguiluz V. (2013). *Systemic delay propagation in the US airport network*. Scientific Reports, 3, 1-6. <https://doi.org/10.1038/srep01159>

- [130] Cai Q., Alam S., Doung V. (2021). *A Spatial–Temporal Network Perspective for the Propagation Dynamics of Air Traffic Delays*. Engineering, 7, 452–464. <https://doi.org/10.1016/j.eng.2020.05.027>
- [131] Zhang M., Zhou X., Zhang Y., Sun L., Dun M., Du W., Cao X. (2019). *Propagation Index on Airport Delays*. Transportation Research Record, 2673(8), 536–543. DOI: <https://doi.org/10.1177/0361198119844240>
- [132] Campanelli B., Fleurquin P., Arranz A., Etxebarria I., Ciruelos C., Eguíluz V., Ramasco, J. (2016). *Comparing the modeling of delay propagation in the US and European air traffic networks*. Journal of Air Transport Management, 56, 12–18. <http://dx.doi.org/10.1016/j.jairtraman.2016.03.017>
- [133] Wang P., Schaefer L., Wojcik L. (2003). *Flight connections and their impacts on delay propagation*. Digital Avionics Systems Conference, 12–16 October 2003, Indianapolis (USA).
- [134] Markovic D., Hauf T., Roehner P., Spehr U. (2008). *A statistical study of the weather impact on punctuality at Frankfurt airport*. Meteorological Applications, 15(2), 293–303.
- [135] Xu N., Sherry L., Laskey K. (2008). *Multifactor model for predicting delays at U.S. airports*. Transportation Research Record, 2052, 62–71. <https://doi.org/10.3141/2052-08>
- [136] Zhang J., Xu X., Wang F., Wei D. (2010). *Airport delay performance evaluation based on fuzzy linear regression model*. Journal of Traffic and Transportation Engineering, 10(4), 109–114.
- [137] Hao L., Hansen M., Zhang Y., Post J. (2014). *New York, New York: Two ways of estimating the delay impact of New York airports*. Transportation Research Part E, 70, 245–260. <https://doi.org/10.1016/j.tre.2014.07.004>
- [138] Boswell S. and Evans J. *Analysis of downstream impacts of air traffic delay*. Project Report, Lincoln Laboratory, Massachusetts Institute of Technology, 1997.

- [139] Wong J., Tsai S. (2012). *A survival model for flight delay propagation*. Journal of Air Transport Management, 23, 5–11.
<https://doi.org/10.1016/j.jairtraman.2012.01.016>
- [140] Xu N., Donohue G., Laskey K., Chen C. (2005). *Estimation of delay propagation in the national aviation system using Bayesian networks*. 6th USA/Europe Air Traffic Management Research and Development Seminar, Citeseer, 2005.
- [141] Wu W., Meng T., Zhang H. (2016). *Flight plan optimization based on airport delay prediction*. Journal of Transportation Systems Engineering and Information Technology, 16(6), 189–195.
- [142] Baspinar B., Koyuncu E. (2016). *A Data-Driven Air Transportation Delay Propagation Model Using Epidemic Process Models*. International Journal of Aerospace Engineering, 2016, Article ID 4836260.
<https://doi.org/10.1155/2016/4836260>
- [143] Pyrgiotis N., Malone K., Odoni A. (2013). *Modelling delay propagation within an airport network*. Transportation Research Part C, 27, 60–75.
<https://doi.org/10.1016/j.trc.2011.05.017>
- [144] Hansen M. (2002). *Micro-level analysis of airport delay externalities using deterministic queuing models: a case study*. Journal of Air Transport Management, 8(2), 73–87. [https://doi.org/10.1016/S0969-6997\(01\)00045-X](https://doi.org/10.1016/S0969-6997(01)00045-X)
- [145] Morrison S., Winston C. (2008). *The effect of FAA expenditures on air travel delays*. Journal of Urban Economics, 63(2), 669–678.
- [146] Zou B., Hansen M. (2012). *Flight delays, capacity investment and social welfare under air transport supply-demand equilibrium*. Transportation Research Part A, 46(6), 965–980. <https://doi.org/10.1016/j.tra.2012.02.015>
- [147] Qin Q. & Yu H. (2014). *A statistical analysis on the periodicity of flight delay rate of the airports in the US*. Advances in Transportation Studies, 3, 93–104.

- [148] Pathomsiri S., Haghani A., Dresner M., Windle R. (2008). *Impact of undesirable outputs on the productivity of US airports*. Transportation Research Part E, 44(2), 235–259. <https://doi.org/10.1016/j.tre.2007.07.002>
- [149] Kotegawa T., De Laurentis D., Noonan K., Post J. (2011). *Impact of commercial airline network evolution on the U.S. air transportation system*. Proceedings of the 9th USA/Europe Air Traffic Management Research and Development Seminar, 572–580.
- [150] Pfeil D., Balakrishnan H. (2012). *Identification of robust terminal-area routes in convective weather*. Transportation Science, 46(1), 56–73. <https://www.jstor.org/stable/41432825>
- [151] Abdelghany K., Shah S., Raina S., Abdelghany A. (2004). *A model for projecting flight delays during irregular operation conditions*. Journal of Air Transport Management, 10(6), 385–394. <https://doi.org/10.1016/j.jairtraman.2004.06.008>
- [152] Ahmadbeygi S., Cohn A., Guan Y., Belobaba P. (2008). *Analysis of the potential for delay propagation in passenger airline networks*. Journal of Air Transport Management, 14(5), 221–236. <https://doi.org/10.1016/j.jairtraman.2008.04.010>
- [153] Soomer M., Franx G. (2008). *Scheduling aircraft landings using airlines' preferences*. European Journal of Operational Research, 190(1), 277–291. <https://doi.org/10.1016/j.ejor.2007.06.017>
- [154] Wu C. (2005). *Inherent delays and operational reliability of airline schedules*. Journal of Air Transport Management, 11(4), 273–282. <https://doi.org/10.1016/j.jairtraman.2005.01.005>
- [155] Mofokeng T., Marnewick, A. (2017). *Factors contributing to delays regarding aircraft during A-check maintenance*. IEEE Technology and Engineering Management Society Conference, TEMSCON 2017. <https://doi.org/10.1109/TEMSCON.2017.7998375>

- [156] Zou B., Hansen M. (2014). *Flight delay impact on air fare and flight frequency: a comprehensive assessment*. Transportation Research Part E, 69, 54–74. <https://doi.org/10.1016/j.tre.2014.05.016>
- [157] Abdel-Aty M., Lee C., Bai Y., Li X., Michalak M. (2007). *Detecting periodic patterns of arrival delay*. Journal of Air Transport Management, 13(6), 355–361. <https://doi.org/10.1016/j.jairtraman.2007.06.002>
- [158] Zhong Z., Varun D., Lin Y. (2017). *Studies for air traffic management R&D in the ASEAN-region context*. Journal of Air Transport Management, 64, 15–20. <https://doi.org/10.1016/j.jairtraman.2017.06.020>
- [159] Ferrer J., Oliveira, P., Parasuraman A. (2012). *The behavioral consequences of repeated flight delays*. Journal of Air Transport Management, 20, 35–38. <https://doi.org/10.1016/j.jairtraman.2011.11.001>
- [160] Lubbe B., Victor C. (2012). *Flight delays: towards measuring the cost to corporations*. Journal of Air Transport Management, 19, 9–12. <https://doi.org/10.1016/j.jairtraman.2011.11.004>
- [161] Li Q., Jing R. (2021). *Characterization of delay propagation in the air traffic network*. Journal of Air Transport Management, 94, 102075. <https://doi.org/10.1016/j.jairtraman.2021.102075>
<https://doi.org/10.1016/j.jairtraman.2020.101996>
- [162] Mohammadian I., Abbasi B., Abareshi A., Goh M. (2019). *Antecedents of flight delays in the Australian domestic aviation market*. Transportation Research Interdisciplinary Perspectives, 1, 100007. <https://doi.org/10.1016/j.trip.2019.100007>
- [163] Du W., Zhang M., Zhang Y., Cao X., Zhang J. (2018). *Delay causality network in air transport systems*. Transportation Research Part E, 118, 466–476. <https://doi.org/10.1016/j.tre.2018.08.014>
- [164] Chen Y., Lin J. (2021). *Determinants of flight delays at East Asian airports from an airport, route, and network perspective*. Journal of Air Transport Management, 94, 102064. <https://doi.org/10.1016/j.jairtraman.2021.102064>

- [165] Tu Y., Ball M., Jank W. (2008). *Estimating flight departure delay distributions, A statistical approach with long-term trend and short-term pattern*. Journal of the American Statistical Association, 103(481), 112–125.
<https://doi.org/10.1198/016214507000000257>
- [166] Mueller E., Chatterji G. (2002). *Analysis of aircraft arrival and departure delay characteristics*. AIAA's Aircraft Technology, Integration and Operations Conference, 1-3 October 2002, Los Angeles (USA).
- [167] Schaefer L. & Millner D. (2001). *Flight delay propagation analysis with the Detailed Policy Assessment Tool*. 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace, 7-10 Oct. 2001, Tucson (USA).
<https://doi.org/10.1109/ICSMC.2001.973100>
- [168] Hunter G., Boisvert B., Ramamoorthy K. (2007). *Advanced national airspace traffic flow management simulation experiments and validation*. Proceedings of the 39th conference on Winter simulation, 1261–1267.
- [169] Bakhshandeh R., Shahgholian K., Shahraki A. (2013). *Model For Reduce Flights Delays Using System Dynamics (Case Study in Iranian Airports Company)*. Interdisciplinary Journal of Contemporary Research in Business, 4(9), 746-757.
- [170] Santos G., Robin M. (2010). *Determinants of delays at European airports*. Transportation Research Part B, 44, 392–403.
<https://doi.org/10.1016/j.trb.2009.10.007>
- [171] Sternberg A., Carvalho D., Murta L., Soares, J., Ogasawara E. (2016). *An analysis of Brazilian flight delays based on frequent patterns*. Transportation Research Part E, 95, 282–298. <http://dx.doi.org/10.1016/j.tre.2016.09.013>
- [172] Wang C., Wang X. (2019). *Airport congestion delays and airline networks*. Transportation Research Part E, 122, 328–349.
<https://doi.org/10.1016/j.tre.2018.12.008>

- [173] Wang Y., Zhu Y., Zhu C., Wu F., Yang H., Yan Y., Hu C. (2019). *Indicator of serious flight delays with the approach of time-delay stability*. Physica A, 518, 363–373. <https://doi.org/10.1016/j.physa.2018.11.038>
- [174] Wesonga R., Nabugoomu F., Jehopio P. (2012). *Parameterized framework for the analysis of probabilities of aircraft delay at an airport*. Journal of Air Transport Management, 23, 1-4. <https://doi.org/10.1016/j.jairtraman.2012.02.001>
- [175] Zhang M., Chen S., Sun L., Du W., Cao X. (2021). *Characterizing Flight Delay Profiles with a Tensor Factorization Framework*. Engineering, 7, 465–472. <https://doi.org/10.1016/j.eng.2020.08.024>
- [176] Alonso H., Loureiro A. (2015). *Predicting Flight Departure Delay at Porto Airport: A Preliminary Study*. Proceedings of the 7th International Joint Conference on Computational Intelligence, Volume 3: NCTA, 93-98.
- [177] Belcastro L., Marozzo F., Talia M., Trunfio P. (2016). *Using Scalable Data Mining for Predicting Flight Delays*. ACM Transactions on Intelligent Systems and Technology, 8(1), Article 5. <http://dx.doi.org/10.1145/2888402>
- [178] Henriques R. & Feiteira I. (2018). *Predictive Modelling: Flight Delays and Associated Factors, Hartsfield–Jackson Atlanta International Airport*. Procedia Computer Science, 138, 638–645.
- [179] Kalliguddi A., Leboulluec A. (2017). *Predictive Modeling of Aircraft Flight Delay*. Universal Journal of Management, 5(10), 485-491. <https://doi.org/10.13189/ujm.2017.051003>
- [180] Pérez-Rodríguez J., Pérez J., Gómez E. (2017). *Modelling the asymmetric probabilistic delay of aircraft arrival*. Journal of Air Transport Management, 62, 90-98. <http://dx.doi.org/10.1016/j.jairtraman.2017.03.001>
- [181] Rebollo J., Balakrishnan H. (2014). *Characterization and prediction of air traffic delays*. Transportation Research Part C, 44, 231–241. <http://dx.doi.org/10.1016/j.trc.2014.04.007>

- [182] Yu B., Guo Z., Asian S., Wang H., Chen G. (2019). *Flight delay prediction for commercial air transport: A deep learning approach*. Transportation Research Part E, 125, 203–221. <https://doi.org/10.1016/j.tre.2019.03.013>
- [183] Díaz O. & Pulido L. (2019). *Measurement of airport efficiency. The case of Colombia*. Transport and Telecommunication, 20(1), 40-51
- [184] Díaz O. & Ávila J. (2015). *Evolution of the airport and air transport industry in Colombia and its impact on the economy*. Journal of Airline and Airport Management, 5(1), 39-66.
- [185] Díaz O. (2021). *The role of regional airports in connectivity and regional development*. Periodica Polytechnica Transportation Engineering, 49(4), 1-13. <https://doi.org/10.3311/PPtr.16557>
- [186] Díaz O. & Zea J.F. (2018). *Influence of the liberalization of the air transport industry on configuration of the traffic in the airport network*. Transportation Research Procedia, 33, 43-50.
- [187] Boletines Operacionales de Aerocivil (2023). <https://www.aerocivil.gov.co/atencion/estadisticas-de-las-actividades-aeronauticas/boletines-operacionales>
- [188] Díaz Olariaga, O. & Carvajal A.F. (2020). *Perspectiva geográfica del desarrollo de la conectividad aérea en Colombia*. Boletín Geográfico, 42(2), 145-168.
- [189] Estadísticas del Banco de la República de Colombia BRC (2023). <https://www.banrep.gov.co/es/-estadisticas>
- [190] Airbus (2016). Global Market Forecast 2016–2035. Blagnac: Airbus.
- [191] Boeing (2016). *World Air Cargo Forecast 2016–2035*. Seattle: Boeing.
- [192] Hakim M.M. & Merkert R. (2016). *The causal relationship between air transport and economic growth: empirical evidence from South Asia*. Journal of Transport Geography, 56, 120–127.
- [193] Morrell P.S. (2011). *Moving Boxes by Air: The Economics of International Air Cargo*. Aldershot: Ashgate.

- [194] Ashford N., Stanton M., Moore C.A., Coutu P., Beasley J. (2013). *Airport Operations*. New York: McGraw-Hill.
- [195] Halpern N. & Graham A. (2018). *Air Transport Management*. New York: Routledge.
- [196] Díaz Olariaga O. & Alonso C. (2021). *Impact of airport policies on regional development. Evidence from the Colombian case*. Regional Science Policy & Practice, 1–26. <https://doi.org/10.1111/rsp3.12483>
- [197] Wensveen J.G. (2016). *Air Transportation*. New York: Routledge.
- [198] DANE - Departamento Administrativo Nacional de Estadística (Colombia) (2022). <https://www.dane.gov.co/>
- [199] Díaz O. & Girón E. (2020). *Análisis de la influencia de la privatización de aeropuertos en el pronóstico de la demanda de pasajeros. El caso de Colombia*. Revista Transporte y Territorio, 22, 94-113.
<https://doi.org/10.34096/rtt.i22.8408>
- [200] Vang-Mata R. (2020). *Multilayer Perceptrons: Theory and Applications*. Hauppauge (NY): Nova Science Pub. Inc.
- [201] Hastie T., Tibshirani R. & Friedman. J. (2013). *The elements of statistical learning. Data mining, inference, and prediction*. New York: Springer.
- [202] Dorffner G. (1996). *Neural Networks for Time Series Processing*. Neural Network World, 4(6), 447-468.
- [203] Aggarwal C. (2018). *Neural Networks and Deep Learning*. Cham: Springer International
- [204] Bianchi F., Maiorino E., Kampffmeyer M., Rizzi A. & Jenssen R. (2017). *Recurrent Neural Networks for Short-Term Load Forecasting*. Cham: Springer.
- [205] Blokdyk G. (2017). *Recurrent neural network*. Brendale (Australia): The Art of Service.
- [206] Mandic D. & Chambers J. (2001). *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. New York: Wiley.

- [207] Brownlee J. (2018). *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery, E-book: Jason Brownlee.
- [208] Trifa A., Sbail A. & Chaari W. (2017). *Enhancing assessment of personalized multi-agent system through convlstm*. *Procedia Computer Science*, 112, 249–259.
- [209] Yan W.Q. (2020). *Computational Methods for Deep Learning*. Cham: Springer.
- [210] Hu W., Li H., Pan L., Li W., Tao R. & Du Q. (2019). *Feature extraction and classification based on spatial-spectral ConvLSTM neural network for hyperspectral images*. arXiv:1905.03577 [cs.CV].
- [211] Donahue J., Hendricks L., Guadarrama S., Rohrbach M., Venugopalan S., Darrell T. & Saenko K. (2015). *Long-term Recurrent Convolutional Networks for Visual Recognition and Description*. IEEE Conference on Computer Vision and Pattern Recognition, 7-12 July 2015, Boston (MA).
- [212] Hochreiter S. & Schmidhuber J. (1997) *Long short-term memory*. *Neural Computation*, 9(8), 1735–1780.
- [213] Shi X., Chen Z., Wang H., Yeung D., Wong W. & Woo W. (2015). *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting*. arXiv:1506.04214 [cs.CV].
- [214] Nair V. & Hinton G. (2010). *Rectified Linear Units Improve Restricted Boltzmann Machines*. 27th International Conference on Machine Learning. June 21-24, 2010, Haifa (Israel).
- [215] Yi Z. & Tan K. (2004). *Convergence Analysis of Recurrent Neural Networks*. Dordrecht: Springer.
- [216] Goodfellow I., Bengio Y. & Courville A. (2016). *Deep learning*. Cambridge (MA): MIT Press.
- [217] Hassoun M. (1995). *Fundamentals of Artificial Neural Networks*. Cambridge (MA): MIT Press.

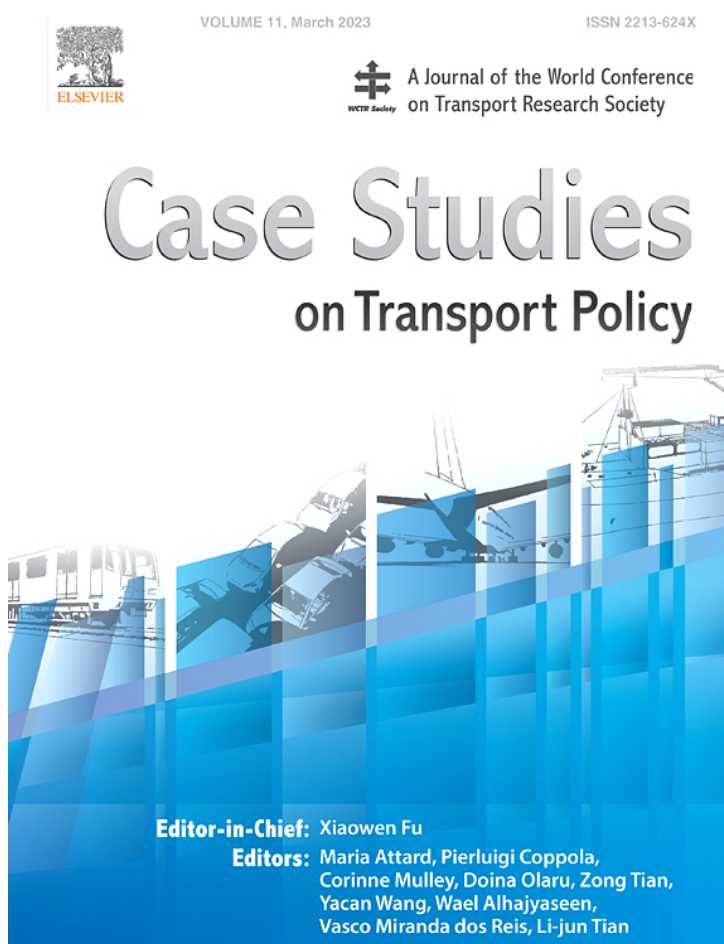
- [218] Kingma D & Ba J. (2014). *Adam: A Method for Stochastic Optimization*.
<https://arxiv.org/abs/1412.6980>
- [219] Patterson J. & Gibson A. (2017). *Deep Learning: A Practitioner's Approach*.
Sebastopol (CA): O'Reilly Media Inc.
- [220] Kim, S. & Kim, H. (2016). *A new metric of absolute percentage error for intermittent demand forecasts*. International Journal of Forecasting, 32(3), 669-679.
- [221] Ren L. & Glasure Y. (2009). *Applicability of the revised mean absolute percentage errors (MAPE) approach to some popular normal and non-normal independent time series*. International Advances in Economic Research, 15(4), 409.
- [222] Guss W. & Salakhutdinov R. (2019). *On universal approximation by neural networks with uniform guarantees on approximation of infinite dimensional maps*. <https://arxiv.org/abs/1910.01545>
- [223] Agarap A. (2019). *Deep Learning using Rectified Linear Units (ReLU)*.
arXiv:1803.08375v2 [cs.NE].
- [224] Arora R., Basu A., Mianjy P., Mukherjee A. (2018). *Understanding deep neural networks with rectified linear units*. arXiv:1611.01491v6 [cs.LG].
- [225] Hara K, Saito D., Shouno H. (2015). *Analysis of function of rectified linear unit used in deep learning*. 2015 International Joint Conference on Neural Networks (IJCNN). <https://doi.org/10.1109/IJCNN.2015.7280578>
- [226] Arora S. & Barak B. (2009). *Computational complexity: a modern approach*.
Cambridge: Cambridge University Press.
- [227] Schonlau M & Zou R. (2020). *The random forest algorithm for statistical learning*. The Stata Journal, 20(1), 3-29.
<https://doi.org/10.1177/1536867X20909688>
- [228] Montgomery D., Peck E., Vining G. (2012). *Introduction to Linear Regression Analysis*. Hoboken (NJ): Wiley.

- [229] Rokach L. (2019). *Ensemble Learning: Pattern Classification Using Ensemble Methods*. Singapore: World Scientific Publishing
- [230] Ganjisaffar Y., Caruana R., Lopes C. (2011). *Bagging Gradient-Boosted Trees for High Precision, Low Variance Ranking Models*. SIGIR'11, July 24–28, 2011, Beijing (China).
- [231] Friedman J. (2000). *Greedy function approximation: A gradient boosting machine*. *Annals of Statistics*, 29, 1189–1232.
- [232] Hanif I. (2019). *Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction*. ICSA 2019, August 02-03, Bogor (Indonesia). <https://doi.org/10.4108/eai.2-8-2019.2290338>
- [233] Chen T. & Guestrin C. (2016). *XGBoost: A Scalable Tree Boosting System*. KDD'16, August 13-17, 2016, San Francisco (USA).
<https://doi.org/10.1145/2939672.2939785>
- [234] Wang L., Zheng C., Zhou W., Zhou W. (2022). *A new principle for tuning-free Huber Regression*. *Statistica Sinica*, 32, 1-25. DOI:
<https://doi.org/10.5705/ss.202019.0045>
- [235] Sun Q., Zhou W., Fan J. (2018). *Adaptive Huber Regression*.
<https://arxiv.org/abs/1706.06991>
- [236] Camana M., Ahmed S., García C., Koo I. (2020). *Extremely Randomized Trees-Based Scheme for Stealthy Cyber-Attack Detection in Smart Grid Networks*. *IEEE*, 20, 19921-19933.
<https://doi.org/10.1109/ACCESS.2020.2968934>
- [237] Geurts P., Ernst D., Wehenkel L. (2006). *Extremely randomized trees*. *Mach Learn*, 63, 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- [238] Izza Y., Ignatiev A., Marques-Silva J. (2020). *On explaining Decision Trees*.
<https://arxiv.org/abs/2010.11034>
- [239] Azadkia M. (2020). *Optimal choice of k for k -nearest neighbor regression*.
<https://arxiv.org/pdf/1909.05495.pdf>

- [240] Jiang X., Wu C., Guo H. (2017). *Forest Pruning Based on Branch Importance*. Computational Intelligence and Neuroscience, Article ID 3162571. <https://doi.org/10.1155/2017/3162571>
- [241] Kulkarni V. & Sinha P. (2012). *Pruning of Random Forest classifiers: A survey and future directions*. International Conference on Data Science & Engineering (ICDSE). DOI: <https://doi.org/10.1109/ICDSE.2012.6282329>
- [242] Rousseeuw P. & van Driessen K. (1999). *Elliptic Envelope. A Fast Algorithm for the Minimum Covariance Determinant Estimator*. Technometrics, 41(3), 212-223.
- [243] global market estimates. <https://www.globalmarketestimates.com/market-report/airline-industry-market-3733>
- [244] *Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models*, by Alejandro Lopez-Lira and Yuehua Tang University of Florida September 8, 2023. <https://arxiv.org/pdf/2304.07619.pdf>
- [245] *Forecast time series like a ChatGPT Prompt Engineer*. Krzysztof Joachimiak. Mayo 2023. <https://medium.com/@joachimiak.krzysztof>
- [246] *5 Best Open Source LLMs*. A. McFarland (2023). <https://www.unite.ai/best-open-source-llms/>

10. LISTADO DE PUBLICACIONES

10.1 Anexo I: Air Cargo Transport Demand Forecasting using ConvLSTM2D, and Artificial Neural Network Architecture Approach



Revista: Case Studies on Transport Policy (NL)

H-Index: 31

Cuartil: Q1

SJR (2022): 0.719

JCR (Factor Impacto): 2.5

Categoría: Transporte

Rank: 34/50

Estado: Publicado

Cita Artículo: Muros Anguita, J. G.; Díaz Olariaga, O. (2023). Air cargo transport demand forecasting using ConvLSTM2D, an artificial neural network architecture approach. *Case Studies on Transport Policy*, Volume 12, 2023, 101009, DOI: <https://doi.org/10.1016/j.cstp.2023.101009>

Air cargo transport demand forecasting using ConvLSTM2D, an Artificial Neural Network architecture approach

Juan Gerardo Muros Anguita¹; Oscar Díaz Olariaga²

¹ Escuela Internacional de Postgrado, Universidad de Granada. Granada, España.

E-mail: e.juangerardo@go.ugr.es

² Facultad de Ingeniería Civil, Universidad Santo Tomas. Bogotá, Colombia.

E-mail: oscardiazolariaga@usta.edu.co

ABSTRACT

The prediction of air traffic demand (passengers and cargo) in a regional/national air transport system is essential. Knowing the behavior of future demand helps, on the one hand, the design and execution of air transport public policies, which, for example, help to focus, guide and prioritize investment (public and private) for the expansion / modernization of airport infrastructures (or development of new airports), act on tariff policies, implement changes in regulatory policy, etc.; on the other hand, it helps airport managers to plan the airport. Therefore, in this paper, a short-term forecast (5 years) of the demand for air cargo transport was carried out, applied to a specific case study (Colombia), considering the most severe pandemic period (the year 2020). To perform the forecast, an approach based on Machine Learning/Deep Learning (ML/DL) method comprising a hybrid of convolutional and recurrent memory neural networks (that allow space-temporal non-linear analysis, such as multi-variable spaces and temporal multi-steps), is presented. The analysis developed here establishes the optimal length of the prediction period; on the other hand, the proposed methodology allows the identification of the most relevant socioeconomic features in the prediction of air cargo demand (domestic and international), i.e., interpreting the ML/DL results obtained through the

variational analysis of different combinations of features. The results show that international air cargo demand is strongly dependent on Gross Domestic Product (GDP) and PCG (Per Capita GDP), while domestic air cargo demand is significantly dependent on PCG. Finally, the results show, for the case study country, very rapid recovery of air cargo demand at pre-pandemic rates (behavior already found in other recent studies and research).

Keywords: air cargo demand, airport, air transport public policy, machine learning, deep learning, convolutional/recurrent neural networks

1. INTRODUCTION

Plans for the development of the various components of the airport system depend to a large extent on the levels of activity expected in the future. To plan the facilities and infrastructure of an airport, or system/set of airports, to meet future needs, it is essential to predict the level and distribution of demand in the various components of the airport system (TRB, 2002). Forecasting demand in an industry as dynamic and sensitive to exogenous factors as aviation is an extremely difficult task. Nevertheless, it is necessary to make air traffic estimates as a prior step to the planning and design of airport facilities, whether of an airport or an airport network (Horonjeff et al., 2010; Wells and Young, 2004).

In Colombia, the country was chosen as the case study for this research, airports (main and regional) have played an essential role in regional connectivity. The country is the seventh largest on the American continent, and has an area of 1.14 million km², with an insular region in the Caribbean Sea 775 km from the Atlantic coast and 42.3% of the continental territory is Amazon rainforest. In addition, the country is crossed, from southwest to northeast, by three mountain ranges of the Andes. Under these conditions, air transport is, in many cases, the only possible alternative for accessibility to several regions of the country, especially those called 'remote, peripheral, and isolated regions' (Díaz Olariaga, 2021). All this without forgetting international connectivity, both for passengers and air cargo (for foreign trade), where the country's main airport, Bogotá-El Dorado (BOG), is one of the first in the Latin American subcontinent by volume of air cargo transported (ACI, 2021; IATA, 2021a). The importance of the national airport network is strengthened if poor geographical coverage of the national road network and a non-existent rail network for the transport of passengers and goods are added to the geographical characteristics of the country (Díaz Olariaga and Carvajal, 2016, 2020). For all these reasons, the Colombian Government, aware of the importance of domestic air transport, has been implementing public and investment policies for nearly three decades (and still in progress) with the aim to improve, expand, and technologically modernize the airport infrastructure in all regions of the country (Díaz Olariaga and Alonso, 2021).

Then, and based on what has just been formulated, understanding the patterns of future demand allows public decision-makers to formulate the most appropriate and focused air public (and investment) policies, and subsequently, recommend consistent airport development programs and project the sources and level of income to support the capital investments to be made (Ashford et al., 2011). And finally, demand forecasting is a basic requirement for policymakers to develop a master plan or strategic plan for an airport system at the local, regional or national level (Janic, 2021, 2009; ICAO, 1987).

To assess the characteristics of future demand, it is necessary to develop reliable forecasts of airport activity. Numerous factors will affect demand, therefore, planners developing demand forecasts must consider, in addition to historical aeronautical data (air traffic), local socioeconomic data, such as national wealth, purchasing power of the inhabitants, demographics (population), industrial production, consumer price index, the exchange rate (of the local currency against the US dollar), etc., as these indicators have a great influence on the behavior of air traffic demand (Rodríguez et al., 2020; Horonjeff et al., 2010; ICAO, 2006),

On the period to forecast, practical experience shows that air traffic forecasts are usually not accurate when forecasts are made in the long (15-20 years) or very long term (25-30 years), but inexorably short to medium term (between 5 and 10 years) forecasts (generally more accurate) are important for the planner and/or decision maker, as they comprise a usual airport (or airport system/network) planning period (ACI, 2016; Kazda and Caves, 2015; de Neufville and Odoni, 2013).

Regarding the main objective of the present research, the demand for air cargo transportation has been experiencing an uninterrupted growth in the last 35 years. More precisely, in the last two decades (up to 2019), and worldwide, the average annual growth rate of the Freight Tonne-Kilometer (FTK) indicator was 4.1%, while that of the Mail Tonne-Kilometer (MTK) was 4.05 %; in 2019 FTK was 225×10^9 , and the Revenue Tonne-Kilometer (RTK) was 1043×10^9 (ICAO, 2020; IATA 2021a, 2021b). The COVID-19 pandemic produced a drop in FTK in 2020, globally, of 18.5% over 2019 (World Bank, 2022; IATA, 2021a, 2021b). Also, several studies are forecasting that air cargo demand (globally), will recover to pre-

pandemic growth rates in the very near term, even before air passenger demand (Gudmundsson et al., 2021; JADC, 2021; Boeing, 2020).

Therefore, the objective of this research is to perform a short-term forecast of air cargo demand, for which the Colombian air market has been used as an application case, with the special feature of including in the analysis the demand data for the year 2020, which has been severely affected by the COVID-19 pandemic, and estimate then, as a complementary result (but of great interest), approximate date of recovery of both the volume of demand and the growth trend of the same to the pre-pandemic period (2019). To achieve such an objective, and as a computational tool, a model derived from artificial neural networks of the ConvLSTM2D type (< Conv> for Convolutional and <LSTM2D> for short-long term memory 2-dimensional) is developed; this type of architecture is a hybrid between convolutional neural networks (CNN), very useful for patterns invariant extraction of the spatial context, and recurrent neural networks (RNN), very appropriate for patterns extraction of features temporal context (Millstein, 2018; Sewak et al., 2018; Yang et al., 2015; Malhotra et al., 2015; Hermans and Schrauwen, 2013). These prediction techniques, based on Machine Learning/Deep Learning (ML/DL), can incorporate more elements of analysis and thus potentially be more effective (Ketkar and Moolayil, 2021). Other advantages of ConvLSTM2D networks, concerning classical methods based on autoregression, are that they support multivariate treatment (several input features) and nonlinear analysis (Pedrycz and Chen, 2020; Calin, 2020). Finally, it should be noted that, in the application to the case study of this research, this type of neural model (ML/DL ConvLSTM2D) represents the abstract knowledge, inferred from the learning of historical patterns of air traffic time series, with which to predict the future evolution of such series.

2. LITERATURE REVIEW

Despite its great importance the analysis or research of air cargo demand forecasting does not present (historically) many publications, or at least not as much as air passenger demand forecasting (Baier et al., 2021). Different approaches for air cargo demand forecasting can be found in the scientific literature, e.g. (the most used in the last decade): classical gravity model, time series models (such as error correction models or ARIMAX)

(Gudmundsson et al. (2021), Madhavan et al. (2020)), Dynamic Linear Models, and approaches using ANN / ML. In the following, we provide a brief overview of these approaches.

Regarding the use of the gravitational model methodology for the forecast of air cargo demand, Alexander and Merkert (2021) propose a model for forecasting international air cargo demand using the U.S. market as an application case; on the other hand, Baier et al. (2021) present a quantitative approach to air cargo forecasting using global airport data (from various regions of the world) in generalized and linearized airport fixed effects gravity models. These models, both at the aggregate and disaggregated levels make it possible to accurately account for certain relevant impacts on the historical development of demand data (in this case, air cargo).

The use of Dynamic Linear Models (DLM) has the following advantages over the usual forecasting methodologies: it detects stochastic trends hidden in the time series, and it also detects structural changes that allow estimating the time-varying effect of exogenous shocks without increasing the number of parameters. In this line, Rodríguez et al. (2020) perform a short-term forecast of air traffic (including air cargo) using DLM.

About approach using ANN / ML, Chen et al. (2012) use backpropagation neural networks (BPNs) to improve the forecast accuracy of passenger and air cargo demand; the authors analyze the factors that influence passenger and air cargo demand. Chou et al. (2011) develop a fuzzy regression forecasting model (FRFM) to forecast air cargo demand (from the current international air cargo market). Li et al. (2020) proposes a new secondary decomposition ensemble (SDE) approach with a 'Cuckoo' search algorithm (CSA). Liu et al. (2020) develop an empirical evaluation of two statistical techniques and three Machine Learning models for air cargo demand forecasting: multiple linear regression (MLR), auto regression integrated moving average (ARIMA), support vector regression (SVR), neural network (NN), and gradient boosting regression tree (GBRT).

The common denominator of the methodologies based on ANN, used for the prediction of air cargo demand (and in general of any air transport indicator), is based on: (a) unlike traditional methods ANNs are data-driven self-adaptive methods in the sense

that there are few a priori assumptions about the models for the problems under study (e.g., not need to perform feature engineering). They learn from examples and capture subtle functional relationships between data, even if the underlying relationships are unknown or difficult to describe. Thus, ANNs are suitable for problems whose solutions require knowledge that is difficult to specify, but for which there is sufficient data or observations. (b) ANNs can generalize, i.e., after knowing the data presented to them, ANNs can often correctly infer the invisible part of a population, even if the sample data contain noisy information. (c) ANNs are universal functional approximators; it has been shown that a network can approximate any continuous function to any desired accuracy. ANNs have more general and flexible functional forms (than those possessed by traditional statistical methods) that they can handle efficiently. (d) Finally, ANNs are nonlinear; prediction has long been the domain of linear statistics. Traditional approaches to time series forecasting, such as the Box-Jenkins or ARIMA method, assume that the time series under study are generated from linear processes, but it is not reasonable to assume. Therefore, it can be stated that ANNs are a more general and flexible modeling tool for forecasting air cargo (or air passenger) demand (Gupta et al., 2019; Dingari et al., 2019; Mostafaeipour et al., 2018; Srisaeng et al., 2015; Bao et al., 2012).

Finally, concerning studies that link or interrelate the development and/or prediction of air cargo demand and the implementation of related air transport public policies (another topic addressed in this research), there are the works cited below. Through a very long-term forecast (30 years), Lakew and Tok (2015) study the relationships between regional economies and air cargo traffic using panel data on airport traffic, employment, wages, and the composition demographics of urban areas in the California region (United States); contributing with this analysis so that decision-makers know about the development and implementation of public policies. Baier et al. (2021) state, in their research, that the reliability of forecast models in the aviation sector is an important factor both for the industry and for the generators of related public policies. The authors also state that airport expansion is an intensive process. in costs and time, and to maintain constant efficiency in the market, attempts are made to accurately anticipate future demand. Therefore, the authors develop a gravity model for air cargo transport demand

forecasting. Suryani et al. (2012) carries out a long-term forecast of air cargo demand (for India as a case study), using System Dynamics as a methodology, to present information to air transport public policy developers, on when and where to start developing plans for investment in airport infrastructure (either expansion or construction of new airports). Li et al. (2020), with their development of air cargo demand forecasting (for the Chinese airport system), using a new secondary decomposition ensemble (SDE) approach with a cuckoo search algorithm (CSA) as a methodology, state that their results contribute to the air transport policy formulation, to overall planning in the air transport environment, and very helpful in acquiring future requirements for strategic air transport options. Rodriguez et al., 2020 carry out a study, based on two approaches, firstly, they evaluate the influence of air transport public policies on the historical development of air traffic (of Colombia, as a country-case study), and secondly, they make a (short-term) forecast of passenger and air cargo demand, using Dynamic Linear Models (DLM) as a methodology, for the main airport of the country-case study, thereby evaluating the (future) impact of current air transport policies. And lastly, Hwang and Shiao (2011) develop a gravity model to study the flow of air cargo transport, and thereby know/identify the factors (socioeconomic, regulatory, geographic, etc.) that could influence international air cargo flows from airports. All of this is in the idea that the demand for air cargo is an important aspect of the planning carried out by the public agencies that manage national air transport and generate *ad hoc* transport policies.

3. APPLICATION CASE

The data for the development of the present research are obtained from the country-case of application (or study) Colombia, currently the third largest air market in the Latin American subcontinent, and fifth in the Americas, by volume of traffic handled (ACI, 2021; IATA, 2021a). In Colombia, the air transport/aviation industry was liberalized in the early 1990s. This brought structural reforms in both the airport sector (leading to the privatization of the main airports in the network (Díaz Olariaga and Pulido, 2019)) and the airline sector, all through an uninterrupted battery of public policies (still in force today) that includes not only normative and regulatory aspects but also aggressive public and

private investment programs in infrastructure and technology (Díaz Olariaga, 2021), and where, on the other hand, airfares were fully liberalized since 2012 (Díaz Olariaga and Zea, 2018). As a result of all these air public policies, since the beginning of the liberalization of the industry (1991) and until 2019, passenger transport (total) grew by almost 800% (led by domestic passenger transport), while air cargo demand (total) in the same period grew by almost 200% (led by international air cargo transport) (Aerocivil, 2022). In terms of infrastructure, it should be mentioned that the Colombian airport network is composed of 58 airports open to commercial traffic (passenger and cargo).

In the economic sphere, to highlight that the GDP growth of the country-case study, in the period 1979-2020, was 850% (Banco de Colombia, 2022); it is thought appropriate to provide this data since several researchers have identified a close and direct relationship between the demand for air cargo transport and the evolution of GDP in the country or region of study (Airbus 2016; Boeing, 2016; Hakim and Merkert, 2016; Morrell, 2011; Ashford et al., 2013; Halpern and Graham, 2018).

About the impact of the COVID-19 pandemic in the country-case study, such circumstance generated, in the worst year of the pandemic, 2020, a resounding drop in air passenger demand (85% in domestic passengers, 75% in international passengers, both concerning 2019); however, air cargo demand (total) only fell by 16% concerning 2019, a concept where the impact of the pandemic was much lower (Díaz Olariaga and Alonso, 2021).

Then, and for the study period 1979-2020 (both inclusive), the historical data of air cargo demand (domestic and international) are obtained from the statistical system of the Colombian Aeronautical Authority (Aerocivil, 2022), and the historical series of socioeconomic indicators (nine in total) from the two related official institutions of the country (Banco de Colombia, 2022; DANE, 2022), both data sets are public and freely available (free of charge).

4. METHODOLOGY

ML/DL techniques such as CNN have been successfully applied to analyze the spatial spectrum of an image in the field of ML/DL computer vision, and ML/DL RNN through network models such as LSTM allows to analyze of features such as the temporal spectrum of video images or just the text sentences in the field of Natural Language Processing (Hu et al., 2019; Tariq et al., 2020). But in the present research, the novelty lies in applying a hybrid combination of both CNN and LSTM, via ConvLSTM2D networks, to extract the behavioral patterns of time series evolution for air cargo prediction, taking into consideration the context of the socioeconomic feature. Statistical methods such as ARIMA have been providing acceptable successes in the problems of time forecasting of a variable, but these classical methods are limited in that they can only perform linear and uni-variable analyses, which does not allow to correctly collect the existing couplings with the other variables of the multi-variable forecast, nor to collect non-linear effects, such as those that occur with the appearance of cycles and sub-cycles within the general trend of the variables (Rodríguez et al., 2020).

This study is based on a dataset composed of eleven input features (two aeronautical and nine socioeconomic) described in historical time series from 1979 to 2020 (inclusive). These eleven variables are listed in Table 1. The variables not only reflect the air transport evolution of the demand for domestic air cargo (DOM) and international air cargo (INT) but are also studied in the context of nine additional socioeconomic features of the country-case study, which aim to provide the historical environment to explain the reasons for the evolution of the demand for air cargo transport. These context variables have been chosen based on their a priori relevance and statistical availability. Then, the method will interpret by assigning the right feature importance to these variables concerning the air cargo transport forecast.

ACRONYM	DESCRIPTION	UNIT
GDP	Gross Domestic Product	US\$
PCG	Per Capita GDP	US\$
POP	Population	No unit
IPI	Industrial Production Index	Dimensionless index
CPI	Consumer Price Index	Dimensionless index
BMR	Benchmark (or Currency) Market Rate	COP (local currency)
FDI	Foreign Direct Investment	US\$
EXP	Exports	US\$
IMP	Imports	US\$
DOM	Domestic Air Cargo	Tonne
INT	International Air Cargo	Tonne

Table 1. Aeronautical and socioeconomic variables of the country-case study were used in the research.

The ML/DL method has been implemented in a proprietary algorithm written in Python and using the open source Keras and Tensorflow libraries, which consists of the process described below.

Step 1 (data preparation). The first thing to do is to normalize the data, that is, to scale the variables on the same scale between 0 and 1 so that all of them are comparable as inputs for the ConvLSTM2D model. In the preparation of the data, a subsequent normalization was implemented, consisting of assigning a value of 0 to the minimum value and 1 to the maximum value of each feature. An additional standardization was applied to this new transformation, which consisted of assigning a standard distribution to the series, with a mean of 0 and a standard deviation of 1. The choice of such transformations of the original dataset is part of heuristic processes within the transformation search spaces, by testing and measuring the target metric to select the best one.

Step 2 (separation of the dataset into tensors of Inputs X and Outputs Y). Since the ConvLSTM2D model is of the supervised type (Millstein, 2018; Sewak et al., 2018; Yang et al., 2015; Malhotra et al., 2015), it is required to separate the historical dataset into 5-dimensional tensors for the inputs X and, 3-dimensional tensors for the searched outputs Y , where the dimensions need to be adapted to the model architecture (e.g., to the network

input layer dimension). The goal of the model is to map or search for a function $f_{model\ forecast}$ (where $\hat{y}$ is denoted with a hat to indicate that it is a forecast or estimate of variable j) from some input variables ($\mathbf{X}$ inputs), as shown in Equation (1).

$$\hat{y}_{j,t+n_outs}, \hat{y}_{j,t+n_outs-1}, \dots, \hat{y}_{j,t} = f_{model\ forecast}(x_{i,t-1}, x_{i,t-2}, \dots, x_{i,t-n_lags}) \quad (1)$$

In Equation (1) the subscript j of $\hat{y}$ denotes the output variable considered, in this work it may take the values DOM or INT corresponding respectively to domestic and international air cargo, but not both at the same time, since a single output variable prediction scenario has been considered for simplicity, although it has been considered a multiple input variable $\mathbf{X}_i$ (multi-variable scenario). The subscript $t+n_outs$ indicate the future prediction steps of the time t , therefore, the prediction $\hat{\mathbf{Y}}$ consists of a vector of n_outs time components. The input variable $\mathbf{X}$ comprises all the input variables considered in each scenario, where each variable is denoted by the subscript i , and for each variable n_lags or previous steps of time t will be considered. In summary, n_outs time steps after t of a single output variable are predicted from the n_lags previous steps at time t of all the input variables considered in each scenario, so that, although 42 historical years from 1979 to 2020 are performed, they are chosen in slots or sequences of fixed time lengths n_outs and n_lags , to pick up the patterns of the time cycles present within the general trends of the time series. That is why the sequences of n_outs and n_lags have been parameterized to be able to vary them. For simplification reasons, in this study, we will choose n_outs equal to n_lags and when parameterizing such sequence values, we will study which is the optimal prediction length value n_outs that provides the minimum MAPE (Mean Absolute Percentage Error). The historical time series starts at the first-time t equal to the first year which may leave behind n_lags in previous years and ends at instant t which may leave $t+n_outs$ future years, so that the $\mathbf{X}$ Input and $\mathbf{Y}$ Output couples necessary for supervised learning, can always be formed.

Also, the temporal data must be structured as 5D $\mathbf{X}$ and 3D dimensional $\mathbf{Y}$ tensors to adapt them to the input and output layers of the ConvLSTM2D neural model respectively

(Ketkar and Moolayil, 2021). The X and Y data are further split up into training and test data. With the training data, the model will be trained, and with the test data, the MAPE model is measured. In the present investigation, a rate of 90% of total data was chosen for the training data versus 10% for the test data, given the scarcity of historical data (Millstein, 2018; Sewak et al., 2018).

Step 3 (definition of the ConvLSTM2D model). The ML/DL model used here is a hybrid model composed of an encoder of the data patterns using CNN and LSTM-type recurrent networks to capture both correlation patterns; on one side the context coupling between all features and on the other side the time series context (Aggarwal, 2018; Bianchi et al., 2017; Blokdyk, 2017; Mandic and Chambers, 2001). The equations describing the context correlations between input variables are those described by Hu et al. (2019) whose feature extraction diagram is shown in Figure 1.

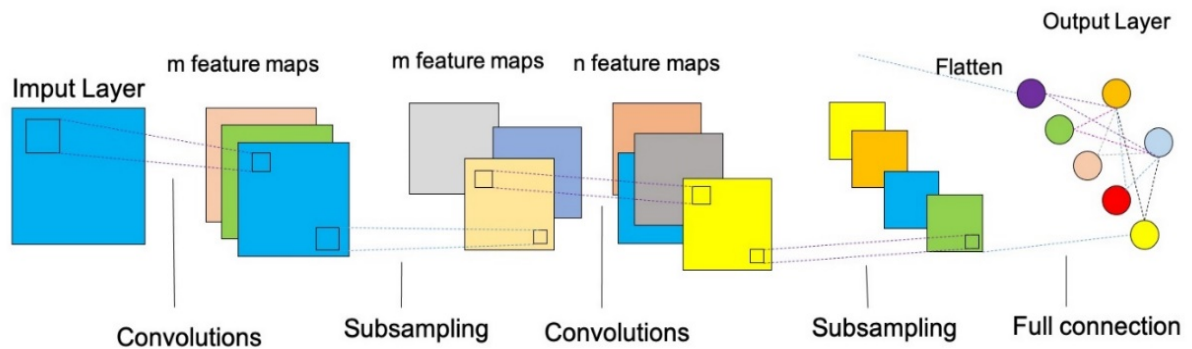


Fig. 1. General diagram to produce feature maps made by convolutional layers. Source: Hu et al., 2019.

On the other hand, LSTM recurrent networks are used to describe the temporal contexts of each feature, where their equations are those described by Donahue et al. (2015), and whose diagram shows the recurrence between short-term and long-term temporal memory is shown in Figure 2.

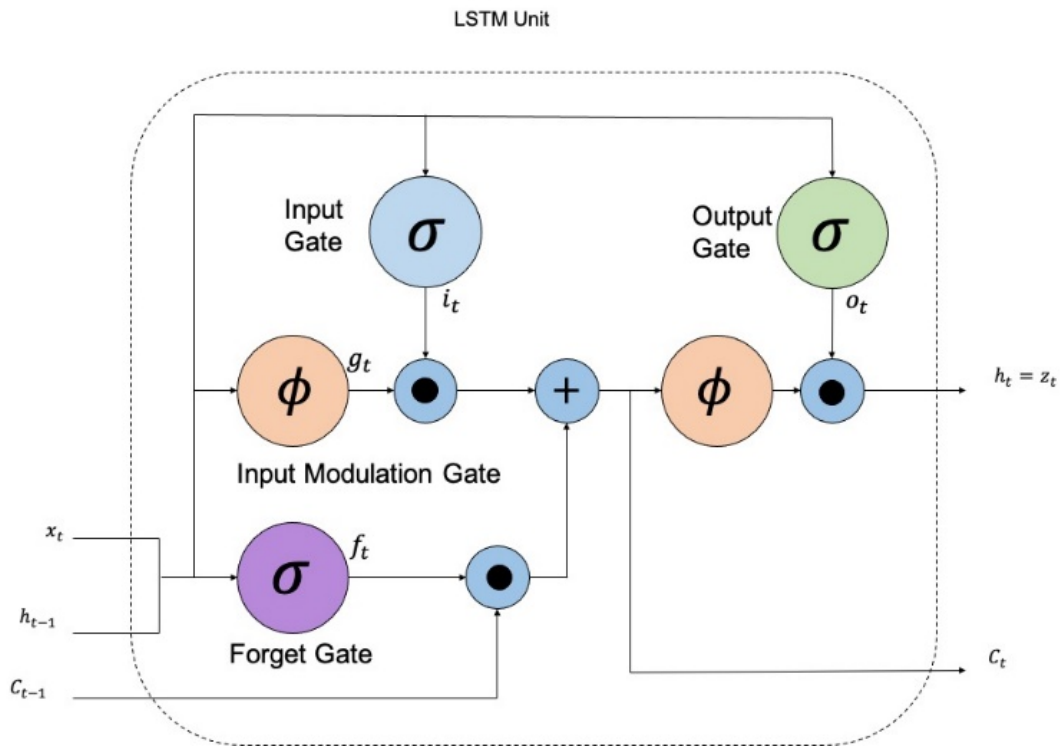


Fig. 2. Schematic of an LSTM long-short-term memory cell or unit. Source: Donahue et al., 2015.

In this study, the ConvLSTM2D model has been implemented in an algorithm written in Python using the open ML/DL libraries of Tensorflow and Keras, parameterizing the architecture, configuration, and quantity of the layers present in the ConvLSTM2D model, to allow later model optimization; in this study, three blocks of hybrid ConvLSTM2D layers have been chosen after performing some heuristic research.

Figure 3 shows the complete and detailed architecture of the ConvLSTM2D model for the particular case of a hyperparameter configuration using only two input features with six previous time steps of n_lags and six predicted time steps n_outs , with two subsequences (or sub-cycles) and three years per subsequence; in this figure the five dimensions of the input tensor $\mathbf{X}$ corresponding respectively to the number of samples, number of subsequences, one variable per column, three years per subsequence and two features. The previous time steps n_lags are the product of the number of subsequences by the years of a subsequence. The output tensor $\mathbf{Y}$ presents only three dimensions corresponding respectively to the number of samples, the time steps to predict n_outs , and

one as the amount of output variable chosen. In Figure 3, it is observed that the inputs of the $\mathbf{X}$ 5D tensor are connected to the input layer which is followed by a ConvLSTM2D layer of 64 convolutional filters, followed by a MaxPooling3D layer, which subsamples or regularizes the output of the previous layer, followed by two more layers of ConvLSTM2D with 128 and 256 convolutional filters respectively, followed by a flatten layer that reduces the multidimensional output into two dimensions, and this is followed by the LSTM decoder block, which needs to be adapted to the three dimensions by including the n_outs dimension, which is done via a previous RepeatVector layer. Finally, two Time Distributed (dense) layers are included to adapt the outputs of the LSTM decoder to the 3D of the $\mathbf{Y}$ output tensor (Calin, 2020). It is important to note that, since the architecture of the ConvLSTM2D model implemented here is parameterized, it is possible to perform a variational study to optimize its architecture by changing the configuration values, quantity, and types of layers in the model.

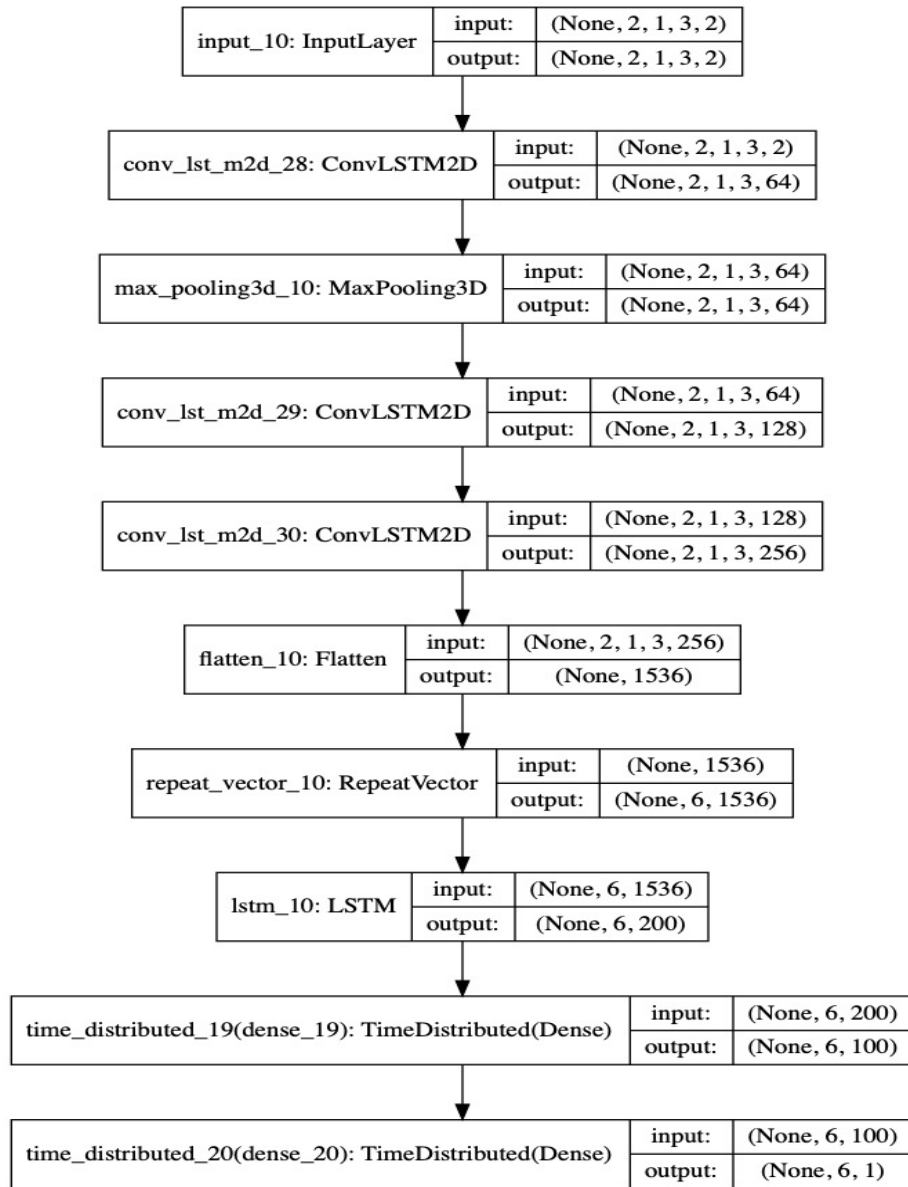


Fig. 3. Architecture implemented in the research, consisting of three blocks of ConLSTM2D layers.

Source: authors.

Step 4 (model compilation, training, and evaluation). Once the LSTM2D model is defined, it is compiled by defining the cost function (to be minimized during training), the metrics with which learning is measured, and the type of gradient optimizer. MSE (Mean Square Error) has been used here as a cost or loss function, together with the metrics RMSE (Root Mean Square Error) and MAPE, whose definitions are given respectively in Equations

2, 3, and 4. Equation (2) defines the MSE error with the norm 2 $\| \cdot \|_2$, where n is the number of samples to obtain the mean.

$$MSE = \frac{1}{n} \sum_x \|y(x) - \hat{y}(x)\|^2 \quad (2)$$

$$RMSE = (MSE)^{1/2} \quad (3)$$

$$MAPE = \frac{100}{n} \sum_x |y(x) - \hat{y}(x)| \quad (4)$$

The training of the model consists of introducing the training tensor $\mathbf{X}$ input in the first input layer to the ConvLSTM model and then going through all the layers up to the output layer, which provides an estimation denoted by $\hat{Y}(x)$, with which to compare the tensor of the expected output $\mathbf{Y}$, which defines the cost function MSE. Using the Backpropagation (BP) process, this MSE is distributed into the different weights of the neurons that make up the model, updating them on each forward-to-back network iteration (Goodfellow et al., 2016). In summary, the BP process calculates the gradient (or derivative) of the total error provided by the cost function by comparing the estimated output with the target output, concerning all the neurons of the model, layer by layer, and applying the chain rule of the derivative to calculate the dependence or sensitivity of the error concerning all the neurons of all the layers, going from front to back, until reaching the first layer. The process is repeated by introducing a new batch of training couples of $\mathbf{X}$ input and $\mathbf{Y}$ output to repeat the forward output estimation cycle, then, the total cost function is calculated, which distributes the error among all the neurons by applying the BP method, with which the neuron weights are updated, and then proceed to load the new batch. When the batches of training couples of $\mathbf{X}$ input and $\mathbf{Y}$ output are finished, the cycle is repeated a certain number of times called epoch (400 cycles were used in this study) or

until a small relative error or a sufficiently accurate level of the model, adjustment is obtained.

Figures 4 and 5 show the change of the MSE during successive training iterations of all batches for both training and test data and with the MSE cost function and the RMSE chosen metric, respectively. Note that the samples were normalized and standardized so their values are not referenced to the original scale of the samples. Note also, how after each iterative cycle is completed, both the error of the training and test samples are obtained, so that the test sample metric error (validation) is very close but not equal to the training data error.

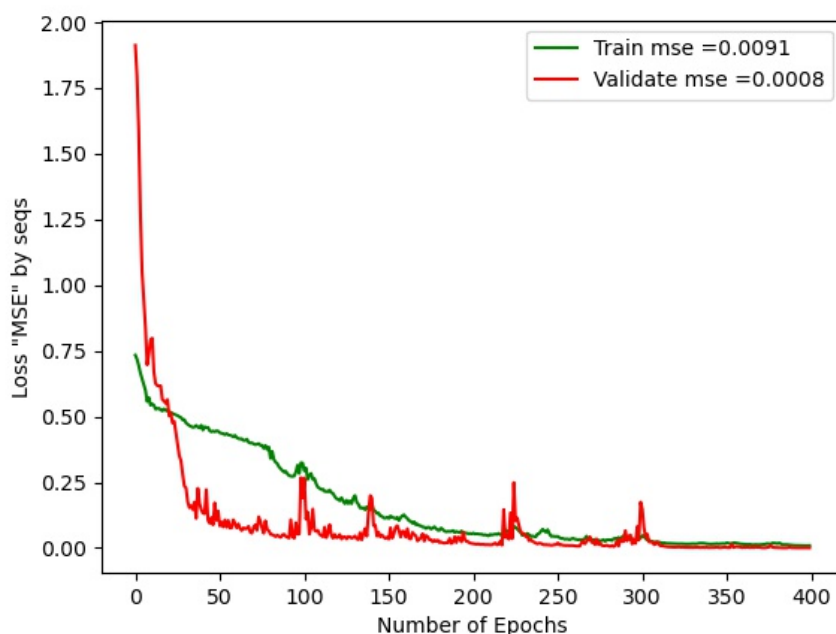


Fig. 4. Evolution of the MSE cost function during training vs. number of epochs (or cycles) for both training and test samples. Source: authors.

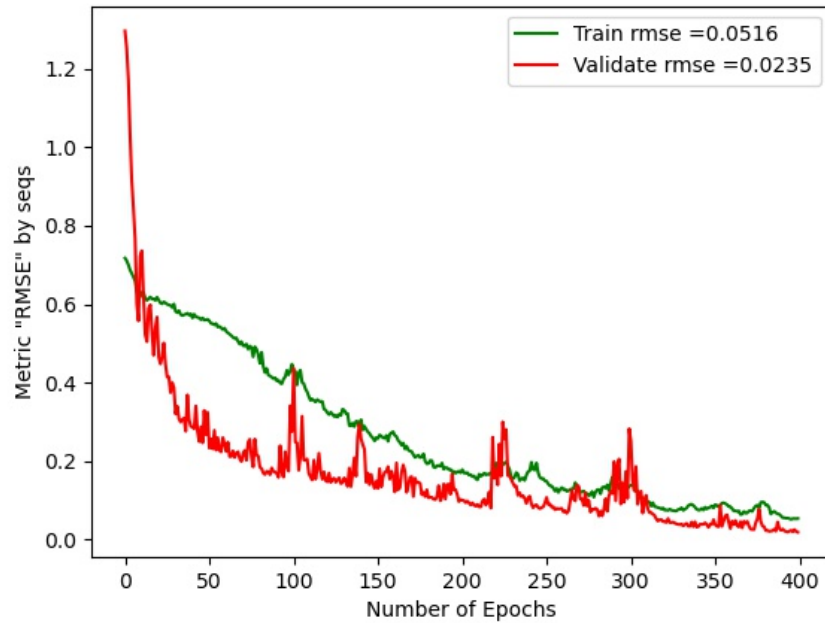


Fig. 5. Evolution of the RMSE metric during training vs. cycle repetition for both training and test samples.

Source: authors.

5. RESULTS

To find the optimal values in the number of years of the sequence to be predicted and to assess the most important features that accompany the best forecast of the variables domestic air cargo (DOM) and international air cargo (INT), a variational study of the model parameters is carried out to choose and forecast the best possible scenario.

5.1. Forecasting international air cargo demand

The first variational analysis implemented consists of searching the sequence length corresponding to the number of n_outs years to be predicted that are optimal in the sense that they provide the lowest MAPE of the international air cargo (INT) variable. This MAPE is measured in the last cycle of n_outs of the historical series. To simplify the problem, we have chosen n_outs (predicted future sequence years) equal to n_lags (previous sequence years), under a scenario with only two input features (GDP and INT itself), but it has been also parameterized the annual sub-sequences into which each sequence of n_lags can be

split of. The choice of the cycles and sub-cycles into which the historical series is grouped is critical to ease how the model extracts the patterns from the historical INT time series and input features context. As ML/DL learning processes are stochastics, the training of each parametric configuration of the model scenarios has been repeated 10 times to provide the statistical result in terms of MAPE mean and standard deviation. Table 2 shows the MAPE results as a function of the different n_outs and the number of sub-sequences, to which corresponds a specific number of years per sub-cycle. It should be noted that these MAPE values refer to a dataset whose values have been scaled, normalized, and standardized and therefore do not correspond to the original scale.

n_lags years	n_Subseqs	subseq years	mean MAPE	std MAPE	Ranking
12	4	3	90.47	30.28	6th best
12	3	4	83.79	47.08	4th best
10	2	5	102.33	38.64	7th best
8	2	4	87.15	53.91	5th best
6	3	2	14.87	4.66	2nd best
6	2	3	8.35	2.51	1st best
4	2	2	16.49	21.36	3rd best

Table 2. Variational analysis to find the optimal lengths of sequences and subsequences within historical series to get the lowest MAPE error in INT prediction.

A first result to highlight (shown in Table 2), is that the optimal number of n_outs , which is equal to that of n_lags , is 6 years, which are taken in 2 sub-sequences of 3 years each. This scenario improves slightly on the 6-year scenario when 3 sub-sequences of 2 years each are taken, which shows the importance of the choice of this sub-sequence or sub-cycle length parameter. The explanation of why the MAPE is higher when fewer years are chosen, i.e., 4 years and 2 sub-cycles of 2 years each, is that when ML/DL models are learning with this shorter length of sequences are not able to appropriately gather all the information that appears in the pattern of 6 years cycles with 2 sub-cycles of 3 years each. Finally, when these cycles are longer, such as 8 or 12 years, the errors are greater, also influenced by the greater number of years to be predicted.

This behavior is imprinted in the structure of the tensors 5D $\mathbf{X}$ inputs and 3D output $\mathbf{Y}$, which is a singular difference between ML/DL time series methodology and statistical methods, where the error, being cumulative, is always higher for a greater number of years to be predicted since they cannot segment their analysis into sequences and sub-sequences as it does the ML/DL ConvLSTM2D approach allowing cycles and subcycles pattern recognition.

The second variational analysis focuses on measuring the importance of the features on INT forecast, also called sensitivity analysis, by determining the combination and features that most impact predicting the demand of INT between the initial set of 11 features, where 9 of them are taken from the socioeconomic context and the other two are domestic (DOM) and international (INT) air cargo. Given the large amount of 11 feature combinations, it has been implemented only a few possible potential scenarios. First, it has considered all possible scenarios of all combinations between 2 features, in addition to considering the simplest scenario of only one feature (INT). The following scenario configurations come from implementing the INT forecast with 3 features, plus 2 more scenarios with 5 features each, one scenario with 10 features, and finally a complete scenario with all 11 features. The complexity of searching into the space of all possible combinations of features is a highlight, so these few have been chosen, but when combinations of several features have been considered, the ones that gave the best forecast for the base case of 2-feature combinations have been chosen. The best mean MAPE and standard deviation (when running each scenario 10 times) is the criterion chosen to select the optimal combination of features. For this purpose, the n_lags and n_outs have been set equal to 6 with 2 sub-sequences of 3 years each, obtained from the previous variational analysis. The results are shown in Table 3.

Input feature scenario	Number of features	Mean MAPE	std MAPE	Ranking
INT	1	13.24	3.84	5th best
GDP, INT	2	8.35	2.51	1st best
PCG, INT	2	9.72	7.18	2nd best
POP, INT	2	22.99	1.5	11th best
IPI, INT	2	23.94	1.76	12th best
CPI, INT	2	26.3	1.6	13th best
BMR, INT	2	44.72	4.43	15th best
FDI, INT	2	12.17	2.51	4th best
EXP, INT	2	20.96	12.52	9th best
IMP, INT	2	11.79	4.6	3rd best
DOM, INT	2	31.14	15.5	14th best
GDP, DOM, INT	3	51.66	17.58	16th best
GDP, EXP, IMP, DOM, INT	5	18,13	2,4	7th best
GDP, PCG, FDI, IMP, INT	5	15,47	1,82	6th best
GDP, PCG, POP, IPI, CPI, BMR, EXP, IMP, DOM, INT	10	21,5	6,96	10th best
GDP, PCG, POP, IPI, CPI, BMR, FDI; EXP, IMP, DOM, INT	11	18,25	3,99	8th best

Table 3. Variational analysis to find the best features and their combinations in terms of better MAPE to forecast international air cargo (INT) demand.

Table 3 shows relevant results, such as that when only one feature is used (the same INT to be predicted) its MAPE is higher than when the INT feature is used together with other features, highlighting the importance of including another feature in the predicting scenario context. It also shows that when all the available features (11 in total) are used, its MAPE prediction is not very good, because some of these features, besides not impacting the international air cargo (INT) feature, introduce noise to the pattern learning, which implies worsening MAPE results. As can be seen in Table 3, the best predictive results are provided by the combinations of two variables (including the one to be predicted), highlighting as a winner the feature GDP (a feature used in other studies (Hakim and Merkert, 2016; Morrell, 2011; Ashford et al., 2013; Halpern and Graham, 2018)), followed by the PCG (Per Capita GDP). The next feature of importance, for determining the

prediction of the INT variable, imports (IMP). The FDI (Foreign Direct Investment) feature is the next in importance. It is again striking, as a finding of the algorithm, that when the 5 best features are combined, the prediction is behind the best predictions made by each one separately. This can be associated with the complexity involved in pattern extraction when several features are considered at the same time.

To conclude this analysis, Figure 6 shows the short-term prediction of international air cargo (INT) demand (in orange), in contrast with the curves of the historical series of international air cargo demand (INT) and GDP (in blue) (as the most relevant feature for the determination of international air cargo demand). The red color identifies the ConvLSTM2D model's predictive line of the historical INT series (and how it learns from less to more, adjusting better in the last years of the historical series). The prediction curve shows a significant recovery but not enough to recover the value existing in the pre-pandemic period, together with a new cycle that will cause international air cargo demand to fall again in the immediate future. This optimal scenario is the result of the previous variational analysis, where n_lags is equal to n_outs and equal to 6 years considering 2 sub-sequences of 3 years each and for the case of 2 features where the most relevant feature for international air cargo is GDP.

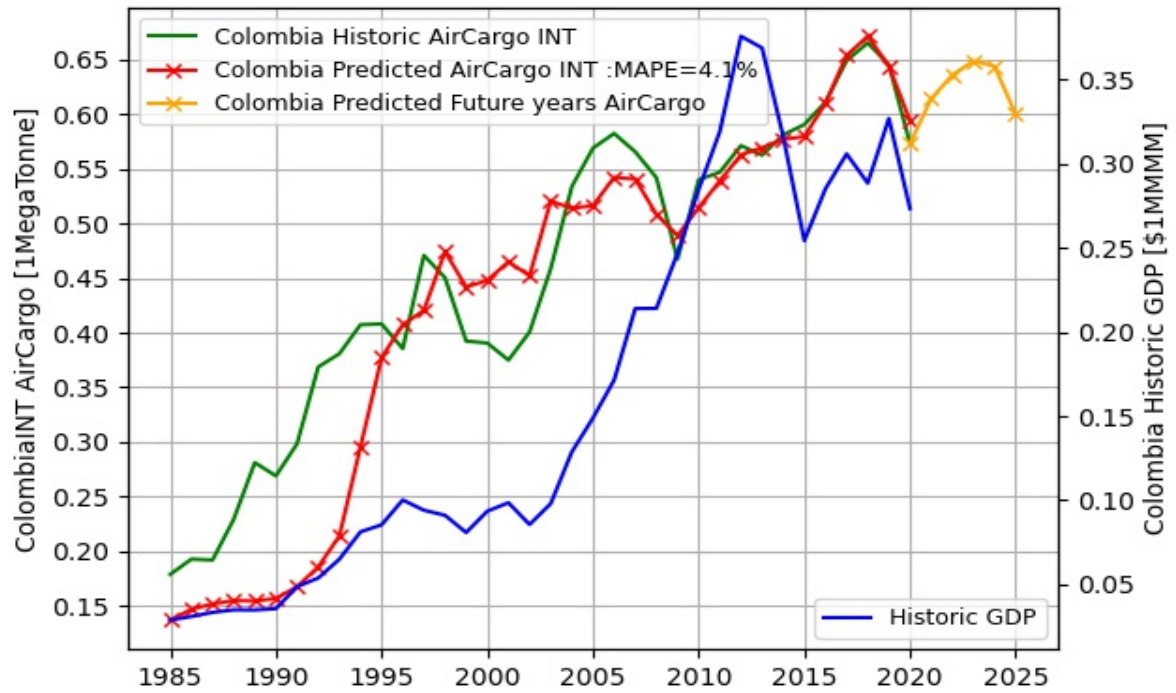


Fig. 6. Forecast of international air cargo demand (in million tons) as a function of the own historic evolution and the history of GDP (in US\$ $\times 10^{12}$).

5.2. Domestic air cargo demand forecasting

The first variational analysis implemented consists of searching the best n_{out} sequence parameter, in the sense of providing the lowest MAPE, measuring this in the last n_{out} cycle of the historical series. To simplify the problem, we have chosen n_{outs} (future sequence years to predict) equal to n_{lags} (previous sequence years), under a scenario with only two features, PCG and the domestic air cargo (DOM) itself, it has also been parameterized the sub-sequences or sub-cycles into which each sequence of n_{lags} can be split of. The choice of the cycles and sub-cycles into which the historical series is grouped is important to ease time series pattern extractions by the ML/DL ConvLSTM model. As the learning process of the ML/DL models is stochastic, the training of each parametric configuration of the scenarios has been repeated 10 times to obtain the statistical mean and standard deviation of the MAPE. Table 4 shows the MAPE results as a function of the

different n_outs and the number of sub-sequences. It should be noted that these MAPE values refer to a dataset whose values have been scaled, normalized, and standardized and therefore do not correspond to the original scale.

n_lags years	$n_Subseqs$	subseq years	mean MAPE	std MAPE	Ranking
4	2	2	68.68	1.73	3rd best
8	2	4	65.83	3.46	2nd best
12	3	4	187.88	62.25	3rd best
12	4	3	209.31	55.05	4th best
6	2	3	42.05	2.14	1st best

Table 4. Variational analysis to find the optimal lengths of sequences and sub-sequences that provides the lowest MAPE in the prediction of domestic air cargo (DOM).

A first highlight in the prediction of DOM scenarios (shown in Table 4) is that MAPE errors are more than five times higher than international air cargo INT prediction. This may be a consequence of the possible high variability (and/or volatility) intrinsic to the historical time series of the domestic air cargo (DOM) feature. This implies that the learning of behavioral patterns is a more complex problem for DOM prediction than for INT. It can be seen from Table 4 that the number of n_outs , which is equal to the optimal n_lags , is also 6 years, which are taken, as in the INT case, in 2 sub-sequences of 3 years each. The explanation of why a higher MAPE has resulted when fewer years are chosen, e.g., 4 years and 2 sub-cycles of 2 years, is that this shorter sequence length is not able to include all the information appearing in the cycles of 6 years with 2 sub-cycles of 3 years. Finally, when these cycles are longer, such as 8 or 12 years, the errors are greater because of a higher cumulative number of years to predict. This behavior is collected in the structure of tensors 5D $\mathbf{X}$ inputs and the 3D output $\mathbf{Y}$ and it is unique compared to statistical methods, where the error, being cumulative, is always greater for a greater number of years to predict since they cannot segment their analysis into sequences and sub-sequences time series as it does ML/DL ConvLSTM2D model.

The second variational scenario implemented, also called sensitivity analysis, aims to measure the importance of the features, and the best combination between the 11 features to ease the DOM prediction, 9 of which belong to socioeconomic context. For this

purpose and given the great multiplicity of possible combinations of features, it has been implemented the following scenarios. First, all possible combinations of 2 features have been carried out, in addition to considering the simplest scenario of one feature alone (DOM itself). Additionally, a scenario with 3 features, another scenario with 4 features, and the last scenario with all the features (11) have been analyzed. The best mean MAPE and standard deviation (when running each scenario 10 times) is the criterion chosen to select the optimal combination of features. For this purpose, the n_lags and n_outs have been set to 6 with 2 sub-sequences of 3 years, obtained from the previous variational analysis. The results are shown in Table 5.

Input feature scenario	Number of features	mean MAPE	std MAPE	Ranking
DOM	1	66.41	10.31	5th best
GDP, DOM	2	51.18	3.70	3rd best
PCG, DOM	2	42.05	2.14	1st best
POP, DOM	2	133.34	17.19	12th best
IPI, DOM	2	135.37	12.77	13th best
CPI, DOM	2	138.62	11.52	14th best
BMR, DOM	2	95.45	10.50	10th best
FDI, DOM	2	69.00	7.32	7th best
EXP, DOM	2	42.14	4.87	2nd best
IMP, DOM	2	71.47	16.55	8th best
INT, DOM	2	121.80	11.00	11th best
PCG, EXP, DOM	3	74.11	14.76	9th best
PCG, EXP, FDI, DOM	4	54.56	2.66	4th best
GDP, PCG, POP, IPI, CPI, BMR, FDI; EXP, IMP, INT, DOM	11	67.96	4.46	6th best

Table 5. Variational analysis to find the best features and their combinations to predict domestic air cargo demand (DOM).

The first highlight result that Table 5 shows is that when using only one feature (the same feature DOM to be predicted) its MAPE is higher (ranking fifth see table 5) than when using the DOM feature together with other features, underlining the importance of applying time series socioeconomic context to improve DOM forecast. The results also

show that, when all the available features (11) are used, its MAPE prediction is not very good, because some of these features, in addition to not impacting the DOM variable, introduce noise to learning patterns extraction, which implies worsening MAPE metric. As shown in Table 5, the best predictive results are provided by combinations of two variables (including the one DOM to be predicted). The winning feature is the per capita GDP income (PCG), with which DOM is most strongly linked, followed by the EXP feature (exports). The next most important feature in determining the prediction of DOM is GDP. The fourth feature in importance is the combination of the 4 features PCG, EXP, FDI, and DOM, which means that the model finds relevant patterns in the combination of these 4 features. The fifth feature in importance is to consider only the (historical) time series of domestic air cargo demand, which evidences the positive contribution of combining them with other features as opposed to considering only the DOM isolated series. Finally, it should be noted that DOM is sensitive to incorporating more features to predict the DOM itself, compared to the coupling to other isolated features such as FDI, IMP, IPI, CPI, etc. This is another differential finding of domestic air cargo demand concerning international air cargo demand predictions.

Finally, Figure 7 shows DOM prediction in the short term (in orange color), in contrast with the curves of the historical series of DOM (in green color) and PCG (in blue color) as the most relevant feature that impacted DOM prediction. The red color identifies the predictive curve made by the ConvLSTM2D model of the DOM historical series (it shows how this curve is learning from less to more, adjusting better in the last years of the historical series).

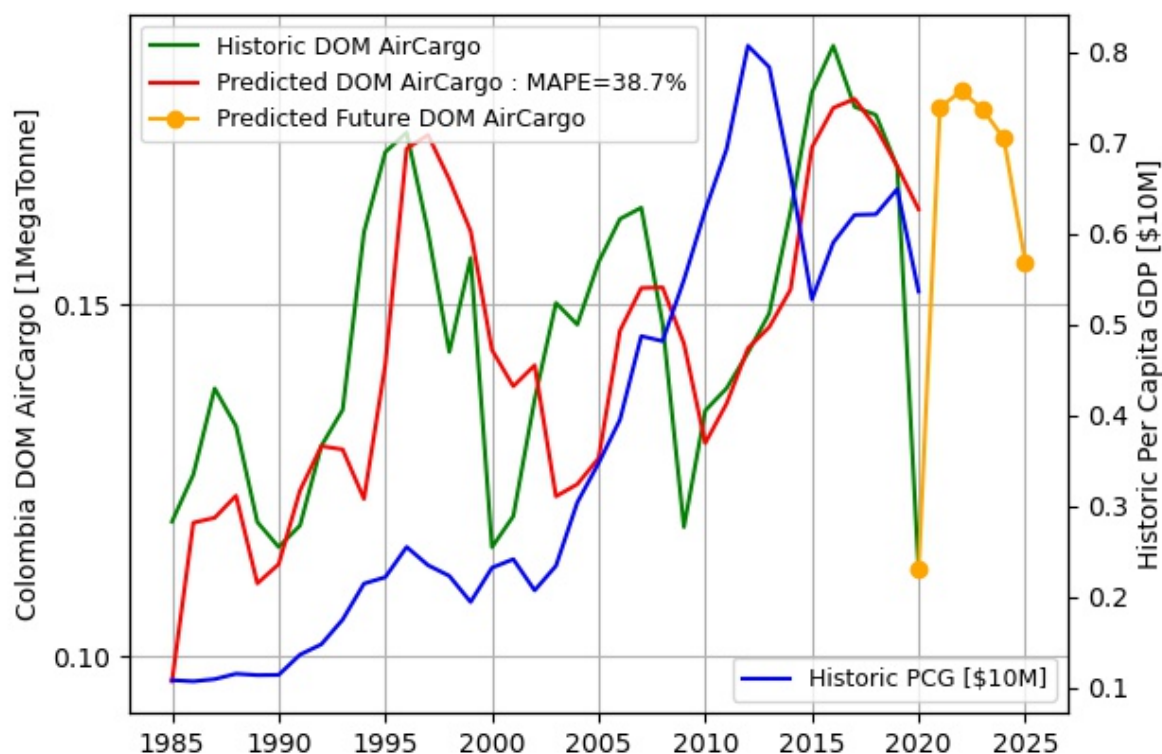


Fig. 7. Forecast of domestic air cargo demand (DOM) (in million tons) as a function of the historical series of the variable itself and the historical evolution of PCG (in US\$ $\times 10^4$).

6. DISCUSSION ON METHODOLOGY

The main objective of this research is to make a short-term forecast of air cargo transport demand, applied to a particular case study (Colombian air transport system), and to estimate the behavior of the demand recovery for the next post-pandemic period. Due to the non-linear characteristics of air cargo demand, classical time series such as econometric-statistical approaches are currently not considered the most convenient methodology, as these approaches are severely criticized due to their poor and limited forecasting capacity (Suryani et al., 2012; Tsui et al., 2014; Rodríguez et al., 2020; Li et al., 2020; Liu et al., 2020; Tascón and Díaz Olariaga, 2021; Ensafi et al., 2022). For this reason, a methodology based on a type of artificial neural network architecture called

ConvLSTM2D is proposed, which, although they have been very successful in areas of Machine Learning (such as computer vision and natural language processing (Alayba and Palade, 2022; Chaiani et al., 2022; Elboushaki et al., 2020; Fang et al., 2021; Kumar et al., 2022; Xingjian et al., 2015)), have not yet been tested in time series forecasting of air traffic (such as demand for air cargo transport). The reason for choosing this methodology lies in the fact that these artificial neural network architectures have shown to be very robust and successful in the fields of deep learning mentioned, from which they come, by automatically extracting the intrinsic patterns of the non-linear relationships between the variables considered, without a priori knowledge about the existing relationships between the input variables among themselves and between these and the output considered, so, in short, it is estimated that they can be a viable and promising tool to be explored for new flexible modeling of the forecast (Gupta et al., 2019; Dingari et al., 2019). Therefore, this research pursues (with a certain specific scope) to demonstrate the feasibility of applying these artificial neural network models defined by the ConvLSTM2D architecture, which are of the Deep Learning (DL) type, and show it can be obtained acceptable and hopefully results when applied to multivariate time series forecasting such as our air cargo demand forecasting (Agga et al., 2022; Ensafi et al., 2022; Huang et al., 2022; Prince, 2022; Shastri et al., 2020).

In another order, it is considered appropriate to make certain observations on the development of the proposed model and/or its operation. In the first place, it should be mentioned that the validation data group can coincide with the test data group when a single model is being tested (as in the present study) and not several models at the same time (which is when there must be three independent training, validation and test sets), since the validation set is used to show the differences in how the evaluation of each Machine Learning model progresses during the training process, while the use of the test dataset remains reserved to measure the final performances of the model (once learning process is finished), and therefore the objectivity of the results obtained with the metrics used to check the single model, such as the MAPE / RMSE errors defined for this study, is guaranteed. Mention that, in any case, the validation and/or test data have never been used to train the model, thus preserving the objectivity of the results obtained with this

method. In other words, for the unique ConvLSTM2D model developed here, it is sufficient to use only two independent training and test sets to perform the training and final evaluation of the model, since no differential information is lost concerning other models (since the model tested was unique) (Bai et al., 2021; Tennenholtz et al., 2018; Lones, 2021).

Finally, it should also be mentioned that the data of the validation set (independent of the test data) are also needed to tune the hyperparameters of the model (which has been defined as an optimization process of the model architecture), but having defined the specific scope of this study to demonstrate uniquely the forecasting feasibility of the model proposed here, it is not necessary to use any other independent set of validation data (Cawley and Talbot, 2010; Wang et al., 2021; Yang and Shami, 2020). Furthermore, as the dataset available for the study is small, it was decided to generate only two independent data sets for training and testing. Where the test one was also used as validation to show the evaluation of the learning unique model, although this evaluation was not necessary, since there was only one model o test and without having other additional objectives (such as tuning hyperparameters or optimizing the architecture of the model used) (Paullada et al., 2020).

7. CONCLUSIONS

Undoubtedly, concerning the development of the demand for air cargo transport for the period pre-pandemic (as shown by the indicators presented here), the liberalization of the air transport sector in Colombia, which began in the early 1990s, and the continuous development of public policies for the sector, as well as public investment policies and in airport infrastructure that involved almost the entire airport network, contributed to the progressive improvement (expansion/modernization) of the structure of the national airport network. This fact benefits connectivity and territorial cohesion by air and the development of the regions where the airports are located. Then, forecasting air cargo demand is extremely important for both air operators and public airport authorities. Forecasts as accurate as possible are essential for carriers for short- and medium-term planning (for route design, fleet, operational, financial, and personnel planning, etc.). For

national aeronautical authorities and/or airport managers (individual or network), having accurate forecasts (which is usually already done through the airport master plan) helps to make timely decisions (in time and form) on investment policies (public or private) for the expansion and modernization of airports (or construction of new ones), as well as contributing to the development of tariff policies that encourage the attraction of new air operators and/or the opening of new routes. This has been a highlight objective of this research.

A second objective of the present study was to estimate the recovery pattern of air cargo demand (domestic and international) in the country-case study, considering the strong impact of the global COVID-19 pandemic in the year 2020 (whose data were included in the research). The results show a rapid recovery of air cargo demand, very similar (although slightly faster) to that of air passenger demand, as compared to recent research on the case study country (Muros and Díaz Olariaga, 2021).

Regarding the methodology proposed for the analysis, the main contribution of this research to the scientific literature is the suitability of the use of ML/DL methods such as ConvLSTM2D on the prediction of time series applied to air transport compared to classical statistical methods, as it can incorporate nonlinear and multivariate analysis with the extraction of patterns by segments of cycles and subcycles of the historical series. Followed by the novelty to apply these successful ML/DL methods to the air transport field, that were originated in the field of computer vision and natural language processing. The variational analysis incorporated aim to determine the optimal lengths of cycles and sub-cycles (to extract better behavioral patterns from historical series) together with the optimal determination of the features and the combinations between them that provides the optimal domestic and international air cargo demand prediction from the best socioeconomic context scenarios. The ML/DL methodology developed here incorporates a singular feature analysis to be able to act on the variables that most influence air cargo transport demand, and where the most relevant singularities of dependencies have been shown.

Future research invites to consider other configurations of ANN architecture using these ConvLSTM2D layers as the core of the ML/DL model, but changing the types of layers, quantity, and their configurations, to improve predictive MAPE. Other lines of future research that are proposed are to carry out comparative studies with other time series prediction methods (for example, with other artificial neural network models and legacy statistical approaches), to fine-tune the hyperparameters to optimize the architecture of the ConvLSTM2D model proposed here, and finally expand and/or vary the air cargo demand dataset (also including other socioeconomic variables), for whose calculations it will be necessary to use the three independent data types related to training, validation and test data.

8. REFERENCES

- ACI (2021). *Annual World Airport Traffic Report*. Montreal: ACI (Airports Council International).
- ACI (2016). *Traffic Forecast*. Montreal: ACI (Airports Council International).
- Aerocivil (2022). <https://www.aerocivil.gov.co/atencion/estadisticas-de-las-actividades-aeronauticas/boletines-operacionales>
- Airbus (2016). *Global Market Forecast 2016–2035*. Blagnac: Airbus.
- Agga, A.; Abbou, A.; Labbadi, M.; El Houm, Y.; Ali, I. (2022). CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electric Power Systems Research*, 208, 107908.
<https://doi.org/10.1016/j.epsr.2022.107908>
- Aggarwal, C. (2018). *Neural Networks and Deep Learning*. Cham: Springer International.
- Alayba, A.; Palade, V. (2022). Leveraging Arabic sentiment classification using an enhanced CNN-LSTM approach and effective Arabic text preparation. *Journal of King Saud University – Computer and Information Sciences*.
<https://doi.org/10.1016/j.jksuci.2021.12.004>

- Alexander, D.; Merkert, R. (2021). Applications of gravity models to evaluate and forecast US international air freight markets post-GFC. *Transport Policy*, 104, 52–62. DOI: 10.1016/j.tranpol.2020.04.004
- Ashford, N.; Stanton, M.; Moore, C.A.; Coutu, P.; Beasley, J. (2013). *Airport Operations*. New York: McGraw-Hill.
- Ashford, N.; Mumayiz, S.; Wright, P. (2011). *Airport Engineering*. New Jersey: John Wiley & Sons.
- Bai, Y.; Chen, M.; Zhou, P.; Zhao, T.; Lee, J.; Kakade, S.; Wang, H.; Xiong, C. (2021). How Important is the Train-Validation Split in Meta-Learning? arXiv:2010.05843v2 [cs.LG].
- Baier, F.; Berster, P.; Gelhausen, M. (2021). Global cargo gravitation model: airports matter for forecasts. *International Economics and Economic Policy*. DOI: 10.1007/s10368-021-00525-2
- Banco de Colombia (2022). Estadísticas. <https://www.banrep.gov.co/es/-estadisticas>
- Bao, Y.; Xiong, T.; Hu, Z. (2012). Forecasting Air Passenger Traffic by Support Vector Machines with Ensemble Empirical Mode Decomposition and Slope-Based Method. *Discrete Dynamics in Nature and Society*, ID 431512, 1-12. doi:10.1155/2012/431512
- Bianchi, F.; Maiorino, E.; Kampffmeyer, M.; Rizzi, A.; Jenssen, R. (2017). *Recurrent Neural Networks for Short-Term Load Forecasting*. Cham: Springer.
- Blokdyk, G. (2017). *Recurrent neural network*. Brendale (Australia): The Art of Service.
- Boeing (2020). *World Air Cargo Forecast 2020-2039*. Seattle: Boeing.
- Boeing (2016). *World Air Cargo Forecast 2016–2035*. Seattle: Boeing.
- Calin, O. (2020). *Deep Learning Architectures. A Mathematical Approach*. Cham: Springer.
- Cawley, G.; Talbot, N. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*, 11, 2079-2107.

Chaiani, M.; Selouani, S.; Boudraa, M.; Yakoub, M. (2022). Voice disorder classification using speech enhancement and deep learning models. *Biocybernetics and Biomedical Engineering*, 42, 463– 480. <https://doi.org/10.1016/j.bbe.2022.03.002>

Chen, S.; Kuo, S.; Chang, K.; Wang, Y. (2012). Improving the forecasting accuracy of air passenger and air cargo demand: the application of back-propagation neural networks. *Transportation Planning and Technology*, 35(3), 373-392. DOI: 10.1080/03081060.2012.673272

Chou, T.; Liang, G.; Han, T. (2011). Application of fuzzy regression on air cargo volume forecast. *Quality and Quantity*, 45, 1539–1550. DOI: 10.1007/s11135-010-9342-8

DANE - Departamento Administrativo Nacional de Estadística (Colombia) (2022). <https://www.dane.gov.co/>

de Neufville, R.; Odoni, A. (2013). *Airport Systems, Planning, Design, and Management*. New York: McGraw-Hill.

Díaz Olariaga, O. (2021). The role of regional airports in connectivity and regional development. *Periodica Polytechnica Transportation Engineering*, 49(4), 1-13. DOI: 10.3311/PPtr.16557

Díaz Olariaga, O.; Alonso, C. (2021). Impact of airport policies on regional development. Evidence from the Colombian case. *Regional Science Policy & Practice*, 1–26. DOI: 10.1111/rsp3.1248326

Díaz Olariaga, O.; Zea, J.F. (2018). Influence of the liberalization of the air transport industry on configuration of the traffic in the airport network. *Transportation Research Procedia*, 33, 43-50.

Díaz Olariaga, O.; Pulido, L. (2019). Measurement of airport efficiency. The case of Colombia. *Transport and Telecommunication*, 20(1), 40-51

Díaz Olariaga, O.; Carvajal, A.F. (2020). Perspectiva geográfica del desarrollo de la conectividad aérea en Colombia. *Boletín Geográfico*, 42(2), 145 - 168

Díaz Olariaga, O.; Carvajal, A.F. (2016). Efectos de la liberalización en la geografía del transporte aéreo en Colombia. *Cuadernos Geográficos*, 55(2), 344-364.

Dingari, M.; Reddy, M.; Sumalatha, V. (2019). Air Traffic Forecasting Using Artificial Neural Networks. *International Journal of Scientific & Technology Research*, 8(10), 556-559.

Donahue, J., Hendricks, L., Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Darrel, T.; Saenko, K. (2015). Long-term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Conference on Computer Vision and Pattern Recognition*, 7-12 July 2015, Boston (MA).

Elboushaki, A.; Hannane, R., Afdel, K.; Koutti, L. (2020). MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences. *Expert Systems With Applications*, 139, 112829.
<https://doi.org/10.1016/j.eswa.2019.112829>

Ensafi, Y.; Amin, S.; Zhang, G.; Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning –A comparative analysis. *International Journal of Information Management Data Insights*, 2, 100058.
<https://doi.org/10.1016/j.jjime.2022.100058>

Fang, Y., Zhang, H., Zuo, Y., Jiang, W.; Huang, H.; Yan, J. (2021). Visual attention prediction for autism spectrum disorder with hierarchical semantic fusion. *Signal Processing: Image Communication*, 93, 116186. <https://doi.org/10.1016/j.image.2021.116186>

Goodfellow, I.; Bengio, Y.; Courville, A. (2016). *Deep learning*. Cambridge (MA): MIT Press.

Gudmundsson, S.V.; Cattaneo, M.; Redondi, R. (2021). Forecasting temporal world recovery in air transport markets in the presence of large economic shocks: The case of COVID-19. *Journal of Air Transport Management*, 91, 102007. DOI: 10.1016/j.jairtraman.2020.102007

Gupta, V.; Sharma, K.; Sangwan, M. (2019). Airlines passenger forecasting using LSTM based recurrent neural networks. *International Journal Information Theories and Applications*, 26(2), 178-187.

Hakim, M.M.; Merkert, R. (2016). The causal relationship between air transport and economic growth: empirical evidence from South Asia. *Journal of Transport Geography*, 56, 120–127.

- Halpern, N.; Graham, A. (2018). *Air Transport Management*. New York: Routledge.
- Hermans, M.; Schrauwen, B. (2013). Training and analyzing deep recurrent neural networks. *Working Paper*, Ghent University.
- Horonjeff, R.; McKelvey, F.; Sproule, W.; Young, S. (2010). *Planning and Design of Airports*. New York: McGraw-Hill.
- Hu, W.; Li, H.; Pan, L.; Li, W.; Tao, R.; Du, Q. (2019). Feature Extraction and Classification Based on Spatial-Spectral ConvLSTM Neural Network for Hyperspectral Images. *IEEE Transactions on Geoscience and remote sensing*. arXiv:1905.03577v1 [cs.CV].
- Huang, X.; Li, Q.; Tai, Y.; Chen, Z.; Liu, J.; Shi, J.; Liu, W. (2022). Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. *Energy*, 246, 123403. <https://doi.org/10.1016/j.energy.2022.123403>
- Hwang, C.; Shiao, G. (2011). Analyzing air cargo flows of international routes: an empirical study of Taiwan Taoyuan International Airport. *Journal of Transport Geography*, 19, 738–744. DOI: 10.1016/j.jtrangeo.2010.09.001
- IATA (2021a). *World Air Transport Statistics 2021*. Geneva: IATA.
- IATA (2021b). *Air Cargo Market Analysis*. Geneva: IATA.
- ICAO (2020). Presentation of 2019 Air Transport Statistical Results. Montreal: ICAO. https://www.icao.int/annual-report-2019/Documents/ARC_2019_Air%20Transport%20Statistics_es.pdf
- ICAO (2006). *Manual on Air Traffic Forecasting. Doc 8991*. Montreal: ICAO.
- ICAO (1987). *Master Planning. Part 1. Doc 9184*. Montreal: ICAO.
- JADC (2021). *Worldwide Market Forecast 2020-2040*. Tokyo: Japan Aircraft Development Corporation.
- Janic (2021). *System Analysis and Modelling in Air Transport*. Boca Raton: CRC Press
- Janic, M. (2009). *Airport analysis, planning, and design: demand, capacity, and congestion*. New York: Nova Science Publishers.

- Kazda, A.; Caves, R. (2015). *Airport design and operations*. Bingley: Emerald.
- Ketkar, N.; Moolayil, J. (2021). *Deep Learning with Python*. New York: Springer.
- Kumar, S.; Rajesh, D.; Pranesh, S., Kollipara, H., Agrawal, G.; Anbarasi, M.; Valarmathi, J. (2022). Classification of Indian Media Titles using Deep Learning Techniques. *International Journal of Cognitive Computing in Engineering*.
<https://doi.org/10.1016/j.ijcce.2022.04.001>
- Lakew, P.; Tok, Y. (2015). Determinants of air cargo traffic in California. *Transportation Research Part A*, 80, 134–150. DOI: 10.1016/j.tra.2015.07.005
- Li, H.; Bai, J.; Cui, X.; Li, Y.; Sun, S. (2020). A new secondary decomposition-ensemble approach with cuckoo search optimization for air cargo forecasting. *Applied Soft Computing Journal*, 90, 106161. DOI: 10.1016/j.asoc.2020.106161
- Liu, J.; Ding, L.; Guan, X.; Gui, J., Xu, J. (2020). Comparative analysis of forecasting for air cargo volume: Statistical techniques vs. machine learning. *Journal of Data, Information and Management*, 2, 243–255. DOI: 10.1007/s42488-020-00031-1
- Lones, M. (2021). How to avoid machine learning pitfalls: a guide for academic researchers. arXiv:2108.02497v1 [cs.LG].
- Madhavan, M.; Sharafuddin, M.; Piboonrungrroj, P., Yang, C. (2020). Short-term Forecasting for Airline Industry: The Case of Indian Air Passenger and Air Cargo. *Global Business Review*. DOI: 10.1177/0972150920923316
- Malhotra, P.; Vig, L.; Shroff, G.; Agarwal, P. (2015). Long-short term memory networks for anomaly detection in time series. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. 22-24 April 2014, Bruges.
- Mandic, D.; Chambers, J. (2001). *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. New York: Wiley.
- Millstein, F. (2018). *Convolutional Neural Networks in Python*. Scotts Valley: CreateSpace Independent Publishing Platform.

Morrell, P.S. (2011). *Moving Boxes by Air: The Economics of International Air Cargo*. Aldershot: Ashgate.

Mostafaeipour, A.; Goli, A.; Qolipour, M. (2018). Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study. *The Journal of Supercomputing*, 74(10), 5461-5484. <https://doi.org/10.1007/s11227-018-2452-0>

Muros Anguita, J.G.; Díaz Olariaga, O. (2021). Utilización de algoritmos de redes neuronales artificiales en el pronóstico de la demanda de pasajeros aéreos. In Serna, E. (Ed.), *Desarrollo e Innovación en Ingeniería Vol. I* (pp. 277-294). Medellín (Colombia): Editorial Instituto Antioqueño de Investigación. DOI: 10.5281/5513899

Paullada, A.; Raji, I.; Bender, E.; Denton, E.; Hanna, A. (2020). Data and its (dis)contents: A survey of dataset development and use in machine learning research. arXiv:2012.05345v1 [cs.LG].

Pedrycz, W.; Chen, S. (2020). *Deep Learning: Concepts and Architectures*. Cham: Springer.

Prince, A. (2022). Convolutional neural network-long short-term memory optimization for accurate prediction of airflow in a ventilation system. *Expert Systems With Applications*, 195, 116618. <https://doi.org/10.1016/j.eswa.2022.116618>

Rodríguez, Y.; Pineda, W.; Díaz Olariaga, O. (2020). Air traffic forecast in post-liberalization context: a Dynamic Linear Models approach. *Aviation*, 24(1), 10-19. DOI: 10.3846/aviation.2020.12273

Sewak, M.; Karim, R.; Pujari, P. (2018). *Practical Convolutional Neural Networks*. Birmingham: Packt Publishing.

Shastri, S.; Singh, K.; Kumar, S.; Kour, P.; Mansotra, V. (2020). Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons and Fractals*, 140, 110227. <https://doi.org/10.1016/j.chaos.2020.110227>

Srisaeng, P.; Baxter, G.; Wild, G. (2015). Using an artificial neural network approach to forecast Australia's domestic passenger air travel demand. *World Review of Intermodal Transportation Research*, 5(3), 281-313.

Suryani, E.; Chou, S.Y.; Chen, C.H. (2012). Dynamic simulation model of air cargo demand forecast and terminal capacity planning. *Simulation Modelling Practice and Theory*, 28, 27–41. <http://dx.doi.org/10.1016/j.simpat.2012.05.012>

Tascón, D.; Díaz Olariaga, O. (2021). Air traffic forecast and its impact on runway capacity. A System Dynamics approach. *Journal of Air Transport Management*, 90, 101946. <https://doi.org/10.1016/j.jairtraman.2020.101946>

Tariq, S.; Lee, S., Woo, S. (2020). A Convolutional LSTM based Residual Network for Deepfake Video Detection. arXiv:2009.07480v1 [cs.CV].

Tennenholtz, G.; Zahavy, T.; Mannor, S. (2018). Train on validation: squeezing the data lemon. arXiv:1802.05846v1 [stat.ML].

TRB (2002). *Aviation Demand Forecast: Survey of Methodologies*. Washington DC: Transportation Research Board.

Tsui, W.; Balli, H.; Gilbey, A.; Gow, H. (2014). Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, 42, 62–76. <https://doi.org/10.1016/j.tourman.2013.10.008>

Wang, X.; Yuan, S.; Wu, C.; Ge, R. (2021). Guarantees for Tuning the Step Size using a Learning-to-Learn Approach. arXiv:2006.16495v2 [stat.ML].

Wells, A.; Young, S. (2004). *Airport Planning & Management*. New York: McGraw-Hill.

World Bank (2022). Air transport.

<https://data.worldbank.org/indicator/IS.AIR.GOOD.MT.K1>

Xingjian, S.; Chen, Z.; Wang, H.; Yeung, D.; Wong, W.; Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 802–810. <https://hdl.handle.net/1783.1/72919>

Yang, J.; Nguyen, M.; San, P.; Li, X.; Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, 3995–4001.

Yang, L.; Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295-316. DOI: 10.1016/j.neucom.2020.07.061

10.2 Anexo II: Air Passenger Demand Forecast Through the use of Artificial Neural Network Algorithms

2022

Air passenger demand forecast through the use of Artificial Neural Network algorithms

Juan Gerardo Muros Anguita
Universidad de Granada, gerardo.muros@gmail.com
Oscar Díaz Olariaga
Universidad Santo Tomas - Colombia, diaz_olariaga@yahoo.es

Revista: International Journal of Aviation, Aeronautics, and Aerospace (USA)

H-Index: 12

Cuartil: Q3

SJR (2022): 0.23

JCR (Factor Impacto): 0.8

Categoría: Ingeniería, Aeroespacial

Rank: 41/52

Estado: Publicado

Cita Artículo: Muros Anguita, J. G.; Díaz Olariaga, O. (2022). Air passenger demand forecast through the use of Artificial Neural Network algorithms. *International Journal of Aviation, Aeronautics, and Aerospace*, 9(3), 7. DOI: <https://doi.org/10.15394/ijaaa.2022.1726>

Air passenger demand forecast through the use of Artificial Neural Network algorithms

Juan Gerardo Muros Anguita

Universidad de Granada (Spain), e.juangerardo@go.ugr.es

Oscar Díaz Olariaga

Universidad Santo Tomás (Colombia), diaz_olariaga@yahoo.es

Abstract

Airport planning depends to a large extent on the levels of activity that are anticipated. To plan the facilities and infrastructures of an airport system and to be able to satisfy future needs, it is essential to predict the level and distribution of demand. This document presents a short- and medium-term forecast of the demand for air passengers carried out through a specific case study (Colombia), in which the impact of the pandemic period due to COVID-19 on air traffic was considered. To make the forecast, an algorithm that implements techniques based on Artificial Neural Networks (ANN) (Machine Learning (ML)) was developed. In particular, for the analysis of the available time series, techniques of encoder-decoder networks of the type ConvLSTM2D have been applied. These architectures are a hybrid between Convolutional Neural Networks (CNN), very useful for the extraction of invariant patterns in their spatial position, and Recurrent Neural Networks (RNN), appropriate for the extraction of patterns within their temporal context (time series). The most relevant result of the present research is that the recovery in demand (volume and trend) to the levels reported before the pandemic is forecast for the period between the end of 2022 and the beginning of 2024 (depending on the type of traffic and scenario considered). Finally, the application of the forecasting model based on Machine Learning/Deep Learning (DL) presents, as a metric performance, a Mean Absolute

Percentage Error (MAPE) value from 3% to 9% (depending on the scenario), which enables predictions of relative precision and introduces a new alternative technical approach to develop reliable air traffic forecasts, at least in the short and medium term.

1. Introduction

Plans for developing different components of an airport system depend to a great extent on the levels of activity that are anticipated. To plan facilities and infrastructures of an airport or system/group of airports, and to be able to satisfy future needs, it is essential to predict the level and distribution of demand in the various components of the airport system (Transportation Research Board [TRB], 2002). Forecasting demand in an industry as dynamic and sensitive to exogenous factors as aviation is an extremely difficult task. However, it is necessary to make air traffic estimates as a preliminary step in planning and designing airport facilities, be it an airport, an airport system, or a network (Horonjeff et al., 2010; Wells & Young, 2004).

Understanding future demand patterns enables the airport planner to evaluate the future performance of the airport and, thereby, recommend consistent development programs so that the costs associated with these development plans are estimated and the sources and level of income to support future capital investments are projected (Ashford et al., 2011). Demand forecasting is a basic requirement to develop an airport master plan or an airport system plan at the regional or national level, thus understanding the entire airport network of a region or country (Janic, 2021, 2009; International Civil Aviation Organization [ICAO], 1987).

To assess the characteristics of future demand, it is necessary to develop reliable predictions of airport activity. Many factors will affect demand, so planners preparing demand forecasts or updating existing forecasts should consider, in addition to historical aeronautical data (air traffic), local socio-economic data (historical series) such as national wealth, purchasing power of the inhabitants, demographics (population), industrial production, consumer price index, an exchange rate (of the local currency against the US dollar), etc. These indicators have a great influence on the behavior of air traffic demand (Rodríguez et al., 2020; García Cruzado, 2013; Horonjeff et al., 2010; ICAO, 2006).

So, the objective of this research is to make a short and medium-term forecast of air passenger demand. For this, Colombia (a complete network of airports as a whole) has been used as a case study, with the particularity that the analysis includes data on demand

for 2020, which has been severely affected by the COVID-19 pandemic. This analysis will serve to estimate as a complementary result (but of great interest approximate date of recovery of both the volume of demand and its growth trend to the pre-pandemic period (2019). To achieve this objective, and as a calculation tool, a model derived from Artificial Neural Networks of the ConvLSTM2D type (<Conv> of Convolutional and <LSTM2D> of long-short-term memory 2-dimensional) is developed. This type of architecture is a hybrid between Convolutional Neural Networks (CNN), very useful for the extraction of invariant patterns in their spatial position, and Recurrent Neural Networks (RNN), very appropriate for the extraction of patterns within their temporal context, as it is the case of time series forecast (Millstein, 2018; Sewak et al., 2018; Yang et al., 2015; Malhotra et al., 2015; Hermans & Schrauwen, 2013). These prediction techniques, based on Machine Learning/Deep Learning (ML/DL), can incorporate more elements of analysis for pattern extraction and, therefore, potentially be more efficient (Ketkar & Moolayil, 2021). Another advantage of ConvLSTM2D networks, compared to classical methods based on autoregression, is that they admit a more complex multivariate treatment (several input features) (Pedrycz & Chen, 2020; Calin, 2020). Finally, when applying this type of neural model (ML/DL ConvLSTM2D) to the case study of the present research, it comes to represent the abstract knowledge model, inferred from the learning of historical patterns in the time series of air traffic, which predicts the future evolution of these time series.

2. Literature Review

In the scientific literature, there are various approaches based on ANN for forecasting passenger demand and air travel. The first is a hybrid neural model with data preprocessing by decomposition of variables, which allows for improving the performance of the network and, thus, optimizing the results in the forecast of demand (Alekseev & Seixas, 2009). The second is the Ensemble Empirical Mode Decomposition (EEMD) based on Support Machines Vector (SMV), with a modeling framework that incorporates slope-based methods to constrain the problem of the final effect that occurs during the process of change of the EEMD; in other words, an EEMD-Slope-SVMs model (Bao et al., 2012).

The third is the use of Backpropagation Neural Network (BNN) to improve the accuracy of the demand forecast (Chen et al., 2012; Kuo & Chen, 2010). The fourth is a hybrid approach VMD-ARMA/KELM-KELM, which consists of the Variational Mode Decomposition (VMD) in an Autoregressive Moving Average (ARMA) model and in a Kernel Extreme Learning Machine (KELM). This means that the VMD is first adopted to break down the original data into various functions, in order to reduce its complexity. Then ARMA and KELM models are used to forecast the stationary and non-stationary components, respectively, and the final result is integrated by another KELM model, which incorporates the forecast results of all components (Jin et al., 2020).

The fifth approach consists of the combined use of Machine Learning and Support Vector Regression (SVR) (Plakandaras et al., 2019; Sulistyowati et al., 2018). The sixth is the Nonlinear Vector Autoregression Neural Network (NVARNN) approach, based on MIV. This approach consists of using: (1) a Mean Impact Value (MIV) method from a neural network to identify and extract input variables, and (2) NVARNN to deal with the irregularity and volatility of the time series (Sun et al., 2019). The seventh is the ensemble forecast modeling based on Singular Spectrum Analysis (SSA), this implies that the original time series is decomposed into three components: trend, seasonal oscillations, and irregular components. The trend is predicted using a Generalized Regression Neural Network (GRNN), while seasonal oscillations are predicted using Radial Basis Function Network (RBF Network) (Xiao et al., 2015).

The eighth approach consists of using a Multilayer Perceptron Architecture (MLP) with a feedback propagation algorithm (Dingari et al., 2019). The ninth entails using evolutionary metaheuristic algorithms (Mostafaeipour et al., 2018); the tenth, uses time-delayed feedback neural networks (Blinova, 2007), and, finally, the eleventh uses recurrent neural networks (RNN) based on long-short-term memory (LSTM) (Gupta et al., 2019).

3. Data and Methodology

In the methodology, before the development of the model, the data (or dataset) related to the annual historical time series of each of the six independent variables (or selected features) were prepared. These are the aeronautical variables (national/domestic passenger or international passenger), and the socioeconomic variables (at the national level) usually used in the calculation of forecasts in air transport (Rodríguez et al., 2020; Díaz Olariaga & Girón, 2020), which are: GDP (Gross Domestic Product), population, IPI (Industrial Production Index), IPC (Consumer Price Index), and TRM (US Dollar-Colombian Peso official exchange rate). The data of this research correspond to the country-case study selected, Colombia; historical data cover a continuous period of 42 years (1979-2020); aeronautical data were obtained from the statistical system of the Colombian Aeronautical Authority (Aerocivil, 2022) and socioeconomic data were obtained from related official sources existing in the country (BRC, 2022; DANE, 2022).

Although the data –both air traffic and socioeconomic– for 2020 are very low compared to previous years due to the reasons explained previously, they are included in the calculation to impose the model to consider the COVID-19 effect.

Finally, as far as data is concerned, an important part of preparing the dataset has consisted in extracting the temporal sequences of input to the model and split up them into selected parametric subsequences relative to the number of previous steps (*lags* or *past*) of each independent variable or feature, necessary to make the predictions (*outs* or *future steps*). Although the code trains the complete historical time series of 42 years, the data are prepared sequentially into temporal subdivisions of the steps of the previous years (called n_lags) before considering future prediction steps (called n_outs). These n_lags grouping can be seen as time context to extract time cycle features. In addition, these values have been hyper-parametrized in the algorithm to study their cost sensitivity to them. Likewise, output data corresponding to a national or international passenger is extracted from the dataset, with the historical subsequences corresponding to the number of steps in future years to be predicted. For simplicity, the same number of steps from previous years has been considered as future steps to be predicted ($n_lags = n_outs$).

The model that is developed in the present research, which is one derived from Artificial Neural Networks, is called ConvLSTM2D (<Conv> of Convolutional and <LSTM2D> of long-short-term memory) (Millstein, 2018; Sewak et al., 2018; Yang et al., 2015; Malhotra et al., 2015; Hermans & Schrauwen, 2013). What the Conv2LSTM2D model aims, as expressed in equation (1), is that once the model has been trained or adjusted in the connections (*weights*) within the neurons of the different layers designed, symbolized by the function $f_{\text{model forecast}}$, this predicts (output variables denoted $\hat{Y}$) the n_outs temporary outputs after time t corresponding to the j feature selected as output ($\hat{y}_{it+n_outs}$), from the input of the n_lags observations of previous times of the whole i input features considered (x_{it-n_lags}).

$$\hat{y}_{jt+n_outs}, \hat{y}_{jt+n_outs-1}, \dots, \hat{y}_{jt} = f_{\text{model forecast}}(x_{it-1}, x_{it-2}, \dots, x_{it-n_lags}) \quad (1)$$

To describe the equations of the ConvLSTM2D model, we begin by outlining the basic architecture of ANN, on which the ConvLSTM2D model is based, called Multilayer Perceptron Networks (MLP) (Vang-Mata, 2020). These are made up of arbitrary several layers of neurons, with a first layer corresponding to the input (associated with the $\mathbf{X}$ tensor) and a last output layer (associated with the $\mathbf{Y}$ tensor), the hidden layers are interspersed between them, to modelized the output variables from the input ones. Each layer contains an arbitrary variable number of neurons (or nodes) that can be activated with different activation functions (Hastie et al., 2013).

The ConvLSTM2D model adopted here is of the Deep Learning (DL) type and has an encoder-decoder structure in which the $\mathbf{X}$ inputs are first encoded, with hybrid layers of convolutional (CNN) and recurrent (RNN) characteristics (Aggarwal, 2018; Bianchi et al., 2017; Blokdyk, 2017; Mandic & Chambers, 2001). On the other hand, the decoding is performed only with the recurring layers of long-short-term memory (LSTM). The coding provided by the convolutional layers facilitates the representation of the invariances of the sequences of the features of the inputs, while the contribution of the LSTM cells in the coding allows memorizing the context of the temporality of the inputs. The decoding part

of the model is performed only by the LSTM layers, which allows extracting and identifying the patterns of the features and the present temporality, such as trend and seasonality (Pedrycz & Chen, 2020). Finally, the output of the LSTM networks is delivered to other dense layers of neurons, MLP type (where all neurons are connected between them) which are the ones that conduct the learning of the feature patterns extracted by the previous encoder-decoder block, until reaching the last layer, where the output tensor $\mathbf{Y}$ is obtained. In summary, the application of encoding-decoding with convolutional CNN and recurring LSTM layers in the ConvLSTM2D model is equivalent to performing double deep learning (Brownlee, 2018; Trifa et al., 2017; Yang et al., 2015).

The group of equations (2) that define the layers of the ConvLSTM2D model, was proposed as an improvement of the traditional RNN (Hu et al., 2019; Donahue, 2015), which managed the hidden states of the cell to map the output from the input. LSTM networks were proposed to solve the problems that appeared by gradient vanishing and exploding (Hochreiter & Schmidhuber, 1997). The superiority of the ConvLSTM2D layered architecture lies in its ability to handle short and long-term memory provided by the LSTM architecture by managing the three types of gates: input, forget and output, plus the new gate coming from the contribution of the convolutional layers with which they are combined to produce the new ConvLSTM2D layer (Shi et al., 2015). Then:

$$i_t = \sigma(W_{xi} * x_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \quad (2a)$$

$$f_t = \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \quad (2b)$$

$$o_t = \sigma(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{co} \circ c_{t-1} + b_o) \quad (2c)$$

$$g_t = \tanh(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c) \quad (2d)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t \quad (2e)$$

$$h_t = o_t \circ \tanh(c_t) \quad (2f)$$

in which: i_t , f_t , o_t represent respectively the input gate, the forget gate and the output gate of the ConvLSTM2D cell; x_t , the temporary inputs of the cell; h_{t-1} and c_{t-1} , the output and the state respectively of the previous cell; $*$ is the convolution operator and W is the convolution filter; with $W_{\cdot i}$ and b_i the equivalent of the weights and biases of the input gate; $W_{\cdot f}$ and b_f , the equivalents to the weights and biases of the forget gate; $W_{\cdot o}$ and b_o , those at the output gate, and W_c and b_c the weights and biases of the cell state. The dimensions and processing of the ConvLSTM2D layers, while being analogous to those of the LSTM cell, are different, since the hybrid ConvLSTM2D cell also performs the convolution operation (Hu et al., 2019). Finally, the operator $\circ$ denotes the Hadamard product and σ the activation function.

It is worth mentioning that, in the algorithm developed for the present research, the activation function σ chosen has been one of the Rectified Linear Unit Function (ReLU) (Nair & Hinton, 2010), because it offers better results for the selected case study than the function of the hyperbolic tangent ($\tanh$) present in the equations described in (2d).

Before training the model, the compilation method is implemented. In the compilation method, the model's arguments are defined, such as: gradient optimizers, cost or error function, and metric evaluation. The next step in the methodology involves training the ConvLSTM2D model. Training the ConvLSTM2D model is under supervised learning, that involves defining the error metric, which measures the 'distance' or norm between the value of the real outputs of the time series used from the $\mathbf{Y}$ tensor, and the predicted or estimated $\hat{\mathbf{Y}}$ value by the ConvLSTM2D model, after applying it to the tensor input $\mathbf{X}$ (Goodfellow et al., 2016). In this research, the mean square error (MSE) has been used for the cost function. Equation (3) defines this error with the norm 2 $\| \cdot \|_2$, where $\mathbf{E}$ is the error and n is the number of samples from the mean.

$$E = \frac{1}{n} \sum_x \| (y(x) - \hat{y}(x)) \|^2 \quad (3)$$

In ML, Backpropagation (BP) is the fundamental process for the automatic adjustment of the model weights going backward from the cost function (Goodfellow et al., 2016). BP calculates the gradient of said total error (in this case the MSE) compared to each of the weights or connections of the neurons of each layer of the model. For this, the chain rule in the derivation applies, to propagate said error backward, to distribute the error among all the weights of the model, neuron by neuron, layer by layer. With this, the new adjustment of weights of the model is sought to produce a smaller error when it is iterated with the next batch of input samples $\mathbf{X}$. If the solution converges, said error becomes smaller or asymptotic, as is the case analyzed here. The group of equations (4) shows the derivation chain rule in ANN networks (Aggarwal, 2018; Hassoun, 1995):

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \quad (4a)$$

$$\frac{\partial net_j}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \left(\sum_{k=1}^n w_{kj} o_k \right) = \frac{\partial}{\partial w_{ij}} w_{ij} o_i = o_i \quad (4b)$$

The problem of minimum searching of the cost function is implemented through different optimizers, generically known as Stochastic Gradient Descent (SGD) (Kingma & Ba, 2014). These optimizers try to avoid that the solution is trapped in some of the local minima that the cost function presents when it is not a convex function. To this end, each of them has implemented different strategies to seek the global optimum.

For the present research, one of the most efficient optimizers has been used, given its high convergence speed and robust stability, known as Adaptive Moment Estimation (ADAM) (Kingma & Ba, 2014). This is an improved version of the RMSProp that implements a variable/adaptive speed (or learning step) and a moment or inertia that tries to avoid getting stuck in the local minima of the cost function. The group of equations (5) describes the update of the weights by the ADAM optimizer (Kingma & Ba, 2014):

$$m_w^{(t+1)} \leftarrow \beta_1 m_w + (1 + \beta_1) \nabla_w l^{(t)} \quad (5a)$$

$$v_w^{(t+1)} \leftarrow \beta_2 v_w^{(t)} + (1 - \beta_2) (\nabla_w l^{(t)})^2 \quad (5b)$$

$$\hat{m}_w = \frac{m_w^{(t+1)}}{1 + \beta_1^{(t+1)}} \quad (5c)$$

$$\hat{v}_w = \frac{v_w^{(t+1)}}{1 + \beta_2^{(t+1)}} \quad (5d)$$

$$w^{(t+1)} \leftarrow w^t - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_w + \epsilon}} \quad (5e)$$

in which $w^{(t+1)}$ represents the weights in iteration $t+1$; η , the learning speed or step; $\nabla_w l^{(t)}$ represents the gradient of the cost function l with respect to the weights; ϵ , a minimum value to avoid dividing by 0 (in Keras 10^{-7} is used); β_1 and β_2 are the forget factors for the gradient or the first and second moments, respectively; $m_w^{(t+1)}$ is the moving mean of the weights or first moments in the iteration $t+1$, and $v_w^{(t+1)}$ is the moving variance (or second moment) of the weights in the iteration $t+1$. It should be noted that in the notation of equations (5) the superscripts with the variable t correspond to the iterations (or learning cycles), while the notation of the subscripts t of the rest of the equations corresponds to the different instants of time dataset features from the time series.

In this work, both the number of iterations and the learning batch size have been hyper-parametrized. A batch is a unit or subset of the dataset from which the error of the cost function is estimated and, therefore, the weights are adjusted. Also, the gradient has the stochastic attribute or name, since it cannot be calculated for the full dataset, due to time, process, and memory computational constraint, then SGD choose a stochastic batch instead. All of this makes the learning process sensitive to the batch size (and the number of epochs or iterations), so these values were hyper-parameterized in the algorithm

(Pedrycz & Chen, 2020; Calin, 2020; Yi & Tan, 2004) to optimize learning process via cost-sensitivity analysis.

The next step in the methodology is the evaluation of the model, in which the goodness of the training or adjustment of the model weights is compared and evaluated. The sensitivity of training variations is considered through performing different hyperparameters configurations defining both the model and the training process itself. In the field of ML/DL, the search for the best solution is a heuristic problem, in which different solutions of architectures and model training processes are proposed and tested, extracted from the infinite space of possible solutions to, then, measure their performances and be able to select among them the best models and configurations (Aggarwal, 2018; Patterson & Gibson, 2017).

For this reason, in the evaluation process it is required, on the one hand, to define the space of the hyperparameters used to configure the model and its training, and, on the other hand, to define the metrics used to evaluate and compare the accuracy of the results. As evaluation metrics for regression problems the Mean Absolute Percentage Error (MAPE), which measures the prediction error as a percentage, has been implemented here. This is an advantage since it provides an intuitive way to evaluate the error of the model; the smaller the MAPE, the better the prognosis (Rodríguez et al., 2020; Kim & Kim, 2016; Ren & Glasure, 2009). Additionally, the Root Mean Square Error (RMSE) was also implemented. To implement this, once the model has been trained with all the data, so that the weights of all the connections of the neurons/layers of the model are finally fixed, the full output values $\mathbf{Y}$ of the dataset are again compared to the predicted or estimated $\hat{\mathbf{Y}}$ values, after applying the model to the full input values $\mathbf{X}$.

Regarding the ConvLSTM2D model, the hyperparameters initially managed to define the possible space of solutions have been the number of filters per convolutional-recurrent layer, the size of the sliding windows of the convolution, the number of times the encoder of ConvLSTM2D layers is repeated, the presence or absence of specific layers such as dropouts and batch normalization (which help to avoid overfitting and be trained robustly) and, finally, varying the number of neurons in the LSTM layer (Donahue et al., 2015). With

all this, a (relative to space research) robust model has finally been obtained regarding its performance.

In relationship to the process of training and splitting up the time series dataset into sequences of n_lags length to predict future sequences n_outs , after performing several internal tests the framework analysis has focused on two forecast scenarios: the first one, 6 years (short-term) and the second, 12 years (medium term), with the two hyperparameters n_lags and n_outs equal. Given that implementing Keras callbacks techniques, such as EarlyStopping and ModelCheckpoint, related to stopping learning when the threshold error is reached and saving the optimal learning model found in training respectively, the algorithm is, to some extent, independent of the number of iterations used. Finally, the hyperparameters that have been considered have been the number of subsequences, and the number of features used as inputs to predict national passengers or international passengers, together with previous steps n_lags and future steps to predict n_outs .

Finally, and as the last stage of the methodology, the prediction of the model is developed, which consists of making the predictions of the future series (n_outs) with the trained model, from the series of the previous years (n_lags) chosen and from the features considered as independent input variables.

For this evaluation, applying said prediction to the historical (real) series $\mathbf{Y}$ and, thus, be able to observe the deviation with the estimated outputs $\hat{\mathbf{Y}}$ or predicted by the ConvLSTM2D model was considered good enough and promising (since the final optimization was not performed due to time and resources constraint). This prediction process is ultimately the final objective of this methodology, since the development of the algorithm to predict the demand for air passengers for future years is what is sought. Although the objective has been focused on two of the features, national (or domestic) passenger and international passenger (annually), the model could predict for the coming years any of the other five additional features included in the historical data set (as are socioeconomic data).

It should be noted that the model weights for each of the possible combinations between any number of the seven possible output features, depending on the seven input features, are different, since the training $\mathbf{Y} = f(\mathbf{X})$, or adjustment of weights of the ConvLSTM2D model, is done only with the appropriate features of the $\mathbf{X}$ and $\mathbf{Y}$ tensors extracted from the historical data set.

4. Results

The results of the prediction of passenger demand are presented, interpreted, and discussed, for the short and medium term, corresponding to the domestic and international cases gathered from the Colombian airport's system dataset, and getting with the defined, developed, trained, and evaluated own specific algorithm.

The two hyperparameters to which the possible prediction scenarios have been reduced are those relative to the number of previous input and prediction years, respectively n_lags and n_outs . On the other hand, said pre-post times are subdivided into subsequences, because results are sensitive to those number of subsequences, among other hyperparameters, since they define the temporality context from which the patterns of the short-term cycles may exist and have to be extracted, within those longer historical periods.

This sensitivity study is common in the analysis of ConvLSTM2D networks and it is derived from the fact that the input data set $\mathbf{X}$ has many (five in this case) dimensions: the number of samples, the number of subsequences, the time series, the number of years per subsequence associated to n_lags , and finally the number of features selected as $\mathbf{X}$ input tensor variables (e.g. GDP, Population, IPI, IPC, TRM). On the contrary, output tensor $\mathbf{Y}$ has 3 dimensions: the number of samples, the number of forecast years (n_outs) and the number of dependent variables to be forecasted (domestic or international passenger). In the scenarios considered, the time series itself of the dependent variable for the next years to be predicted is always, taken also as an input variable (of previous years), to provide greater robustness and predictability to the model.

Before describing the hyperparameters and results achieved in each of the most relevant scenarios, Figure 1 shows the learning curve of the model, relative to the iterations of learning cycles (where the robustness of convergence towards asymptotic values of MAPE is observed, with small fluctuations). While Figure 2 shows the prediction of the historical series with the model already trained where we can realize how the decrease in passenger traffic in 2020 (full pandemic situation) was not predicted, as a result of pre-pandemic values. Both curves have been taken as examples of scenario 1, which is described later.

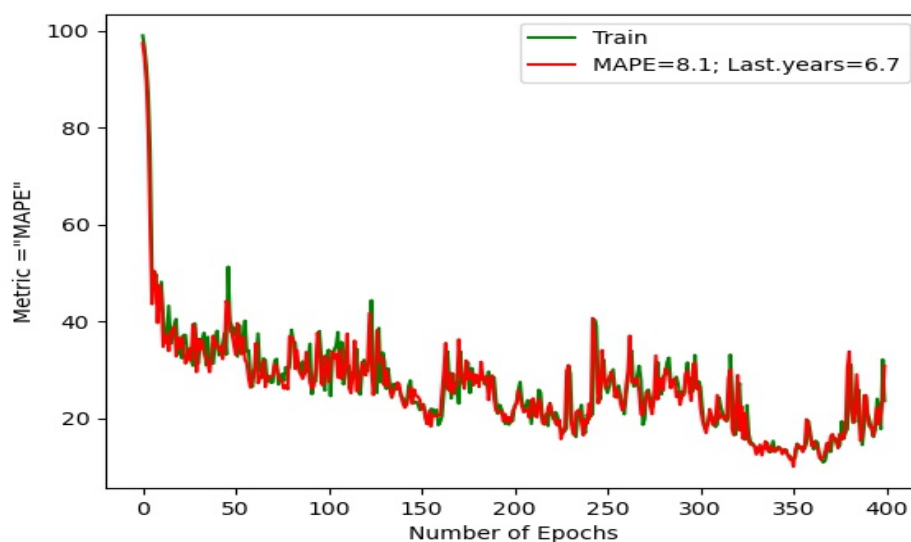


Figure 1 Evolution of MAPE according to the learning curve of the trained model described in scenario 1.

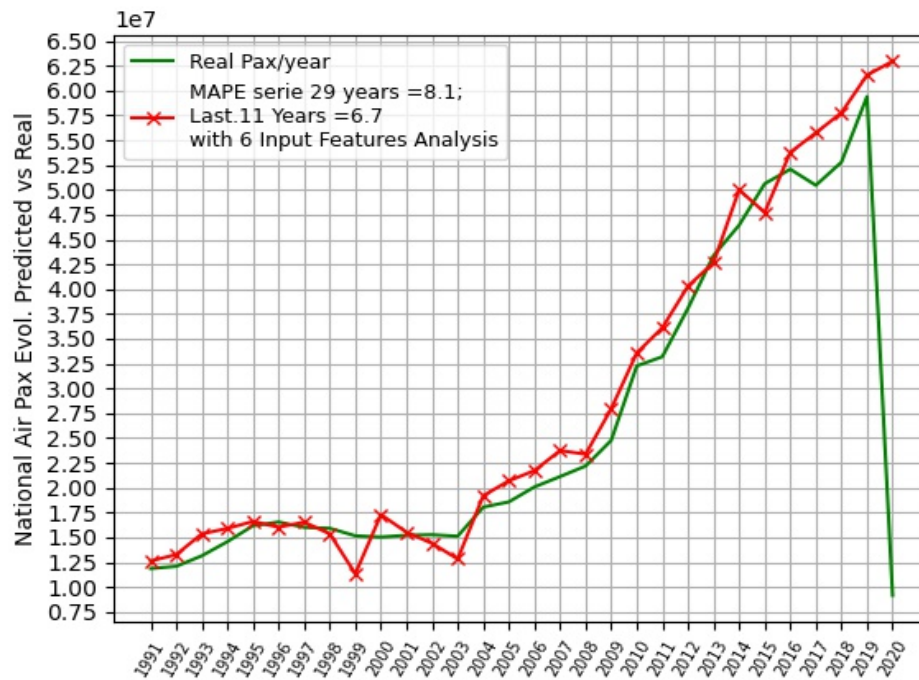


Figure 2 Evolution of passenger prediction vs. the real historical series, started n_lags years after 1979 of the model already trained and described in scenario 1.

Results are generated and presented in various scenarios to deal with the fact that ML/DL models are stochastic in nature. It means, results are sensitive to many external factors, such as dataset quality (veracity on historical values collected without noise and biases), or the number of inputs (features) and output selected, but also to the internal factors associated with the own ConvLSTM2D model such as the hyperparameters deeply described here as, the number of subsequences, length of years considered as previous inputs (n_lags) and predicted outputs (n_outs), number of model training epochs (or iterations), training batch size, training, and test dataset size, optimizers and cost function chosen, number and type of layers implemented, number of neurons per layer, to mention the import ones.

Scenario 1. A first prediction is presented, based on sequences of the previous 12 years and the following 12 years, from the total historical series, with the time series of six

features chosen as input variables (national passenger, GDP, population, IPI, IPC, and TRM) and with national passengers as the only output variable to be predicted. The previous 12 years' n_lags are grouped into three subsequences of 4 years each. Figure 3 shows this medium-term prediction with a MAPE value of 6.7% for the last 11 years and 8.1% for the whole historical series of 29 years, where 12 years after 1979 are counted as previous n_lags steps to predict the following 12 years n_outs . In Figure 3 it can be observed that by the beginning of 2024 the traffic (or demand) of domestic passengers at the country level, would recover the existing level of the pre-pandemic period (2019).

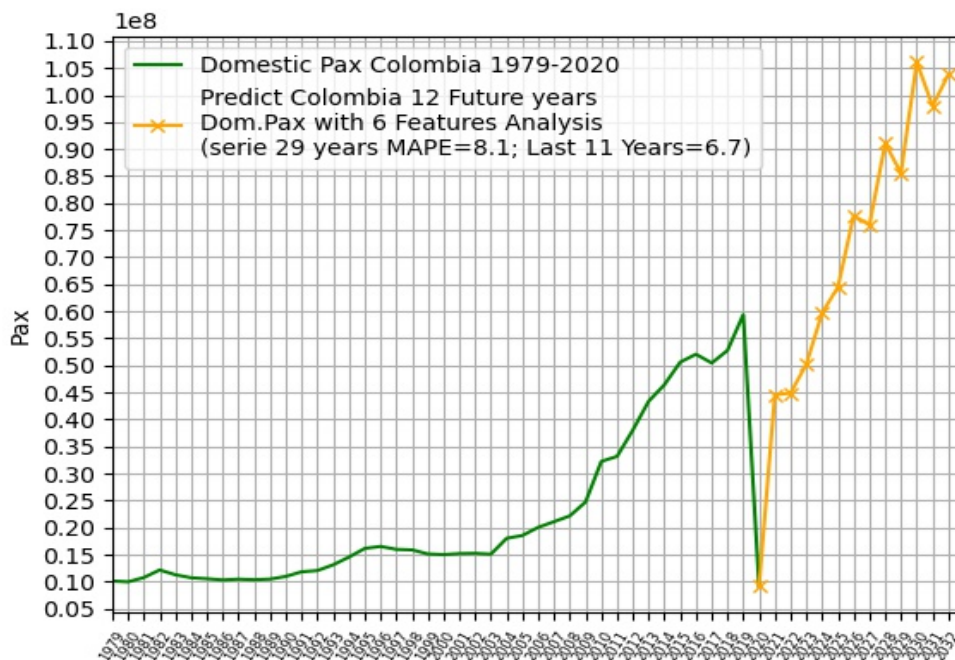


Figure 3 Forecast of domestic passenger demand at the country level, period 2021-2032; scenario 1.

Scenario 2. A second prediction is presented considering sequences of the previous 12 years and the following 12 years, from the total historical series, with the time series but with only two features chosen as input variables (international passenger and GDP) and with the international passenger as the only output variable to be predicted. The previous 12 years n_lags are grouped into three subsequences (or possible sub-cycles) of 4 years each. Figure 4 shows this medium-term prediction with a MAPE value of 9.3% for the last

11 years and 9.7% for the total historical series of 29 years (12 years after 1979 are counted as previous n_lags steps to predict the following 12 years n_outs). In Figure 4 it can be observed that by the beginning of the year 2024 the traffic (or demand) of an international passenger in Colombia would recover to the level existing in the pre-pandemic period (2019) and with a recovery of the trend in demand very similar to that of the year 2019.

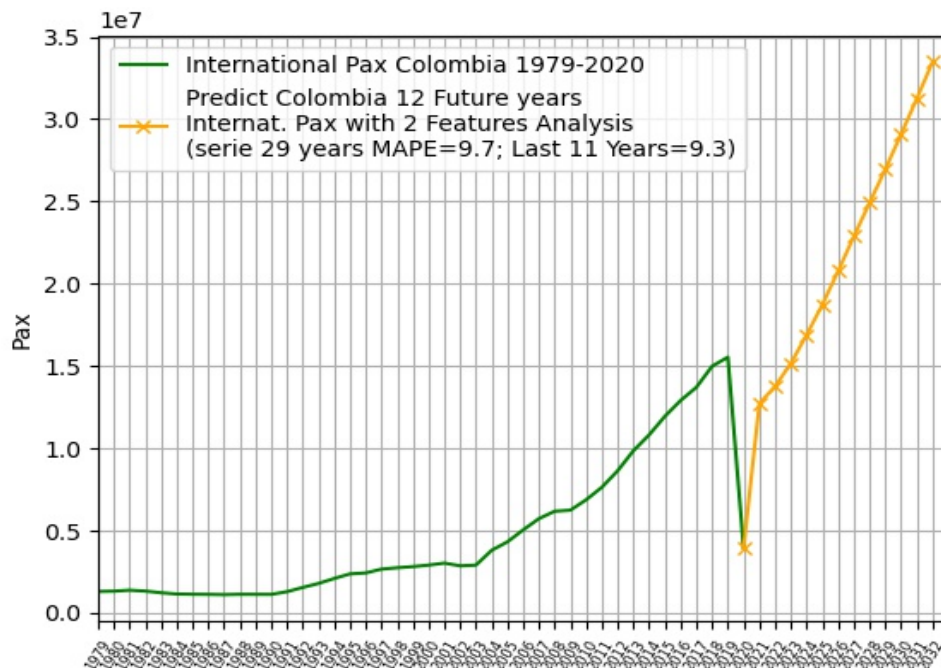


Figure 4 Forecast of the demand for international passengers, period 2021-2032; scenario 2.

Scenario 3. A third prediction is presented based on sequences of six previous years and the six following years, out of the total historical series, with the time series for only two features chosen input variables (domestic passenger and GDP) and domestic passenger as the only output variable to be predicted. The previous six years n_lags are grouped into two subsequences (or possible sub-cycles), therefore, of three years each. Figure 5 represents this short-term prediction with a MAPE value of 4.5% for the last five years (2020 is excluded) and 6.0% for the total 35-year historical series (starting six years after 1979, the initial year of the available history, to predict the following n_outs years). The results show that by the end of 2022 or the beginning of 2023, the traffic (or volume

of demand) of national passengers at the country level, would have recovered to the previous pre-pandemic level (2019).

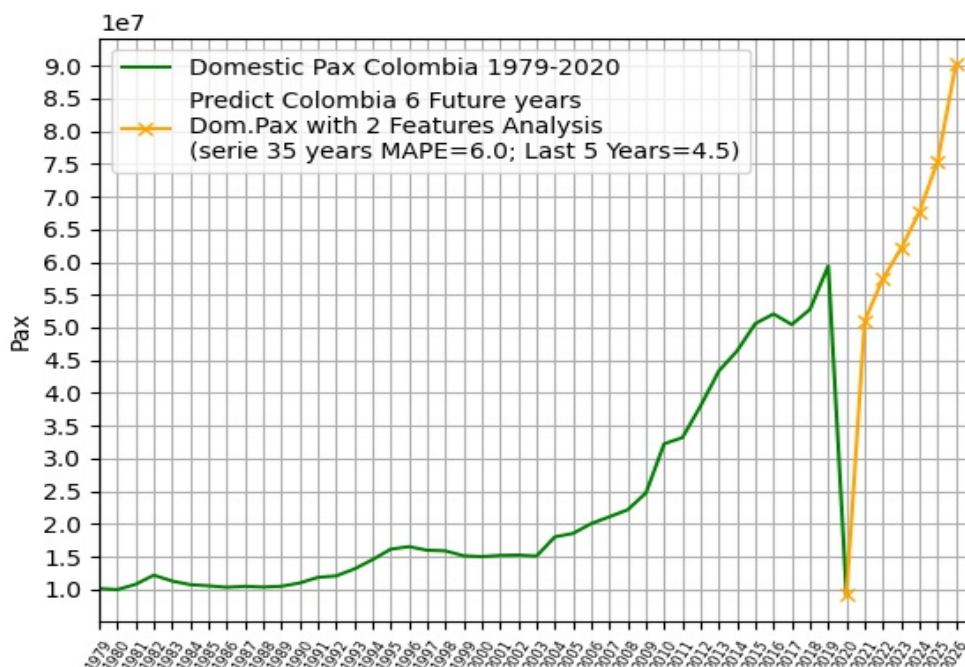


Figure 5 Forecast of the demand of domestic passenger at the country level, period 2021-2026; scenario 3.

Scenario 4. A fourth prediction scenario is presented, calculated with the trends of the time series of domestic and international passengers used as input features. To analyze its impact the number of features used as inputs is left at three. Sequences of the previous six years and the following six years have been considered from the total historical series, with the time series for only 3 features chosen as input variables (intl. pax, domestic pax, GDP) and with the international passenger as the unique output variable to be predicted. The previous six years n_lags are also grouped into two subsequences (or possible sub-cycles), therefore, of three years one. The results of this calculation are shown in Figure 6. The resulting MAPE values are 7.6% for the last five years (2020 is excluded) and 5.8% for the total 35-year historical series (starting six years after 1979, the initial year of the available data history, to predict the following n_outs years). The results indicate by mid-

2022 the demand for international passengers, at the country level, would recover the volume of pre-pandemic demand and even with a higher trend (in demand growth).

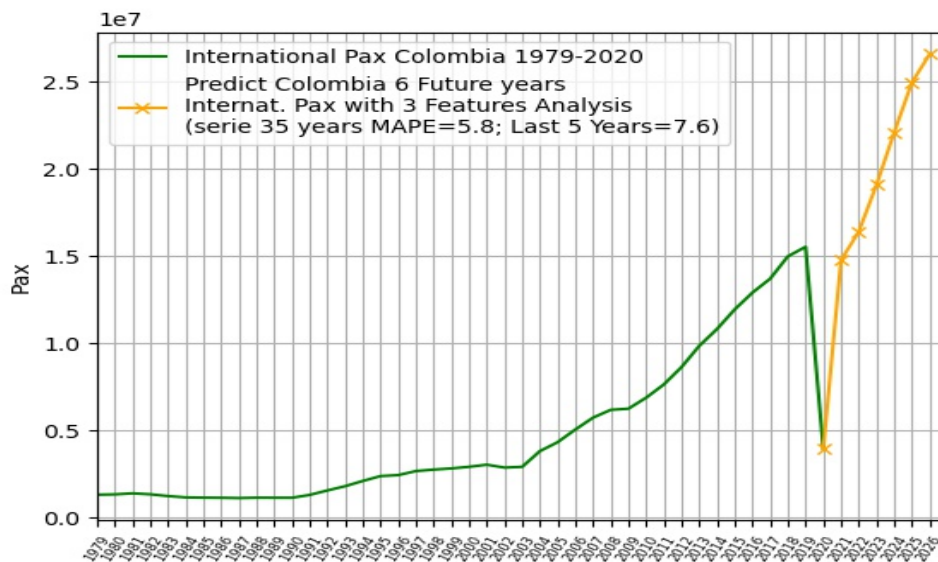


Figure 6 Forecast of the demand for the intl. passenger, period 2021-2026; scenario 4.

Scenario 5. Finally, a fifth prediction is presented with sequences of six previous years and the six following years, out of the total historical series, with the time series of only two features chosen as input variables (international passenger, GDP) and with the international passenger as the only output variable to predict. The previous six years n_lags were also grouped into two subsequences (or possible sub-cycles), therefore, of three years each. The results are shown in Figure 7, the MAPE values are 3% for the last five years (2020 is excluded) and 4.8% for the total 35-year historical series (starting six years after 1979, the initial year of the available data history, and predicting the following n_outs years). Under this fifth scenario, the results indicate that by mid-2022, Colombia's international passenger demand would have recovered to the level of pre-pandemic demand (2019).

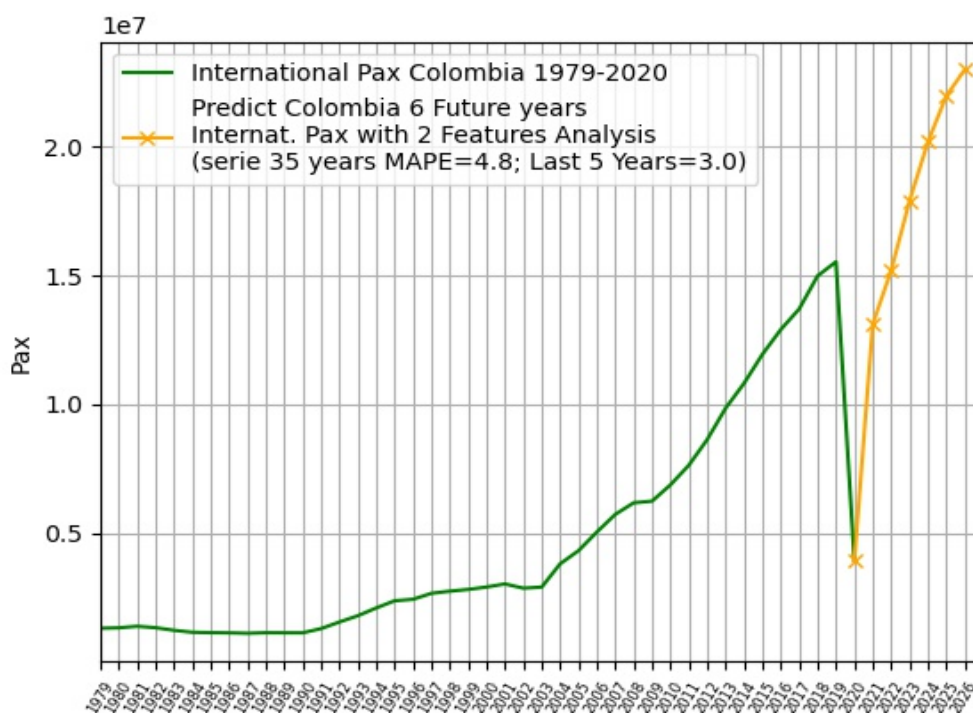


Figure 7 Forecast of the demand of intl. passengers, period 2021-2026; scenario 5.

As a conclusion to the analysis of the results, three observations are made. In the first place, the algorithm developed here has a basic-medium complexity, characterized by the presence of 1,920,009 neural connections, which must be adjusted as unknowns in each iteration during the learning process. This value is specified for scenario 5 (with two features and six years of forecast) but grows to 3,149,833 neural connections when six features and 12 years of forecast are considered for scenario 1. Second, as has been exhaustively shown in the description of the ML/DL model of ConvLSTM2D, these are stochastic processes, that implicitly originated in the same process of calculating the Stochastic Gradient Descent (SGD), which has been improved here by choosing ADAM as loss (or cost function) optimizer (Kingma & Ba, 2014) and, therefore, subject to uncertainties or errors. Third, to constraint the uncertainties associated with the stochastic processes the algorithm has been executed several times, with the same hyperparameters –that process is known as cross-validation techniques (K-fold cross-validation) (Pedrycz & Chen, 2020; Chang & Lin, 2011).

Finally, the results obtained here coincide with the forecasts made in a recent study (Gudmundsson et al., 2021) on the recovery of air transport worldwide, which predicts that the demand for air passengers will recover to pre-pandemic levels at the end of 2022 (optimistic scenario). And the results also coincide with the IATA study (2020) which foresees the recovery of worldwide air passenger traffic at the 2019 level by the end of 2022 for domestic passenger and the beginning of 2024 for the international passenger.

5. Discussion on Methodology

Due to the non-linear characteristics of air traffic demand, classical time series such as econometric-statistical approaches are currently not considered the most convenient methodology, as these approaches are severely criticized due to their poor and limited forecasting capacity (Suryani et al., 2012; Tsui et al. al., 2014; Rodríguez et al., 2020; Li et al., 2020; Liu et al., 2020; Tascón & Díaz Olariaga, 2021; Ensafi et al., 2022). For this reason, a methodology based on a type of Artificial Neural Network architecture called ConvLSTM2D is proposed, which, although they have been very successful in areas of Machine Learning (such as computer vision and natural language processing (Alayba & Palade, 2022; Chaiani et al., 2022; Elboushaki et al., 2020; Fang et al., 2021; Kumar et al., 2022; Xingjian et al., 2015)), have not yet been tested in time series forecasting of air passenger demand until now. The reason for choosing this methodology lies in the fact that these Artificial Neural Network architectures have shown to be very robust and successful in the fields of Deep Learning (DL) mentioned, which they come, by automatically extracting the intrinsic patterns of the non-linear relationships between the variables considered, without a priori knowledge about the existing relationships between the input variables among themselves and between these and the output variables considered, so, in short, it is estimated that this model can be a viable and promising tool to be explored for new flexible modeling of the forecast (Gupta et al., 2019; Dingari et al., 2019). Therefore, this research pursues (with a certain specific scope) to demonstrate the feasibility of applying these Artificial Neural Network models defined by the ConvLSTM2D architecture, which are of the Deep Learning (DL) type, and show it can be obtained acceptable and hopefully results when applied to multivariate time series forecasting such

as our air traffic demand forecasting (Agga et al., 2022; Ensafi et al., 2022; Huang et al., 2022; Prince, 2022; Shastri et al., 2020).

As mentioned before, this promise ConvLSTM2D-ML/DL model needs to be optimized in terms of tuning the hyperparameters and architecture design of the model (number and type of layers, number, and types of neurons in each layer, batch size, number of training epochs, different input features, number of subsequences, etc.). Therefore, this study has been limited to performing certain variability analyses considering different scenarios (number of subsequences, number of years to be predicted, number and types of input features), due to time and resource constraints to work within the unlimited search ML/DL hyperspace. But on the contrary, the different and similar MAPE obtained for the variability analysis performed on the model vs. different scenarios seems to work sufficiently robust and stable, concluding that this model is appropriate and convenient to keep on developing for air passenger demand.

On the other hand, it is considered appropriate to make certain observations and constraints on the development of the proposed model and/or its operativity. In the first place, 42 years could be not considered the ideal amount of data to gather all possible air passenger demand patterns. Even to have only one year (2020) of COVID-19 time series included, is not enough information about COVID-19 patterns, but due to currently available dataset constraints on Colombia air passenger demand at the moment to perform the study, we must accept this (not as extensive as we would have liked) dataset and study limitation.

6. Conclusions

Firstly, estimating the demand for air passengers at the national level in the short and medium term provides valuable information so that the aviation/air transport planners of a country can well in advance design, plan, and implement: (a) development strategies (infrastructures, facilities, equipment, technological modernization, training of technical personnel, etc.); (b) a capital investment calendar (to address the proposed development programs); and (c) related public policies (to consolidate and reinforce the development of the local air transport industry).

Secondly, academics, analysts, planners, and decision-makers in the international civil aviation industry around the world are currently developing forecast studies that allow them to know when and how air traffic will recover (mainly the volume of the air passenger demand) to pre-pandemic levels (the year 2019), due to the importance of the aviation industry, not only in terms of local/regional/global connectivity but also because of its contribution to the global economy. The present work had the objective of contributing from the academic point of view to these two mentioned points. On the other hand, this work provides a novel and original geographic character, as one of the first studies (of an academic nature) in the Latin American subcontinent on air traffic forecasting for the post-COVID-19 era.

Regarding the approach, it was decided to develop a model based on an Artificial Neural Network architecture compared to other possible classical statistical techniques, considering that ML/DL prediction techniques can not only incorporate more elements of analysis but also extract more complex patterns from historical time series dataset, without any previous feature engineering and, therefore, be more potentially effective and accurate. In particular, for the analysis of the time series, techniques of encoder-decoder networks of the ConvLSTM2D type have been applied. The ConvLSTM2D model developed here admits a multivariate treatment (the same does not happen with the classic methods based on autoregression). On the other hand, the ConvLSTM2D model developed is of the supervised learning type, which means that the model is trained with tensor inputs and tensor outputs extracted from the historical series of the dataset. The possibility of hybridization of the convolutional CNN networks with the recurrent networks of long-short-term memory LSTM has allowed the implementation of the ConvLSTM2D model in the present research (novel in terms of its application for air traffic prediction time series). The application of encoder-decoder blocks with hybrid convolutional-recurrent networks represents a novelty for the treatment of time series. Finally, the results of the application of the model based on ML/DL present very acceptable MAPE values (in the order of 3% to 9%, depending on the scenario), which makes the model developed here a feasible alternative to develop reliable air traffic forecasts, at least in the short and medium term.

7. References

Aerocivil (2022). Aeronautical Statistics.

<https://www.aerocivil.gov.co/atencion/estadisticas-de-las-actividades-aeronauticas/boletines-operacionales>

Agga, A., Abbou, A., Labbadi, M., El Houm, Y., & Ali, I. (2022). CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electric Power Systems Research*, 208, 107908.

<https://doi.org/10.1016/j.epsr.2022.107908>

Aggarwal, C. (2018). *Neural Networks and Deep Learning*. Cham: Springer International

Alayba, A., & Palade, V. (2022). Leveraging Arabic sentiment classification using an enhanced CNN-LSTM approach and effective Arabic text preparation. *Journal of King Saud University – Computer and Information Sciences*. DOI: 10.1016/j.jksuci.2021.12.004

Alekseev, K.P.G., & Seixas, J.M. (2009). A multivariate neural forecasting modeling for air transport – Preprocessed by decomposition: A Brazilian application. *Journal of Air Transport Management*, 15, 212–216.

Ashford, N., Mumayiz, S., & Wright, P. (2011). *Airport Engineering*. New Jersey: John Wiley & Sons.

Bao, Y., Xiong, T., & Hu, Z. (2012). Forecasting Air Passenger Traffic by Support Vector Machines with Ensemble Empirical Mode Decomposition and Slope-Based Method. *Discrete Dynamics in Nature and Society*, ID 431512, 1-12. doi:10.1155/2012/431512

BCR – Banco de la República de Colombia (2022). Estadísticas.

<https://www.banrep.gov.co/es/-estadisticas>

Bianchi, F., Maiorino, E., Kampffmeyer, M., Rizzi, A., & Jenssen, R. (2017). *Recurrent Neural Networks for Short-Term Load Forecasting*. Cham: Springer.

Blinova, T.O. (2007). Analysis of possibility of using neural network to forecast passenger traffic flows in Russia. *Aviation*, 11(1), 28-34.

- Blokdyk, G. (2017). *Recurrent neural network*. Brendale: The Art of Service.
- Brownlee, J. (2018). *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery, E-book: Jason Brownlee.
- Calin, O. (2020). *Deep Learning Architectures. A Mathematical Approach*. Cham: Springer.
- Chaiani, M., Selouani, S., Boudraa, M., & Yakoub, M. (2022). Voice disorder classification using speech enhancement and deep learning models. *Biocybernetics and Biomedical Engineering*, 42, 463– 480. DOI: 10.1016/j.bbe.2022.03.002
- Chang, C., & Lin, C. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 1-27.
- Chen, S., Kuo, S., Chang, K., & Wang, Y. (2012). Improving the forecasting accuracy of air passenger and air cargo demand: the application of back-propagation neural networks. *Transportation Planning and Technology*, 35(3), 373-392.
<http://dx.doi.org/10.1080/03081060.2012.673272>
- DANE - Departamento Administrativo Nacional de Estadística (Colombia) (2022).
<https://www.dane.gov.co/>
- Díaz Olariaga, O., & Girón, E. (2020). Análisis de la influencia de la privatización de aeropuertos en el pronóstico de la demanda de pasajeros. El caso de Colombia. *Revista Transporte y Territorio*, 22, 94-113. doi: 10.34096/rtt.i22.8408
- Dingari, M., Reddy, M., & Sumalatha, V. (2019). Air Traffic Forecasting Using Artificial Neural Networks. *International Journal of Scientific & Technology Research*, 8(10), 556-559.
- Donahue, J., Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Darrel, T., & Saenko, K. (2015). Long-term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Conference on Computer Vision and Pattern Recognition*, 7-12 July 2015, Boston (MA).
- Elboushaki, A., Hannane, R., Afdel, K., & Koutti, L. (2020). MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture

recognition in RGB-D image sequences. *Expert Systems with Applications*, 139, 112829.

<https://doi.org/10.1016/j.eswa.2019.112829>

Ensafi, Y., Amin, S., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning –A comparative analysis. *International Journal of Information Management Data Insights*, 2, 100058.

<https://doi.org/10.1016/j.jjime.2022.100058>

Fang, Y., Zhang, H., Zuo, Y., Jiang, W., Huang, H., & Yan, J. (2021). Visual attention prediction for Autism Spectrum Disorder with hierarchical semantic fusion. *Signal Processing: Image Communication*, 93, 116186. DOI: 10.1016/j.image.2021.116186

García Cruzado, M. (2013). *Aeropuertos. Planificación, Diseño y Medio Ambiente*. Madrid: Ibergarceta Publicaciones.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge (MA): MIT Press.

Gudmundsson, S.V., Cattaneo, M., & Redondi, R. (2021). Forecasting temporal world recovery in air transport markets in the presence of large economic shocks: The case of COVID-19. *Journal of Air Transport Management*, 91, 102007.

Gupta, V., Sharma, K., & Sangwan, M. (2019). Airlines passenger forecasting using LSTM based recurrent neural networks. *International Journal Information Theories and Applications*, 26(2), 178-187.

Hassoun, M. (1995). *Fundamentals of Artificial Neural Networks*. Cambridge (MA): MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The elements of statistical learning. Data mining, inference, and prediction*. New York: Springer.

Hermans, M., & Schrauwen, B. (2013). Training and analyzing deep recurrent neural networks. *Working Paper*, Ghent University.

Hochreiter, S., & Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

- Horonjeff, R., McKelvey, F., Sproule, W., & Young, S. (2010). *Planning and Design of Airports*. New York: McGraw-Hill.
- Hu, W., Li, H., Pan, L., Li, W., Tao, R., & Du, Q. (2019). Feature extraction and classification based on spatial-spectral ConvLSTM neural network for hyperspectral images. arXiv:1905.03577 [cs.CV].
- Huang, X., Li, Q., Tai, Y., Chen, Z., Liu, J., Shi, J., & Liu, W. (2022). Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. *Energy*, 246, 123403. DOI: 10.1016/j.energy.2022.123403
- IATA (2020). *Air Passenger Forecasts: Potential Paths for Recovery into the Medium and Long run*. Geneva: IATA.
- International Civil Aviation Organization. (2006). *Manual on Air Traffic Forecasting. Doc 8991*. Montreal: ICAO.
- International Civil Aviation Organization (1987). *Master Planning. Part 1. Doc 9184*. Montreal: ICAO.
- Janic, M. (2021). *System Analysis and Modelling in Air Transport*. Boca Raton: CRC Press.
- Janic, M. (2009). *Airport analysis, planning and design: demand, capacity and congestion*. New York: Nova Science Publishers.
- Jin, F., Li, Y., Sun, S., & Li, H. (2020). Forecasting air passenger demand with a new hybrid ensemble approach. *Journal of Air Transport Management*, 83, 101744. <https://doi.org/10.1016/j.jairtraman.2019.101744>
- Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG].
- Ketkar, N., & Moolayil, J. (2021). *Deep Learning with Python*. New York: Springer.
- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669-679.
- Kumar, S., Rajesh, D., Pranesh, S., Kollipara, H., Agrawal, G., Anbarasi, M., & Valarmathi, J. (2022). Classification of Indian Media Titles using Deep Learning Techniques. *International*

Journal of Cognitive Computing in Engineering.

<https://doi.org/10.1016/j.ijcce.2022.04.001>

Kuo, S., & Chen, S. (2010). Air passenger and air cargo demand forecasting: applying artificial neural networks to evaluating input variables. *12th WCTR*, July 11-15, 2010. Lisbon (Portugal).

Li, H., Bai, J., Cui, X., Li, Y., & Sun, S. (2020). A new secondary decomposition-ensemble approach with cuckoo search optimization for air cargo forecasting. *Applied Soft Computing Journal*, 90, 106161. DOI: 10.1016/j.asoc.2020.106161

Liu, J., Ding, L., Guan, X., Gui, J., & Xu, J. (2020). Comparative analysis of forecasting for air cargo volume: Statistical techniques vs. machine learning. *Journal of Data, Information and Management*, 2, 243–255. DOI: 10.1007/s42488-020-00031-1

Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long-short term memory networks for anomaly detection in time series. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. 22-24 April 2014, Bruges.

Mandic, D., & Chambers, J. (2001). *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. New York: Wiley.

Millstein, F. (2018). *Convolutional Neural Networks in Python*. Scotts Valley: CreateSpace Independent Publishing Platform.

Mostafaeipour, A., Goli, A., & Qolipour, M. (2018). Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study. *The Journal of Supercomputing*, 74(10), 5461-5484. <https://doi.org/10.1007/s11227-018-2452-0>

Nair, V., & Hinton, G. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. *27th International Conference on Machine Learning*. June 21-24, 2010, Haifa (Israel).

Patterson, J., & Gibson, A. (2017). *Deep Learning: A Practitioner's Approach*. Sebastopol (CA): O'Reilly Media Inc.

Pedrycz, W., & Chen, S. (2020). *Deep Learning: Concepts and Architectures*. Cham: Springer.

Prince, A. (2022). Convolutional neural network-long short-term memory optimization for accurate prediction of airflow in a ventilation system. *Expert Systems with Applications*, 195, 116618. DOI: 10.1016/j.eswa.2022.116618

Plakandaras, V., Papadimitriou, T., & Gogas, P. (2019). Forecasting transportation demand for the U.S. market. *Transportation Research Part A*, 126, 195–214.
<https://doi.org/10.1016/j.tra.2019.06.008>

Ren, L., & Glasure, Y. (2009). Applicability of the revised mean absolute percentage errors (MAPE) approach to some popular normal and non-normal independent time series. *International Advances in Economic Research*, 15(4), 409.

Rodríguez, Y., Pineda, W., & Díaz Olariaga, O. (2020). Air traffic forecast in post-liberalization context: a Dynamic Linear Models approach. *Aviation*, 24(1), 10-19.
<https://doi.org/10.3846/aviation.2020.12273>

Sewak, M., Karim, R., & Pujari, P. (2018). *Practical Convolutional Neural Networks*. Birmingham: Packt Publishing.

Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons and Fractals*, 110227. DOI: 10.1016/j.chaos.2020.110227

Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., & Woo, W. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. arXiv:1506.04214 [cs.CV].

Sulistyowati, R., Suhartono, H., Kuswanto, H., Setiawan, K., & Astuti, E. (2018). Hybrid forecasting model to predict air passenger and cargo in Indonesia. *International Conference on Information and Communications Technology (ICOIACT)*, 6-7 March 2018. Yogyakarta, Indonesia.

- Sun, S., Lu, H., Tsui, K., & Wang, S. (2019). Nonlinear vector auto-regression neural network for forecasting air passenger flow. *Journal of Air Transport Management*, 78, 54–62. <https://doi.org/10.1016/j.jairtraman.2019.04.005>
- Suryani, E., Chou, S.Y., & Chen, C.H. (2012). Dynamic simulation model of air cargo demand forecast and terminal capacity planning. *Simulation Modelling Practice and Theory*, 28, 27–41. <http://dx.doi.org/10.1016/j.simpat.2012.05.012>
- Tascón, D., & Díaz Olariaga, O. (2021). Air traffic forecast and its impact on runway capacity. A System Dynamics approach. *Journal of Air Transport Management*, 90, 101946. <https://doi.org/10.1016/j.jairtraman.2020.101946>
- Transportation Research Board. (2002). *Aviation Demand Forecast: Survey of Methodologies*. Washington DC: TRB.
- Trifa, A., Sbair, A., & Chaari, W. (2017). Enhancing assessment of personalized multi-agent system through convlstm. *Procedia Computer Science*, 112, 249–259.
- Tsui, W., Balli, H., Gilbey, A., & Gow, H. (2014). Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, 42, 62–76. DOI: 10.1016/j.tourman.2013.10.008
- Vang-Mata, R. (2020). *Multilayer Perceptrons: Theory and Applications*. Hauppauge (NY): Nova Science Pub. Inc.
- Wells, A., & Young, S. (2004). *Airport Planning & Management*. New York: McGraw-Hill.
- Xiao, Y., Liu, J., Xiao, J., Hu, Y., Bu, H., & Wang, S. (2015). Application of multiscale analysis-based intelligent ensemble modeling on airport traffic forecast. *Transportation Letters: International Journal of Transportation Research*, 7(2), 73-79. doi: 10.1179/1942787514Y.0000000035
- Xingjian, S., Chen, Z., Wang, H., Yeung, D., Wong, W., & Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 802–810. <https://hdl.handle.net/1783.1/72919>

Yang, J., Nguyen, M., San, P., Li, X., & Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, 3995–4001.

Yi, Z., & Tan. K. (2004). *Convergence Analysis of Recurrent Neural Networks*. Dordrecht: Springer.

10.3 Anexo III: Prediction of Departure Flight Delays Through the use of Predictive Tools based on Machine Learning/Deep Learning Algorithms.



Revista: The Aeronautical Journal (UK)

H-Index: 48

Cuartil: Q2

SJR (2022): 0.402

JCR (Factor Impacto): 1.4

Categoría: Ingeniería /Aeroespacial

Rank: 25/34

Estado: Publicado

Cita Artículo: Muros Anguita, J.G; Díaz Olariaga, O. (2024). Prediction of departure flight delays through the use of predictive tools based on machine learning/deep learning algorithms. *The Aeronautical Journal*, 128 (1319), 111 – 133.

DOI: <https://doi.org/10.1017/aer.2023.41>

Prediction of departure flight delays through the use of predictive tools based on Machine Learning / Deep Learning algorithms

Juan Gerardo Muros Anguita¹; Oscar Díaz Olariaga²

¹ Escuela Internacional de Postgrado, Universidad de Granada. Granada, España.

E-mail: e.juangerardo@go.ugr.es

² Facultad de Ingeniería Civil, Universidad Santo Tomas. Bogotá, Colombia.

E-mail: oscardiazolariaga@usta.edu.co

ABSTRACT

The objective of this research is to predict the delays in the departure of scheduled commercial flights through a methodology that uses predictive tools based on Machine Learning/Deep Learning (ML/DL), with supervised training in regression, based on the available flight datasets. Since the novel contribution of this work is, first, to make the comparison of the predictions in terms of means and statistical variance of the different ML/DL models implemented (10 in total) and, second, to determine the coefficients of the importance of the features or flight attributes, using ML methods known as Permutation Importance, it is possible to rank the importance of flight attributes by their influence in determining the delay time and reduce the problem of selecting the most important flight attributes. From the results obtained, it is worth mentioning that the model that presents the best performance is the Ensemble or combinatorial method of Random Forest Regressor models, with an acceptable prediction range (measured with the Root-Mean-Square-Error (RMSE)).

Keywords: flight delay, air traffic, air transport, airport, Machine Learning, Deep Learning

1. INTRODUCTION

Demand for air transport worldwide has grown vertiginously in the last three decades [1, 2]. Airlines require more capacity and more available from airports (in infrastructure) to operate routes and frequencies designed to respond to this demand. The global airport industry has also reacted by permanently expanding its infrastructure and even building new airports with a vast volume of public and private investment [3]. However, the demand and the response have not grown equally fast: the capacity offered or installed in many airports in the world cannot fully satisfy the needs of airlines. This situation has resulted, first, in airport congestion and, consequently, in the generation of delays [4, 5, 6]. For this reason, many studies, both technical and academic, state that delay is a concept that is closely linked to the capacity of airports [7, 8, 9].

The volume of delays has been increasing in the last two decades, not only in its magnitude but also in its negative economic impact, mainly in terms of costs [10, 11, 12, 13, 14]. In the aviation system of the United States in 2019, 18.87% of all flights in the system had a delay [15]; in Europe, in 2019, 46.4% of all takeoffs departed late (whatever the reason for the delay), and the average (time) of these delays (of all late departed flights, concerning time scheduled departure) was 28.4 minutes [16]; in Colombia, the country used as a case for this research, in 2019, 28.60% and 22.54% of domestic and international flights, respectively, departed with delay [17]. Regarding the economic impact of delays, according to FAA estimates [18], the cost generated in the United States aviation system in 2019 was USD 33 billion (including all items); in Europe, the total cost of delays for 2012 was estimated at EUR 11.2 billion (in all respects) [19]; in Colombia, the cost of delays, only for airlines (and only for passenger compensation), in 2019 was USD 6.5 million [17].

The objective of the present research is to predict the delays in the departures of scheduled commercial flights, using the data of the complete airport network of Colombia (country-case study: 58 airports) observing the data of 357,595 flight departures for the entire year 2018 (study period), including both domestic and international departures. For this, and as a methodology, predictive tools based on Machine Learning/Deep Learning (ML/DL) were used, with supervised training (that is, the delays of the dataset flights used

in training and testing are known) in prediction in regression (which will result in an actual expected delay time for a flight departure). The novel contribution of this work is, on the one hand, that it compares the predictions in terms of mean and statistical variance according to the different ML/DL models implemented (10 in total), so that the optimum model can be chosen from among them and, on the other hand, that it determines the coefficients of importance of the flight attributes, using ML methods known as Permutation Importance [20, 21], which allows ranking the importance of flight attributes by their influence on the delay time and filtering out the less important attributes of the flight. Finally, all ML/DL models are retrained to analyze their new features by using the reduced dataset with the most important attributes selected by the models to simplify dataset gathering and ML/DL model training.

2. LITERATURE REVIEW

The two most frequent perspectives in which the delay of commercial flights (mainly scheduled flights) is studied are its prediction and its propagation. Rich literature can be found on this second topic [22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38]. Other lines of research review the relationship of delays with a) flight efficiency and other economic (for example, costs) and logistical aspects of delays [12, 14, 39, 40, 41, 42, 43, 44, 45, 46, 35]; b) aircraft ground operations [47]; c) airport capacity [48, 49, 50]; d) aspects related to clients [51, 52]; and finally, e) its generation, characterization, behavior and structure [53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 41, 64, 10, 65, 66, 67, 68].

On the other hand, prediction of scheduled flight delays has been analyzed from different methodologies that can be grouped into at least six preponderant groups [69, 54, 70, 71, 57]: 1) statistical analysis (regression models, correlation analysis, econometric models, Monte-Carlo, parametric tests, non-parametric tests and multivariate analysis [MVA]); 2) probabilistic models (which encompass or include analysis tools that estimate the probability of an event based on historical data); 3) network analysis (mainly graph theory, although studies related to Bayesian network analysis are also included in this category); 4) operations research (comprising advanced analytical methods such as optimization, simulations, and queuing theory); 5) Machine Learning with regression

prediction (in which the most used methods include: k-Nearest Neighbor, Recurrent NN, Cascade NN, Deep Learning [CNN and LSTM], SVM, fuzzy logic, and Random Forests); and finally 6) Machine Learning in prediction in binary classification (delayed/not delayed) and multiclass (different levels of delay), in which two commonly used classification methods include Decision Trees and Naive Bayes. Most ensemble models, such as the Random Forest, fall into the latter category.

3. APPLICATION CASE

The data for the development of the present research are obtained from the country-case of application (or study) Colombia, currently the third largest air market in the Latin American subcontinent, and fifth in the Americas, by volume of traffic handled [108] [109].

In Colombia, the air transport industry was liberalized in the early 1990s. This liberalization brought structural reforms in both the airport and airline sectors, through an uninterrupted battery of public policies (still in force today), which include not only normative and regulatory aspects, but also aggressive public and private investment programs in infrastructure and technology [110]. Regarding the management of airport infrastructure, Colombia has followed the regional trend of granting the administration of said infrastructures to the private sector. Since the mid-1990s, and in various temporary phases called generations, the Colombian government has granted concessions to 19 airports, which manage the bulk of air traffic in the entire network, including the largest and most important in the country [111]. As a result of public policies –both privatization and public and private investment in airport infrastructure— and deregulation policies of the commercial aviation sector –airfares have been fully liberalized since 2012 [112]— since the beginning of the liberalization of the industry in 1991 and until 2019, the total number of passengers transported grew almost 800% [17].

On the other hand, the entry into the market in Colombia of aero-private operators with a full-service business model (Full-Service Carrier, FSC), occurred very soon after the liberalization of the sector (mid-1990s). But the entry into the market of low-cost airlines (Low-Cost Carrier, LCC) was many years after liberalization (2012). In another order, it can be affirmed that air transport in Colombia has practically no competition (at the domestic

level) with other means of transport, especially for medium and long distances, due to two determining situations: first, the complex geography of the country (crossed from southwest to northeast by three mountain ranges of the Andes mountain range), and, second, to the deficiency (in infrastructure, coverage, capacity and technology) of the terrestrial (road) and rail communication systems [113]. Figure 1 shows the geographical location of the most important airports (by volume of managed traffic) of the Colombian airport network (and participants in this research). Figure 2 shows the development of passenger air traffic in the last four decades.

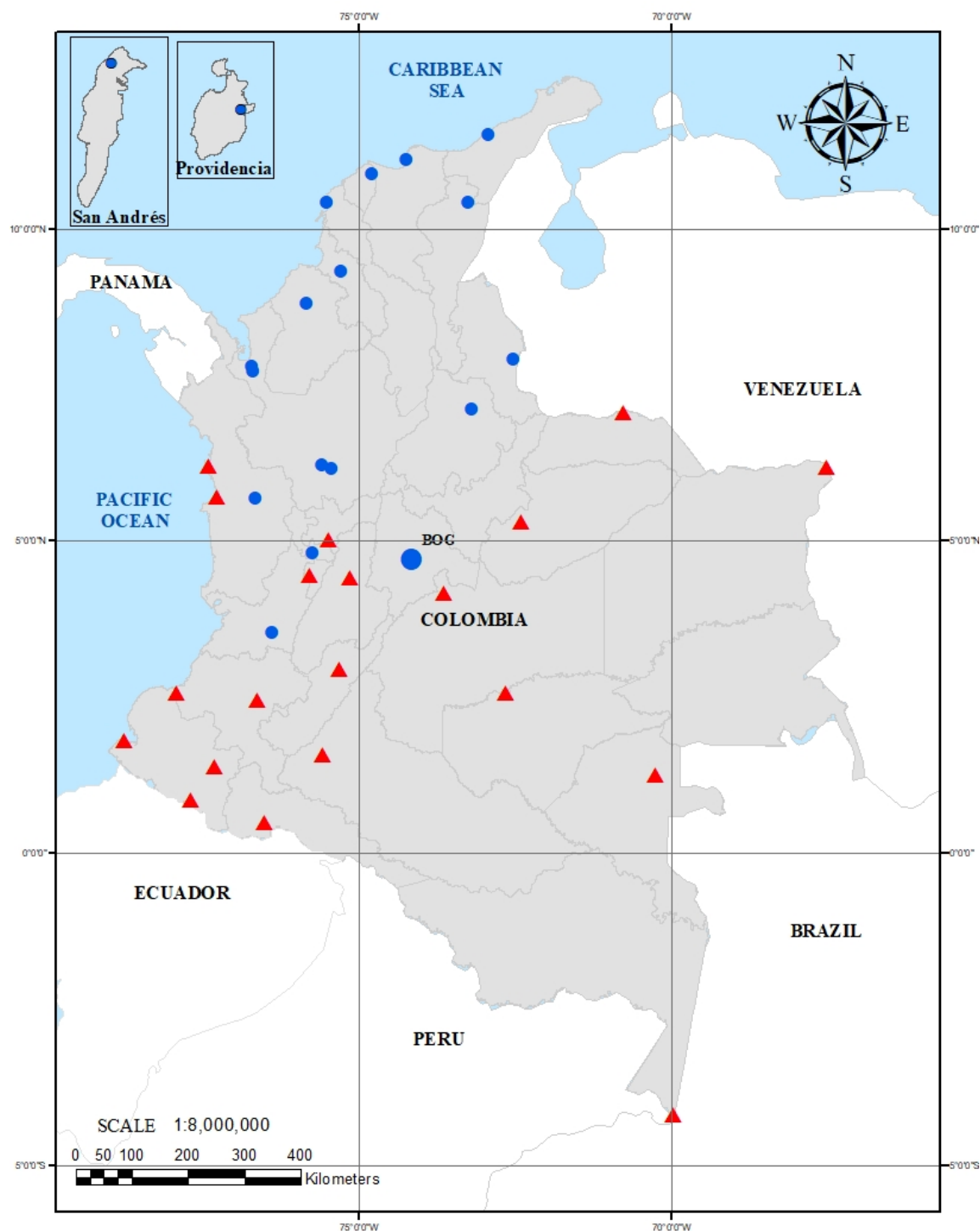


Figure 1. Geographical location of the most important airports (by volume of managed traffic) of the Colombian airport network (and participants in this research). Description: blue circle, privately governed airports; red triangle, airports with public governance. BOG is the IATA code for the country's main airport, Bogotá-El Dorado, located in the capital city of Colombia. Source: [17].

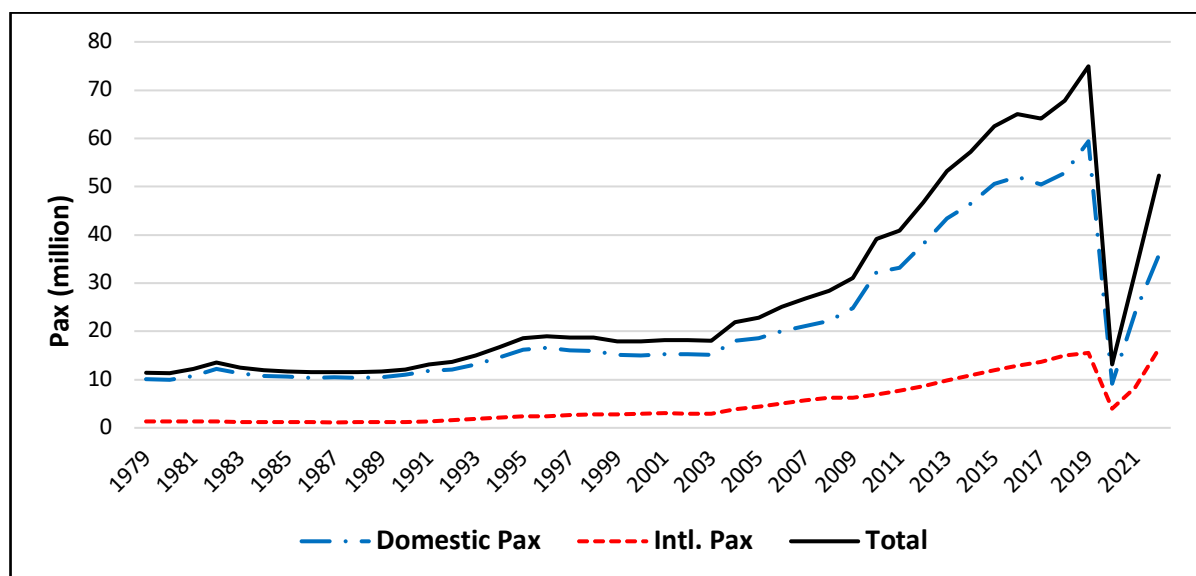


Figure 2. Development of passenger air traffic in Colombia (complete airport network), period 1979-2022.

Source: [17].

The resounding drop in air traffic verified in 2020 (see Figure 2) is due exclusively to the complete suspension of commercial passenger air traffic (throughout the country's airport network) between mid-March and early August 2020, due to the health emergency generated by the COVID-19 pandemic. The reopening and reactivation of the commercial aviation sector was very slow, in stages and by airports, and developed throughout the last four months of 2020.

Finally, and exclusively for the present study, the Colombian Airport Authority uses the Standard IATA Delay Codes-AHM 730 [72] to encode and record the causes of all the delays that occur in the departures of all scheduled flights in the entire national network of airports open to commercial traffic every day of the year. In Colombia, the standard accepted by the national Airport Authority to measure the punctuality of the departure (or arrival) is up to 15 minutes 59 seconds (positive or negative) compared to the scheduled time. The exact moment to measure the real time of departure is when towing and push-back of the aircraft is initiated from its parking position on the apron.

4. OPERATIONAL FUNDAMENTALS OF FLIGHT DELAYS

What exactly is meant by delay and how is it measured? Before providing an answer to this question, it is worth noting that the air transport system has several actors or stakeholders. The three most important (from an operational point of view) are: airports, airlines, and the air traffic management system (ATM). Each one of these actors has its own policies and operational strategies [114] [115]. For this reason, the problem of delay cannot be attributed solely to airport operation (or management of its capacity). Furthermore, the problem of delay cannot be attributed only to the operation of these three actors, since there are other external factors that influence the appearance or aggravation of delays, one of which (the most important) is adverse weather [116] [117] [62].

On the other hand, each country in the world has its own regulatory policy (to control and monitor activity) of its civil aviation system, as well as management policies for its air transport infrastructure. All of this makes it difficult, if not impossible, to count on a universal standard that defines and regulates the concept of delay equally throughout the world [118]. Likewise, and out of necessity, each actor in the air transport system (in each country) manages the problem of delays according to their own operational strategies. This has led to the adoption of a globally coordinated approach on how to manage delay, following certain examples or initiatives. This way, delay can be scalable between airports, both locally/nationally and internationally. In addition, national regulators of commercial aviation (or control and monitoring bodies) increasingly monitor the quality of air service. It is at this point that the concept of punctuality is introduced and serves as a performance indicator [69].

Adopting recommendations or suggestions from national or supranational industry reference organizations, airport authorities (or similar bodies) in many countries have implemented the criterion that a commercial (scheduled) flight is punctual if it starts (or ends) in a range of 15 minutes (positive or negative) compared to the programmed time. If not, that flight is considered to depart or arrive late [119] [16] [120]. Some countries may apply this standard slightly differently in terms of the exact interpretation of “15 minutes”. That is, in some cases punctuality is accepted up to 15 minutes 59 seconds and in others,

punctuality is limited to 14 minutes and 59 seconds. Early departures (sometimes called *negative delays*) cannot be considered a delay but depending on the timing of such anticipation they can cause certain problems in airport operations. Unexpected departures can disrupt the departure flight sequence and early arrivals can influence apron parking allocation [61].

Finally, and to complete the answer to the initial question, it remains to be determined from which event or operation the time that leads to a flight being considered punctual or delayed is counted. Again, globally, there are several criteria for when counting starts. Each national regulatory body and even each specific airport authority defines its policy in this regard. A widely used first standard is that, for departures, the scheduled (reference) time is compared with the exact (and real) time of towing or push-back. That is, at which point the tractor or trailer starts the withdrawal of the aircraft from its parking position on the apron. Another standard is the comparison of the scheduled time with the actual off-block time (AOBT). On the other hand, the arrival delay is usually calculated as the difference between the scheduled arrival time and the actual on-block time (AIBT, literally in-block) at the parking space on apron [121] [122] [123].

As mentioned, delays can have different origins, so it is very important for airport planners and managers, as well as for national regulators of the air transport service, to know the reasons for the delays. This knowledge helps, on the one hand, airport managers to generate operational strategies for their mitigation and/or minimization and, on the other, regulators or control agents of the air transport service to identify those responsible for the delays and the actors harmed and, where appropriate, apply suitable penalties or mandatory financial compensation (or similar) to the affected actors. For this, it has been common for many countries to adopt (or adapt) the IATA Delay Codes-AHM 730 Standard [72], in which all the possible reasons that generate a delay are specified and codified. All the possible causes related to a delay are coded there, associated (by groups) with the operations of the airline, the airport, and Air Traffic Flow Management (ATFM), as well as the meteorological and other diverse causes (security, immigration, customs, government guidelines, etc.) [72].

5. DATA AND METHODOLOGY

5.1. Data

For this research, a database with departures of scheduled commercial flights, domestic and international, of the entire Colombian airport network (58 airports) for the whole year 2018 was used (of the data provided by the Colombian aeronautical authority, only delays in flight departures are available). Such database includes statistical information related to 357,595 flights, with a total of 5,721,520 data entries (each flight data contains 16 associated flight attributes). Each flight attribute is associated with the following relevant information: a) type of traffic (domestic or international), b) airline, c) origin airport (always national), d) destination airport (national or international), e) flight number (linked to the airline, route and frequency of said flight), f) scheduled departure date, g) scheduled departure time, h) towing date (push-back), i) towing time (push-back), j) delay expressed in fraction of an hour, k) delay expressed in minutes, l) flight situation, m) departure status (according to the standard of the country's airport authority), n) delay code (according to the Standard IATA Delay Codes-AHM 730 [72]), in which each code (89 possible) is associated with the reason that generates or causes said delay, o) reason for the delay, p) observations associated with the flight and/or its delay (by the airport authority).

Machine learning models require all input and output variables to be numeric. Encoding is a required pre-processing step when working with categorical data, as it is the case of our different flight's attributes. Then, all the flight attributes or inputs must be converted into number via ordinal encoders functions. In addition, data scaling is a recommended pre-processing step when working with many machine learning algorithms, as it is this case with the 10 models' analysis. Unscaled target variables on regression problems can result in exploding gradients causing the learning process to fail. Also, data scaling is a recommended pre-processing step when working with deep learning neural networks as it is the case of MLP model. Therefore, the target flight attribute data called here time (flight) delay (DEM) must be standardized, that involve rescaling the distribution of data values (of DEM), then the mean of observed values is 0 and the standard deviation

is 1. It is sometimes referred to as “whitening”. This can be thought of as subtracting the mean value or centering the data. Like normalization, standardization can be useful, and even required in some machine learning algorithms when the input data has values with differing scales. Standardization assumes the data fit a Gaussian distribution with a well-behaved mean and standard deviation, as it is in this case. But even this standardized Gaussian pre-processing it is recommended to be applied, if this expectation is not met, to get more reliable results.

5.2. Typology of implemented models

ML/DL (Machine Learning/Deep Learning) models mathematically represent some types of universal approximators [73] of the function f , such that it fulfills: $Y = f(X)$. So, in the present work ten ML/DL models are proposed and used to map the Y output (delay time) with a finite and arbitrary (available attributes) set of X inputs (flight attributes, which can be up to ten in its most extensive version of the dataset), although of these ten models dare defined in Table 1, only a single DL-type model is developed and used, the so-called Multi-Layer Perceptron (MLP) [74].

The MLP model, whose architecture is shown in Figure 1, has been generated using the keras/tensorflow libraries, which consist of two dense layers (connections of all neurons in the previous layer with all neurons in the posterior layer) of 9 neurons in the first layer and 36 neurons in the second, using the activation function ReLU (Rectified Linear Unit) [75, 76, 77, 78], so that neurons can learn the non-linear relationships between the output and the input (of each neuron). The ReLU function is defined as: $y = \max(0, x)$, in which y represents the neuron output and x the neuron input. The ReLU function has several advantages: its computation is easy and, therefore, training times are short, it also converges quickly and does not present saturations (a problem that does occur when other activation functions such as sigmoid and hyperbolic tangent are used by having a linear output); then ReLU neurons activation do not present the vanishing gradient problem (which prevent changing or updating the model weights because the derivatives of weights sensitivity are nearly zero); they are sparsely activated (sparsity feature), which simplifies

the existing connectivity or couplings between all neurons and, finally, allows the ML/DL model to learn the non-linear relationships between the input x and the output y .

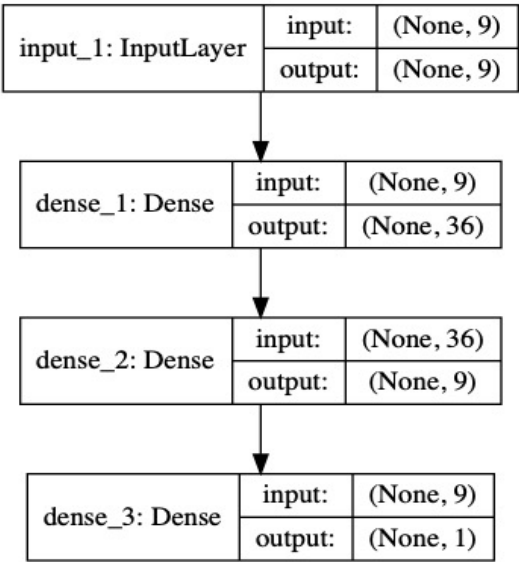


Figure 3. Layers and neurons type architecture of the developed MLP model. Source: authors.

The rest of the models developed in this work, and referred in the Table 1, are the following: the Support Vector Machine (SVR) model, a very robust model in both linear and non-linear predictions because the dimensional space of the features has been increased [79]; the Random Forest (RFR) model, which is an ensemble-type method that improves the results of any of its simplest constituent algorithms by combinations of these [80, 21]; the Simple Linear Regression (LIR) model [81]; the Bagging Gradient (BGR) model, which improves predictions by producing multigroup by combinations and repetitions [82, 83, 84]; the Extreme Gradient Boosting (XGR) model, which from weaker models generates stronger models using the descending gradient as an optimizing method [85, 86]; the Huber Regressor (HBR) model, which is also very robust in training the data with outliers (or atypical values) by penalizing them through *ad hoc* cost functions [87, 88]; the Extra Trees (ETR) model, which implements many more decision trees to better fit learning than simple DTR decision tree methods [89, 90]; the Decision Trees (DTR) model, which is the simple model of decision trees, which is reused due to its self-explanatory capacity of the patterns

learned by the ML [91]; the K-Nearest Neighbors (KNN) model, which predicts closest neighbors based on its K parameter [92]. The “R” at the end of model names indicates that the model has been implemented for Regression analysis. Table 1 lists, at the summary level, the models used.

Model	Acronym	Main characteristic
Multi-Layer Perceptron	MLP	Supervised Machine Learning model consisting of fully connected neurons of feedforward Artificial Neural Network (ANN) for regression analysis of Deep Learning model type.
Support Vector Machine	SVR	Supervised Machine Learning model based on constructing a set of hyperplanes in a high dimensional space searching to maximize different data distance for regression analysis.
Random Forest	RFR	Supervised Machine Learning model as an ensemble learning method that operates by constructing a multitude of decision trees as conditional for regression analysis.
Simple Linear Regression	LIR	Supervised Machine Learning model approaching linearly the relationship between a scalar response and several explanatory variables (also known as independent variables) for regression analysis.
Bagging Gradient	BGR	Supervised Machine Learning model for regression by bootstrap aggregation, also known as bagging, is an ensemble method works by training multiple models independently and combining later (using some deterministic averaging method) resulting in a strong model. This final model is more robust than those individual weak learners' components.
Extreme Gradient Boosting	XGR	Supervised Machine Learning model for regression is an ensemble constructed from weak decision trees model. Trees are added one at a time to the ensemble and fit to correct the prediction errors made by prior models. It is scalable.
Huber Regressor	HBR	Supervised Machine Learning model for a robust regression using the Huber loss function that is less sensitive to outliers in data, via penalizing the outlier data training and learning.
Extra Trees	ETR	Supervised Machine Learning model for regression consisting of extremely randomized trees, or extra trees for short, is an ensemble method that combines the predictions from many unpruned decision trees. Predictions are made by averaging the prediction of the decision trees
Decision Trees	DTR	Supervised Machine Learning model for regression, based on decision support hierarchical models that uses a tree-like model of decision and their possible consequences,

		including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.
K-Nearest Neighbors	KNR	Supervised Machine Learning model for regression is a non-parametric method (not making any assumption on underlying data) and one of the simplest algorithms. The target or output is predicted by local interpolation of the targets associated of the K-Nearest Neighbors in the training set.

Table 1. Models used in this research.

5.3. Training ML/DL models

Training the models is part of the core ML process. In this work, each of the 10 ML/DL models has been trained separately with the dataset under four possible scenarios: the first scenario with the 10 full flight attributes (9 inputs + 1 output), the second scenario reducing scenario one to only six features (5 inputs + 1 output), the third scenario considering only flight features that could be known in advance (5 inputs + 1 output), and finally the fourth scenario reducing the third one to only 4 features (3 inputs + 1 output). The reduced scenarios are aimed at evaluating whether the initial flight attributes of the dataset can be reduced. Another aspect to highlight from this paper is having carried out the statistical model training by implementing the *k-fold* and *cross-validation* techniques, to ensure that there are no biases in the training due to having chosen special subsets from the dataset for training/test, resulting in a more robust evaluation of the performance of ML/DL models.

The training has been implemented by dividing the dataset into batches of 128 samples each, subsequently, the ML/DL model processes the data forward (from the inputs to the output layer), to predict or estimate the model output values. At this point, with the error obtained by measuring the predicted value versus the target value, the training method goes backward (from the output layer to the input layer) to update the model weights (or connections between neurons) of all neurons of all layers of the model, through the Backpropagation method [93, 94, 95]. This method of backward propagation of the error starts obtaining the total error through the appropriate defined cost function. Here, the RMSE applied metric results from comparing the values (of the batch) estimated by the

model, with the real or target values collected from the known supervised features of the dataset. This core process of ML can be explained as the process distribution of the total error between the different contributions made by all the weights or nodes of the model, which is conducted in turn by the stochastic gradient descent (SGD) method [96, 97]. The SGD method calculates the partial derivatives of the error versus each of the weights, which is a way of measuring the model sensitivity of the error compared to each of the weights of the artificial neural network. The adjective “stochastic” of the SGD method comes from the fact that the gradient is calculated by randomly chosen batches and not for the entire dataset.

The equations corresponding to multilayer perceptron networks (MLP) [74, 98] are those associated with neuron layers, in which a first layer corresponds to the input (represented by the $\mathbf{X}$ tensor), and the last output layer (associated with the $\mathbf{Y}$ tensor). Between them, the hidden layers to be implemented are inserted. Each layer contains a variable number of neurons (or nodes), which can be activated with the activation functions, which can be of various types [99]. The input and output a neuron are shown schematically in Figure 4.

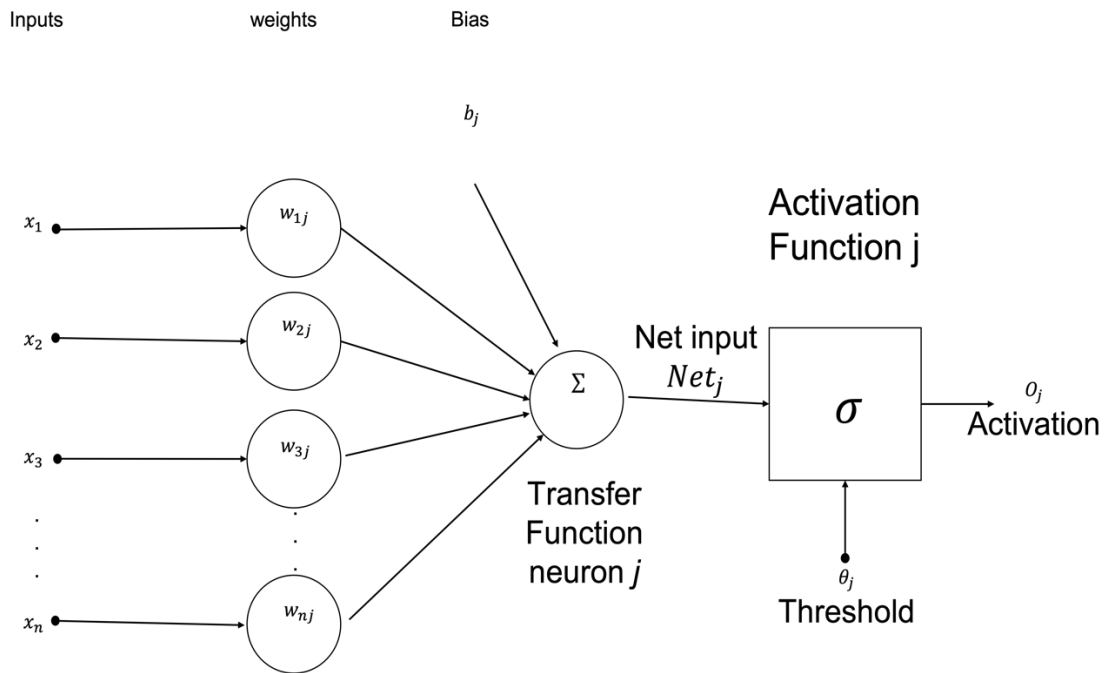


Figure 4. Neuron connections of MLP networks. Source: [100].

In Figure 4, x_i represents the inputs of feature i of the input layer, w_{ij} represents the weight of feature i of the previous layer in connection with neuron j of the posterior layer, Σ represents the transfer function (sum of all the inputs coming from the outputs of all the neurons of the previous layer with neuron j corrected with the b_j bias of layer j , which produces the net output (transfer function + bias) Net_j . The chosen activation function σ is then applied to each neuron (for example, *tanh* or *sigmoid*, although *ReLU* has been chosen here), which in turn produces the output O_j . From here, the connections of the output of this neuron are established again, together with all the neurons of its layer, with the neurons of the next layer, until reaching the last output layer Y_j (which can be several neurons of output, although in the case of regression prediction it is constituted by a single neuron). This inflow-outflow is called *forward flow* or *propagation* (from the input layer through all the neurons in the intermediate layers, to the final output layer). θ_j is the threshold value of neuron j , to activate its output function (if it exists).

In this work, the RMSE has been used as the cost function. Equation (1) defines this error with the norm 2 $\| \cdot \|_2$, in which E is the error, and n is the number of samples in the dataset, $y(x)$ is the actual output (provided by the dataset under supervised learning method) and $\hat{y}(x)$ is the output estimated by the model.

$$E = \sqrt{\frac{1}{n} \sum_x \|y(x) - \hat{y}(x)\|^2} \quad (1)$$

In ML, once the cost function is obtained, the fundamental process for the automatic adjustment of the weights of the model is implemented via backpropagation (BP) [93]. Such method calculates the gradient of said total error (in this case the RMSE) compared to each of the weights or connections of the neurons of each layer of the model, for which the rule of chain derivation applies, to propagate said error backward or to distribute the error between all the model weights, neuron by neuron, layer by layer. If the method converges, the new adjustment of weights of the model is sought to produce a smaller error when it is iterated with the next batch of input samples X . This is the essence of ML, minimizing some loss function (cost or error), adjusting the contributions of the weights of the model neurons connections per each dataset batch and repeating the BP process adjustment to all dataset iterations implemented. Learning that is sufficiently guaranteed when using the SGD method by the theorems of the existence of the null derivation at the points of minimum error. The following group of equations (2) shows the derivation chain rule in ANN networks [101, 102].

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \quad (2a)$$

$$\frac{\partial net_j}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \left(\sum_{k=1}^n w_{kj} o_k \right) = \frac{\partial}{\partial w_{ij}} w_{ij} o_i = o_i \quad (2b)$$

The problems of searching the cost function minimums are solved through different optimizers, generically known as the SGD [99]. These optimizers are in charge of avoiding the solution from getting trapped in any of the local minimums that the cost function presents when it is not a convex function, for which each of them has implemented different strategies to find the global optimum. In the present work, the adaptive moment estimation (ADAM) optimizer has been used because it is one of the most efficient optimizers in terms of its high convergence speed and robust stability [100], which is an improved version of the RMSProp that implements a variable/adaptive speed (or learning step) and a moment or inertia that tries to avoid getting caught in the local minima of the cost function. The following group of equations (3) describes the update of the weights by the ADAM optimizer [100].

$$m_w^{(t+1)} \leftarrow \beta_1 m_w + (1 - \beta_1) \nabla_w l^{(t)} \quad (3a)$$

$$v_w^{(t+1)} \leftarrow \beta_2 v_w^{(t)} + (1 - \beta_2) (\nabla_w l^{(t)})^2 \quad (3b)$$

$$\hat{m}_w = \frac{m_w^{(t+1)}}{1 + \beta_1^{(t+1)}} \quad (3c)$$

$$\hat{v}_w = \frac{v_w^{(t+1)}}{1 + \beta_2^{(t+1)}} \quad (3d)$$

$$w^{(t+1)} \leftarrow w^t - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_w + \epsilon}} \quad (3e)$$

In the equations (3), $w^{(t+1)}$ represents the weights in iteration $t + 1$; η represents learning speed or learning step; $\nabla_w l^{(t)}$ represents the gradient of the cost function l with

respect to the weights; ϵ is a minimum value to avoid dividing by 0 (in keras 10^{-7} is used); β_1 y β_2 are the forgetting factors for the gradient or first moment and the second moment respectively; $m_w^{(t+1)}$ is the moving mean of the weights or first moments in the iteration $t + 1$; $v_w^{(t+1)}$ is the moving variance (or second moment) of the weights in the iteration $t + 1$. The variables with *hat* indicate that they are estimators and the superscripts with the variable t correspond to each of the iterations (or learning cycles). This is the process of fitting the MLP/DL model. On the other hand, in models based on decision trees, the adjustment is made mainly by induction and pruning methods of the decision trees [101, 102].

The dataset learning process is repeated in this work 150 times (iterations) for the entire dataset and only for the MLP DL model (considered sufficient to achieve an asymptotic or stable RMSE value), which marks the maximum fine adjustment possible of the model in the learning of the dataset, or until a sufficient adjustment of the error is obtained for the rest of the non-DL models, characterized by belonging to the ensemble or boosting or stacking type, in which they improve the simplest models of decision trees by their appropriate combinations [82].

To give more robustness to the value of the metrics obtained from the RMSE, *k-folds* techniques have been implemented [103, 104]. Specifically, three groups have been applied to divide the dataset between three different ways of grouping the training and test data subsets. The training data, as its name implies, is used to train the model and the test data to evaluate the learning performance of the model. However, the model can never see the test data during the learning process to make an objective evaluation. In addition, with cross-validation techniques this training is repeated a variable number of times. Because target delay flight attribute was standardized as necessary step of data pre-processing, the loss or cost function RMSE of each ML/DL model it is also characterized statistically, in terms of a mean and a standard deviation, resulting all of this in a more robust way performance measured compared to the case when a single configuration of training and test subset is used. It should be emphasized that all these methods are stochastic, due both to the use of the SGD method, and to the own stochastic nature of the dataset, so their results are not deterministic, but rather are probabilistic.

5.4. Statistical examination

Before training, different examinations or statistical controls were applied to have a feature insight. For this purpose, different tests has been implemented to determine if flight attribute distributions are Gaussians (or normal) (Shapiro-Wilk, D'Agostino and Anderson-Darling tests), or whether flight attributes are not correlated with the output (linear Pearson, Spearman Rank, Kendall Rank, and Chi-Squared tests), or whether the flight attributes are not stationary (ADF Fuller test), or whether flight attributes have equal distribution as the output (Student's test, Paired Student, Anova, Mann-Whitney U, Kruskal-Wallis H, and Friedman tests). These tests, of statistical characterization of flight attributes, serve to know among other things, whether their learning process will be easy or which data transformation must be required to facilitate it. These techniques are an enhancement of dataset preprocessing to facilitate ML training.

5.5. Determination of the importance of the flight attributes

To rank the importance of flight attributes in determining flight delay time, the Permutation Importance (PI) method [20, 21] has been implemented. This method consists of applying inspection techniques to trainable estimators; each model proposed and configured for this work offers a different response regarding the selection of the most significant attributes of the flights. The PI method consists of randomly permuting the features among themselves and measuring in each case how the output (delay time) varies through the training of the new dataset (permuted or sometimes called corrupted in the scientific literature) with each ML/DL model, which allows establishing the coefficient of importance or the sensitivity of the output (time delay) to each feature j called c_j (see equation (4)).

$$c_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j} \quad (4)$$

In equation (4) s is the result of training the ML model when it is trained with the D dataset without permuting. Next, each feature j associated with a column of dataset D is randomly permuted with the other features to produce a permuted version of the dataset $\tilde{D}(j, k)$, in which the subscript k represents the repetitions used to manage the features, K is the maximum number of permutations, and $s_{k,j}$ is the new result of training the ML model for each feature j repeated k times with the permuted (or corrupted) dataset $\tilde{D}(j, k)$.

6. RESULTS

Before training the 10 ML/DL models implemented, the outliers (or atypical data values) are removed from the dataset. If they are retained, the models are forced to learn marginal patterns of said flight attributes, which is detrimental to a good generalization in the prediction of the flight delay time. In this work, four methods have been used to identify and remove outliers, two of which are properly unsupervised ML methods, such as the Local Outlier Factor (LOF) and One Class SVM (OCS) methods, which cluster normal data and mark as outliers those that are outside these perimeters of normal values. The third method is Isolation Forest (IF), which applies binary decision tree theory that does not require pre-grouping within standard profiles, making it faster and more efficient in detecting outliers. The fourth method, the Elliptic Envelope (EE) [105], calculates the outliers by the method of the minimum covariance determinant, which must meet the applicability condition that the number of samples (flights) must be higher than the square of flight attributes, which is true in this case.

Subsequently, the learning of the 10 ML/DL models has been evaluated for four flight attributes selection scenarios, to show the prediction sensitivity of the models depending on which flight attributes are chosen. Table 2 shows the list of flight attributes and indicates in which model scenario each flight attribute participates.

Flight attribute	Description	Scenario in which the attribute participates
TRF	Type of traffic (domestic/international)	1, 3
AEL	Airline	1, 2, 3, 4

ORI	Origin airport name (departure)	1, 2, 3, 4
DES	Destination airport (name)	1, 3
NVU	Flight number	1, 2, 3, 4
FPS	Scheduled departure date	does not participate
HPS	Scheduled departure time	does not participate
FRM	Push-back date	does not participate
HRM	Push-back time	does not participate
DEA	Time delay (according to Standard IATA Delay Codes-AHM 730)	does not participate
DEM	Actual time delay (in minute)	1, 2, 3, 4
EVU	Flight status (can be on time or delayed)	does not participate
CDE	Delay code (according to Standard IATA Delay Codes-AHM 730)	1, 2
MDE	Reason for delay (according to Standard IATA Delay Codes-AHM 730)	1
OBS	Observations (from the airport authority)	1
EAC	Flight status (can be on time, or delayed, or early, or canceled)	1, 2

Table 2. List of flight attributes, and their participation in the different scenarios of the model.

The models' comparative results of the RMSE performances are shown for full scenario 1. This full scenario 1 removes 6 of the 16 initial flight attributes (because some of the them are redundant), resulting in the 10 full following flight attributes: TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC and DEM, in which the first 9 are inputs and the last one is the output to be predicted (DEM is the delay time measured in minutes) (Figure 5).

Figure 5 shows the performances, measured with the RMSE metric, are very similar for all models and their value is below 0.5, except for the behavior of two more dispersed models (MLR and SVR). The RFR model is the one with the highest performance with an RMSE of 0.449 and a standard deviation value of 0.0013, which evidence that the model is robust.

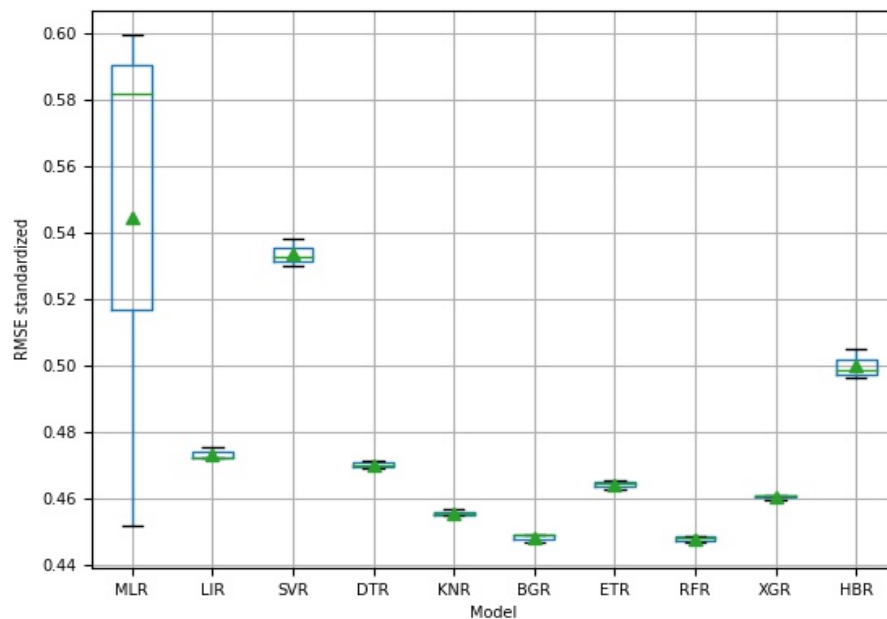


Figure 5. RMSE comparison resulting from the learning obtained for each of the 10 models, transforming the delay time into a standard distribution (0 mean and 1 for standard deviation), referred to scenario 1 resulting from choosing 10 flight attributes, 9 inputs (TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC) and 1 output (DEM). Source: authors.

It is already explained on subsection 5.1 (Data), that the DEM flight attribute has been standardized to facilitate the learning of the ML/DL models. For this reason, Figure 6 shows the results of the RMSE ML models performance, but in its original scale of delay times in minutes (which is done by reversing the standardization process applied to the DEM), and evidence that the prediction errors (on their natural minute scale) are less than 16 minutes (for the best RFR model). That is, when introducing the 9 aforementioned flight attributes, a delay time is predicted that has an error of fewer than 16 minutes, which is a remarkable result for the features considered.

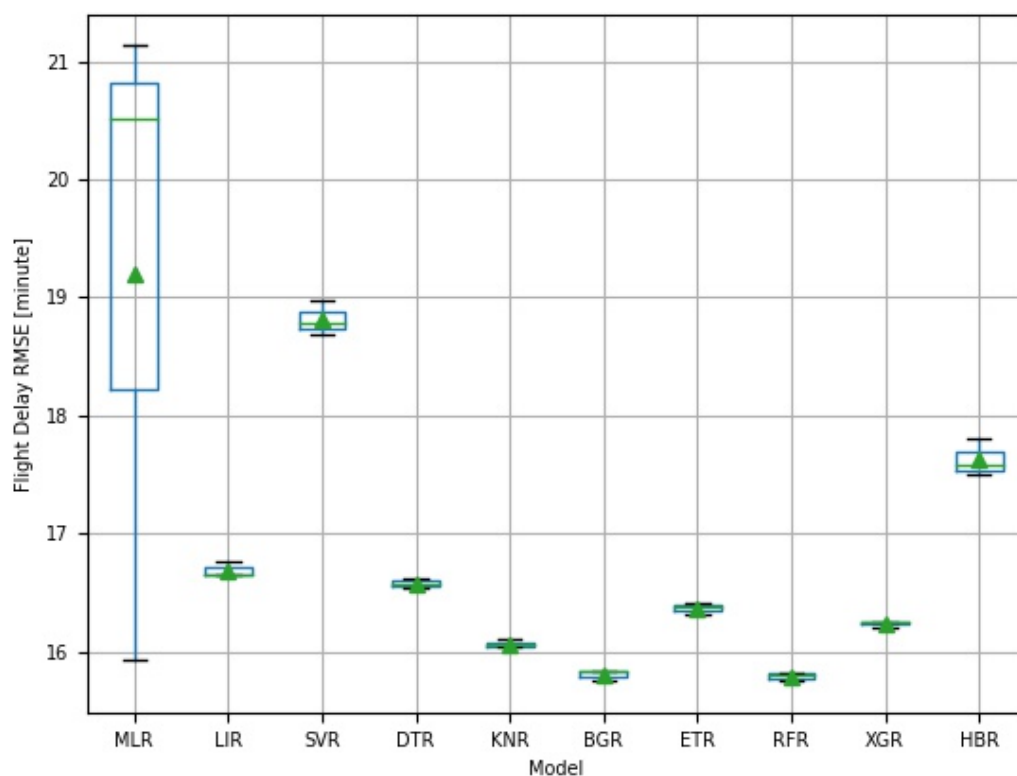


Figure 6. Comparative analysis of the 10 models' performance (RMSE), in its original scale of minutes of delay, referred to as scenario 1, choosing 10 flight attributes, 9 inputs (TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC) and 1 output (DEM). Source: authors.

Next, reduced scenario 2 is generated, after applying Permutation Importance method whereby each of the 10 ML/DL models auto-selects the 5 best flight attributes that better determine the delay time, from among the 9 full flight attributes inputs of the scenario (the tenth flight attribute is the output). That is, each model ranks the flight attributes inputs by their importance in determining the delay time, assigning them a (relative) importance coefficient.

Figure 7 shows the ranking of importance assigned by the RFR model to the 9 input features, in which the last 5 are the most important (ORI, AEL, NVU, CDE, EAC) in increasing order.

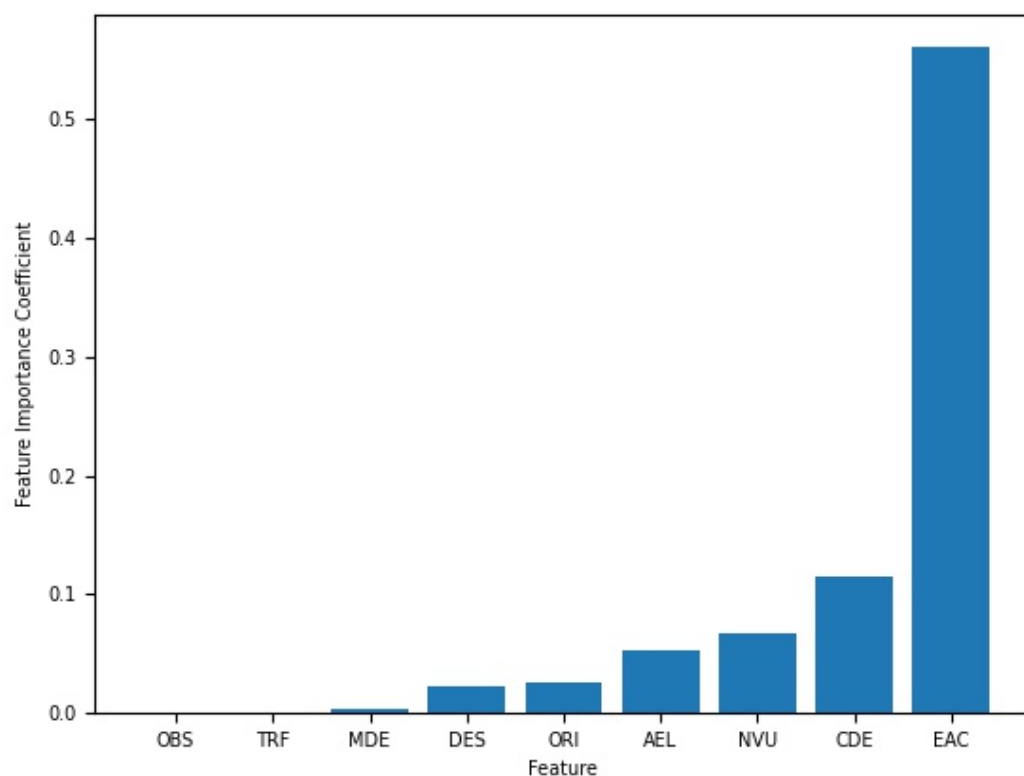


Figure 7. Relative importance coefficients in increasing order, of the 9 full flight attributes considered as inputs of scenario 1, when determining the flight delay time (DEM) as output. Source: authors.

To evaluate the quality selection of the 5 most important or significant flight attributes (EAC, CDE, NVU, AEL, ORI), now arranged in descending order, the models are retrained with this new reduced scenario 2, considering as output the flight attribute DEM, to compare the learning models performance between the two scenarios. The results of the RMSE obtained in this reduced scenario 2 are shown in Figure 8, where it can be seen they roughly get the same results when using the 9 input flight attributes of full scenario 1. In fact, even the RMSE improves slightly for reduced scenario 2, which shows that the learning patterns are better grasped by the models when they are trained on this reduced scenario 2 (using only the 5 more significant flight attributes) than when they are trained using the full scenario 1.

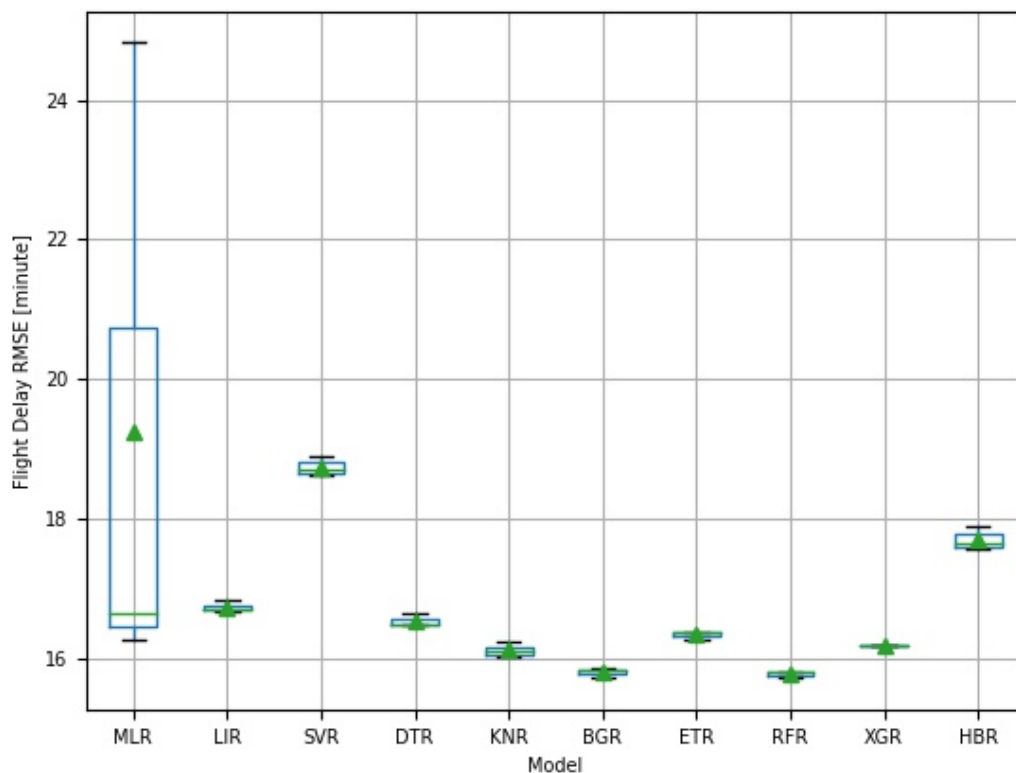


Figure 8. Comparative results of the RMSE obtained by the 10 ML/DL models when trained under reduced scenario 2, with only 5 flight attributes. It can be seen how the model learning performance (in terms of RMSE) is roughly identical to full scenario 1. Source: authors.

Since both scenarios full 1 and reduced 2 involve a series of flight attributes that are measured *a posteriori* and, therefore, are not known in advance, such as the EAC (Flight status according to airport authority standard), CDE (Delay code according to Standard IATA Delay Codes-AHM 730), MDE (Reason for delay according to Standard IATA Delay Codes-AHM 730) and OBS (Observation of flight from the airport authority), in this paper, the 4 aforementioned flight attributes from the dataset have been removed to make the predictions of flight delay departures more realistic (according to real data known before the flight departure). Particularly, a new realistic scenario 3 is now defined with only 6 total flight attributes, 5 of which are inputs (TRF, AEL, ORI, DES, NVU) and 1 output (DEM).

After models have been conveniently trained, an RMSE of 33.8 minutes is obtained by the best RFR model (approximately double that when the 4 eliminated flight attributes were retained), results are shown in Figure 9.

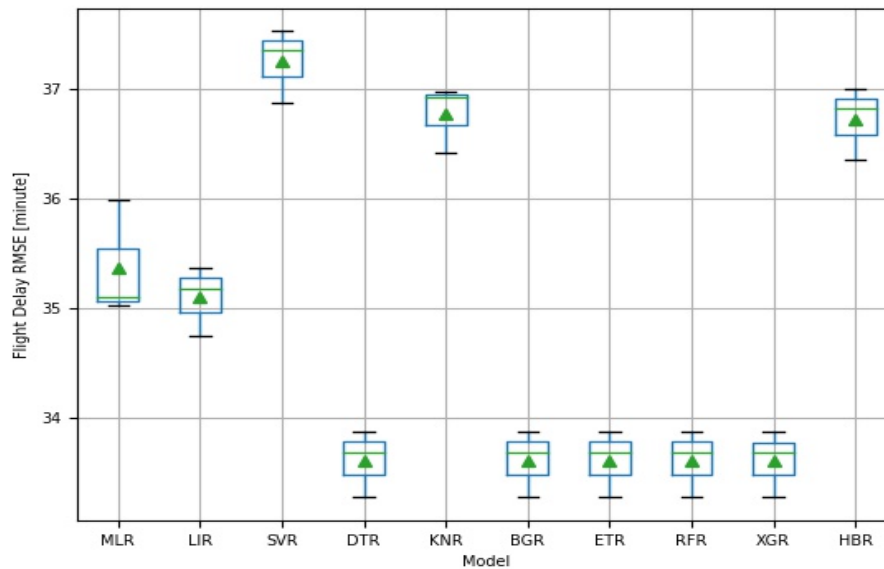


Figure 9. RMSE comparative analysis of the 10 models under a more realistic dataset scenario 3 (5 inputs: TRF, AEL, ORI, DES, NVU and 1 output: predicting delay (DEM)), in which the best RFR model obtains an RMSE in predicting flight delay of 33.8 minutes. Source: authors.

The reason for this loss of precision when eliminating these 4 keys flight attributes (EAC, CDE, MDE, OBS) is that other variables that are not registered in the dataset may influence the level of delay, such as the type of aircraft (mainly its size), the level of congestion at the airport, weather conditions, ATM conditions, etc. When Permutation Importance method is applied to scenario 3 of 6 flight attributes (5 inputs +1 output) to get the importance coefficients of the flight attributes in determining the flight delay time, the significance ranking is shown in Figure 10 (flight attributes selected by the best RFR model).

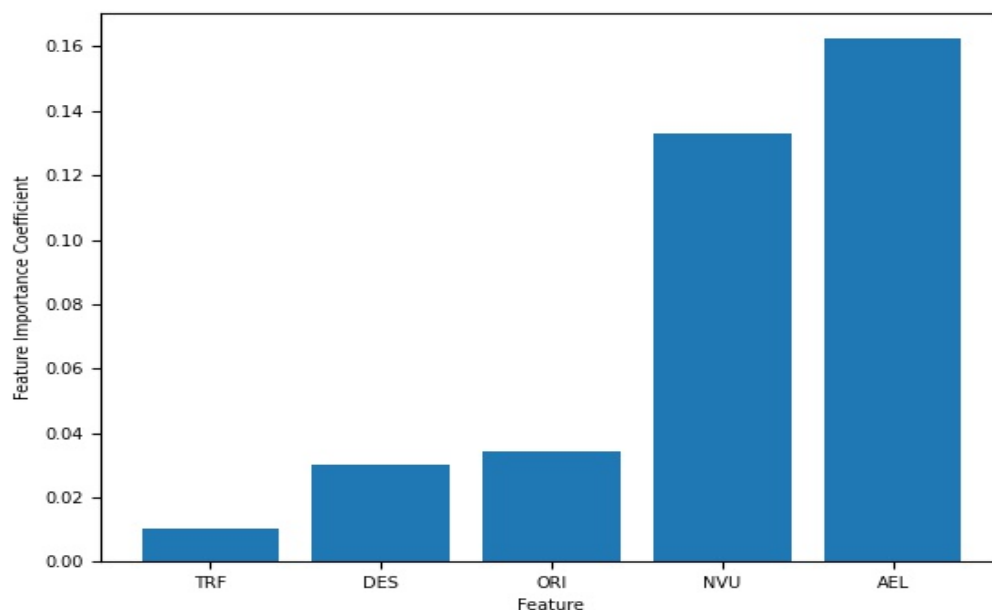


Figure 10. Rank of the importance of the 5 input flight attributes considered in scenario 3 (TRF, DES, ORI, NVU, AEL) in increasing order. Source: authors.

Reduced scenario 4 (similar process to get the reduce scenario 2) is generated by applying reduction to scenario 3, to keep only the best 3 flight attributes that auto select the best (RFR) models when applying the Permutation Importance method.

Figure 11 shows the comparison of the RMSE for the 10 ML/DL models of scenario 4 consisting of 4 flight attributes (3 input: ORI, NVU, AEL, and 1 output: DEM). The results show that, despite the reduction of features (to only the 3 most important), the same results are obtained as in scenario 3 when 5 input flight attributes are used. This shows that the feature auto-reductions conducted by the Permutation Importance method have been carried out correctly and, most importantly, the dataset can significantly be reduced or simplified, thereby saving model training time and dataset memory, without any loss in the accuracy of the time delay predictability (DEM).

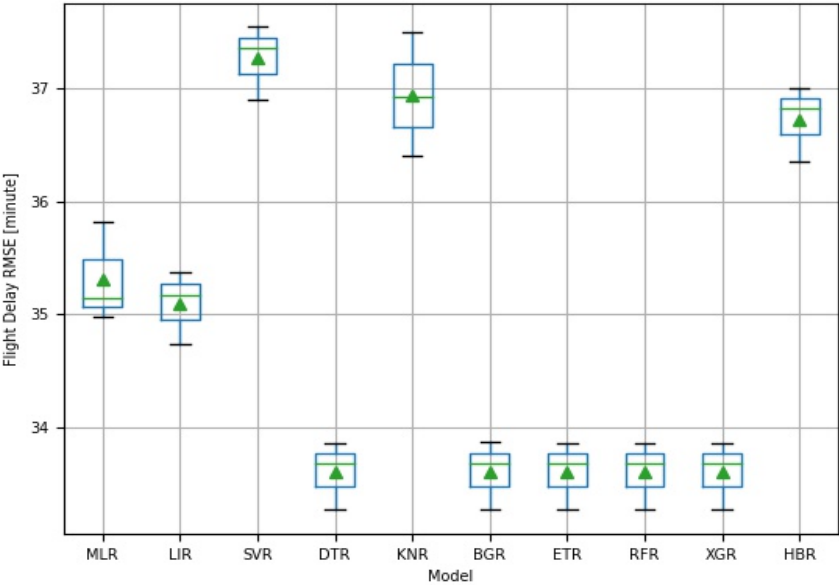


Figure 11. RMSE comparison of models for reduced scenario 4, with 3 inputs (ORI, NVU, AEL) and DEM as time delay to be predicted, in which the best RFR model obtains the same RMSE of the prediction of flight delay of 33.8 minutes than on scenario 3. Source: authors.

Figure 12 shows the distribution of the flight delay time (DEM) dataset flight attribute in terms of minimum (-15 minutes) and maximum (240 minutes), with the quantiles 25% and 75% and mean (15 minutes).

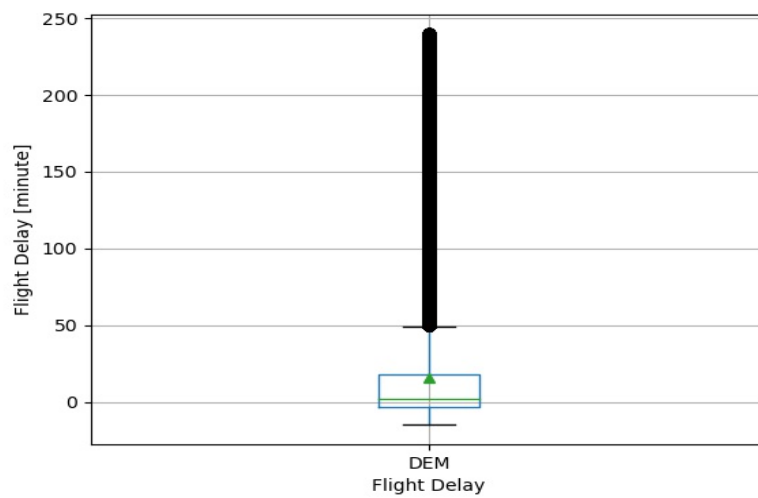


Figure 12. Time distribution of the flight departure delay. Source: authors.

Finally, Figure 13 shows that the continuous distribution function (designated as Density in the figure) of the flight delay time is unimodal, corresponding to the 15 minutes mean of the distribution.

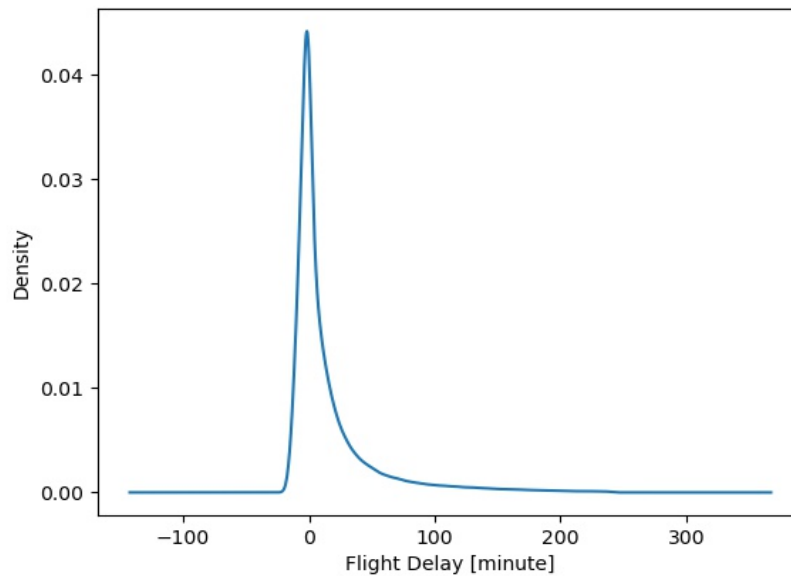


Figure 13. Unimodal distribution (density) function of a flight delay (at departure). Source: authors.

In summary, it can be stated that with the specific dataset flight attributes provided to the models for training purposes, the predictions of the delay time made by the ML/DL models are acceptable. Depending on the scenario considered, such models include error ranges between 16 and 33 minutes, which implies acceptable range values that are similar or even better than those obtained in the related research literature [71, 70, 106, 107].

7. CONCLUSIONS

As the periodic reports of the main international civil aviation organizations (IATA, ACI, ICAO, etc.) show, the demand for air transport has evidenced continuous and relevant growth for almost three decades. Airport authorities around the world are trying to respond to this demand by expanding air infrastructure, building new airports, optimizing the use of existing capacity, improving ATM procedures, etc. Even so, many airports (mainly large and international ones) continue to present congestion (often severe in certain periods of the day), a situation that inevitably leads to delays, which can be aggravated by adverse weather. Therefore, the most accurate prediction methods have become

indispensable, as they contribute to formulating strategies (efficient use of capacity, better management of operations, optimization of ATM procedures, etc.) to mitigate the proliferation of delays and prevent their spread into other airports in the network.

Therefore, the present research proposes a comparative analysis of the performance (measured in RMSE) of flight time delay prediction (at the departure of a flight) of 10 ML/DL models under different scenarios configured by the known flight attributes registered on the dataset for each flight. The ML models are in their most advanced variants (combinatorial models), such as ensemble techniques in their different methodologies such as voting, bagging algorithms (e.g., segmenting the dataset to obtain the advantages of training each model with different subsets), stacking (using outputs from some models to train others) and boosting (specifically penalizing unwanted errors). All these strategies serve to compensate or balance the errors of each model, so that the ensemble ML models focus on generalizing better the learning process. The DL/MLP model has also been included in the comparison (although it was not the winning model because it was not adjusted in its design parameters and hyperparameters). The comparative RMSE results for the transformed time delay (using standardized output data) are of the order of 0.45 and 0.96 (for the winning RFR model) depending on the flight attributes, or scenarios considered in the training process, which translates into (acceptable) prediction range errors between 16 and 33 minutes in time delay flights departures.

On the other hand, in this work the Permutation Importance method is implemented, which allows the input flight attribute to be auto ranked by their importance in determining the time delay by each model. Said flight attribute selection is re-evaluated on the corresponding reduced dataset scenarios, verifying that said auto reductions do not affect the precision (RMSE) of the time delay forecast and, in addition, allow reducing the size of the dataset.

Finally, although for the study case implemented here there is an extensive database of flight departures for a whole year associated with the entire national airport network and with many flight attributes (linked to each flight), there are constraints that prevent better ML performance. That is, certain important flight attributes are missing (e.g.,

weather conditions, aircraft size, airport congestion level, ATM condition, runway system capacity, aircraft turnaround time, etc.), that if included in the dataset could significantly improve the time delay ML prediction. Another possible line of research would be to compare the behavior of the prediction models developed here for the current case study (Colombian air transport system) (and using the same flight attributes) with another case study (air transport system of another country) with the same or similar characteristics.

8. REFERENCES

- [1] ICAO. Presentation of 2019 Air Transport Statistical Results. Montreal: ICAO, 2019.
- [2] IATA. World Air Transport Statistics. Geneva: IATA, 2021.
- [3] Graham, A., and Morrell, P. Airport finance and investment in the global economy. London: Routledge, 2017.
- [4] de Neufville, R., and Odoni, A. Airport Systems, Planning, Design, and Management. New York: McGraw-Hill, 2013.
- [5] Horonjeff, R., McKelvey, F., Sproule, W., and Young, S. Planning and Design of Airports. New York: McGraw-Hill, 2010.
- [6] Ashford, N., Mumayiz, S., and Wright, P. Airport Engineering. New Jersey: John Wiley, 2011.
- [7] ACRP. Defining and measuring aircraft delay and airport capacity thresholds. ACRP Report 104. Washington, D.C.: Transportation Research Board, 2014.
- [8] Gelhausen, M., Berster, P., and Wilken, D. Airport capacity constraints and strategies for mitigation. London: Academic Press, 2020.
- [9] Janic, M. Airport Analysis, Planning and Design: Demand, Capacity and Congestion. New York: Nova Science Publishers, 2009.
- [10] Borsky, S., and Unterberger, C. Bad weather and flight delays: The impact of sudden and slow onset weather events. *Economics of Transportation*, 2019, 18, pp 10–26. DOI: 10.1016/j.ecotra.2019.02.002

-
- [11] University of Westminster. European airline delay cost reference values. London: University of Westminster, 2015.
- [12] Peterson, E., Neels, K., Barczi, N., and Graham, T. The economic cost of airline flight delay. *Journal of Transport Economics and Policy*, 2013, 47(1), pp 107–121.
- [13] NEXTOR. Total Delay Impact Study. Washington D.C.: National Center of Excellence for Aviation Operations Research (Federal Aviation Administration), 2010.
- [14] Cook, A., Tanner, G., Jovanović, R., and Lawes, A. The cost of delay to air transport in Europe - Quantification and management. 13th Air Transport Research Society World Conference, June 27-30, 2009, Abu Dhabi.
- [15] BTS. Bureau of Transportation Statistic (U.S. Department of Transportation), 2021. <https://www.transtats.bts.gov/HomeDrillChart.asp>
- [16] EUROCONTROL. CODA Digest. All-causes delay and cancellations to air transport in Europe. Brussels: EUROCONTROL, 2020.
- [17] Aerocivil. Aeronautical statistics of Colombia, 2023. <https://www.aerocivil.gov.co/atencion/estadisticas-de-las-actividades-aeronauticas>
- [18] FAA. Cost of Delay Estimates. FAA APO-100. Washington D.C.: Federal Aviation Administration, 2020.
- [19] IATA. Inefficiency in European Airspace. Geneva: IATA, 2013.
- [20] Altmann, A., Tolosi, L., Sander, O., and Lengauer, T. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 2010, 26(10), pp 1340–1347. DOI: 10.1093/bioinformatics/btq134
- [21] Breiman, L. Random forests. *Machine Learning*, 2001, 45, pp 5-32.
- [22] Tan, X., Jia, X., Yan, J., Wang, K., and Bian, L. An Exploratory analysis of flight delay propagation in China. *Journal of Air Transport Management*, 2021, 92, 102025. DOI: 10.1016/j.jairtraman.2021.102025

- [23] Kim, M., and Park, S. Airport and route classification by modelling flight delay propagation. *Journal of Air Transport Management*, 2021, 93, 102045. DOI: 10.1016/j.jairtraman.2021.102045
- [24] Kafle, N., and Zou, B. Modeling flight delay propagation: A new analytical-econometric approach. *Transportation Research Part B*, 2016, 93, pp 520–542. DOI: 10.1016/j.trb.2016.08.012
- [25] Fleurquin, P., Ramasco, J., and Eguiluz, V. Systemic delay propagation in the US airport network. *Scientific Reports*, 2013, 3, pp 1-6. DOI: 10.1038/srep01159
- [26] Cai, Q., Alam, S., and Doung, V. A Spatial–Temporal Network Perspective for the Propagation Dynamics of Air Traffic Delays. *Engineering*, 2021, 7, pp 452–464. DOI: 10.1016/j.eng.2020.05.027
- [27] Zhang, M., Zhou, X., Zhang, Y., Sun, L., Dun, M., Du, W., and Cao, X. Propagation Index on Airport Delays. *Transportation Research Record*, 2019, 2673(8), pp 536–543. DOI: 10.1177/0361198119844240
- [28] Wang, P., Schaefer, L., and Wojcik, L. Flight connections and their impacts on delay propagation. *Digital Avionics Systems Conference*, 12-16 October 2003, Indianapolis (USA).
- [29] Markovic, D., Hauf, T., Roehner, P., and Spehr, U. A statistical study of the weather impact on punctuality at Frankfurt airport. *Meteorological Applications*, 2008, 15(2), pp 293–303.
- [30] Xu, N., Sherry, L., and Laskey, K. Multifactor model for predicting delays at U.S. airports. *Transportation Research Record*, 2008, 2052, pp 62–71. DOI: 10.3141/2052-08
- [31] Zhang, J., Xu, X., Wang, F. Airport delay performance evaluation based on fuzzy linear regression model. *Journal of Traffic and Transportation Engineering*, 2010, 10(4), pp 109-114.

-
- [32] Hao, L., Hansen, M., Zhang, Y., Post, J. New York, New York: Two ways of estimating the delay impact of New York airports. *Transportation Research Part E*, 2014, 70, pp 245–260. DOI: 10.1016/j.tre.2014.07.004
- [33] Boswell, S., and Evans, J. Analysis of downstream impacts of air traffic delay. Project Report, Lincoln Laboratory, Massachusetts Institute of Technology, 1997.
- [34] Wong, J., and Tsai, S. A survival model for flight delay propagation. *Journal of Air Transport Management*, 2012, 23, pp 5–11. DOI: 10.1016/j.jairtraman.2012.01.016
- [35] Wu, C. Inherent delays and operational reliability of airline schedules. *Journal of Air Transport Management*, 2005, 11(4), pp 273–282. DOI: 10.1016/j.jairtraman.2005.01.005
- [36] Baspinar, B., and Koyuncu, E. A Data-Driven Air Transportation Delay Propagation Model Using Epidemic Process Models. *International Journal of Aerospace Engineering*, 2016, Article ID 4836260. DOI: 10.1155/2016/4836260.
- [37] Pyrgiotis, N., Malone, K., and Odoni, A. Modelling delay propagation within an airport network. *Transportation Research Part C*, 2013, 27, pp 60–75. DOI: 10.1016/j.trc.2011.05.017
- [38] Hansen, M. Micro-level analysis of airport delay externalities using deterministic queuing models: a case study. *Journal of Air Transport Management*, 2002, 8(2), pp 73–87. DOI: 10.1016/S0969-6997(01)00045-X
- [39] Morrison, S., and Winston, C. The effect of FAA expenditures on air travel delays. *Journal of Urban Economics*, 2008, 63(2), pp 669–678.
- [40] Zou, B., and Hansen, M. Flight delays, capacity investment and social welfare under air transport supply-demand equilibrium. *Transportation Research Part A*, 2012, 46(6), pp 965–980. DOI: 10.1016/j.tra.2012.02.015
- [41] Qin, Q., and Yu, H. A statistical analysis on the periodicity of flight delay rate of the airports in the US. *Advances in Transportation Studies*, 2014, 3, pp 93–104.

- [42] Pathomsiri, S., Haghani, A., Dresner, M., and Windle, R. Impact of undesirable outputs on the productivity of US airports. *Transportation Research Part E*, 2008, 44(2), pp 235–259. DOI: 10.1016/j.tre.2007.07.002
- [43] Kotegawa, T., De Laurentis, D., Noonan, K., and Post, J. Impact of commercial airline network evolution on the U.S. air transportation system. *Proceedings of the 9th USA/Europe Air Traffic Management Research and Development Seminar*, 2011, pp 572–580.
- [44] Pfeil, D., and Balakrishnan, H. Identification of robust terminal-area routes in convective weather. *Transportation Science*, 2012, 46(1), pp 56–73.
- [45] Ahmadbeygi, S., Cohn, A., Guan, Y., and Belobaba, P. Analysis of the potential for delay propagation in passenger airline networks. *Journal of Air Transport Management*, 2008, 14(5), pp 221–236. DOI: 10.1016/j.jairtraman.2008.04.010
- [46] Soomer, M., and Franx, G. Scheduling aircraft landings using airlines' preferences. *European Journal of Operational Research*, 2008, 190(1), pp 277–291. DOI: 10.1016/j.ejor.2007.06.017
- [47] Mofokeng, T., and Marnewick, A. Factors contributing to delays regarding aircraft during A-check maintenance. *IEEE Technology and Engineering Management Society Conference, TEMSCON 2017*. DOI: 10.1109/TEMSCON.2017.7998375
- [48] Zou, B., and Hansen, M. Flight delay impact on air fare and flight frequency: a comprehensive assessment. *Transportation Research Part E*, 2014, 69, pp 54–74. DOI: Xu10.1016/j.tre.2014.05.016
- [49] Abdel-Aty, M., Lee, C., Bai, Y., Li, X., and Michalak, M. Detecting periodic patterns of arrival delay. *Journal of Air Transport Management*, 2007, 13(6), pp 355–361. DOI: 10.1016/j.jairtraman.2007.06.002
- [50] Zhong, Z., Varun, D., and Lin, Y. Studies for air traffic management R&D in the ASEAN-region context. *Journal of Air Transport Management*, 2017, 64, pp 15–20. DOI: 10.1016/j.jairtraman.2017.06.020

- [51] Ferrer, J., Oliveira, P., and Parasuraman, A. The behavioral consequences of repeated flight delays. *Journal of Air Transport Management*, 2012, 20, pp 35–38. DOI: 10.1016/j.jairtraman.2011.11.001
- [52] Lubbe, B., and Victor, C. Flight delays: towards measuring the cost to corporations. *Journal of Air Transport Management*, 2012, 19, pp 9–12. DOI: 10.1016/j.jairtraman.2011.11.004
- [53] Li, Q., and Jing, R. Characterization of delay propagation in the air traffic network. *Journal of Air Transport Management*, 2021, 94, 102075. DOI: 10.1016/j.jairtraman.2021.102075
- [54] Kim, M., and Bae, J. Modeling the flight departure delay using survival analysis in South Korea. *Journal of Air Transport Management*, 2021, 91, 101996. DOI: 10.1016/j.jairtraman.2020.101996
- [55] Mohammadian, I., Abbasi, B., Abareshi, A., and Goh, M. Antecedents of flight delays in the Australian domestic aviation market. *Transportation Research Interdisciplinary Perspectives*, 2019, 1, 100007. DOI: 10.1016/j.trip.2019.100007
- [56] Du, W., Zhang, M., Zhang, Y., Cao, X., and Zhang, J. Delay causality network in air transport systems. *Transportation Research Part E*, 2018, 118, pp 466–476.
- [57] Lambelho, M., Mitici, M., Pickup, S., and Marsden, A. Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions. *Journal of Air Transport Management*, 2020, 82, 101737. DOI: 10.1016/j.jairtraman.2019.101737
- [58] Chen, Y., and Lin, J. Determinants of flight delays at East Asian airports from an airport, route and network perspective. *Journal of Air Transport Management*, 2021, 94, 102064. <https://doi.org/10.1016/j.jairtraman.2021.102064>
- [59] Tu, Y., Ball, M., and Jank, W. Estimating flight departure delay distributions — A statistical approach with long-term trend and short-term pattern. *Journal of the American Statistical Association*, 2008, 103(481), pp 112–125. DOI: 10.1198/016214507000000257

- [60] Mueller, E., and Chatterji, G. Analysis of aircraft arrival and departure delay characteristics. AIAA's Aircraft Technology, Integration and Operations Conference, 1-3 October 2002, Los Angeles (USA).
- [61] Pamplona, D., and Alves, C.J. An overview of air delay: A case study of the Brazilian scenario. *Transportation Research Interdisciplinary Perspectives*, 2020, 7, 100189. DOI: 10.1016/j.trip.2020.100189
- [62] Fernandes, N., Moro, S., Costa, C., and Aparicio, M. Factors influencing charter flight departure delay. *Research in Transportation Business & Management*, 2020, 34, 100413.
- [63] Hunter, G., Boisvert, B., and Ramamoorthy, K. Advanced national airspace traffic flow management simulation experiments and validation. *Proceedings of the 39th conference on Winter simulation*, 2007, pp 1261–1267.
- [64] Bakhshandeh, R., Shahgholian, K., and Shahraki, A. Model for Reduce Flights Delays Using System Dynamics (Case Study in Iranian Airports Company). *Interdisciplinary Journal of Contemporary Research in Business*, 2013, 4(9), pp 746-757.
- [65] Santos, G., and Robin, M. Determinants of delays at European airports. *Transportation Research Part B*, 2010, 44, pp 392–403. DOI: 10.1016/j.trb.2009.10.007
- [66] Wang, C., and Wang, X. Airport congestion delays and airline networks. *Transportation Research Part E*, 2019, 122, pp 328–349. DOI: 10.1016/j.tre.2018.12.008
- [67] Wang, Y., Zhu, Y., Zhu, C., Wu, F., Yang, H., Yan, Y., and Hu, C. Indicator of serious flight delays with the approach of time-delay stability. *Physica A*, 2019, 518, pp 363–373. DOI: 10.1016/j.physa.2018.11.038
- [68] Zhang, M., Chen, S., Sun, L., Du, W., and Cao, X. Characterizing Flight Delay Profiles with a Tensor Factorization Framework. *Engineering*, 2021, 7, pp 465–472. DOI: 10.1016/j.eng.2020.08.024

- [69] Carvalho, L., Sternberg, A., Goncalves, L., Cruz, A., Soares, J., Brandão, D., Carvalho, G., and Ogasawara, E. On the relevance of data science for flight delay research: a systematic review. *Transport Reviews*, 2021, 41(4), pp 499–528. DOI: 10.1080/01441647.2020.1861123
- [70] Truong, D. Using causal machine learning for predicting the risk of flight delays in air transportation. *Journal of Air Transport Management*, 2021, 91, 101993. DOI: 10.1016/j.jairtraman.2020.101993
- [71] Guo, Z., Yu, B., Hao, M., Wang, W., Jiang, Y., and Zong, F. A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient. *Aerospace Science and Technology*. 2021. DOI: 10.1016/j.ast.2021.106822
- [72] IATA. *Airport Handling Manual*. Geneva: IATA, 2021.
- [73] Guss, W., and Salakhutdinov, R. On universal approximation by neural networks with uniform guarantees on approximation of infinite dimensional maps. 2019, arXiv:1910.01545v1.
- [74] Vang-Mata, R. *Multilayer Perceptrons: Theory and Applications*. Hauppauge (NY): Nova Science Publication, 2020.
- [75] Agarap, A. *Deep Learning using Rectified Linear Units (ReLU)*. 2019, arXiv:1803.08375v2.
- [76] Arora, R., Basu, A., Mianjy, P., and Mukherjee, A. Understanding deep neural networks with rectified linear units. 2018, arXiv:1611.01491v6.
- [77] Hara, K., Saito, D., and Shouno, H. Analysis of function of rectified linear unit used in deep learning. 2015 International Joint Conference on Neural Networks (IJCNN). DOI: 10.1109/IJCNN.2015.7280578
- [78] Arora, S., and Barak, B. *Computational complexity: a modern approach*. Cambridge: Cambridge University Press, 2009.
- [79] Bao, Y., Xiong, T., and Hu, Z. Forecasting Air Passenger Traffic by Support Vector Machines with Ensemble Empirical Mode Decomposition and Slope-Based Method.

- Discrete Dynamics in Nature and Society, 2012, ID 431512, 1-12. DOI: 10.1155/2012/431512
- [80] Schonlau, M., and Zou, R. The random forest algorithm for statistical learning. The Stata Journal, 2020, 20(1), pp 3-29. DOI: 10.1177/1536867X20909688
- [81] Montgomery, D., Peck, E., and Vining, G. Introduction to Linear Regression Analysis. Hoboken (NJ): Wiley, 2012.
- [82] Rokach, L. Ensemble Learning: Pattern Classification Using Ensemble Methods. Singapore: World Scientific Publishing, 2019.
- [83] Ganjisaffar, Y., Caruana, R., and Lopes, C. Bagging Gradient-Boosted Trees for High Precision, Low Variance Ranking Models. SIGIR'11, July 24–28, 2011, Beijing (China).
- [84] Friedman, J. Greedy function approximation: A gradient boosting machine. Annals of Statistics, 2000, 29, pp 1189–1232.
- [85] Hanif, I. Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction. ICSA 2019, August 02-03, Bogor (Indonesia). DOI: 10.4108/eai.2-8-2019.2290338
- [86] Chen, T., and Guestrin, C. XGBoost: A Scalable Tree Boosting System. KDD'16, August 13-17, 2016, San Francisco (USA). DOI: 10.1145/2939672.2939785
- [87] Wang, L., Zheng, C., Zhou, W., and Zhou, W. A new principle for tuning-free Huber Regression. Statistica Sinica, 2022, 32, pp 1-25. DOI: 10.5705/ss.202019.0045
- [88] Sun, Q., Zhou, W., and Fan, J. Adaptive Huber Regression. 2018, arXiv:1706.06991v2 [math.ST].
- [89] Camana, M., Ahmed, S., Garcia, C., and Koo, I. Extremely Randomized Trees-Based Scheme for Stealthy Cyber-Attack Detection in Smart Grid Networks. IEEE, 2020, 20, 19921-19933. DOI: 10.1109/ACCESS.2020.2968934
- [90] Geurts, P., Ernst, D., and Wehenkel, L. Extremely randomized trees. Mach Learn, 2006, 63, pp 3–42. DOI: 10.1007/s10994-006-6226-1

-
- [91] Izza, Y., Ignatiev, A., and Marques-Silva, J. On explaining Decision Trees. 2020, arXiv:2010.11034v1 [cs.LG].
- [92] Azadkia, M. Optimal choice of k for k-nearest neighbor regression. 2020, arXiv:1909.05495v4 [math.ST].
- [93] Goodfellow, I., and Bengio, Y.; Courville, A. Deep learning. Cambridge (MA): MIT Press, 2016.
- [94] Chen, S., Kuo, S., Chang, K., and Wang, Y. Improving the forecasting accuracy of air passenger and air cargo demand: the application of back-propagation neural networks. *Transportation Planning and Technology*, 2012, 35(3), pp 373-392.
- [95] Kuo, S., and Chen, S. Air passenger and air cargo demand forecasting: applying artificial neural networks to evaluating input variables. 12th WCTR, July 11-15, 2010. Lisbon (Portugal).
- [96] Liu, Y., Gao, Y., and Yin, W. An Improved Analysis of Stochastic Gradient Descent with Momentum. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver (Canada).
- [97] Gower, R., Loizou, N., Qian, X., Sailanbayev, A., Shulgin, E., and Richtarik, P. SGD: General Analysis and Improved Rates. 2019, arXiv:1901.09401v4 [cs. LG].
- [98] Dingari, M., Reddy, M., and Sumalatha, V. Air Traffic Forecasting Using Artificial Neural Networks. *International Journal of Scientific & Technology Research*, 2019, 8(10), pp 556-559.
- [99] Hastie, T., Tibshirani, R., and Friedman, J. The elements of statistical learning. Data mining, inference, and prediction. New York: Springer, 2013.
- [100] Kingma, D., and Ba, J. Adam: A Method for Stochastic Optimization. 2014, arXiv:1412.6980.
- [101] Aggarwal, C. Neural Networks and Deep Learning. Cham: Springer International, 2018.

- [102] Hassoun, M. *Fundamentals of Artificial Neural Networks*. Cambridge: MIT Press, 1995.
- [103] Pedrycz, W., and Chen, S. *Deep Learning: Concepts and Architectures*. Cham: Springer, 2020.
- [104] Chang, C., and Lin, C. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2(3), pp 1-27.
- [105] Rousseeuw, P., and van Driessen, K. Elliptic Envelope. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 1999, 41(3), pp 212-223.
- [106] Henriques, R., and Feiteira, I. Predictive Modelling: Flight Delays and Associated Factors, Hartsfield–Jackson Atlanta International Airport. *Procedia Computer Science*, 2018, 138, pp 638–645.
- [107] Belcastro, L., Marozzo, F., Talia, M., and Trunfio, P. Using Scalable Data Mining for Predicting Flight Delays. *ACM Transactions on Intelligent Systems and Technology*, 2016, 8(1), Article 5. DOI: 10.1145/2888402
- [108] ACI. *Annual World Airport Traffic Report*. Montreal: ACI (Airports Council International), 2021.
- [109] IATA. *World Air Transport Statistics 2021*. Geneva: IATA, 2021.
- [110] Díaz Olariaga, O. The role of regional airports in connectivity and regional development. *Periodica Polytechnica Transportation Engineering*, 2021, 49(4), pp 1-13. DOI: 10.3311/PPtr.16557
- [111] Díaz Olariaga, O., Pulido, L. Measurement of airport efficiency. The case of Colombia. *Transport and Telecommunication*, 2019, 20(1), pp 40-51. DOI: 10.2478/ttj-2019-0004
- [112] Díaz Olariaga, O., Zea, J.F. Influence of the liberalization of the air transport industry on configuration of the traffic in the airport network. *Transportation Research Procedia*, 2018, 33, pp 43-50.

- [113] Díaz Olariaga, O., Carvajal, A. Perspectiva geográfica del desarrollo de la conectividad aérea en Colombia. *Boletín Geográfico*, 2020, 42(2), pp 145 – 168.
- [114] Kearns, S. *Fundamentals of International Aviation*. London: Routledge, 2021.
- [115] Budd, L., Ison, S. *Air Transport Management*. London: Routledge, 2017.
- [116] Borsky, S., Unterberger, C. Bad weather and flight delays: The impact of sudden and slow onset weather events. *Economics of Transportation*, 2019, 18, pp 10–26. DOI: 10.1016/j.ecotra.2019.02.002
- [117] Choi, S., Kim, Y., Briceno, S., Mavris, D. Prediction of Weather-induced Airline Delays Based on Machine Learning Algorithms. *IEEE/AIAA 35th Digital Avionics Systems Conference Proceeding*, 2016. DOI:10.1109/DASC.2016.7777956
- [118] ACI. *Guide to Airport Performance Measures*. Montreal: ACI World, 2012.
- [119] FAA - Federal Aviation Administration. Definitions of variables Available, 2022, https://aspmhelp.faa.gov/index.php/APM:_Analysis:_Definitions_of_Variables
- [120] OAG. *OAG Punctuality League*. Chicago: OAG Aviation Worldwide, 2017.
- [121] ICAO. *Manual on Collaborative Air Traffic Flow Management (ATFM)*. Doc 9971. Montreal: ICAO, 2018.
- [122] European Observatory on Airport Capacity & Quality. *Delays to air transport in Europe: Methods of measuring, reporting, and analyzing*. Brussels: European Commission, 2015.
- [123] EUROCONTROL. *A Matter of Time: Air Traffic Delay in Europe*. Brussels: EUROCONTROL, 2007.